UNIVERSIDADE ESTADUAL DE CAMPINAS

INSTITUTO DE MATEMÁTICA, ESTATÍSTICA E COMPUTAÇÃO CIENTÍFICA

CESAR ANDREY GOMES FERREIRA

COMPRESSÃO DE IMAGENS DIGITAIS E A PSEUDO-INVERSA VIA DECOMPOSIÇÃO EM VALORES SINGULARES

Campinas

2021

CESAR ANDREY GOMES FERREIRA

COMPRESSÃO DE IMAGENS DIGITAIS E A PSEUDO-INVERSA VIA DECOMPOSIÇÃO EM VALORES SINGULARES

Dissertação apresentada ao Instituto de Matemática, Estatística e Computação Científica da Universidade Estadual de Campinas como parte dos requisitos exigidos para a obtenção do título de **Mestre em Matemática Aplicada e Computacional**

Orientador: Simão Nicolau Stelmastchuk

Este trabalho corresponde à versão final da Dissertação defendida pelo aluno Cesar Andrey Gomes Ferreira e orientada pelo Prof. Dr. Simão Nicolau Stelmastchuk.

Campinas

2021

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca do Instituto de Matemática, Estatística e Computação Científica
Ana Regina Machado - CRB 8/5467

F413c Ferreira, Cesar Andrey Gomes, 1967-
Compressão de imagens digitais e a pseudo-inversa via decomposição em valores singulares / Cesar Andrey Gomes Ferreira. – Campinas, SP : [s.n.], 2021.

Orientador: Simão Nicolau Stelmastchuk.
Dissertação (mestrado) – Universidade Estadual de Campinas, Instituto de Matemática, Estatística e Computação Científica.

1. Decomposição em valores singulares. 2. Fatoração (Matemática). 3. Matrizes (Matemática). 4. Imagens digitais. 5. Compressão de imagens. I. Stelmastchuk, Simão Nicolau, 1977-. II. Universidade Estadual de Campinas. Instituto de Matemática, Estatística e Computação Científica. III. Título.

Informações para Biblioteca Digital

Título em outro idioma: Digital images compression and the pseudoinverse by the singular value decomposition
Palavras-chave em inglês:
Singular value decomposition
Factorization (Mathematics)
Matrices
Digital images
Image compression
Área de concentração: Matemática Aplicada e Computacional
Titulação: Mestre em Matemática Aplicada e Computacional
Banca examinadora:
Simão Nicolau Stelmastchuk [Orientador]
Priscila Cristina Berbert Rampazzo
Marcos André Verdi
Data de defesa: 30-04-2021
Programa de Pós-Graduação: Matemática Aplicada e Computacional

Identificação e informações acadêmicas do(a) aluno(a)
- ORCID do autor: https://orcid.org/0000-0002-0528-8475
- Currículo Lattes do autor: http://lattes.cnpq.br/5804555877158499

Dissertação de Mestrado Profissional defendida em 30 de abril de 2021 e aprovada pela banca examinadora composta pelos Profs. Drs.

Prof(a). Dr(a). SIMÃO NICOLAU STELMASTCHUK

Prof(a). Dr(a). PRISCILA CRISTINA BERBERT RAMPAZZO

Prof(a). Dr(a). MARCOS ANDRÉ VERDI

A Ata da Defesa, assinada pelos membros da Comissão Examinadora, consta no SIGA/Sistema de Fluxo de Dissertação/Tese e na Secretaria de Pós-Graduação do Instituto de Matemática, Estatística e Computação Científica.

Agradecimentos

Inicialmente agredeço a Deus por ter colocado a Unicamp e seus espetaculares professores no meu caminho, motivo de extrema felicidade.

A minha mamãe, Dona Raimunhdinha e minhas tias Cleide e Terezinha pelo imenso carinho e dedicação. Agradeço também aos meus 7 irmãos pelo apoio e carinho.

Àos servidores da Universidade Estadual de Campinas, pela qualidade e dedicação no atendimento e pelo carinho que recebi de todos, em particular aos servidores da Secretaria de Pósgraduação e da Biblioteca.

Ao IFBA Campus Camaçari na pessoa da então Diretora de Ensino Profª Elisa Casaes por ter me dado condições de trabalho compativéis com meu status de professor estudante.

À meu ilustre e dedicado orientador, Profº Dr. Simão Nicolau. O rigor e o cuidado na orientação do meu trabalho contribuíram de modo decisivo para meu crescimento acadêmico. Foi uma honra trabalhar com o professor que tenho a maior admiração.

A meus colegas de trabalho e amigos, professores Dr. Jarbas Cordeiro pelo apoio constante, Dr. Ives Lima pela orientação valiosa a respeito da área de conhecimento referente ao trabalho, Dr. Alexandre do Nascimento Silva pela determinante contribuição para a estética e organização tipográfica do trabalho. Dr. Marcos Figueredo que praticamente me alfabetizou em Python e Dra. Andrea Brito pela atenção, paciência, carinho e boas ideias.

Ao casal de amigos Albérico e Nadja que foi meu esteio em São Paulo.

A todos os meus queridos colegas do Mestrado, em particular a Camila que mudou a direção do meu trabalho, a Makson que colaborou de modo determinante com o referencial teórico deste trabalho, a Saris e Suellen pelo apoio constante.

Por fim, ao amor da minha vida José Tiago Lino de Souza, razão de tudo.

Resumo

Esse trabalho pretende traçar uma estratégia para diagonalizar matrizes através de alguns tipos de fatorações, dentre elas, fatoração de matrizes quadradas simétricas e não simétricas e fatoração de matrizes quaisquer. Essa última fatoraçao diz respeito a um resultado notável chamado Decomposição em Valores Singulares cuja a sigla é SVD (Singular Value Decomposition). Para isso foram abordados alguns conceitos da Álgebra Linear como por exemplo, espaços e subespaços vetoriais, operadores lineares, autovalores, autovetores, formas quadráticas e o problema dos quadrados mínimos. O resultado pretendido é um material didático que auxilie professores e estudantes a trabalhar esses conteúdos em sala de aula, principalmente a Decomposição em Valores Singulares que geralmente não é abordado em cursos de graduação e possui aplicações poderosas. Serão apresentadas duas aplicações utilizando a Decomposição em Valores Singulares como instrumento, a Pseudo-Inversa e uma das mais relevantes, a compressão de imagens digitais.

Palavras-chaves: diagonalização de Matrizes; decomposição em valores singulares; compressão de imagens digitais.

Abstract

This work intends to outline a strategy to diagonalize matrices through some types of factoring, among them, factoring symmetric and non-symmetric square matrices and factoring any matrices. This last factorization concerns a remarkable result called Singular Value Decomposition whose acronym is SVD. For that, some concepts of Linear Algebra were approached, for example, vector spaces and subspaces, linear operators, eigenvalues, eigenvectors, quadratic forms, and the minimum squares problem. The intended result is a didactic material that helps teachers and students to work on these contents in the classroom, mainly the Singular Value Decomposition that is generally not addressed in undergraduate courses and has powerful applications. Two applications will be presented using Singular Value Decomposition as an instrument, the Pseudo-Inverse and one of the most relevant, the digital image compression.

Key-words: diagonalization of matrices; singular values decomposition; digital image compression.

Lista de ilustrações

Lista de Algoritmos

Sumário

1 Introdução

O Mestrado Profissional em Matemática Aplicada e Computacional, no qual esta dissertação está inserida, tem como objetivo capacitar pessoal para a prática profissional avançada no exercício do magistério superior. Neste contexto, o presente trabalho, foi pensado e produzido para ser um material didático complementar do componente curricular Álgebra Linear para cursos de graduação nas áreas das ciências exatas, como matemática, física e engenharias, ou pós graduações em que são realizadas pesquisas com suporte nessa área da Matemática.

Este trabalho foi dividido em três capítulos. O primeiro capítulo apresenta os conceitos preliminares com um referencial teórico que nos permita obter o conhecimento adequado para um satisfatório entendimento da decomposição em valores singulares, objeto principal desse trabalho. São apresentados definições, exemplos, teoremas com suas respectivas demonstrações com o intuito de tornar esse trabalho agradável e enriquecedor aos olhos do leitor, cujos conteúdos abordados para alcançar esse objetivo são vetores do espaço Euclidiano, o conceito de dependência linear de vetores, base e dimensão de espaços vetoriais, transformações lineares. Veremos também o significado de autovalores e autovetores, e estabelecer uma estratégia para diagonalizar matrizes invertíveis e matrizes simétricas. Além de dar um enfoque matricial ao problema de quadrados mínimos.

O segundo capítulo trata propriamente da teoria da decomposição em valores singulares conhecida também como SVD que é a abreviatura do seu nome em inglês, Singular Value Decomposition. A SVD será apresentada por meio de conceitos e propriedades. Antes se faz necessário apresentar um breve histórico do surgimento e desenvolvimento da SVD.

A teoria da SVD pode ser traçada até o trabalho de cinco matemáticos: o italiano Eugenio Beltrami, o francês Camille Jordan, o inglês James Sylvester, e os alemães Erhard Schimidt e Herman Weyl. Mais recentemente, os esforços pioneiros do matemático norte-americano Gene Golub produziram um algoritmo estável e eficaz para calculá-la. Beltrami e Jordan foram os pais da decomposição, sendo que, em 1873 Beltrami deu uma prova para o caso de matrizes reais invertíveis com valores singulares distintos. Subsequentemente, Jordan refinou a teoria e eliminou as restrições desnecessárias impostas por Beltrami. Sylvester, aparentemente desconhecendo o trabalho de Beltrami e Jordan, redescobriu o resultado em 1889 e indicou sua importância. Schimidt foi o primeiro a mostrar que a decomposição em valores singulares poderia ser usada para aproximar uma matriz por outra de posto menor e, ao fazer isso, ele transformou-a de uma curiosidade matemática numa importante ferramenta prática. Weyl mostrou como encontrar a aproximação de posto menor na presença de erro (STRANG, 2009; LIMA, 2011; ANTON; RORRES, 2012).

Este capítulo foi desenvolvido e estruturado para deixar claro o que significa uma SVD por meio de sua definição e suas propriedades apresentadas como teoremas, proposições e suas respectivas demonstrações. Além disso, é apresentada uma interpretação geométrica com o intuito de facilitar a visualização dessa decomposição acompanhada de algumas aplicações, entre elas, a norma euclidiana, a norma de Frobenius e o número de condição de uma matriz que permite classificar a matriz em mal-condicionada e bem-condicionada.

Por fim, no último capítulo são apresentadas duas aplicações da SVD, a primeira aplicação é a pseudo-inversa, de extrema importância na resolução de sitemas lineares quando estes têm infinitas soluções ou não têm solução alguma. A segunda é compactar imagens digitais com o objetivo de transmití-las a um custo computacional bem mais baixo do que transmitir a imagem original.

Por exemplo, se colocarmos uma figura, uma foto, no software MatLab, através de um determinado programa é possível transformar a referida figura em uma matriz numérica e aplicar a SVD nesta matriz aproximado-a assim de matrizes de posto inferior, obviamente perdendo algumas informações sobre a figura, mas mantendo as informações essenciais para tansmiti-la digitalmente com custo computacional mais barato. Contudo o software MatLab não é livre, então apresentaremos este trabalho com auxílio do software Python. Assim a proposta deste trabalho é fomentar uma discussão a cerca dessas aplicações ressaltando que existem outras inúmeras aplicações da decomposição em valores singulares.

2 Conceitos Preliminares

Nesse capítulo vamos apresentar e desenvolver as ferramentas que nos auxiliarão a entender a estrutura dos operadores lineares, o significado de autovalores e autovetores. Além disso, estabelecer uma estratégia para diagonalizar matrizes e resolver sistemas lineares pelo método dos mínimos quadrados (BOLDRINI *et al.*, 1980; AXLER, 1997; TREFETHEN; III, 1997; LEON, 2011; WATKINS, 2002; CALLIOLI; DOMINGUES; COSTA, 2003; SANTOS, 2007; LIPSCHUTZ; LIPSON, 2011; WATKINS, 2010; POOLE, 2011; ANTON; RORRES, 2012; GOLUB; LOAN, 2013; COELHO; LOURENçO, 2013).

2.1 Vetores de $\mathbb{R}^n$

A álgebra de vetores e a álgebra de matrizes possuem similaridades em vários aspectos. Em particular, podemos efetuar duas importantes operações, adição e multiplicação por escalar tanto de vetores quanto de matrizes. As propriedades resultantes dessas duas operações são as mesmas nesses dois espaços. Agora vamos definir vetores suas propriedades e estender para outros ambientes, nesse caso, para o ambiente das matrizes.

O conjunto de todas as ênuplas de números reais, denotada por $\mathbb{R}^n$, é chamado espaço Euclidiano. Uma ênupla específica de $\mathbb{R}^n$, digamos $v = (v_1, v_2, \cdots, v_n)$, é denominada ponto ou vetor. Os números v_i com $i = 1, \cdots, n$, são denominados coordenadas, componentes ou entradas de v. Além disso, quando trabalhamos com o espaço $\mathbb{R}^n$ usamos o termo escalar para os elementos de $\mathbb{R}$.

Exemplo 2.1.1. *Os elementos $v = (2, 5)$ e $u = (3, 8)$ são elementos de $\mathbb{R}^2$. Os vetores $t = (0, 0, 0)$ e $w = (1, 2, 6)$ são vetores de $\mathbb{R}^3$ em que t é o vetor nulo de $\mathbb{R}^3$.*

As vezes é conveniente escrever um vetor de $\mathbb{R}^n$ verticalmente em vez de horizontalmente. Dizemos que um vetor escrito dessa forma é um vetor coluna. Nesse contexto, os vetores do exemplo anterior estão escritos na forma horizontal, e são denominados vetores linhas. Por exemplo, seguem os vetores coluna, v, u, t e w, respectivamente,

$$v = \begin{bmatrix} 2 \\ 5 \end{bmatrix} \quad u = \begin{bmatrix} 3 \\ 8 \end{bmatrix} \quad t = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \text{ e } w = \begin{bmatrix} 1 \\ 2 \\ 6 \end{bmatrix}.$$

O que é relevante para a Álgebra Linear é que a extensão para dimensão n seja natural e imediata. Por exemplo, para um vetor de $\mathbb{R}^5$, precisamos de cinco componentes, mesmo que a visão geométrica seja perdida. Algumas características importantes dos vetores em

$\mathbb{R}^n$ são as propriedades de somar dois vetores e multiplicar um vetor por escalar. De forma clara, considere os vetores $u = (u_1, u_2, \cdots, u_n)$ e $v = (v_1, v_2, \cdots, v_n)$ de $\mathbb{R}^n$ e um escalar c de $\mathbb{R}$. A **soma** de $u + v$ dos vetores u e v é o vetor obtido somando os componentes correspondentes de u e v, ou seja,

$$u + v = (u_1 + v_1, u_2 + v_2, \cdots, u_n + v_n).$$

A **multiplicação por escalar** cv de vetor v pelo número real c é o vetor obtido pela multiplicação de cada componente de v por c, ou seja,

$$cv = (cv_1, cv_2, \cdots, cv_n).$$

Considere os vetores $v_1, v_2, \cdots, v_r$ de $\mathbb{R}^n$ e os escalares $c_1, c_2, \cdots, c_r$ de $\mathbb{R}$. Podemos multiplicar os vetores pelos escalares correspondentes e então somar os múltiplos escalares resultantes para obter o vetor

$$v = c_1v_1 + c_2v_2 + \cdots + c_rv_r.$$

Um vetor v, assim constituido, é denominado **combinação linear** dos vetores $v_1, v_2, \cdots, v_r$.

O conjunto $\mathbb{R}^n$ e o conjunto de todas as matrizes $m \times n$ têm uma estrutura comum chamada Espaço Vetorial. A motivação para a definição de um espaço vetorial vem das propriedades de adição e multiplicação escalar em $\mathbb{R}^n$. A adição é comutativa, associativa, e tem uma identidade. Cada elemento possui um inverso aditivo. A multiplicação por escalar é associativa, tem elemento neutro. Adição e multiplicação por escalar são conectadas por propriedades distributivas. Agora vamos definir formalmente espaço vetorial.

Definição 2.1.2. *Um espaço vetorial real é um conjunto V munido com uma adição em V e uma multiplicação por escalar em V que satisfazem as seguintes propriedades:*

a) *Comutatividade:* $u + v = v + u$, *para todo* $u, v \in V$*;*

b) *Associatividade:* $(u + v) + w = u + (v + w)$, *para todo* $u, v, w \in V$*;*

c) *Elemento Neutro da Adição: Existe* $0 \in V$ *tal que* $v + 0 = v$, *para todo* $v \in V$*;*

d) *Inverso Aditivo: Para todo* $v \in V$ *existe um elemento* $-v \in V$, *tal que* $v + (-v) = 0$*;*

e) *Elemento Neutro da Multiplicação: dado* $1 \in \mathbb{R}$ *temos que* $1 \cdot v = v$, *para todo* $v \in V$*;*

f) *Associatividade da Multiplicação por Escalar:* $a \cdot (b \cdot u) = (a \cdot b) \cdot u$, *para todo* $u \in V$ *e* $a, b \in \mathbb{R}$;

g) *Distributivas:* $a \cdot (u + v) = a \cdot u + a \cdot v$ *e* $(a + b) \cdot v = a \cdot v + b \cdot v$, *para todo* $a, b \in \mathbb{R}$ *e todo* $u, v \in V$.

Exemplo 2.1.3. *O conjunto* $\mathbb{R}^n$ *é um espaço vetorial. Para verificarmos as propriedades da definição de espaço vetorial considere os vetores* $u = (x_1, \cdots, x_n)$, $v = (y_1, \cdots, y_n)$ *e* $w = (z_1, \cdots, z_n)$ *de* $\mathbb{R}^n$, *e os escalares* a *e* b *de* $\mathbb{R}$.

a) *Comutativa – pela propriedade comutativa da soma dos números reais temos:*

$$\begin{aligned} u + v &= (x_1, \cdots, x_n) + (y_1, \cdots, y_n) \\ &= (x_1 + y_1, \cdots, x_n + y_n) \\ &= (y_1 + x_1, \cdots, y_n + x_n) \\ &= (y_1, \cdots, y_n) + (x_1, \cdots, x_n) \\ &= v + u. \end{aligned}$$

b) *Associatividade:*

Usando a propriedade associativa da soma entre números reais obtemos:

$$\begin{aligned} (u + v) + w &= (x_1 + y_1, \cdots, x_n + y_n) + (z_1, \cdots, z_n) \\ &= (x_1 + y_1 + z_1, \cdots, x_n + y_n + z_n) \\ &= (x_1 + (y_1 + z_1), \cdots, x_n + (y_n + z_n)) \\ &= (x_1, \cdots, x_n) + (y_1 + z_1, \cdots, y_n + z_n) \\ &= u + (v + w). \end{aligned}$$

c) *Elemento Neutro da Adição:*

Dado que existe um elemento neutro no corpo dos números reais, segue que

$$\begin{aligned} u + 0 &= (x_1, \cdots, x_n) + (0, \cdots, 0) \\ &= (x_1 + 0, \cdots, x_n + 0) \\ &= (x_1, \cdots, x_n) \\ &= u. \end{aligned}$$

Ou seja, $(0, 0, \cdots, 0)$ *é o elemento neutro em* $\mathbb{R}^n$.

d) *Inverso Aditivo:*

Dado que $\mathbb{R}$ *é um corpo, é possível mostrar que*

$$\begin{aligned} u + (-u) &= (x_1, \cdots, x_n) + (-(x_1, \cdots, x_n)) \\ &= (x_1, \cdots, x_n) + (-x_1, \cdots, -x_n) \\ &= (x_1 - x_1, \cdots, x_n - x_n) \\ &= (0, \cdots, 0) \end{aligned}$$

e) *Elemento Neutro da Multiplicação:*

Aqui, usaremos o fato de que o número real 1 *é o elemento neutro para a multiplicação no corpo dos números reais. De fato,*

$$\begin{aligned} 1 \cdot u &= 1 \cdot (x_1, \cdots, x_n) \\ &= (1 \cdot x_1, \cdots, 1 \cdot x_n) \\ &= (x_1, \cdots, x_n) \\ &= u. \end{aligned}$$

f) *Associatividade da Multiplicação por Escalar:*

Para mostrar esta propriedade usaremos o fato da multiplicação no corpo dos números reais ser associativa.

$$\begin{aligned} a(bu) &= a(b(x_1, \cdots, x_n)) \\ &= a(bx_1, \cdots, bx_n) \\ &= (abx_1, \cdots, abx_n) \\ &= ab(x_1, \cdots, x_n) \\ &= (ab)u. \end{aligned}$$

g) *Distributivas:*

Para provar ambas propriedades associativas, usaremos a propriedade distributiva do corpo dos números reais.

i)

$$\begin{aligned} a(u + v) &= a((x_1, \cdots, x_n) + (y_1, \cdots, y_n)) \\ &= a(x_1, \cdots, x_n) + a(y_1, \cdots, y_n) \\ &= au + av \end{aligned}$$

ii)

$$
\begin{aligned}
(a+b)u &= (a+b)(x_1,\cdots,x_n)\\
&= ((a+b)x_1,\cdots,(a+b)x_n)\\
&= (ax_1+bx_1,\cdots,ax_n+bx_n)\\
&= (ax_1,\cdots,ax_n)+(bx_1,\cdots,bx_n)\\
&= a(x_1,\cdots,x_n)+b(x_1,\cdots,x_n)\\
&= au+bu.
\end{aligned}
$$

Definição 2.1.4. *Um* ***subespaço*** *de um um espaço vetorial V é um subconjunto S de V não vazio que satisfaz as propriedades de um espaço vetorial. As combinações lineares e o vetor nulo permanecem no subespaço, isto é,*

i) *O vetor nulo pertence a S;*

ii) *S é fechado para adição, isto é, se $u, v \in S$ então $u+v \in S$;*

iii) *S é fechado para multiplicação por escalar, isto é, se $\alpha \in \mathbb{R}$ e $v \in S$ então $\alpha \cdot v \in S$.*

As propriedades (ii) e (iii) podem ser resumidas numa única afirmação equivalente que diz que para quaisquer $u, v \in S$ e $\alpha \in \mathbb{R}$, a combinação linear $\alpha u + v \in S$.

Exemplo 2.1.5. *O plano $\mathbb{R}^2$ pode ser definido como um subconjunto W do $\mathbb{R}^3$ onde a terceira coordenada é nula, isto é,*

$$W = \{(x,y,z) \in \mathbb{R}^3 \mid z = 0\}.$$

Nesse caso, o conjunto W é um subespaço vetorial do $\mathbb{R}^3$. De fato, sejam $u = (x_1, y_1, 0)$ e $v = (x_2, y_2, 0)$ vetores de W, e $\alpha \in \mathbb{R}$ um número qualquer. É imediato que o vetor nulo $(0,0,0)$ pertence a W, pois sua terceira coordenada é nula. Agora, como $\mathbb{R}^3$ é um espaço vetorial temos

$$\alpha \cdot u + v = \alpha(x_1, y_1, 0) + (x_2, y_2, 0) = (\alpha \cdot x_1 + x_2, \alpha \cdot y_1 + y_2, 0).$$

O que implica que $\alpha \cdot u + v \in W$. Assim, por definição, concluímos que W é um subespaço do $\mathbb{R}^3$.

Além do espaço vetorial $\mathbb{R}^n$, em nosso trabalho, precisaremos do espaço vetorial das matrizes. Para tal iniciamos por lembrar a definição de matriz.

Definição 2.1.6. *Sejam $m \geq 1$ e $n \geq 1$ dois números inteiros. Uma matriz é uma dupla sequência de números reais, distribuídas em m linhas e n colunas, formando uma tabela como a seguir:*

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

Abreviadamente esta matriz pode ser expressa por $[a_{ij}]$ com $1 \leq i \leq m$ e $1 \leq j \leq n$, ou apenas $[a_{ij}]$ quando não houver possibilidade de confusão quanto à variação dos índices. As matrizes são denotadas por letras maiúsculas do nosso alfabeto. Indicaremos por $M_{m\times n}(\mathbb{R})$ ou por $\mathbb{R}^{m\times n}$ o conjunto das matrizes reais $m \times n$.

A **soma** de duas matrizes $A = [a_{ij}]$ e $B = [b_{ij}]$, ambas de mesma ordem $m \times n$, é uma matriz denotada por $A + B$ também de ordem $m \times n$, cujos elementos são somas dos elementos correspondentes de A e B, isto é,

$$A + B = [a_{ij} + b_{ij}]_{m\times n}.$$

Dadas as matrizes $A = [a_{ij}]$, $B = [b_{ij}]$ e $C = [c_{ij}]$ de mesma ordem $m \times n$, a soma de matrizes goza das seguintes propriedades:

A1) Comutatividade: $A + B = B + A$;
Pela definição de soma de matrizes,

$$A + B = [a_{ij} + b_{ij}] = [b_{ij} + a_{ij}] = B + A,$$

onde na segunda igualdade usamos a comutatividade entre dois números reais.

A2) Associatividade: $A + (B + C) = (A + B) + C$;
Por definição,

$$\begin{aligned}(A + B) + C =&[a_{ij} + b_{ij}] + [c_{ij}] = [(a_{ij} + b_{ij}) + (c_{ij})]\\ =&[a_{ij} + (b_{ij} + c_{ij})] = [a_{ij}] + [b_{ij} + c_{ij}]\\ =&A + (B + C),\end{aligned}$$

onde na terceira igualdade usamos a associatividade entre três números reais.

A3) Elemento Neutro: $A + 0 = A$, em que 0 é a matriz nula $m \times n$;
Pela definição de soma de matrizes,

$$A + O = [a_{ij} + o_{ij}] = [a_{ij}] = A,$$

onde a última igualdade decorre do fato de zero ser um elemento neutro para a soma entre dois números reais.

A4) Elemento Oposto: existe uma matriz $(-A)$, também $m \times n$, tal que $A + (-A) = 0$.
Se $A = [a_{ij}]$, é evidente que $-A = [-a_{ij}]$. Por definição,

$$A + (-A) = [a_{ij} + (-a_{ij})] = 0,$$

onde na última igualdade usamos o fato da soma de dois números reais opostos ser nula.

Considere uma matriz $A = [a_{ij}]$ de ordem $m \times n$ e um número real k, a **multiplicação** de uma matriz por um escalar também é uma matriz $m \times n$ dada por

$$k \cdot A = [k \cdot a_{ij}].$$

Dadas as matrizes $A = [a_{ij}]$, $B = [b_{ij}]$ de mesma ordem $m \times n$, e os escalares $c, d \in \mathbb{R}$, o produto de uma matriz por um escalar goza das seguintes propriedades:

M1) Associatividade: $(c \cdot d) \cdot A = c \cdot (d \cdot A)$;
Pela definição de multiplicação de matrizes por um escalar,

$$(cd)A = [(cd)(a_{ij})] = [c(d \cdot a_{ij})] = c[d \cdot a_{ij}] = c(dA),$$

onde na segunda igualdade usamos a associatividade de números reais.

M2) Distributivas: $(c + d) \cdot A = c \cdot A + d \cdot A$ e $c \cdot (A + B) = c \cdot A + c \cdot B$;
Pela definição de multiplicação de matrizes por um escalar,

$$(c + d)A = [(c + d)a_{ij}] = [c \cdot a_{ij} + d \cdot a_{ij}] = [c \cdot a_{ij}] + [d \cdot a_{ij}] = cA + dA,$$

onde na segunda igualdade usamos a distributividade de números reais. Da mesma forma,

$$c(A + B) = [c(a_{ij} + b_{ij})] = [c \cdot (a_{ij})] + [c \cdot (b_{ij})] = cA + cB.$$

onde a segunda igualdade decorre também da distributividade de números reais.

M3) Elemento Neutro: $1 \cdot A = A$.
Pela definição de multiplicação de matrizes por um escalar,

$$1 \cdot A = [1 \cdot (a_{ij})] = [(a_{ij})] = A.$$

onde a segunda igualdade decorre do fato de 1 ser neutro na multiplicação de números reais.

As propriedades A1-A4 e M1-M3 demonstradas acima nos mostram que o espaço das matrizes de ordem $m \times n$ é um espaço vetorial.

Exemplo 2.1.7. *Dadas as matrizes* $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ *e* $B = \begin{bmatrix} -1 & 0 \\ 1 & -2 \end{bmatrix}$ *pela definição de soma de matrizes e multiplicação de uma matriz por um escalar temos*

$$2A + 3B = 2\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} + 3\begin{bmatrix} -1 & 0 \\ 1 & -2 \end{bmatrix} = \begin{bmatrix} -1 & 4 \\ 9 & 2 \end{bmatrix}$$

Além da soma e de multiplicação por um escalar no espaço de matrizes, o produto entre duas matrizes faz um importante papel para nós.

Definição 2.1.8. *Seja* $A = [a_{ij}]$ *uma matriz* $m \times r$ *e* $B = [b_{ij}]$ *uma matriz* $r \times n$*, o produto* $A \cdot B$ *é uma matriz* $m \times n$ *definida por*

$$A \cdot B = \left[\sum_{k=1}^{r} a_{ik} b_{kj} \right].$$

De outra forma, para obter a entrada da linha i e coluna j de $A \cdot B$, destacamos a linha i de A e a coluna j de B, daí multiplicamos as entradas correspondentes da linha e da coluna destacadas e então somamos os produtos resultantes.

Exemplo 2.1.9. *Considere as matrizes*

$$A = \begin{bmatrix} 1 & 0 & -1 \\ 3 & 2 & 1 \end{bmatrix} \quad \text{e} \quad B = \begin{bmatrix} -1 & 0 \\ 1 & -2 \\ 3 & 4 \end{bmatrix}$$

Como A *é* 2×3 *e* B *é* 3×2*, o produto* AB *é uma matriz quadrada* 2×2*. Para determinar a entrada na linha 1 e coluna 2 de* AB*, destacamos a linha 1 de* A *e a coluna 2 de* B *e procedemos conforme a definição, ou seja,*

$$\begin{bmatrix} 1 & 0 & -1 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ -2 \\ 4 \end{bmatrix} = 1 \cdot 0 + 0 \cdot (-2) + (-1) \cdot 4 = -4.$$

Procedendo de modo análogo obtemos todas as entradas do produto AB*, isto é,*

$$AB = \begin{bmatrix} -4 & -4 \\ 2 & 0 \end{bmatrix}$$

Definição 2.1.10. *Dada uma matriz* $A = [a_{ij}]$ *de ordem* $m \times n$ *, podemos obter uma outra matriz* $A^T = [b_{ji}]$ *de ordem* $n \times m$*, cujas linhas são as colunas de* A*, isto é,* $b_{ji} = a_{ij}$*.* A^T *é denominada transposta de* A*.*

A partir da definição de transposta, podemos mostrar que a transposta do produto de duas matrizes é dado por $(A \cdot B)^T = B^T \cdot A^T$.

Lembramos que o produto escalar de dois vetores $u = (u_1, u_2, \cdots, u_n)$ e $v = (v_1, v_2, \cdots, v_n)$ em $\mathbb{R}^n$ é dado por

$$u \cdot v = u_1 v_1 + u_2 v_2 + \cdots u_n v_n.$$

Exemplo 2.1.11. *Considererando os vetores $u = (1, 3, 5)$ e $v = (4, 5, 0)$ de $\mathbb{R}^3$, vamos calcular o produto escalar $u \cdot v$ e os produtos de matrizes $u^T \cdot v$ e $v^T \cdot u$.*

- $u \cdot v = (1, 3, 5) \cdot (4, 5, 0) = 1 \cdot 4 + 3 \cdot 5 + 5 \cdot 0 = 19.$
- $u^T \cdot v = \begin{bmatrix} 1 \\ 3 \\ 5 \end{bmatrix} \cdot \begin{bmatrix} 4 & 5 & 0 \end{bmatrix} = 1 \cdot 4 + 3 \cdot 5 + 5 \cdot 0 = 19.$
- $v^T \cdot u = \begin{bmatrix} 4 \\ 5 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 1 & 3 & 5 \end{bmatrix} = 4 \cdot 1 + 5 \cdot 3 + 0 \cdot 5 = 19.$

Esse exemplo destaca um caso especial de um produto de um vetor linha por um vetor coluna como os mesmo número de entradas. Em verdade, é fácil mostrar que

$$u \cdot v = u^T \cdot v = (u^T \cdot v)^T = v^T \cdot u.$$

2.2 Sistemas de Equações Lineares

Em muitas ciências como, por exemplo, Matemática, Física, Química, Administração, entre outras, é comum a informação ser organizada em linhas e colunas formando o que conhecemos como matrizes e os sistemas lineares tem uma relação íntima com as matrizes.

Toda informação necessária para classificar um sistema linear com respeito ao número de soluções ou a resolução desse sistema podem ser tiradas de uma matriz através de operações apropriadas nessa matriz. Por exemplo, as informações sobre o sistema linear

$$\begin{cases} 2x + 3y = 4 \\ x - 4y = 1 \end{cases}$$

estão localizadas na matriz

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

Essa relação íntima entre sistemas lineares e matrizes é bastante relevante no desenvolvimento programas computacionais para resolver sistemas, isso porque os programas computacionais são apropriados para manipular tabelas numéricas. As matrizes não servem apenas para auxiliar na resolução de sistemas, existe uma vasta e rica teoria em torno das matrizes, além disso, as matrizes têm uma imensa variedade de aplicações que veremos a seguir.

Definição 2.2.1. *Uma equação linear nas incógnitas $x_1, x_2, \cdots, x_n$ é uma equação da forma*

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = b.$$

Cada n-upla $(x_1, x_2, \cdots, x_n)$ que torna essa equação uma setença verdadeira denominamos solução da equação.

Ao conjunto formado por duas ou mais equações lineares damos o nome de Sistema Linear, ou seja,

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

Para um melhor tratamento computacional, é possivel escrever este sistema na forma matricial da seguinte maneira

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

ou $Ax = b$ em que $A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$ é a matriz dos coeficientes $m \times n$, $x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ é a matriz das incógnitas $n \times 1$ e $b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$ é a matriz dos termos independentes $m \times 1$.

Para um sistema de equações lineares, podemos encontrar três possibilidades quanto as soluções: existe uma única solução, existem inúmeras soluções ou não existe solução. Para cada caso temos uma nomenclatura: dizemos que um sistema é **consistente** se tiver uma ou mais soluções; se não possuir solução alguma dizemos que o sistema é **inconsistente**. Dois sistemas lineares são ditos **equivalentes** se possuem as mesmas soluções.

Exemplo 2.2.2. *Seja o sistema de equações lineares*

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

e $W = \{X = (x_1, x_2, \cdots, x_n)\} \subset \mathbb{R}^n$ *o conjunto solução do sistema* S*. Afirmamos que o conjunto* W *é um subspaço vetorial de* $\mathbb{R}^n$ *se, e somente se,* S *é um sistema homogêneo, ou seja,* $b_i = 0$ *para* $i = 1, \cdots m$*. De fato, supondo que* W *é subspaço vetorial de* $\mathbb{R}^n$*, então* $(0, 0, \cdots, 0) \in W$*. Considerando a forma matricial do sistema temos* $A \cdot 0 = B$*, o que implica que* $B = 0$*. Logo,* S *é homogêneo. Reciprocamente, se* S *é homogêneo, isto é,* $A \cdot X = 0$*, vamos verificar que* W *é subspaço de* $\mathbb{R}^n$*.*

i) *O vetor nulo pertence a* W *pois* $A \cdot 0 = 0$*.*

ii) *Sejam* $X, Y \in W$ *e* $c \in \mathbb{R}$*, então* $A(cX+Y) = A(cX)+AY = c(AX)+AY = c\cdot 0+0 = 0$*, o que implica que* $cX + Y \in W$*.*

Agora, por i) e ii), concluímos que W *é subspaço de* $\mathbb{R}^n$*.*

2.3 Base, Dimensão e Posto

A Geometria Analítica nos dá um alicerce adequado para nos apropriarmos de certos conceitos da Álgebra Linear. Observe que a noção de um subespaço é simplesmente uma generalização algébrica dos exemplos geométricos de retas e planos que passam pela origem. O conceito fundamental de base para um subespaço, tendo em vista essa analogia, é derivado da ideia de vetores diretores para tais retas e planos. O conceito de base nos dará instrumento para definir precisamente dimensão. Por fim, essas ideias trazem luz ao conhecimento que já trazemos conosco sobre matrizes e soluções de sistemas lineares.

Analisar se existe uma relação de dependência entre os vetores de um conjunto

$$S = \{v_1, v_2, \cdots, v_n\}$$

é fundamental para resolver o problema de decidir se um ou mais vetores de S podem ser expressos como combinação linear dos outros. Então vamos definir geradores e dependência linear.

Definição 2.3.1. *Seja $S = \{v_1, v_2, \cdots, v_n\}$ um subconjunto do espaço vetorial V. O conjunto de todas as combinações lineares $c_1v_1 + c_2v_2 + \cdots + c_nv_n$ de $v_1, v_2, \cdots, v_n$, é chamado de conjunto gerado por $v_1, v_2, \cdots, v_n$ e é denotado por $span(v_1, v_2, \cdots, v_n)$. Se $V = span(v_1, v_2, \cdots, v_n)$, dizemos que S é um conjunto de geradores de V, e diz-se que V é gerado por S.*

Definição 2.3.2. *Um conjunto $S = \{v_1, v_2, \cdots, v_n\}$ em um espaço vetorial V é linearmente dependente se existirem escalares $c_1, c_2, \cdots, c_n$ não todos nulos tais que a combinação linear dos vetores de S seja nula, isto é, $c_1v_1 + c_2v_2 + \cdots + c_nv_n = 0$. Caso contrário, dizemos que o conjunto S é linearmente independente.*

Agora considere um espaço vetorial V e pense em um subconjunto finito de vetores tais que, qualquer outro vetor de V seja uma combinação linear deles. Ou seja, queremos obter um conjunto que gere o espaço V de modo que seus elementos sejam absolutamente necessários para alcançar tal objetivo.

Exemplo 2.3.3. *Da Geometria Analítica temos que qualquer vetor (x, y) do plano $\mathbb{R}^2$ pode ser expresso como combinação linear dos vetores $i = (1, 0)$, $j = (0, 1)$ e $w = (2, 0)$ da seguinte forma: $a(1, 0) + b(0, 1) + 0(2, 0)$, para todo $a, b \in \mathbb{R}$. Observe que não é necessário esses três vetores para gerar o plano $\mathbb{R}^2$, mais ainda, o conjunto formado por esses três vetores são linearmente dependentes pois $w = 2i$. Bastam os vetores $i = (1, 0)$ e $j = (0, 1)$ para gerar o plano $\mathbb{R}^2$. Além disso, esses vetores são linearmente independentes.*

Conjuntos de vetores com essas características são os menores conjuntos que geram um espaço vetorial. Denominamos os conjuntos de vetores desse tipo de base.

Definição 2.3.4. *Uma base de V é uma lista de vetores $\{v_1, v_2, \cdots, v_n\}$ em V que é linearmente independente e gera V.*

Teorema 2.3.5. *Seja V um espaço vetorial gerado por um conjunto finito de vetores e $B = \{v_1, v_2, \cdots, v_n\}$ uma base de V. Então, qualquer conjunto com mais de n vetores é linearmente dependente.*

Demonstração. Seja $S = \{w_1, w_2, \cdots, w_m\}$ um conjunto qualquer de m vetores em V, com $m > n$. Como $B = \{v_1, v_2, \cdots, v_n\}$ é uma base, cada w_i, $i = 1, \cdots, m$, pode ser expresso como combinação linear dos vetores de B. Temos então

$$\begin{aligned} w_1 = a_{11}v_1 + a_{21}v_2 \quad & \cdots \quad a_{n1}v_n \\ w = a_{12}v_1 + a_{22}v_2 \quad & \cdots \quad a_{n1}v_n \\ \vdots \qquad\qquad & \ddots \qquad \vdots \\ w = a_{1m}v_1 + a_{2m}v_2 \quad & \cdots \quad a_{nm}v_n. \end{aligned}$$

Agora, devemos encontrar escalares $c_1, c_2, \cdots, c_m$ não todos nulos, tais que

$$c_1w_1 + \cdots c_mw_m = 0$$

. Podemos escrever então

$$(a_{11}c_1 + \cdots + a_{1m}c_m)v_1 + \cdots + (a_{n1}c_1 + \cdots + a_{nm}c_m)v_m = 0.$$

Pela independência linear da base $B = \{v_1, v_2, \cdots, v_n\}$ temos o seguinte sistema

$$\begin{cases} a_{11}c_1 + \cdots + a_{1m}c_m = 0 \\ \vdots \\ a_{n1}c_1 + \cdots + a_{nm}c_m = 0. \end{cases}$$

Observe que esse sistema tem mais incógnitas que equações já que $m > n$. Logo o sistema admite pelo menos uma solução não trivial, o que implica que os vetores $w_1, w_2, \cdots, w_m$ são linearmente dependentes. □

Corolário 2.3.6. *Qualquer base de um espaço vetorial V tem sempre o mesmo número de elementos.*

Demonstração. Sejam $B_1 = \{v_1, v_2, \cdots, v_n\}$ e $B_2 = \{w_1, w_2, \cdots, w_m\}$ duas bases de V. Como $B_1 = \{v_1, v_2, \cdots, v_n\}$ é uma base de V e $B_2 = \{w_1, w_2, \cdots, w_m\}$ é linearmente independente então, pelo Teorema 2.3.5, $m \leq n$. Por outro lado, $B_2 = \{w_1, w_2, \cdots, w_m\}$ é uma base e $B_1 = \{v_1, v_2, \cdots, v_n\}$ é linearmente independente, então $n \leq m$. Disso concluímos que $m = n$. □

O número encontrado no Corolário acima é chamado **dimensão** de V, e é denotado por $dimV$.

Como muitos exemplos de subespaços surgem no contexto de matrizes nada mais natural que associar conceitos vistos até o momento às matrizes. Dito isso vamos definir a seguir importantes subespaços associados à matrizes.

Definição 2.3.7. *Seja A uma matriz $m \times n$.*

i) *O espaço linha de A, denotado por $Lin(A)$, é o subespaço de $\mathbb{R}^n$ gerado pelas linhas de A.*

ii) *O espaço coluna de A, denotado por $Col(A)$ é o subespaço de $\mathbb{R}^m$ gerado pelas colunas de A.*

iii) *O posto de uma matriz A, denotado por posto(A) é a dimensão dos seus espaços linha e coluna.*

iv) *O espaço nulo de A é o subespaço $N(A)$ de $\mathbb{R}^n$ que consiste nas soluções do sistema linear homogêneo $Ax = 0$. Ou seja, $N(A) = \{x \in \mathbb{R}^n | Ax = 0\}$.*

v) *A nulidade de uma matriz A, denotada por $nul(A)$ é a dimensão do espaço nulo de A. Em outras palavras, é a dimensão do espaço solução de $Ax = 0$, que é o mesmo número de variáveis livres na solução do sistema.*

2.4 Transformações Lineares

Em um determinado sentido o estudo das transformações lineares tem uma relação íntima com o estudo de matrizes como veremos mais adiante. Desta forma, vamos apresentar sua definição e uma associação do conceito de transformações lineares com o espaço das matrizes. Além disso, estabelecer um paralelo entre espaço nulo de uma matriz e o núcleo de uma transformação linear, o espaço coluna de uma matriz e a imagem de uma transformação linear.

Definição 2.4.1. *Uma transformação linear é uma função $T : V \to W$ com as seguintes propriedades:*

- *aditividade: $T(u + v) = T(u) + T(v)$ para todos $u, v \in V$;*
- *homogeneidade: $T(c \cdot v) = c \cdot T(v)$ para todo $c \in \mathbb{R}$ e para todo $v \in V$.*

Quando $V = W$ dizemos que a transformação linear $T : V \to V$ é um operador linear.

Observação 2.4.2. *Por definição, a aplicação $T : V \to W$ é linear se preserva a soma de vetores e a multiplicação por escalar, as duas operações básicas de um espaço vetorial. As duas propriedades da definição podem ser resumidas em uma única propriedade,*

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

para todos $u, v \in V$ e $\alpha \in \mathbb{R}$.

Conforme foi dito, podemos associar o estudo das transformações lineares ao estudo de matrizes. Seja A uma matriz $m \times n$. Definimos

$T_A : \mathbb{R}^n \to \mathbb{R}^m$ por $v \mapsto y = Av$ em que $v = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$, então

$$T_A(v) = Av = A \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}$$

.

Das propriedades de operações de matrizes temos $T_A(c \cdot u + v) = A(c \cdot u + v) = c \cdot Au + Av$, portanto T_A é uma transformação linear. Esse resultado que, como vimos, vale para espaços Euclidianos pode ser naturalmente estendido para espaços vetoriais V e W quaisquer.

Seja $T : V \to W$ uma transformação linear. Suponhamos que o conjunto $B_1 = \{v_1, \cdots, v_n\}$ seja uma base de V e $B_2 = \{w_1, \cdots, w_m\}$ de W. Então $T(v_1), \cdots, T(v_n)$ são vetores de W e portanto podem ser expressos como combinação linear dos vetores da base $B_2 = \{w_1, \cdots, w_m\}$. Desta forma temos

$$\begin{aligned}
T(v_1) &= a_{11}.w_1 + a_{21}.w_2 + ... + a_{m1}.w_n \\
T(v_2) &= a_{12}.w_1 + a_{22}.w_2 + ... + a_{m2}.w_n \\
&\vdots \\
T(v_n) &= a_{1n}.w_1 + a_{2n}.w_2 + ... + a_{mn}.w_n.
\end{aligned}$$

A transposta da matriz dos coeficientes deste sistema, denotada por $A = [T]_{B_2}^{B_1}$, é chamada matriz de T em relação às bases B_1 e B_2 , ou seja,

$$A = [T]_{B_2}^{B_1} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{mn} \end{bmatrix}.$$

Observe que T se torna a transformação linear associada à matriz A e bases B_1 e B_2, isto é, $T = T_A$.

Definição 2.4.3. *Seja $T : V \to W$ uma transformação linear. O conjunto de todos os vetores $v \in V$ tais que $T(v) = 0$ é chamado núcleo de T e é denotado por Ker(T) ou N(T). Isto é*

$$N(T) = \{v \in V \mid T(v) = 0\}.$$

Note, pela definição, que o núcleo de T equivale ao **espaço nulo** da matriz da transformação $T : V \to W$. Neste caso, se A for a matriz da transformação linear T, isto é, $T = T_A$, então se v é um vetor tal que $T(v) = 0$ se, somente se, $T_A \cdot [v] = 0$.

O núcleo de $T : V \to W$ é subespaço vetorial de V. De fato, como $T(0) = 0$ temos $0 \in Ker(T)$. Além disso, se $u, v \in Ker(T)$ e $\alpha \in \mathbb{R}$, então

$$T(\alpha u + v) = \alpha T(u) + T(v) = 0 + 0 = 0.$$

Portanto, $\alpha u + v \in Ker(T)$. Diante do exposto concluímos que o núcleo de T é subespaço de V.

Definição 2.4.4. *Seja $T : V \to W$ uma transformação linear. A imagem de T, denotada por $R(T)$ ou $Im(T)$ é o conjunto dos vetores $w \in W$ tais que existe um vetor $v \in V$, que satisfaz $T(v) = w$. Ou seja,*

$$R(T) = \{T(v); v \in V\}.$$

Observe, pela definição, que a imagem equivale ao espaço coluna, espaço imagem ou simplesmente imagem da matriz da transformação $T : V \to W$. Ou seja, a imagem de A de ordem $m \times n$ é formada pelos vetores que são combinações lineares das colunas da matriz A. Assim $b \in R(A)$ se, e somente se, existe um $x \in \mathbb{R}^m$ tal que $Ax = b$. Como consequência da definição, temos também que o sistema linear $Ax = b$ é consistente se o vetor b está na imagem de A.

A imagem de $T : V \to W$ é subespaço vetorial de W. De fato, como $T(0) = 0$ temos $0 \in Im(T)$. Além disso, se $w_1, w_2 \in Im(T)$ e $\alpha \in \mathbb{R}$, existem $u, v \in V$ tais que $T(u) = w_1$ e $T(u) = w_2$. Então

$$T(\alpha u + v) = \alpha T(u) + T(v) = \alpha w_1 + w_2 \in Im(T)$$

. Assim concluímos que a imagem de T é subespaço de W.

Exemplo 2.4.5. *Vamos encontrar a matriz $A = [T]^\alpha_\beta$ da transformação linear $T : \mathbb{R}^3 \to \mathbb{R}^2$ definida por $T(x, y, z) = (2x + y - z, 3x - 2y + 4z)$ em relação às bases*

$$\alpha = \{(1, 1, 1), (1, 1, 0), (1, 0, 0)\} \text{ e } \beta = \{(1, 3), (1, 4)\}$$

de $\mathbb{R}^3$ e $\mathbb{R}^2$ respectivamente.

Primeiro vamos calcular A com respeito aos elementos da base α. Assim temos

$$\begin{aligned} T(1, 1, 1) &= (2, 5) = a(1, 3) + b(1, 4) \\ T(1, 1, 0) &= (3, 1) = c(1, 3) + d(1, 4) \\ T(1, 1, 1) &= (2, 3) = e(1, 3) + f(1, 4), \end{aligned}$$

que resulta nos sistemas

$$\begin{cases} a + b = 2 \\ 3a + 4b = 5 \end{cases}, \quad \begin{cases} c + d = 3 \\ 3c + 4d = 1 \end{cases} \quad e \begin{cases} e + f = 2 \\ 3e + 4f = 3, \end{cases}$$

cujas soluções são dadas por

$$a = 3,\ b = -1,\ c = 11,\ d = -8,\ e = 5 \text{ e } f = -3.$$

Desta forma a matriz procurada é

$$A = \begin{bmatrix} a & c & e \\ b & d & f \end{bmatrix} = \begin{bmatrix} 3 & 11 & 5 \\ -1 & -8 & -3 \end{bmatrix}.$$

Agora se resolvermos esse mesmo problema tomando as bases canônicas

$$\alpha = \{(1,0,0),(0,1,0),(0,0,1)\} \text{ e } \beta = \{(1,0),(0,1)\}$$

de $\mathbb{R}^3$ *e* $\mathbb{R}^2$ *respectivamente. Assim teremos,*

$$\begin{aligned} T(1,0,0) &= (2,3) = 2(1,0) + 3(0,1) \\ T(0,1,0) &= (1,-2) = 1(1,0) - 2(0,1) \\ T(0,0,1) &= (-1,4) = -1(1,0) + 4(0,1). \end{aligned}$$

Observe que os cálculos aritméticos são bem mais simples quando usamos bases padrões ou canônicas. Assim a matriz A *da referida transformação linear é dada por*

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

2.5 Autovalores e Autovetores

Iniciamos destacando que o adjetivo germânico "eigen" significa "próprio" ou "característico de". Valores e vetores próprios, ou autovalores e autovetores, são característicos de uma matriz no sentido de conterem informações importantes sobre a natureza da matriz. A letra λ (lambda), letra grega equivalente ao L em português, é utilizada para designar autovalores porque anteriormente esses números eram chamados de valores latentes.

Dada uma transformação linear $T : V \to V$, considere o problema de determinar quais vetores de V são levados em um múltiplo de si mesmo. Ou seja, devemos procurar um vetor $v \in V$ e um escalar $\lambda \in \mathbb{R}$ tais que $T(v) = \lambda v$.

Neste caso, o vetor $T(v)$ tem a mesma direção que v. Como $v = 0$ satisfaz a equação $T(v) = \lambda v$ para todo $\lambda \in \mathbb{R}$, então vamos nos concentrar em determinar vetores $v \neq 0$ que satisfaçam essa equação. O escalar $\lambda \in \mathbb{R}$ será chamado de autovalor ou valor característico e o vetor v será chamado de autovetor ou vetor característico do operador T. Agora vamos formalizar este conceito.

Definição 2.5.1. *Seja* $T : V \to V$ *um operador. Dizemos que um escalar* $\lambda \in \mathbb{R}$ *é um autovalor de* T *se existir um vetor* $v \in V$, $v \neq 0$ *tal que* $T(v) = \lambda.v$. *Todo vetor* v *que satisfaz essa relação é chamado de autovetor de* T *associado ao autovalor* λ.

Em termos geométricos, a matriz A do operador T tem o efeito de esticar ou contrair o vetor próprio v por um quantidade especificada pelo valor próprio λ.

Proposição 2.5.2. *Se* v *é autovetor associado à* λ , *isto é,* $Av = \lambda v$, *então, qualquer múltiplo de* v, cv, *com* $c \neq 0$, *será também autovetor associado a este mesmo autovalor.*

Demonstração. Com efeito, $A(cv) = c(Av) = c(\lambda v) = \lambda(cv)$. O que implica que cv também é um autovetor associado ao autovalor λ. $\square$

Proposição 2.5.3. *Seja λ um autovalor do operador $T : V \to V$. O conjunto $V_\lambda = \{v \in V; T(v) = \lambda v\}$, chamado autoespaço formado pelos autovetores associados a λ, é um subespaço vetorial de V.*

Demonstração. Vamos mostrar que o autoespaço V_λ verifica as propriedades de um subespaço. De fato, o vetor nulo pertence a V_λ pois $T(0) = 0 = \lambda \cdot 0$. Sejam $u, v \in V_\lambda$ e $c \in \mathbb{R}$, então

$$T(cu + v) = T(cu) + T(v) = cT(u) + T(v) = c\lambda u + \lambda v = \lambda cu + \lambda v = \lambda(cu + v).$$

Isso implica que $cu + v \in V_\lambda$. Assim V_λ é subespaço vetorial de V. $\square$

Seja a matriz A associada ao operador $T : V \to V$ e observe que a equação $Av = \lambda.v = (\lambda I)v$ pode ser escrita de forma equivalente a $(A - \lambda.I)v = 0$, cuja forma matricial é dada por

$$\begin{bmatrix} a_{11} - \lambda & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} - \lambda & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} - \lambda \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

Então se $det(A - \lambda.I) \neq 0$ o posto da matriz $A - \lambda.I$ é completo, isto é, a matriz $A - \lambda.I$ tem posto n e portanto o sistema homogêneo acima tem uma única solução. Por ser homogêneo essa única solução será o vetor nulo $v = (x_1, \cdots, x_n) = (0, \cdots, 0)$, ou $v = 0$. Assim, a única maneira de encontrar autovetores de A é se $det(A - \lambda.I) = 0$. Desta forma, os autovalores de A são as raízes reais do polinômio $p(\lambda) = det(A - \lambda.I)$. Esse polinômio recebe o nome de **polinômio característico** do operador T.

Exemplo 2.5.4. *Vamos calcular os autovalores e autovetores do operador $T : \mathbb{R}^2 \to \mathbb{R}^2$ dado por $T(x, y) = (2x + y, 3y)$, seguindo o roteiro de resolução.*

1. Encontrar a matriz $A = [T]_B$, sendo B a base canônica do $\mathbb{R}^2$, ou seja, $B = \{(1, 0), (0, 1)\}$.

Temos que $T(1, 0) = (2, 0)$ e $T(0, 1) = (1, 3)$. Desta forma $A = [T]_B = \begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix}$.

2. Calcular os autovalores da matriz A, que é equivalente a calcular as raízes do polinômio característico $p(\lambda) = det(A - \lambda.I)$ de T. Assim,

$$det(A - \lambda.I) = \begin{vmatrix} 2 - \lambda & 0 \\ -1 & 3 - \lambda \end{vmatrix} = (2 - \lambda)(3 - \lambda) = 0.$$

Logo os autovalores de A são $\lambda_1 = 2$ e $\lambda_2 = 3$.

3. Calcular os autovetores.

i) Autovetores associados ao autovalor $\lambda_1 = 2$.

Por definição temos

$$Av = \lambda v \Longleftrightarrow \begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} = 2 \begin{bmatrix} x \\ y \end{bmatrix}.$$

Resolvendo a equação matricial obtemos o sistema

$$\begin{cases} 2x + y = 2x \\ 3y = 2y, \end{cases}.$$

cuja solução é $y = 0$ e x um número real qualquer. Então os autovetores são do tipo $v_1 = \{(x, 0); x \in \mathbb{R}\} = span((1, 0))$. Em outras palavras, o autoespaço $V_{\lambda_1} = span((1, 0))$.

ii) Autovetores associados ao autovalor $\lambda_2 = 3$.

Por definição temos também

$$Av = \lambda v \Longleftrightarrow \begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} = 3 \begin{bmatrix} x \\ y \end{bmatrix}.$$

Resolvendo a equação matricial obtemos o sistema

$$\begin{cases} 2x + y = 3x \\ 3y = 3y \end{cases}.$$

cuja solução é $y = x$. Então os autovetores são do tipo $v_2 = \{(x, x); x \in \mathbb{R}\} = span((1, 1))$, em outras palavras, o autoespaço $V_{\lambda_2} = span((1, 1))$.

2.6 Matrizes Diagonalizáveis

Nessa seção vamos discutir a diagonalização de uma matriz quadrada A de ordem n ofecerecendo um algorítimo para diagonalizar A, e para tanto vamos apresentar algumas definições e propriedades relacionadas com autovalores e autovetores, iniciando com dois questionamentos, com respeito a referida matriz A, aparentemente diferentes mas que são equivalentes.

Problema 1: Existe alguma matriz invertível P tal que $B = P^{-1} \cdot A \cdot P$ é diagonal?

Problema 2: Existem n autovetores de A linearmente independentes ?

Teorema 2.6.1. *Autovetores associados a autovalores distintos são linearmente independentes.*

Demonstração. Vamos provar de modo indireto supondo que $\{v_1, ..., v_n\}$ é linearmente dependente, com o objetivo de mostrar que essa hipótese nos leva a uma contradição.

Se $v_1, v_2 ..., v_n$ são linearmente dependentes, então um desses vetores pode ser expresso como uma combinação linear dos vetores precedentes. Seja v_{k+1} o primeiro dos vetores v_i, que podem assim ser expresso. Em outras palavras $v_1, v_2, ..., v_k$ são linearmente independentes. Desta forma $v_{k+1} \in span(v_1, ..., v_k)$ e assim existem escalares $a_1, ..., a_k \in \mathbb{R}$ tais que

$$v_{k+1} = a_1 v_1 + ... + a_k v_k. \tag{2.1}$$

Aplicando o operador T nos dois membros da equação acima obtemos

$$T(v_{k+1}) = T(a_1 v_1 + ..., +a v_k) = a_1 T(v_1) + ..., +a_k T(v_k).$$

Por definição de autovetores e autovalores temos

$$\lambda_{k+1} v_{k+1} = a_1 \lambda_1 v_1 + ... + a_k \lambda_k v_k \tag{2.2}$$

Multiplicando a Equação 2.1 por λ_{k+1} obtemos

$$\lambda_{k+1} v_{k+1} = a_1 \lambda_{k+1} v_1 + ... + a_k \lambda_{k+1} v_k \tag{2.3}$$

Subtraindo a Equação 2.2 da Equação 2.3 resulta

$$a_1(\lambda_1 - \lambda_{k+1}) v_1 + ... + a_k(\lambda_k - \lambda_{k+1}) v_k = 0.$$

A independência linear de $v_1, v_2, ..., v_k$ implica que

$$a_1(\lambda_1 - \lambda_k) = 0, ..., a_{k-1}(\lambda_{k-1} - \lambda_k) = 0.$$

Como os autovalores λ_i, com $i = 1, \cdots, k$ são todos distintos, então

$$(\lambda_1 - \lambda_{k+1}) \neq 0, ..., (\lambda_k - \lambda_{k-1}) \neq 0.$$

Logo $a_1 =, ..., a_k = 0$, o que implica que $v_{k+1} = 0$, contradizendo nossa hipótese, pois o autovetor v_{k+1} não pode ser zero. Portanto, temos uma contradição da hipótese assumida. Daí segue que $\{v_1, ..., v_n\}$ tem que ser linearmente independente. □

Definição 2.6.2. *Sejam A e B matrizes quadradas de ordem n. Dizemos que A é semelhante a B se existir uma matriz invertível P também de ordem n tal que $B = P^{-1} \cdot A \cdot P$, ou equivalentemente $AP = PB$ e escrevemos $A \sim B$.*

Teorema 2.6.3. *Sejam A e B matrizes $n \times n$. Se B é semelhante a A, então as duas matrizes têm o mesmo polinômio característico e, consequentemente, os mesmos autovalores.*

Demonstração. Sejam $p_A(\lambda)$ e $p_B(\lambda)$ os respectivos polinômios característicos das matrizes A e B. Por definição, se B é semelhante a A, existe uma matriz invertível P tal que $B = P^{-1} \cdot A \cdot P$. Desta forma,

$$\begin{aligned} p_B(\lambda) &= det(B - \lambda \cdot I) \\ &= det(P^{-1} \cdot A \cdot P - \lambda \cdot P^{-1} I P) \\ &= det(P^{-1} \cdot A \cdot P - P^{-1} \cdot (\lambda \cdot I) \cdot P) \\ &= det[P^{-1}(A - \lambda I)P] \\ &= det(A - \lambda I) \\ &= p_A(\lambda). \end{aligned}$$

Os autovalores de uma matriz são as raízes do seu polinômio característico. Como as duas matrizes têm o mesmo polinômio característico, consequentemente, elas têm os mesmos autovalores. □

2.6.1 Diagonalização

Considerando uma matriz A quadrada de ordem n, vamos definir a seguir as matrizes diagonais e as matrizes diagonalizáveis.

Definição 2.6.4. *Uma **matriz diagonal** é uma matriz quadrada cujos elementos fora da diagonal principal são todos nulos.*

Definição 2.6.5. *Uma matriz A é dita **diagonalizával** se for semelhante a uma matriz diagonal D, ou seja, se existir uma matriz não singular (invertível) P tal que a matriz $D = P^{-1}.A.P$ é diagonal.*

Exemplo 2.6.6. *A matriz $A = \begin{bmatrix} 1 & 3 \\ 2 & 2 \end{bmatrix}$ é diagonalízável. De fato, $P = \begin{bmatrix} 1 & 3 \\ 1 & -2 \end{bmatrix}$ é invertível já que $det(P) = -5 \neq 0$. Agora seja a matriz diagonal $D = \begin{bmatrix} 4 & 0 \\ 0 & -1 \end{bmatrix}$. Temos,*

$$AP = \begin{bmatrix} 1 & 3 \\ 2 & 2 \end{bmatrix} \cdot \begin{bmatrix} 1 & 3 \\ 1 & -2 \end{bmatrix} = \begin{bmatrix} 4 & -3 \\ 4 & 2 \end{bmatrix} \text{ e } PD = \begin{bmatrix} 1 & 3 \\ 1 & -2 \end{bmatrix} \cdot \begin{bmatrix} 4 & 0 \\ 0 & -1 \end{bmatrix} = \begin{bmatrix} 4 & -3 \\ 4 & 2 \end{bmatrix}.$$

Ou seja, $AP = PD$ ou, equivalentemente, $D = P^{-1}.A.P$, o que implica que A é semelhante a uma matriz diagonal e, por definição, A é diagonalizável.

O exemplo anterior nos induz a seguinte questão: de onde vieram as matrizes P e D ?

Note que os elementos 4 e -1 da diagonal de D são os autovalores de A, já a origem de P não é tão óbvia assim, mas tem uma relação íntima com os autovetores de A. O teorema a seguir desvenda esse mistério.

Teorema 2.6.7. *Se A é uma matriz $n \times n$, então A será diagonalizável se, e somente se, A tiver n autovetores linearmente independentes.*

Mais precisamente, existem uma matriz invertível P e uma matriz diagonal D de modo que $P^{-1}.A.P = D$ se, e somente se, as colunas de P forem os n autovetores de A, linearmente independentes, e os elementos da diagonal de D forem os autovalores correspondentes aqueles colocados na mesma ordem.

Demonstração. Suponha que A seja diagonalizável, isto é, A é semelhante à matriz diagonal D, ou seja, $P^{-1}AP = D$ ou $AP = PD$. Sejam $v_1, v_2..., v_n$ os vetores coluna de P, e $\lambda_1, \lambda_2..., \lambda_n$, os elementos da diagonal de D. Então

$$A \cdot [v_1 \ v_2 \ \cdots \ v_n] = [v_1 \ v_2 \ \cdots \ v_n] \cdot \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}$$

ou

$$[A \cdot v_1, \cdots, A \cdot v_n] = [\lambda_1 v_1, \cdots, \lambda_n v_n]. \tag{2.4}$$

Da Equação 2.4 temos n equações, uma para cada coluna

$$Av_1 = \lambda_1 v_1, \cdots, Av_n = \lambda_n v_n.$$

O que mostra que os vetores colunas de P são os autovetores de A cujos os autovalores são os elementos da diagonal de D na mesma ordem. Como P é invertível seus vetores coluna são linearmente independentes.

Reciprocamente suponha que A tem n autovetores linearmente independentes, $v_1, v_2..., v_n$, associados aos autovalores $\lambda_1, \lambda_2..., \lambda_n$, respectivamente. Ou seja,

$$Av_1 = \lambda_1 v_1, \cdots, Av_n = \lambda_n v_n.$$

Consequentemente, se chamarmos de P a matriz $n \times n$ com colunas $v_1, v_2, \cdots, v_n$, então a equação pode ser escrita como $AP = PD$. Como as colunas de P são linearmente independentes, implica que P é invetível, logo $P^{-1} \cdot A \cdot P = D$ e assim A é diagonalizável. □

2.6.1.1 Algorítmo da Diagonalização

Agora vamos estabelecer um algorítmo para calcular autovalores e autovetores de uma matriz quadrada A de ordem n, e a partir daí decidir se existe ou não uma matriz invertível P tal que $P^{-1} \cdot A \cdot P$ seja diagonal, e assim decidir também se A é ou não diagonalizável. Para tanto devemos seguir os seguintes passos:

Passo 1: Forme a matriz $M = A - \lambda \cdot I$. A partir da matriz M, encontre o polinômio característico de A e suas raízes obtendo assim seus autovalores.

Passo 2: Encontre uma base para o espaço solução do sistema homogêneo $M \cdot X = 0$. Esses vetores da base são autovetores linearmente independentes de A associados a λ.

Passo 3: Considere o conjunto $S = \{v_1, v_2..., v_k\}$ de todos os autovetores obtidos no passo 2.

i) Se $k \neq n$, então A não é diagonalizável;

ii) Se $k = n$, então A é diagonalizável.

Mais precisamente, seja P a matriz cujas as colunas são os autovetores $v_1, v_2..., v_k$. Então

$$P^{-1} \cdot A \cdot P = D = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix},$$

em que λ_i é o autovalor associado ao autovetor v_i para qualquer $i = 1, \cdots, n$.

Exemplo 2.6.8. *Vamos aplicar o algorítmo da diagonalização à matriz* $A = \begin{bmatrix} 4 & 2 \\ 3 & -1 \end{bmatrix}$. *De acordo com os passos do algorítmo.*

Passo 1: *Considere a matriz* $M = A - \lambda \cdot I$. *Calculemos o polinômio característico* $p(\lambda)$ *de* A *e suas raízes, que são os autovalores de* A.

$$M = A - \lambda I = \begin{bmatrix} 4-\lambda & 2 \\ 3 & -1-\lambda \end{bmatrix}.$$

O polinômio característico é dado por

$$p(\lambda) = det(A - \lambda I) = \begin{vmatrix} 4-\lambda & 2 \\ 3 & -1-\lambda \end{vmatrix} = (4-\lambda)(-1-\lambda) - 6 = \lambda^2 - 3\lambda - 10.$$

As raízes do polinõmio $p(\lambda) = \lambda^2 - 3\lambda - 10$ *são* $\lambda_1 = 5$ *e* $\lambda_2 = -2$.

Passo 2: *Vamos encontrar uma base para o espaço solução do sistema homogêneo* $MX = 0$, *onde já sabemos que os vetores dessa base são os autovetores linearmente independentes de* A *associados a* λ.

i) Para $\lambda_1 = 5$ *temos*

$$M = A - \lambda I = \begin{bmatrix} 4-\lambda & 2 \\ 3 & -1-\lambda \end{bmatrix} = \begin{bmatrix} 4-5 & 2 \\ 3 & -1-5 \end{bmatrix} = \begin{bmatrix} -1 & 2 \\ 3 & -6 \end{bmatrix}.$$

Então o sistema homogêneo $MX = 0$ é dado por

$$\begin{bmatrix} -1 & 2 \\ 3 & -6 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \iff \begin{cases} -x + 2y = 0 \\ 3x - 6y = 0 \end{cases} \iff x = 2y.$$

Assim o autoespaço de $\lambda_1 = 5$ é o conjunto $V_{\lambda_1} = \{(2y, y); y \in \mathbb{R}\} = \{y(2,1); y \in \mathbb{R}\}$. Desta forma $V_{\lambda_1} = span((2,1))$, logo $v_1 = (2,1)$ é um autovetor associado ao autovalor $\lambda_1 = 5$.

ii) Para $\lambda_2 = -2$ temos

$$M = A - \lambda I = \begin{bmatrix} 4-\lambda & 2 \\ 3 & -1-\lambda \end{bmatrix} = \begin{bmatrix} 4+2 & 2 \\ 3 & -1+2 \end{bmatrix} = \begin{bmatrix} 6 & 2 \\ 3 & 1 \end{bmatrix}.$$

Então o sistema homogêneo $MX = 0$ é dado por

$$\begin{bmatrix} 6 & 2 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \iff \begin{cases} 6x + 2y = 0 \\ 3x + y = 0 \end{cases} \iff y = -3x.$$

Então o autoespaço de $\lambda_2 = -2$ é o conjunto $V_{\lambda_2} = \{(x, -3x); x \in \mathbb{R}\} = \{x(1,-3); x \in \mathbb{R}\}$. Dessa forma $V_{\lambda_2} = span((1,-3))$ e assim $v_2 = (1,-3)$ é um autovetor associado ao autovalor $\lambda_2 = -2$.

Passo 3: *O conjunto $S = \{(2,1), (1,-3)\}$ é o conjunto de todos os autovetores de A. Como, pelo Teorema 2.6.1, autovetores associados a autovalores distintos são linearmente independentes, então S é base do $\mathbb{R}^2$ com dois vetores, então $dim S = 2 = dim \mathbb{R}^2$. Logo $\mathbb{R}^2 = span((2,1),(1,-3))$. Portanto A é diagonalizável.*

Seja P a matriz cujas colunas são os autovetores v_1 e v_2, ou seja, $P = \begin{bmatrix} 2 & 1 \\ 1 & -3 \end{bmatrix}$. Calculando a inversa de P obtemos

$$P^{-1} = -\frac{1}{7} \cdot \begin{bmatrix} -3 & -1 \\ -1 & 2 \end{bmatrix} = \begin{bmatrix} \frac{3}{7} & \frac{1}{7} \\ \frac{1}{7} & -\frac{2}{7} \end{bmatrix}.$$

Dessa forma, a matriz diagonal D cujas entradas são os autovalores de A é dada por

$$D = P^{-1} \cdot A \cdot .P = \begin{bmatrix} \frac{3}{7} & \frac{1}{7} \\ \frac{1}{7} & -\frac{2}{7} \end{bmatrix} \cdot \begin{bmatrix} 4 & 2 \\ 3 & -1 \end{bmatrix} \cdot \begin{bmatrix} 2 & 1 \\ 1 & -3 \end{bmatrix} = \begin{bmatrix} 5 & 0 \\ 0 & -2 \end{bmatrix}.$$

2.6.2 Potência de uma Matriz

Definição 2.6.9. *Uma matriz quadrada $A = [a_{ij}]$ pode ser multiplicada n vezes por si mesma. A matriz que resulta dessas operações, representada por A^n, é chamada potência n da matriz A.*

Note, pela definição, que calcular potência de matrizes é exaustivo e enfadonho, porém o cálculo da matriz A^n é muito simples quando A é uma matriz diagonal de ordem k, isto é, se

$$A = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_k \end{bmatrix}, \text{ então } A^n = \begin{bmatrix} \lambda_1^n & 0 & \cdots & 0 \\ 0 & \lambda_2^n & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_k^n \end{bmatrix}.$$

Dessa forma, para calcular uma potência de uma matriz A basta encontrar uma matriz diagonal D semelhante à matriz A, ou seja, a matriz D é tal que $P^{-1} \cdot A \cdot P = D$.

Teorema 2.6.10. *Seja A uma matriz diagonalízável com $P^{-1} \cdot A \cdot P = D$ ou $A = P \cdot D \cdot P^{-1}$ e $n \in \mathbb{N}$. Então $A^n = P \cdot D^n \cdot P^{-1}$.*

Demonstração. Vamos fazer a prova por indução sobre n. Para $n = 1$ é imediato, pois $A = P \cdot D \cdot P^{-1}$. Para $n = 2$ temos

$$A^2 = (P \cdot D \cdot P^{-1}) \cdot (P \cdot D \cdot P^{-1}) = P \cdot D \cdot \underbrace{(P^{-1} \cdot P)}_{I} \cdot D \cdot P^{-1} = P \cdot D^2 \cdot P^{-1}.$$

Suponha agora que a afirmação é verdadeira para $n = k$, isto é, $A^k = P \cdot D^k \cdot P^{-1}$ e mostraremos que vale para

$$A^{k+1} = P \cdot D^{k+1} \cdot P^{-1}.$$

Multiplicando a equação $A^k = P \cdot D^k \cdot P^{-1}$ por A à direita obtemos $A^k \cdot A = P \cdot D^k \cdot P^{-1} \cdot A$. Disso,

$$A^{k+1} = P \cdot D^k \cdot P^{-1} \cdot P \cdot D \cdot P^{-1} = P \cdot D^k \cdot \underbrace{(P^{-1} \cdot P)}_{I} \cdot D \cdot P^{-1} = P \cdot D^{k+1} \cdot P^{-1},$$

o que conclui nossa demonstração. □

Observação 2.6.11. *No caso da matriz A não ser diagonalizável também é possível calcular potências de matrizes, mas os cálculos de potências de matrizes não diagonalizáveis requerem a forma canônica de Jordan que não é objeto do nosso estudo.*

Exemplo 2.6.12. *Dada a matriz $A = \begin{bmatrix} 0 & 1 \\ 2 & 1 \end{bmatrix}$, vamos calcular A^{10} utilizando o Teorema 2.6.10.*

Para calcular a 10ª potência de A é necessário encontrar os autovalores e autovetores. Efetuando os cálculos necessários, obtemos os autovalores $\lambda_1 = -1$ e $\lambda_2 = 2$ com os respectivos autovetores $v_1 = (1, -1)$ e $v_2 = (1, 2)$. A partir disso, obtemos as matrizes

$$P = \begin{bmatrix} 1 & 1 \\ -1 & 2 \end{bmatrix}, \quad P^{-1} = \begin{bmatrix} \frac{2}{3} & -\frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} \end{bmatrix} \text{ e } D = \begin{bmatrix} -1 & 0 \\ 0 & 2 \end{bmatrix}.$$

Dessa forma,

$$A^{10} = P \cdot D^{10} \cdot P^{-1} = \begin{bmatrix} 1 & 1 \\ -1 & 2 \end{bmatrix} \cdot \begin{bmatrix} (-1)^{10} & 0 \\ 0 & 2^{10} \end{bmatrix} \cdot \begin{bmatrix} \frac{2}{3} & -\frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} \end{bmatrix} = \begin{bmatrix} 342 & 341 \\ 682 & 683 \end{bmatrix}.$$

2.7 Produto Interno e Norma

Nós vivemos em um mundo Euclidiano tridimensional e, portanto, conceitos de Geometria Euclidiana influenciam de modo determinante nossa maneira de ver o mundo. Em particular, é natural dizer que a menor distância entre dois pontos é uma reta, contudo veremos a seguir outra maneira de medir distância que é tão real quanto a distância que estamos acostumados da Geometria Euclidiana, conseqüência do Teorema de Pitágoras, permitindo-nos, assim, pensar em distância de uma forma mais forma flexível, e que nos dará possibilidade de calcular distância, por exemplo, entre polinômios, funções e matrizes.

Para motivar o conceito de **produto interno**, vamos pensar nos vetores no $\mathbb{R}^2$ e $\mathbb{R}^3$ como setas com ponto inicial na origem. O comprimento de um vetor v no $\mathbb{R}^2$ e $\mathbb{R}^3$ é chamado de norma de v, denotada por $\|v\|$. Assim para $v = (x, y)$, sua norma Euclidiana é dada por $\|v\| = \sqrt{x^2 + y^2}$. Se pensarmos nos vetores como pontos em vez de setas, então $\|v\|$ deve ser interpretado como a distância do ponto v a origem.

Exemplo 2.7.1. *Sejam $u = (x_1, \cdots, x_n)$ e $v = (y_1, \cdots, y_n)$ vetores genéricos do $\mathbb{R}^n$. Definimos o seguinte produto no espaço $\mathbb{R}^n$:*

$$< u, v >= u^T \cdot v = v^T \cdot u = \sum_{i=1}^{n} x_i y_i = x_1 \cdot y_1 + x_2 \cdot y_2 + \cdots + x_n \cdot y_n.$$

Tal produto satisfaz a comutatividade, a aditividade, a homogeneidade e a positividade. Essas propriedades estão verificadas a seguir.

a) *Comutatividade:*

$$< u, v >= x_1 \cdot y_1 + \cdots + x_n \cdot y_n = y_1 \cdot x_1 + \cdots + y_n \cdot x_n =< v, u > .$$

b) *Aditividade:*

$$\begin{aligned} < u, v + w > &=< (x_1, \cdots, x_n), (y_1 + z_1, \cdots, y_n + z_n) > \\ &= x_1(y_1 + z_1) + \cdots + x_n(y_n + z_n) \\ &= x_1 \cdot y_1 + x_1 \cdot z_1 + \cdots + x_n \cdot y_n + x_n \cdot z_n \\ &= (x_1 \cdot y_1 + \cdots + x_n \cdot y_n) + (x_1 \cdot z_1 + \cdots + x_n \cdot z_n) \\ &=< u, v > + < u, w > . \end{aligned}$$

c) *Homogeneidade:*

$$\begin{aligned} < \alpha \cdot u, v > &=< (\alpha x_1, \alpha x_2, \cdots, \alpha x_n), (y_1, y_2, \cdots, y_n) > \\ &= \alpha \cdot x_1 \cdot y_1 + \cdots + \alpha \cdot x_n \cdot y_n \\ &= \alpha \cdot (x_1 \cdot y_1 + \cdots + x_n \cdot y_n) \\ &= \alpha \cdot < u, v > . \end{aligned}$$

d) *Positividade:*

Pela definição desse produto temos $< u, u >= x_1^2 + \cdots + x_n^2$*, que obviamente é não negativo, e só será nulo se* $u = 0$*.*

Esse produto pode ser generalizado para outros espaços vetoriais conforme veremos na seguinte definição.

Definição 2.7.2. *Seja V um espaço vetorial real. Um produto interno sobre V é uma função que a cada par de vetores u e v em V, associa um número real* $<u,v>$*, satisfazendo as seguintes propriedades:*

a) *Comutatividade:* $< u, v >=< v, u >$, $\forall u, v \in V$*;*

b) *Aditividade:* $< u, v + w >=< u, v > + < u, w >$,$\forall u, v, w \in V$*;*

c) *Homogeneidade:* $< \alpha \cdot u, v >= \alpha \cdot < u, v >$, $\forall u, v \in V$ *e* $\alpha \in \mathbb{R}$*;*

d) *Positividade:* $< v, v > \geq 0$ *e* $< v, v >= 0$ *se, e somente se,* $v = 0, \forall v \in V$*.*

Observação 2.7.3. *É muito útil observar que se representarmos u e v por vetores coluna, como foi feito no Exemplo 2.1.11, então seu produto interno euclidiano é dado por*

$$< u, v >= u^T \cdot v = v^T \cdot u.$$

Em um espaço com produto interno, podemos definir o comprimento de um vetor e a distância entre dois vetores da seguinte forma.

Definição 2.7.4. *Sejam* V *um espaço vetorial com produto interno e os vetores* $u, v \in V$*.*

i) *O comprimento ou norma de v é dado por* $\|v\| = \sqrt{< v, v >}$*;*

ii) *A distância d entre os vetores u e v é dada por* $d(u, v) = \|u - v\|$*.*

Teorema 2.7.5 (Desigualdade de Cauchy - Schwarz)**.** *Seja* V *um espaço vetorial com produto interno. Para quaisquer* $v, w \in V$*, é verdade que*

$$|< v, w >|^2 \leq < v, v > \cdot < w, w > .$$

Demonstração. Seja o vetor $u = v + t \cdot w$ para qualquer $t \in R$, e consideremos o polinômio

$$p(t) =< u, u >=< v + t \cdot w, v + t \cdot w > .$$

que é não negativo para qualquer $t \in \mathbb{R}$. Então,

$$p(t) =< w, w > t^2 + 2t < v, w > + < v, v > \geq 0.$$

Assim temos um polinômio do segundo grau positivo para qualquer valor de t. Da positividade do produto interno o coeficiente $< w, w >$ de t^2 é sempre positivo, logo o polinômio $p(t)$ não admite raízes reais distintas. Desta forma o discriminante do polinômio é não positivo. Portanto,

$$(2 < v, w >)^2 - 4 < w, w >< v, v > \leq 0.$$

O que implica que $|< v, w >|^2 \leq < w, w >< v, v >$. $\square$

Corolário 2.7.6 (Desigualdade Triangular)**.** *Sejam u e v vetores em um espaço V com produto interno. Então $\|u + v\| \leq \|u\| + \|v\|$.*

Demonstração. Calculando $\|u + v\|^2$ obtemos

$$\|u + v\|^2 =< u + v, u + v >=< u, u > +2 < u, v > + < v, v > .$$

Pela desigualdade de Cauchy - Schwarz $< u, v >\leq \|u\|\|v\|$. Então

$$\|u + v\|^2 =< u, u > +2 < u, v > + < v, v >\leq \|u\|^2 + 2\|u\|\|v\| + \|v\|^2 = (\|u\| + \|v\|)^2.$$

Concluímos assim que $\|u + v\| \leq \|u\| + \|v\|$. $\square$

Como vimos o comprimento de um vetor $v = (x, y)$ é dado por $\|v\| = \sqrt{x^2 + y^2}$. Tal número é conhecido como **norma** do vetor euclidiano. A norma é um conceito que podemos generalizar para diferentes espaços vetoriais. A seguir apresentamos a definição de norma de um espaço vetorial.

Definição 2.7.7. *Uma **norma** em um espaço vetorial V é uma aplicação que associa a cada vetor v um número real $\|v\|$, chamado norma de v, e satisfaz as seguintes propriedades para todos os vetores u e v em V e todos os escalares $\alpha \in \mathbb{R}$:*

a) *Positividade: $\|v\| \geq 0$ e $\|v\| = 0 \iff v = 0$;*

b) *Homogeneidade: $\|\alpha \cdot v\| =| \alpha | \cdot\|v\|$;*

c) *Desigualdade Triangular: $\|u + v\| \leq \|u\| + \|v\|$.*

*Um espaço vetorial com uma norma é chamado de **espaço vetorial normado**.*

Teorema 2.7.8. *Em um espaço vetorial com produto interno, a função $\|v\| = \sqrt{< v, v >}$ define uma norma.*

Demonstração. Como a função raíz é uma função crescente, concluímos, da positividade do produto interno, que $\|v\| = \sqrt{< v, v >} \geq 0$. Além disso, $\|v\| = \sqrt{< v, v >} = 0$ se, somente se $v = 0$. Agora, da homogeneidade do produto interno segue que

$$\|\alpha \cdot v\| = \sqrt{< \alpha v, \alpha v >} = \sqrt{\alpha^2 < v, v >} =| \alpha | \sqrt{< v, v >} =| \alpha | \cdot\|v\|.$$

A desigualdade triangular segue do Corolário 2.7.6. $\square$

Observação 2.7.9. *Cada produto interno dá origem a uma norma que pode ser usada para medir a distância ou comprimento dos elementos de um certo espaço vetorial. No entanto, nem toda norma que é usada em aplicações provém de um produto interno.*

Exemplo 2.7.10. *Uma classe importante de normas vetoriais são as normas-p, as quais são definidas por,*

$$\| v \|_p = (\sum_1^n |v_i|^p)^{\frac{1}{p}} = (| v_1 |^p + \cdots + | v_n |^p)^{\frac{1}{p}}, \text{ com } p \geq 1.$$

A positividade e homogeneidade são propriedades da norma-p que não são difíceis de estabelecer conforme mostraremos a seguir. A desigualdade triangular, no entanto, não é trivial pois sua prova necessita da definição de função convexa e das desigualdades de Young, Holder e Minkowski, as quais não são objeto deste trabalho. Assim vamos verificar as duas primeiras propriedades de norma e no próximo exemplo faremos a verificação completa para a norma-p mais usual, que é a norma Euclidiana.

a) Positividade: é verdade que $\| v \|_p \geq 0$, *pois* $| v_i | \geq 0$ *para todo* $i = 1, \cdots, n$. *Agora,*

$$\|v\| = 0 \iff \| v \|_p = (| v_1 |^p + \cdots + | v_n |^p)^{\frac{1}{p}} = 0,$$

dessa igualdade temos

$$| v_1 |^p + \cdots + | v_n |^p = 0 \iff v_i = 0,\ i = 1, \cdots, n.$$

b) Homogeneidade:

$$\begin{aligned}
\|\alpha \cdot v\|_p &= (| \alpha \cdot v_1 |^p + \cdots + | \alpha \cdot v_n |^p)^{\frac{1}{p}} \\
&= [(| \alpha |)^p \cdot (| v_1 |^p + \cdots + | v_n |^p)]^{\frac{1}{p}} \\
&= | \alpha | \, (| v_1 |^p + \cdots + | v_n |^p)^{\frac{1}{p}} \\
&= | \alpha | \cdot \|v\|_p.
\end{aligned}$$

Exemplo 2.7.11. *A normas-p mais usual em* $\mathbb{R}^n$ *é a norma Euclidiana conhecida como norma-2. A norma Euclidiana de um vetor* v *em* $\mathbb{R}^n$, *denotada por* $\| v \|_2$, *é dada por*

$$\| v \|_2 = (| v_1 |^2 + \cdots + | v_n |^2)^{\frac{1}{2}} = \sqrt{| v_1 |^2 + \cdots + | v_n |^2},$$

que é obtida através do produto interno usual, uma vez que $\| v \|_2 = \sqrt{v^T \cdot v}$ *ou, equivalentemente,* $\| v \|_2^2 = v^T \cdot v$. *Agora vamos mostrar que a norma Euclidiana satisfaz as propriedades da definição de norma.*

É claro que $\| v \|_2^2 = \sum_1^n v_i^2 = < v, v > \geq 0$. *Além disso, pela positividade da definição de produto interno,*

$$\| v \|_2^2 = \sum_1^n v_i^2 = 0 \iff v_i = 0, i = 1, \cdots, n \iff v = 0$$

Isso mostra a propriedade (a) de norma em um espaço vetorial. Para mostrar a propriedade (b) basta notar que

$$\| c \cdot v \|_2^2 = < cv, cv > = c^2 < v, v > = c^2 \| v \|_2^2$$

Portanto $\| c \cdot v \|_2 = | c | \cdot \| v \|_2$, *o que verifica a propriedade (b) de norma. Para mostrar a desiguladade triangular considere, para* $i = 1, \cdots, n$,

$$\|u + v \|_2^2 = \sum_1^n (u_i + v_i)^2 = \sum u_i^2 + 2 \sum_1^n u_i \cdot v_i + \sum_1^n v_i^2.$$

Pela Desigualdade de Cauchy - Schwarz,

$$\begin{aligned} \|u + v \|_2^2 &\leq \| u \|_2^2 + 2u^T \cdot v + \| v \|_2^2 \leq \| u \|_2^2 + 2 \mid u^T \cdot v \mid + \| v \|_2^2 \\ &\leq \| u \|_2^2 + 2 \| u \|_2 \cdot \| v \|_2 + \| v \|_2^2 \\ &= (\| u \|_2 + \| v \|_2)^2, \end{aligned}$$

o que comprova a desigualdade triangular.

Podemos definir normas para matrizes exatamente como definimos normas para vetores em $\mathbb{R}^n$. Acontece que, para matrizes, as normas mais úteis satisfazem uma propriedade adicional, $\| A \cdot B \| \leq \| A \| \cdot \| B \|$. Uma norma de matriz no espaço das matrizes de ordem $n \times n$ é compatível com uma norma de vetor $\| v \|$ em $\mathbb{R}^n$ se, para toda matriz $A_{n \times n}$ e todo vetor v em $\mathbb{R}^n$, tivermos $\| Av \| \leq \| A \| \cdot \| v \|$.

Exemplo 2.7.12. *A* ***norma de Frobenius*** $\| A \|_F$ *de uma matriz A é obtida quando colocamos os elementos da matriz em um vetor e então calculamos a norma-2 (Euclidiana). Em outras palavras,* $\| A \|_F$ *é a raíz quadrada da soma dos quadrados dos elementos de A. Então, se* $A = [a_{ij}]$ *é uma matriz* $n \times n$, *então*

$$\| A \|_F^2 = \sum_{i,j=1}^n a_{ij}^2.$$

Por exemplo, anorma de Frobenius da matriz

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

é dada por

$$\| A \|_F^2 = 3^2 + (-1)^2 + (-2)^2 + 4^2 = 9 + 1 + 4 + 16 = 30$$

Portanto $\| A \|_F = \sqrt{30}$.

Exemplo 2.7.13. *Considere a matriz* $A = \begin{bmatrix} 1 & 1 \\ 1 & 1,0005 \end{bmatrix}$ *e seja* $b = \begin{bmatrix} 3 \\ 3,001 \end{bmatrix}$*, vamos resolver o sistema* $Ax = b$*. Com efeito,*

$$\begin{bmatrix} 1 & 1 \\ 1 & 1,0005 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 3 \\ 3,001 \end{bmatrix} \Longleftrightarrow \begin{cases} x_1 + x_2 = 3 \\ x_1 + 1,0005\, x_2 = 3,001. \end{cases}$$

Subtraindo uma equação da outra obtemos $0,0005x_2 = 0,001$*, logo* $x_2 = 2$ *e* $x_1 = 1$*. Então* $x = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$*. Se alterarmos a matriz* A *para* $\tilde{A} = \begin{bmatrix} 1 & 1 \\ 1 & 1,001 \end{bmatrix}$ *obtemos*

$$\begin{bmatrix} 1 & 1 \\ 1 & 1,001 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 3 \\ 3,001 \end{bmatrix} \Longleftrightarrow \begin{cases} x_1 + x_2 = 3 \\ x_1 + 1,001\, x_2 = 3,001 \end{cases}$$

Procedendo de modo análogo, a nova solução $\tilde{x}$ *do sitema será* $\tilde{x} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$*. Assim, uma mudança relativa de* $\frac{1{,}001-1{,}0005}{1{,}0005} \approx 0,0005$ *ou* $0,05\%$ *acarreta uma mudança de* $\frac{2-1}{1} = 1$ *ou* 100% *em* x_1 *e* $\frac{1-2}{2} = -0,5$ *ou* -50% *em* x_2*.*

Neste caso dizemos que a matriz A é **mal-condicionada** pois uma pequena perturbação nos valores de seus elementos produziu grandes perturbações na solução do sistema $Ax = b$. Caso contrário dizemos que a matriz A é **bem-condicionada**.

É possível usar normas de matriz para oferecer uma maneira mais precisa de determinar quando uma matriz é mal-condionada. Agora vamos pensar na variação da matriz A para a matriz $\tilde{A}$ com erro ΔA que provoca um erro Δx na solução x do sistema $Ax = b$. Desta forma $\tilde{A} = A + \Delta A$ e $\tilde{x} = x + \Delta x$. Como $Ax = b$ e $\tilde{A}\tilde{x} = b$, temos $(A + \Delta A)(x + \Delta x) = b$ ou

$$Ax + A \cdot \Delta x + \Delta A x + \Delta A \cdot \Delta x = b \Longleftrightarrow A \cdot \Delta x + \Delta A x + \Delta A \cdot \Delta x = 0.$$

Logo, $A \cdot \Delta x = -\Delta A(x + \Delta x)$. Como estamos supondo que $Ax = b$ tem uma solução então A é invertível. Portanto, podemos reescrever a última equação como

$$\Delta x = -A^{-1}[\Delta A(x + \Delta x)] = -A^{-1}(\Delta A)\tilde{x}.$$

Calculando as normas de ambos os membros desta última equação temos

$$\| \Delta x \| = \| -A^{-1}(\Delta A)\tilde{x} \| \leq \| A^{-1}(\Delta A) \| \cdot \| \tilde{x} \| \leq \| A^{-1} \| \cdot \| \Delta A \| \cdot \| \tilde{x} \|,$$

o que implica que

$$\frac{\| \Delta x \|}{\| \tilde{x} \|} \leq \| A \| \cdot \| A^{-1} \| \cdot \frac{\| \Delta A \|}{\| A \|}.$$

O número $\| A \| \cdot \| A^{-1} \|$ é definido como número de condição de A, e é denotada por $cond(A) = \| A \| \cdot \| A^{-1} \|$. Se este número é grande dizemos que a matriz A é mal-condicionada, e neste caso, pequenas perturbações no vetor b podem produzir grandes perturbações na solução do sistema original $Ax = b$.

2.8 Ortogonalidade

Generalizando a noção de ortogonalidade de dois vetores em $\mathbb{R}^n$ para um conjunto de vetores, veremos que duas propriedades tornam a base canônica

$$\{(1,0,\cdots,0),(0,1,\cdots,0),\cdots,(0,0,\cdots,1)\} \text{ do } \mathbb{R}^n$$

mais agradável de manipular. A primeira é o fato de dois vetores distintos quaisquer serem ortogonais e a segunda consiste no fato de cada vetor ser unitário. Essas duas propriedades nos remetem às noções de bases ortogonais e ortonormais, cujos conceitos usaremos mais adiante. Em particular, na teoria e aplicações de Decomposição em Valores Singulares. Agora vamos formalizar essa teoria.

Definição 2.8.1. *Dizemos que dois vetores u e v são ortogonais se $< u, v >= 0$.*

Definição 2.8.2. *Seja V um espaço vetorial com produto interno e um conjunto $S = \{v_1, \cdots, v_r\}$ de vetores não nulos em V. Dizemos que S é um conjunto ortogonal se seus elementos são, dois a dois, ortogonais, e dizemos que S é ortonormal se for ortogonal e cada um de seus vetores for unitário. Da mesma forma, seja B uma base de um espaço vetorial V com produto interno, dizemos que B é uma base ortogonal (ortonormal) se B é um conjunto ortogonal (ortonormal).*

Definição 2.8.3. *Sejam V um espaço vetorial com produto interno e W um subconjunto não nulo de V. Chama-se complemento ortogonal de W o conjunto formado por todos os vetores do espaço vetorial V que são ortogonais aos vetores de W. Ou seja,*

$$W^{\perp} = \{v \in V \ ; \ v \cdot w = 0\,,\ \forall\ w \in W\}$$

.

O complemento ortogonal $W^{\perp}$ definido acima é um subespaço vetorial. De fato, se considerarmos W um subconjunto não nulo de um espaço veorial V, um escalar c e os vetores $u, v \in W^{\perp}$ temos

a) O vetor nulo pertence a $W^{\perp}$ pois $0 \cdot w = 0$ para todo $w \in W$.

b) Mostremos agora que $c \cdot u + v \in W^{\perp}$, isto é, $(c \cdot u + v) \cdot w = 0$. De fato,

$$(c \cdot u + v) \cdot w = c \cdot (u \cdot w) + v \cdot w = c \cdot 0 + 0 = 0$$

.

De (a) e (b) concluímos que $W^{\perp}$ é subespaço vetorial de V.

Proposição 2.8.4. *O complemento ortogonal do espaço imagem da matriz de uma transformação linear é igual ao espaço nulo de sua transposta. Isto é, dada uma matriz A, $R(A)^{\perp} = N(A^T)$.*

Demonstração. Para provar a igualdade $R(A)^{\perp} = N(A^T)$, vamos mostrar que $R(A)^{\perp} \subset N(A^T)$ e $N(A^T) \subset R(A)^{\perp}$.

Inicialmente vamos mostrar que $R(A)^{\perp} \subset N(A^T)$. Assim, pelo Exemplo 2.1.11, se $x \in R(A)^{\perp}$ então $x \cdot y = x^T \cdot y = 0$ para todo $y \in R(A)$. Agora, se $y \in R(A)$, existe $w \in \mathbb{R}^n$ tal que $y = A \cdot w$. Desta forma, $x^T \cdot y = x^T \cdot (A \cdot w)$. Usando a associatividade de produto de matrizes, segue que

$$(x^T \cdot A) \cdot w = 0 \Longleftrightarrow (A^T \cdot x)^T \cdot w = 0.$$

Como $y \neq 0$, $w \neq 0$ temos $(A^T \cdot x)^T = 0$. Em consequência, $A^T \cdot x = 0$. Portanto, $x \in N(A^T)$.

Mostremos agora que $N(A^T) \subset R(A)^{\perp}$. Considere $x \in N(A^T)$, logo $A^T x = 0$. Considere também $y \in R(A)$, logo existe $w \in \mathbb{R}^n$ tal que $y = Aw$. Portanto,

$$x \cdot y = x^T \cdot y = x^T \cdot (Aw) = (x^T \cdot A)w = (A^T x)^T w = 0 \cdot w = 0.$$

Como $x \cdot y = 0$ temos que x e y são ortogonais. Assim, se $y \in R(A)$, então $x \in R(A)^{\perp}$. $\square$

Teorema 2.8.5. *Todo conjunto ortogonal de vetores é linearmente independente.*

Demonstração. Considere o conjunto ortogonal $S = \{v_1, \cdots, v_r\}$, com $v_i \neq 0$, para $i = 1, 2, \cdots, r$. Para mostrar que S é linearmente independente, considere a combinação linear

$$c_1 v_1 + c_2 v_2 + \cdots + c_r v_r = 0.$$

Temos então que

$$0 =< c_1 v_1 + c_2 v_2 + \cdots + c_r v_r, v_1 >= c_1 < v_1, v_1 > + c_2 < v_2, v_1 > + \cdots + c_r < v_r, v_1 > .$$

Como S é ortogonal $< v_i, v_j >= 0$ com $i \neq j$. Então a equação anterior reduz-se a $c_1 < v_1, v_1 >= 0$, mas $v_1 \neq 0$, então $c_1 = 0$. Procedendo de modo análogo, mostramos também que $c_2 = \cdots = c_r = 0$ e portanto S é linearmente independente. $\square$

Teorema 2.8.6. *Seja V um espaço vetorial com produto interno e $B = \{v_1, \cdots, v_r\}$ uma base ortogonal de V. Então, para todo $w \in V$ temos $w = \frac{<w,v_1>}{\|v_1\|^2} v_1 + \frac{<w,v_2>}{\|v_2\|^2} v_2 + \cdots + \frac{<w,v_n>}{\|v_n\|^2} v_n$*

Demonstração. Como B é uma base ortogonal em V, então para todo $w \in V$ temos

$$w = c_1 v_1 + c_2 v_2 + \cdots + c_n v_n.$$

Então,

$< w, v_1 >=< c_1v_1+c_2v_2+\cdots+c_nv_n, v_1 >= c_1 < v_1, v_1 > +c_2 < v_2, v_1 > + \cdots +c_n < v_n, v_1 > .$

Como S é ortogonal a equação acima reduz-se a $< w, v_1 >= c_1 < v_1, v_1 >$ o que implica que

$$c_1 = \frac{< w, v_1 >}{< v_1, v_1 >} = \frac{< w, v_1 >}{\|v_1\|^2}.$$

Analogamente, mostramos também que $c_i = \frac{<w,v_i>}{\|v_i\|^2}$, para $i = 2, \cdots, n$. □

Os coeficientes c_i são chamados de Coeficiente de Fourier. O vetor $w \in V$ nada mais é que a soma das projeções ortogonais de w sobre os vetores da base B.

Proposição 2.8.7. *Sejam $w_1, w_2, \cdots, w_r$ vetores ortogonais de V. Dado um vetor v qualquer de V, denotamos por $u = v - (c_1 \cdot w_1 + \cdots + c_r \cdot w_r)$, em que $c_i = \frac{<v,w_i>}{<w_i,w_i>} = \frac{<v,w_i>}{\|w_i\|^2}$ e $i = 1, \cdots, r$. Então u é ortogonal a $w_1, w_2, \cdots, w_r$.*

Demonstração. Dado $i = 1, \cdots, r$ e usando o fato de $< w_i, w_j >= 0$, para $i \neq j$, temos

$$< u, w_i >=< v - (c_1 \cdot w_1 + \cdots + c_r \cdot w_r), w_i > .$$

Mas,

$$< u, w_i >=< v, w_i > -c_1 < w_1, w_i > - \cdots - c_i \cdot < w_i, w_i > - \cdots - c_r < w_r w_i > .$$

Disso

$$< u, w_i >=< v, w_i > -c_1 \cdot 0 - \cdots - c_i \cdot < w_i, w_i > - \cdots - c_r \cdot 0 =< v, w_i > -c_i \cdot < w_i, w_i > .$$

Substituindo o valor de c_i na última equação obtemos

$$\begin{aligned} < u, w_i > &=< v, w_i > -c_i \cdot < w_i, w_i > \\ &=< v, w_i > - \frac{< v, w_i >}{< w_i, w_i >} \cdot < w_i, w_i > \\ &=< v, w_i > - < v, w_i >= 0, \end{aligned}$$

o que gostaríamos de mostrar. □

Vamos assumir que temos uma base $w_1, w_2, \cdots, w_n$ de W, em que $dimW = n$. O objetivo é, a partir da base conhecida, construir uma base ortogonal $v_1, \cdots, v_n$. Vamos construir os elementos da base ortogonal um por um. Como inicialmente não vamos nos preocupar com a normalidade, então não há prejuízo em escolher o primeiro vetor fazendo $v_1 = w_1$.

Começando com w_2, o segundo vetor v_2 deve ser ortogonal ao primeiro vetor, $< v_2, v_1 >= 0$. Vamos arranjar isso subtraindo w_2 de um múltiplo adequado de v_1, e definir $v_2 = w_2 - cv_1$, em que c é o escalar a ser determinado. Pela condição de ortogonalidade temos

$$0 =< v_2, v_1 >=< w_2 - cv_1, v_1 >=< w_2, v_1 > -c < v_1, v_1 >=< w_2, v_1 > -c\|v_1\|^2.$$

Assim $c = \frac{<w_2,v_1>}{\|v_1\|^2}$. A independência linear de $v_1 = w_1$ e w_2 garante que $v_2 \neq 0$.

Agora construimos $v_3 = w_3 - c_1 v_1 - c_2 v_2$. Subtraindo w_3 dos múltiplos adequados dos dois primeiros vetores da base ortogonal, queremos que v_3 seja ortogonal a ambos v_1 e v_2. Como já tínhamos $< v_1, v_2 >= 0$, isso implica que devemos ter

$$0 =< v_3, v_1 >=< w_3 - c_1 v_1, v_1 >=< w_3, v_1 > -c_1 < v_1, v_1 >$$

e

$$0 =< v_3, v_2 >=< w_3 - c_2 v_2, v_2 >=< w_3, v_2 > -c_2 < v_2, v_2 >.$$

Logo $c_1 = \frac{<w_3,v_1>}{\|v_1\|^2}$ e $c_2 = \frac{<w_3,v_2>}{\|v_2\|^2}$. Como v_1 e v_2 são combinações lineares de w_1 e w_2, então devemos ter $v_3 \neq 0$, pois, caso contrário, w_1, w_2 e w_3 seriam linearmente dependentes e não formariam uma base. Continuando da mesma maneira, suponha que já tenhamos construído os vetores ortogonais $v_1, \cdots, v_k$ como combinações lineares de $w_1, \cdots, w_k$. O próximo elemento, v_{k+1}, da base ortogonal, será obtido subtraindo w_{k+1} de uma combinação linear adequada dos elementos anteriores da base ortogonal,

$$v_{k+1} = w_{k+1} - c_1 v_1 - \cdots - c_k v_k.$$

Como $v_1, \cdots, v_k$ já são ortogonais, a restrição de ortogonalidade

$$0 =< v_{k+1}, v_j >=< w_{k+1} - c_j v_j, v_j >=< w_{k+1}, v_j > -c_j < v_j, v_j >$$

requer que $c_j = \frac{<w_{k+1},v_j>}{\|v_j\|^2}$ para $j = 1, \cdots, k$. Ou equivalentemente

$$v_{k+1} = w_{k+1} - \frac{<w_{k+1},v_1>}{\|v_1\|^2} \cdot v_1 - \cdots - \frac{<w_{k+1},v_k>}{\|v_k\|^2} \cdot v_k.$$

O processo iterativo de Gram-Schmidt em que começamos com $v_1 = w_1$ e sucessivamente construímos $v_2, \cdots, v_n$, define um procedimento recursivo explícito para construir a desejada base ortogonal. Se quisermos obter uma base ortonormal, basta normalizar os vetores da base ortogonal.

Matrizes cujas colunas e linhas formam um conjunto ortonormal surgem frequentemente em aplicações. Essas matrizes têm várias propriedades úteis.

Definição 2.8.8. *Uma matriz quadrada A de ordem $n \times n$ é dita **ortogonal** se seus vetores colunas formam um conjunto ortonormal em $\mathbb{R}^n$. Alternativamente, uma matriz quadrada A é dita ortogonal se sua inversa é igual a sua transposta. Ou seja, $A^{-1} = A^T$ ou, equivalentemente, $AA^T = A^T A = I$.*

Proposição 2.8.9. *Dentre algumas propriedades de matrizes ortogonais A de ordem $n \times n$ destacamos as seguintes:*

P1) *O produto entre duas matrizes ortogonais, $n \times n$, resulta em uma matriz ortogonal;*

P2) *Se a matriz A é ortogonal então A preserva o produto interno entre dois vetores, ou seja, $< Av, Au > = < v, u >$ para quaisquer $u, v \in \mathbb{R}^n$;*

P3) *Se a matriz A é ortogonal então A preserva norma Euclidiana entre dois vetores, ou seja, $\| Av \|_2 = \| v \|_2$ para qualquer $v \in \mathbb{R}^n$;*

P4) *Se a matriz Q de ordem $n \times n$ é ortogonal então, $\| QA \|_2 = \| A \|_2$ para qualquer matriz A de ordem $n \times n$.*

Demonstração. Dadas as matrizes ortogonais A e B temos

$AB \cdot (AB)^T = A \cdot B \cdot B^T \cdot A^T = A \cdot A^T = I_n$ e $(AB)^T \cdot AB = B^T \cdot A^T \cdot A \cdot B = B^T \cdot B = I_n$.

Portanto o produto AB é ortogonal.

Considerando a Observação 2.7.3 e empregando a definição de matriz ortogonal de uma matriz A, isto é, $A^{-1} = A^T$ temos

$$< Au, Av > = (Av)^T \cdot (Au) = v^T \cdot A^T \cdot A \cdot u = v^T \cdot u = < u, v >$$

Mostremos agora a propriedade P3. De fato,

$$\| Av \|_2^2 = < Av, Av > = < v, v > = \| v \|_2^2,$$

o que implica que $\| Av \|_2 = \| v \|_2$.

Da definição de norma-2 matricial temos $\| QA \|_2 = \max\limits_{v \neq 0} \frac{\|QAv\|_2}{\|v\|_2}$. Analisando o numerador observamos que para qualquer vetor $v \in \mathbb{R}^n$, $(QAv) = Q(Av)$, com $Av \in \mathbb{R}^m$. Da propriedade P3 da referida proposição, temos,

$$\| (QA)v \|_2 = \| Q(Av) \|_2 = \| Av \|_2 \ .$$

Assim,

$$\| QA \|_2 = \max_{v \neq 0} \frac{\| QAv \|_2}{\| v \|_2} = \max_{v \neq 0} \frac{\| Q(Av) \|_2}{\| v \|_2} = \max_{v \neq 0} \frac{\| Av \|_2}{\| v \|_2} = \| A \|_2 \ .$$

□

2.9 Diagonalização Ortogonal

Os conceitos de matrizes semelhantes e diagonalização de matrizes podem ser analogamente associados à matrizes ortogonais. Neste caso, dadas as matrizes quadradas A e B podemos dizer que elas são **ortogonalmente semelhantes** se existir alguma matriz ortogonal P tal que

$$B = P^{-1} \cdot A \cdot P = P^T \cdot A \cdot P.$$

Da mesma forma, se a matriz A for ortogonalmente semelhante a alguma matriz diagonal D, digamos $D = P^T \cdot A \cdot P$, dizemos que A é **ortogonalmente diagonalizável** e que P diagonaliza A ortogonalmente.

Vimos também que nem todas as matrizes quadradas são diagonalizáveis. Quando se trata de matrizes que são iguais à sua transposta, ou seja, matrizes simétricas reais, a situação muda de maneira determinante pois os autovalores das matrizes simétricas reais são todos reais e elas são sempre diagonalizáveis, conforme veremos a seguir. Além disso, existe uma relação íntima entre matrizes ortogonalmente diagonalizáveis e matrizes simétricas.

Teorema 2.9.1 (Teorema Espectral)**.** *Seja A uma matriz de ordem $n \times n$. As seguintes afirmações são equivalentes.*

a) *A é ortogonalmente diagonalizável;*

b) *A possui um conjunto ortonormal de n autovetores;*

c) *A é simétrica.*

Demonstração. $(a \Rightarrow b)$. Supondo que A seja ortogonalmente diagonalizável, existe alguma matriz ortogonal P tal que $P^{-1} \cdot A \cdot P$ é diagonal. Como mostramos na demonstração do Teorema 2.6.7, os n vetores coluna de P são os n autovetores de A. Como P é ortogonal esses n vetores são ortonormais. Logo A tem n autovetores ortonormais.

$(b \Rightarrow a)$. Suponha que A tenha um conjunto ortonormal $\{v_1, v_2, \cdots, v_n\}$ de n autovetores. A matriz P que tem esses autovetores como colunas diagonaliza A. Como esses autovetores são ortonormais, P é ortogonal e, assim, diagonaliza A ortogonalmente.

$(a \Rightarrow c)$. Devemos mostrar que $A = A^T$. Como A é ortogonalmente diagonalizável então existem uma matriz ortogonal P e uma matriz diagonal D tais que $P^T \cdot A \cdot P = D$. Como $P^{-1} = P^T$ temos $P^T \cdot P = P \cdot P^T = I$ e assim, $A = P \cdot D \cdot P^T$. Agora calculando a transposta de A obtemos

$$A^T = (P \cdot D \cdot P^T)^T = (P^T)^T \cdot D^T \cdot P^T = P \cdot D \cdot P^T = A.$$

Portanto A é simétrica.

$(c \Rightarrow a)$. Seja $A \in \mathbb{R}^{n \times n}$ uma matriz simétrica. Vamos mostrar que A é ortogonalmente diagonalizável por indução sobre n, ou seja, queremos mostrar que existe uma matriz ortogonal $Q \in \mathbb{R}^{n \times n}$ e uma matriz diagonal $D \in \mathbb{R}^{n \times n}$ tais que $D = Q^T \cdot A \cdot Q$. Para $n = 1$, não há nada a fazer já que a matriz A de ordem 1×1 já está na forma diagonal. Agora mostrar que é válido para $n = k$ dado que é válido $n = k - 1$. Seja $A \in \mathbb{R}^{k \times k}$ e λ um autovalor de A associado ao autovetor unitário v. Podemos estender o vetor v para uma base ortonormal de $\mathbb{R}^k$ cujo primeiro elemento é v, e seja Q_1 a matriz cujas colunas são os vetores desta base. Seja $W \in \mathbb{R}^{k\times(k-1)}$ uma submatriz de Q_1, formada pelas colunas de Q_1 exceto a primeira, ou seja, as colunas de 2 até k, de modo que $Q_1 = \begin{bmatrix} v & W \end{bmatrix}$. Como as colunas de W são ortogonais a v, $W^T v = 0$. Seja $A_1 = Q_1^T \cdot A \cdot Q_1$. Então,

$$A_1 = \begin{bmatrix} v^T \\ W^T \end{bmatrix} \cdot A \cdot \begin{bmatrix} v & W \end{bmatrix} = \begin{bmatrix} v^T Av & v^T AW \\ W^T Av & W^T AW \end{bmatrix}$$

Como $Av = \lambda v$, segue que

$$v^T Av = v^T \lambda v = \lambda(v^T v) = \lambda \cdot v^2 = \lambda \ \text{e} W^T Av = W^T \lambda v = \lambda W^T v = 0$$

.

Então A_1 tem a forma em blocos $A_1 = \begin{bmatrix} \lambda & * \\ 0 & B \end{bmatrix}$. Observe que $B \in \mathbb{R}^{(k-1)\times(k-1)}$, então pela hipótese de indução existe uma matriz ortogonal $\hat{Q}_2$ e uma matriz diagonal $\hat{D}$ tais que $\hat{D} = \hat{Q}_2^T \cdot B \cdot \hat{Q}_2$. Agora vamos definir $Q_2 \in \mathbb{R}^{k\times k}$ por $Q_2 = \begin{bmatrix} 1 & 0 \\ 0 & \hat{Q}_2 \end{bmatrix}$, então Q_2 é ortogonal, e

$$Q_2^T \cdot A_1 \cdot Q_2 = \begin{bmatrix} \lambda & 0 \\ 0 & \hat{Q}_2^T \cdot B \cdot \hat{Q}_2 \end{bmatrix} = \begin{bmatrix} \lambda & 0 \\ 0 & \hat{D} \end{bmatrix}$$

que é uma matriz diagonal. Vamos chamar essa matriz de D e seja $Q = Q_1 \cdot Q_2$ que é ortogonal já que é produto de duas matrizes ortogonais. Então,

$$D = Q_2^T \cdot A_1 \cdot Q_2 = Q_2^T \cdot Q_1^T \cdot A \cdot Q_1 \cdot Q_2 = Q^T \cdot A \cdot Q$$

□

Teorema 2.9.2. *Se A é uma matriz real simétrica então:*

a) Os autovalores de A são reais;

b) Autovetores de autoespaços diferentes são ortogonais.

Demonstração. a) Inicialmente vale lembrar que dado um número complexo $a + bi$ seu conjugado complexo é $a - bi$ e que para mostrar que um número λ é real basta mostrar que ele é igual ao seu conjugado complexo. Neste caso, queremos mostrar que $\lambda = \overline{\lambda}$. Suponha que λ seja um autovalor de A associado ao autovetor v. Então $Av = \lambda \cdot v$, e tomando conjugados complexos temos $\overline{Av} = \overline{\lambda \cdot v}$. Como A é uma matriz real

$$A \cdot \overline{v} = \overline{A} \cdot \overline{v} = \overline{A \cdot v} = \overline{\lambda \cdot v} = \overline{\lambda} \cdot \overline{v}.$$

Tomando transpostas e usando o fato de A ser simétrica ($A^T = A$), temos

$$\overline{v}^T \cdot A = \overline{v}^T \cdot A^T = (A \cdot \overline{v})^T = (\overline{\lambda} \cdot \overline{v})^T = \overline{\lambda} \cdot \overline{v}^T.$$

Logo
$$\lambda \cdot (\overline{v}^T \cdot v) = \overline{v}^T \cdot (\lambda v) = \overline{v}^T \cdot (Av) = (\overline{v}^T \cdot A)v = (\overline{\lambda} \cdot \overline{v}^T)v = \overline{\lambda}(\overline{v}^T \cdot v),$$
ou
$$(\lambda - \overline{\lambda})(\overline{v}^T \cdot v) = 0.$$
Agora,
$$v = \begin{bmatrix} a_1 + b_1 i \\ \vdots \\ a_n + b_n i \end{bmatrix} \to \overline{v} = \begin{bmatrix} a_1 - b_1 i \\ \vdots \\ a_n - b_n i \end{bmatrix}.$$
Disso,
$$\overline{v}^T \cdot v = (a_1^2 + b_1^2) + \cdots + (a_n^2 + b_n^2) \neq 0$$
já que $v \neq 0$. Desta forma $\lambda - \overline{\lambda} = 0$, logo $\lambda = \overline{\lambda}$, o que mostra que λ é real.

b) Sejam u e v autovetores correspondentes a autovalores distintos λ_1 e λ_2, de modo que $Au = \lambda_1 u$ e $Av = \lambda_2 \cdot v$. Devemos mostrar que $u \cdot v = 0$. Usando o fato de A ser simétrica e o fato de que $Au \cdot v = (Au)^T \cdot v$, para dois vetores Au e v quaisquer em $\mathbb{R}^n$, temos,
$$Au \cdot v = (Au)^T v = (u^T A^T)v = u^T (A^T v) = u^T (Av) = u \cdot (Av).$$
Como u é autovetor de A associado a λ_1 e v é autovetor associado a λ_2 segue que
$$\lambda_1 \cdot u \cdot v = u \cdot \lambda_2 \cdot v,$$
que pode ser reescrito como
$$(\lambda_1 - \lambda_2)(u \cdot v) = 0.$$
Mas $\lambda_1 \neq \lambda_2$, logo $\lambda_1 - \lambda_2 \neq 0$. Assim $u \cdot v = 0$. □

Exemplo 2.9.3. *Seja* $A = \begin{bmatrix} 2 & -2 \\ -2 & 5 \end{bmatrix}$ *uma matriz real simétrica. Vamos encontrar uma matriz ortogonal P tal que $D = P^T \cdot A \cdot P$ seja diagonal. Lembre-se que os autovalores de A são as raízes do polinômio característico, ou seja,*
$$P(\lambda) = det(A - \lambda \cdot I) = \begin{vmatrix} 2-\lambda & -2 \\ -2 & 5-\lambda \end{vmatrix} = \lambda^2 - 7\lambda + 6 = 0.$$
Resolvendo essa equação encontramos $\lambda_1 = 6$ e $\lambda_2 = 1$. Após os cálculos adequados temos os autovetores $u_1 = (-1, 2)$ e $u_2 = (2, 1)$ que são claramente ortogonais.

Por definição as colunas da matriz ortogonal P são formadas por vetores ortonormais, desta forma devemos normalizar os vetores u_1 e u_2. Com efeito,
$$v_1 = \frac{u_1}{u_1} = \frac{(-1,2)}{\sqrt{(-1)^2 + 2^2}} = \left(-\frac{1}{\sqrt{5}}, \frac{2}{\sqrt{5}}\right) \quad \text{e} \quad v_2 = \frac{u_2}{u_2} = \frac{(2,1)}{\sqrt{(-1)^2 + 2^2}} = \left(\frac{2}{\sqrt{5}}, \frac{1}{\sqrt{5}}\right).$$
Portanto, P é a matriz cujas colunas são v_1 e v_2, ou seja,

$$P = \begin{bmatrix} -\frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{bmatrix} \quad e \quad P^T = \begin{bmatrix} -\frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{bmatrix}.$$

Efetuando o produto de matrizes $P^T \cdot A \cdot P$ obtemos

$$D = \begin{bmatrix} -\frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{bmatrix} \cdot \begin{bmatrix} 2 & -2 \\ -2 & 5 \end{bmatrix} \cdot \begin{bmatrix} -\frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{bmatrix} = \begin{bmatrix} 6 & 0 \\ 0 & 1 \end{bmatrix}.$$

Como esperado as entradas diagonais de D são os autovalores associados às colunas de P.

O Teorema 2.9.2 ilustrado pelo procedimento do exemplo anterior pode ser formalizado no seguinte algorítmo, o qual fornece uma matriz ortogonal P tal que $P^T \cdot A \cdot P$ é diagonal.

2.9.0.1 Algorítmo de Diagonalização Ortogonal

Considere uma matriz simétrica real A e observe os seguintes passos.

Passo 1: Encontre o polinômio caractarístico $P(\lambda)$ de A e seus autovalores que são as raízes do polinômio caractarístico

Passo 2: Para cada autovalor λ de A, encontre os autovetores associados que formam uma base ortogonal de um espaço vetorial V e normalize todos os autovetores formando assim uma base ortonormal de $\mathbb{R}^n$.

Passo 3: Então P é a matriz cujas colunas são os autovetores normalizados e $D = P^T \cdot A \cdot P$ a matriz cujas entradas diagonais são os autovalores de A.

Exemplo 2.9.4. *Dada a matriz $A = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}$, vamos utilizar o Algorítmo de Diagonalização Ortogonal para encontrar uma matriz ortogonal P tal que $D = P^T \cdot A \cdot P$ seja diagonal seguindo os passos sugeridos.*

Passo 1*: O polinômio característico de A é*

$$P(\lambda) = det(A - \lambda \cdot I) = \begin{vmatrix} 3-\lambda & 1 \\ 1 & 3-\lambda \end{vmatrix} = (3-\lambda)^2 - 1.$$

Resolvendo a equação $(3-\lambda)^2 - 1 = 0$ ou $(3-\lambda)^2 = 1$ o que implica que $\lambda_1 = 4$ e $\lambda_2 = 2$.

Passo 2: *Agora calculemos os autovetores ortonormais de A.*

a) Autovetores associados a $\lambda_1 = 4$.

$$Av_1 = \lambda_1 \cdot v_1 \Longleftrightarrow \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} = 4 \begin{bmatrix} x \\ y \end{bmatrix} \Longleftrightarrow \begin{cases} 3x + y = 4x \\ x + 3y = 4y. \end{cases}.$$

Portanto $x = y$. *Assim o autoespaço associado é* $V_{\lambda_1} = \{(x,x), x \in \mathbb{R}\}$. *Logo* $v_1 = (1,1)$.

b) Autovetores associados a $\lambda_2 = 2$

$$Av_2 = \lambda_2 \cdot v_2 \Longleftrightarrow \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} = 2 \begin{bmatrix} x \\ y \end{bmatrix} \Longleftrightarrow \begin{cases} 3x + y = 2x \\ x + 3y = 2y. \end{cases}$$

Portanto $x = -y$. *Assim o autoespaço associado é* $V_{\lambda_2} = \{(-x,x), x \in \mathbb{R}\}$.
Logo $v_2 = (-1,1)$.

Normalizando os vetores v_1 *e* v_2 *obtemos*

$$u_1 = \tfrac{v_1}{v_1} = \tfrac{(1,1)}{\sqrt{1^2+1^2}} = \left(\tfrac{1}{\sqrt{2}}, \tfrac{1}{\sqrt{2}}\right) \ e \ u_2 = \tfrac{v_2}{v_2} = \tfrac{(-1,1)}{\sqrt{(-1)^2+1^2}} = \left(-\tfrac{1}{\sqrt{2}}, \tfrac{1}{\sqrt{2}}\right).$$

Agora note que u_1 *e* u_2 *formam uma base ortonormal de* $\mathbb{R}^2$ *já que*

$$u_1 \cdot u_2 = 0 \qquad e \qquad \|u_i\| = 1, \ com \ i = 1,2.$$

Passo 3: *Desta forma* $P = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$ *e* $D = \begin{bmatrix} 4 & 0 \\ 0 & 2 \end{bmatrix}$. *Portanto a matriz A pode ser escrita na seguinte forma fatorada,*

$$A = PDP^T = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \cdot \begin{bmatrix} 4 & 0 \\ 0 & 2 \end{bmatrix} \cdot \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}.$$

Como percebemos nos exemplos anteriores, os cálculos manuais determinados pelo algorítmo de diagonalização são bastante trabalhosos, por isso os exemplos apresentados utilizam matrizes de ordem 2 ou 3.

2.10 Formas Quadráticas

Uma equação quadrática em duas variáveis x e y é uma equação da forma

$$ax^2 + 2bxy + cy^2 + dx + ey + f = 0$$

A equação acima pode ser escrita em sua forma matricial

$$\begin{bmatrix} x & y \end{bmatrix} \cdot \begin{bmatrix} a & b \\ b & c \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} d & e \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} + f = 0.$$

Considere o vetor $X = \begin{bmatrix} x \\ y \end{bmatrix}$ e a matriz simétrica $A = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$. O termo

$$f(x) = x^T Ax = ax^2 + 2bxy + cy^2$$

é chamado forma quadrática associada à equação quadrática. Esse conceito pode ser estendido para n variáveis. Poderíamos ter equações do segundo grau e formas quadráticas com qualquer número de variáveis. De fato, uma equação quadrática em n variáveis $x_1, \cdots, x_n$ é uma equação da forma

$$x^T Ax + Bx + \alpha = 0$$

em que $x = (x_1, \cdots, x_n)^T$, A é uma matriz simétrica $n \times n$, B é uma matriz $1 \times n$ e α um escalar. Assim, definimos a seguir uma forma quadrática de modo mais geral.

Definição 2.10.1. *Uma forma quadrática em n variáveis é uma função $f : \mathbb{R}^n \to \mathbb{R}$ da forma $f(x) = x^T Ax$, em que A é uma matriz simétrica de ordem $n \times n$ e $x \in \mathbb{R}^n$. A matriz A é denominada matriz associada à forma quadrática $f(x) = x^T Ax$.*

Exemplo 2.10.2. *Seja $f : \mathbb{R}^2 \to \mathbb{R}$ definida por $f(x, y) = x^2 - 10xy + y^2$. Sabemos que*

$$f(x, y) = \begin{bmatrix} x & y \end{bmatrix} \cdot \begin{bmatrix} a & b \\ b & c \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix}.$$

Assim,

$$f(x, y) = ax^2 + 2bxy + cy^2 = x^2 - 10xy + y^2.$$

Logo $a = 1$, $b = -5$ e $c = 1$. Substituindo os valores de a, b e c em $f(x, y)$ obtemos

$$f(v) = \begin{bmatrix} x & y \end{bmatrix} \cdot \begin{bmatrix} 1 & -5 \\ -5 & 1 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix}.$$

Note que a matriz $A = \begin{bmatrix} 1 & -5 \\ -5 & 1 \end{bmatrix}$ é simétrica.

Como foi visto nesta seção, a matriz de uma forma quadrática é simétrica, e essas matrizes podem sempre ser diagonalizadas. Usaremos agora esse fato para mostrar que, em toda forma quadrática, podemos eliminar os termos mistos por intermédio de uma mudança de variáveis conforme o teorema seguinte.

Teorema 2.10.3 (Teorema dos Eixos Principais)**.** *Toda forma quadrática pode ser diagonalizada. Especificamente, se $A_{n\times n}$ é a matriz associada à forma quadrática $v^T \cdot A \cdot v$ e se P é uma matriz ortogonal que faz $P^T \cdot A \cdot P = D$ ser uma matriz diagonal, então a mudança de variavéis $v = Pu$ transforma a forma quadrática $v^T \cdot A \cdot v$ em $u^T \cdot D \cdot u$ que não tem termos mistos.*

Se os autovalores de A são $\lambda_1, \cdots, \lambda_n$ e se $u = [u_1, \cdots, u_n]^T$, então

$$v^T \cdot A \cdot v = u^T \cdot D \cdot u = \lambda_1 u_1^2 + \cdots + \lambda_n u_n^2$$

Demonstração. Seja $Q(v) = v^T \cdot A \cdot v$ uma forma quadrática em n variáveis. Sendo $A_{n\times n}$ uma matriz simétrica, existe uma matriz ortogonal P que diagonaliza A, isto é, $P^T \cdot A \cdot P = D$, em que D é diagonal e os autovalores de A são as entradas diagonais de D.

Agora consideremos $v = Pu$ ou, equivalentemente, $u = P^{-1}v = P^T v$. A substituição na forma quadrática implica que

$$v^T \cdot A \cdot v = (Pu)^T A(Pu) = u^T \cdot \underbrace{P^T \cdot A \cdot P}_{D} \cdot u = u^T \cdot D \cdot u,$$

a qual é uma forma quadrática sem termos mistos já que D é diagonal. Além disso, se os autovalores de A são $\lambda_1, \cdots, \lambda_n$ então D pode ser escolhida de modo que

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

Se $u = [u_1, \cdots, u_n]^T$, então com respeito as novas variáveis, a forma quadrática passa a ser

$$v^T \cdot A \cdot v = u^T \cdot D \cdot u = \lambda_1 u_1^2 + \cdots + \lambda_n u_n^2.$$

$\square$

Esse processo é chamado diagonalização de formas quadráticas.

Exemplo 2.10.4. *Utilizando mudança de variáveis vamos transformar a forma quadrática $Q(x, y) = 3x^2 + 2xy + 3y^2$ e identificar o lugar geométrico representado pela equação $3x^2 + 2xy + 3y^2 = 4$. Escrevendo a forma quadrática na forma matricial obtemos*

$$Q(x, y) = v^T \cdot A \cdot v = \begin{bmatrix} x & y \end{bmatrix} \cdot \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix},$$

em que

$$v = (x, y) \text{ e } A = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}.$$

Note que A é a matriz do Exemplo 2.9.4 cujos autovalores e autovetores normalizados já foram calculados. Assim,

$$P = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \quad e \quad D = P^T \cdot A \cdot P = \begin{bmatrix} 4 & 0 \\ 0 & 2 \end{bmatrix}.$$

Após a mudança de variáveis que elimina os termos mistos $v = Pu$, temos,

$$u = (x_1, y_1) \quad \text{e} \quad \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ y_1 \end{bmatrix}.$$

Isso produz a nova forma quadrática

$$Q = \begin{bmatrix} x_1 & y_1 \end{bmatrix} \cdot \begin{bmatrix} 4 & 0 \\ 0 & 2 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = 4x_1^2 + 2y_1^2,$$

a qual não tem termos mistos. Desta forma temos

$$4x_1^2 + 2y_1^2 = 4 \Longleftrightarrow x_1^2 + \frac{y_1^2}{2} = 1 \quad ou \quad \frac{x_1^2}{1^2} + \frac{y_1^2}{(\sqrt{2})^2} = 1$$

que corresponde a uma elipse, na qual os vértices nos novos eixos são $(1, 0)$ e $(\sqrt{2}, 0)$.

Definição 2.10.5. *Uma matriz simétrica A é dita **definida positiva** se a forma quadrática $Q(v) = v^T \cdot A \cdot v$ for positiva, para todo $v \neq 0$.*

Exemplo 2.10.6. *Considere a matriz $A = \begin{bmatrix} 3 & 0 \\ 0 & 5 \end{bmatrix}$. Para qualquer vetor $v = (x, y)$, $v \neq 0$, é fácil verificar que*

$$Q(v) = v^T \cdot A \cdot v = \begin{bmatrix} x \\ y \end{bmatrix} \cdot \begin{bmatrix} 3 & 0 \\ 0 & 5 \end{bmatrix} \cdot \begin{bmatrix} x & y \end{bmatrix} = 3x^2 + 5y^2 > 0.$$

Porém para o caso geral, $A_{n \times n}$, não é simples verificar se uma matriz é ou não definida positiva a partir da definição. O teorema seguinte auxilia na resolução desse problema.

Teorema 2.10.7. *Seja A uma matriz $n \times n$ simétrica. A forma quadrática $Q(v) = v^T \cdot A \cdot v$ é positiva se, e somente se, todos os autovalores de A são positivos.*

Demonstração. Inicialmente ortogonalizando a matriz A, obtemos uma matriz ortogonal P tal que $D = P^T \cdot A \cdot P$ é a matriz diagonal

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

Então, pelo Teorema 2.10.3 dos Eixos Principais, a mudança de variável $v = Pu$ nos da

$$v^T \cdot A \cdot v = u^T \cdot D \cdot u = \lambda_1 u_1^2 + \cdots + \lambda_n u_n^2,$$

em que os λ_i, $i = 1, \cdots, n$, são os autovalores de A. Agora note que $u = P^T v$ implica que

$$u^T \cdot u = (P^T v)^T \cdot P^T v = v^T (P^T)^T P^T v = v^T (P \cdot P^T) v = v^T v$$

já que $P^T = P^{-1}$.

Desta forma $v^T \cdot A \cdot v > 0$ com $v \neq 0$ se, e só se, todos os valores de λ na equação

$$v^T \cdot A \cdot v = u^T \cdot D \cdot u = \lambda_1 u_1^2 + \cdots + \lambda_n u_n^2$$

são positivos. □

O Teorema acima é de extrema importância para auxiliar na demonstração do Teorema dos Valores Singulares que veremos no próximo capítulo.

Teorema 2.10.8 (Teorema dos Extremos Condicionados)**.** *Seja a forma quadrática*

$$f(x) = x^T \cdot A \cdot x$$

cuja a matriz associada A é de ordem $n \times n$. Sejam $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ os autovalores de A. Então, sujeitas à restrição $\|x\| = 1$, valem as seguintes afirmações:

a) $\lambda_n \leq f(x) \leq \lambda_1$

b) *O valor máximo de $f(x)$ é λ_1, e ocorre quando x é um autovetor unitário correspondente a λ_1.*

c) *O valor mínimo de $f(x)$ é λ_n, e ocorre quando x é um autovetor unitário correspondente a λ_n.*

Demonstração. Vamos começar diagonalizando A ortogonalmente já que A é simétrica. Obtemos assim uma matriz ortogonal Q tal que $Q^T \cdot A \cdot Q = D$ é a matriz diagonal

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

Então, pelo Teorema 2.10.3 dos Eixos Principais, a mudança de variável $x = Qy$ nos da $x^T \cdot A \cdot x = y^T \cdot D \cdot y$. Note que $y = Q^T x$ implica que

$$y^T y = (Q^T x)^T (Q^T x) = x^T Q Q^T x = x^T x.$$

Usando o fato de $x \cdot x = x^T x$ vemos que

$$\|y\|^2 = y \cdot y = y^T y = x^T x = x \cdot x = \|x\|^2 = 1.$$

Assim, se x é um vetor unitário y também é, e os valores de $x^T \cdot A \cdot x$ e $y^T \cdot D \cdot y$ são os mesmos.

Para provar o ítem (a) observe que se $y = (y_1, \cdots y_n)$ então

$$\begin{aligned} f(x) &= x^T \cdot A \cdot x = y^T \cdot D \cdot y = \lambda_1 \cdot y_1^2 + \cdots + \lambda_n \cdot y_n^2 \leq \lambda_1 \cdot y_1^2 + \cdots + \lambda_1 \cdot y_n^2 \\ &= \lambda_1(y_1^2 + \cdots + y_n^2) = \lambda_1 \cdot \|y\|^2 = \lambda_1. \end{aligned}$$

Assim, $f(x) = x^T \cdot A \cdot x \leq \lambda_1$, para todo vetor x, tal que $\|x\| = 1$. De modo análogo mostra-se que $f(x) = x^T \cdot A \cdot x \geq \lambda_n$. Isso mostra que todos os valores da forma quadrática $x^T \cdot A \cdot x$ estão entre o maior e o menor autovalor de A.

Para provar o ítem (b), considere x um autovetor unitário de A correspondente a λ_1. Então,

$$f(x) = x^T \cdot A \cdot x = x^T(\lambda_1 \cdot x) = \lambda_1 \cdot (x^T \cdot x) = \lambda_1 \cdot (x \cdot x) = \lambda_1 \cdot \|x\|^2 = \lambda_1.$$

Isto mostra que λ_1 é máximo de f e ocorre quando x é um autovetor unitário correspondente a λ_1. A prova da propriedade (c) é análoga à prova da propriedade (b). □

2.11 Quadrados Mínimos

O problema dos quadrados mínimos pode ser descrito da seguinte forma: Seja o sistema linear $Ax = b$. Dados uma matriz A, $m \times n$ e um vetor b, $m \times 1$, o problema de quadrados mínimos é encontrar um vetor coluna x, de ordem $n \times 1$ que minimiza a distância entre o vetor b e Ax, ou seja, que minimiza $\|b - Ax\|_2$, ou equivalentemente, minimizar $\|b - Ax\|_2^2$.

Se $m = n$ e A é não singular, a resposta é simples, $x = A^{-1}b$. Mas se $m > n$, então temos mais equações do que incógnitas, este problema é dito sobredeterminado. Tais sistemas, muitas vezes, são incompatíveis. Então, geralmente, não podemos esperar encontrar um vetor $x \in \mathbb{R}^n$ que satisfaz $Ax = b$ exatamente. Em vez disso, vamos procurar um vetor x para o qual Ax está o mais próximo possível de b. A ortogonalidade tem um papel importante na escolha desse x.

Resumindo, quando uma solução é procurada e não existe, o melhor a fazer é encontrar um x que torna Ax o mais próximo possível de b. Um vetor $\bar{x}$ que faz isso é dito uma **solução de quadrados mínimos** para o sistema $Ax = b$. Ocasionalmente encontramos o problema subdeterminado, onde $m < n$, mas vamos nos concentrar no caso sobredeterminado (WATKINS, 2010; POOLE, 2011; OLVER; SHAKIBAN; SHAKIBAN, 2018).

Definição 2.11.1. *Se W é um subespaço de um espaço linear normado V e se v é um vetor em V, então a melhor aproximação em W para v é o vetor u em W tal que $\|v - u\| < \|v - w\|$ para todo w em W diferente de v.*

Supondo que queiramos aproximar o vetor b em $\mathbb{R}^3$ fixado por algum vetor w de algum subespaço $W \subset \mathbb{R}^3$. A menos que b esteja em W, qualquer aproximação dessas resulta num vetor erro $b - w$, independente do vetor $w \neq 0$ escolhido. No entanto, escolhendo $w = proj_W b$ podemos tornar o comprimento do vetor erro

$$\|b - w\| = \|b - proj_W b\|$$

o menor possível.

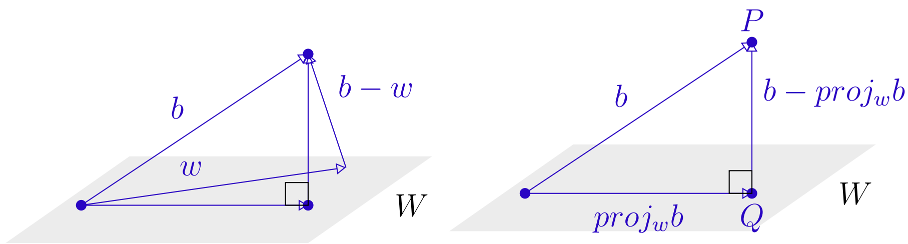

Figura 1 – Melhor aproximação, adaptado de Anton & Rorres (2012, p. 367)

Teorema 2.11.2. *Assuma que V é um espaço com produto interno. Sejam W um subespaço vetorial de V e b um vetor em V. Então a projeção ortogonal do vetor b sobre o subespaço W, denotada por $p = proj_W b$, é a melhor aproximação para b em W. De outra forma, $\|b - p\| < \|b - w\|$ para todo $w \neq p$ em W.*

Demonstração. Sejam $w \in W$ qualquer e escreva $p = proj_W b$. Então $b - w = (b-p)+(p-w)$. Sendo uma diferença de vetores em W, o vetor $p - w \in W$. Dado que $b - p$ é ortogonal a W, segue que $b-p$ e $p-w$ são ortogonais. Assim, por uma versão do Teorema de Pitágoras para espaços vetoriais com produto interno vemos que

$$\|b - w\|_2^2 = \|b - p\|_2^2 + \|p - w\|_2^2 .$$

Entretanto, $\|p - w\|_2^2 > 0$, já que $w \neq p$. Então,

$$\|b - p\|_2^2 < \|b - p\|_2^2 + \|p - w\|_2^2 = \|b - w\|_2^2 .$$

Ou, equivalentemente, $\|b - p\|_2 < \|b - w\|_2$. Assim, por definição, $p = proj_W b$ é a melhor aproximação de b em W. $\square$

Dado uma matriz A de ordem $m \times n$ com $m \geq n$, consideramos o sistema linear $Ax = b$, o qual é um sistema linear sobredeterminado. Para cada $x \in \mathbb{R}^n$, podemos formar o vetor **resíduo** $r = b - Ax$ com $r \in \mathbb{R}^m$. Uma maneira de encontrar alguma solução por quadrados mínimos, $\bar{x}$, de $Ax = b$ é: calcular a projeção ortogonal $p = proj_{R(A)} b$ e, depois, resolver a equação $A\bar{x} = p$. Contudo, podemos evitar o cálculo da projeção escrevendo $A\bar{x} = p$ como $b - A\bar{x} = b - p$. Então multiplicamos ambos os membros dessa equação por A^T, obtendo

$$A^T(b - A\bar{x}) = A^T(b - p).$$

Sendo $b - p$ é ortogonal ao subespaço $R(A)$, concluímos pela Proposição 2.8.4 que

$$b - p \in N(A^T).$$

Isto implica que $A^T(b - p) = 0$. Dessa forma

$$A^T(b - A\bar{x}) = A^T(b - p) = 0 \iff A^Tb - A^TA\bar{x} = 0 \iff A^TA\bar{x} = A^Tb. \tag{2.5}$$

Isso representa um sistema de equações conhecido como **equações normais** para $\bar{x}$.

Observação 2.11.3. *Supondo que A^TA seja invertível, ou seja, que A possua colunas linearmente independentes, a solução por quadrados mínimos é dada por*

$$\bar{x} = (A^TA)^{-1}A^Tb.$$

*Desta forma a matriz $(A^TA)^{-1}A^T$ funciona como uma matriz inversa de A para esse problema. Ela é chamada de **pseudo-inversa** de A e denotada por A^+. Isto é,*

$$A^+ = (A^TA)^{-1}A^T.$$

Um problema comum no trabalho experimental é obter uma relação matemática $y = f(x)$ entre as variáveis x e y através de um ajuste de uma curva aos pontos

$$(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n).$$

determinados experimentalmente.

Se esses dados consistem em $n + 1$ pontos do plano, é possível encontrar um polinômio de grau menor ou igual a n cujo gráfico contém todos os pontos. Tal polinômio é chamado **polinômio interpolador**. Como os dados geralmente envolvem erros experimentais, não há necessidade de exigir que o gráfico contenha todos os pontos. A ideia é escolher a curva que melhor ajusta os dados, que pode ser uma reta, ou uma parábola, ou um polinômio cúbico, entre outras.

Vamos começar definindo o que significa "melhor ajuste". Considere, por exemplo, $(1, 2)$, $(2, 2)$ e $(3, 4)$ como pontos que surgiram de medidas feitas durante algum experimento. Esperamos também que os pontos estejam sobre alguma reta $y = a + bx$. Desta forma, todos os três pontos vão satisfazer essa equação e teremos

$$2 = a + b \cdot 1, \quad 2 = a + b \cdot 2 \text{ e } \quad 4 = a + b \cdot 3.$$

Esse é um sistema com três equações lineares com duas variáveis

$$\begin{cases} a+b=2 \\ a+2b=2 \\ a+3b=4 \end{cases} \iff \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \cdot \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \\ 4 \end{bmatrix}.$$

Esse sistema é inconsistente já que os três pontos não estão sobre uma reta. Então, vamos buscar uma reta que esteja o mais próximo possível de passar pelos pontos dados. Para qualquer reta mediremos a distância vertical de cada ponto dado até a reta, e então tentaremos escolher a reta que minimize o erro total. Se os erros forem denotados por r_1, r_2 e r_3, o vetor erro é $r = (r_1, r_2, r_3)$.

Queremos que o vetor r seja o menor possível. Isto significa que $\| r \|$ deve estar o mais próximo possível de zero. Neste caso, por questão de conveniência, vamos usar a norma Euclidiana. Então, vamos minimizar

$$\| r \|_2 = \sqrt{r_1^2 + r_2^2 + r_3^2} \quad \text{ou} \quad \| r \|_2^2 = r_1^2 + r_2^2 + r_3^2.$$

É daí que vem o termo quadrados mínimos. O número $\| r \|_2$ é chamado de resíduo ou erro quadrático mínimo da aproximação.

Por que, ao invés de minimizar a norma Euclidiana, não minimizamos apenas a soma dos valores absolutos dos erros, isto é, $\| r \|_1 = | r_1 | + \cdots + | r_n |$ do vetor resíduo, ou minimizar o erro máximo, isto é, $\| r \|_\infty = max\{| r_1 |, \cdots, | r_n |\}$? Embora cada um desses critérios alternativos de minimização seja interessante e potencialmente útil, todos eles levam a problemas de minimização não-lineares e, portanto, são muito mais difíceis de resolver.

No exemplo em questão obtemos as seguintes fórmulas para os erros:

$$r_1 = 2 - (a + b \cdot 1), \quad r_2 = 2 - (a + b \cdot 2) \text{ e } \quad r_3 = 4 - (a + b \cdot 3).$$

Em geral, suponha que tenhamos n pontos dados $(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$ e uma reta $y = a + bx$. O vetor erro é $r = (r_1, r_2, \cdots, r_n)$ em que $r_i = b - (a + bx_i)$ com $i = 1, \cdots, n$. A reta que minimiza $r_1^2 + r_2^2 + \cdots + r_n^2$ é chamada **reta de quadrados mínimos** ou **reta de melhor ajuste** para os pontos $(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$.

Desta forma podemos expressar esse problema em forma de matriz. Se os pontos dados estiverem na reta $y = a + bx$, as n equações lineares

$$\begin{aligned} a + bx_1 &= y_1 \\ &\vdots \\ a + bx_n &= y_n \end{aligned}$$

seriam todas verdadeiras. O interessante está no caso em que os pontos não são colineares, no qual o sistema é indeterminado. Em forma de matriz temos

$$\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \cdot \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix},$$

o qual poder representado por $Ax = b$, em que

$$A = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}, \quad x = \begin{bmatrix} a \\ b \end{bmatrix} \text{ e } b = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}.$$

O vetor resíduo é dado por $r = b - Ax$, e queremos minimizar $\| \ r \ \|_2^2$. Assim, podemos reescrever nosso problema em forma de matrizes, como já tínhamos visto acima

$$\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \cdot \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \\ 4 \end{bmatrix},$$

em que

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}, \bar{x} = \begin{bmatrix} a \\ b \end{bmatrix} \text{ e } b = \begin{bmatrix} 2 \\ 2 \\ 4 \end{bmatrix}$$

Suas equações normais são dadas por

$$A^T A\bar{x} = A^T b \Longleftrightarrow \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{bmatrix} \cdot \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \cdot \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ 2 \\ 4 \end{bmatrix}$$

O que implica que

$$A^T A\bar{x} = A^T b \Longleftrightarrow \begin{bmatrix} 3 & 6 \\ 6 & 14 \end{bmatrix} \cdot \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 8 \\ 18 \end{bmatrix}$$

Desta última equação resulta o sistema

$$\begin{cases} 3a + 6b = 8 \\ 6a + 14b = 18, \end{cases}$$

cuja solução é $a = \frac{2}{3}$ e $b = 1$. Isto siginifica que $\bar{x} = (\frac{2}{3}, 1)$. Logo, o polinômio interpolador de primeiro grau que corresponde ao melhor ajuste de quadrados mínimos é $y = \frac{2}{3} + x$. A Figura 2 ilustra o ajuste desenvolvido neste exemplo.

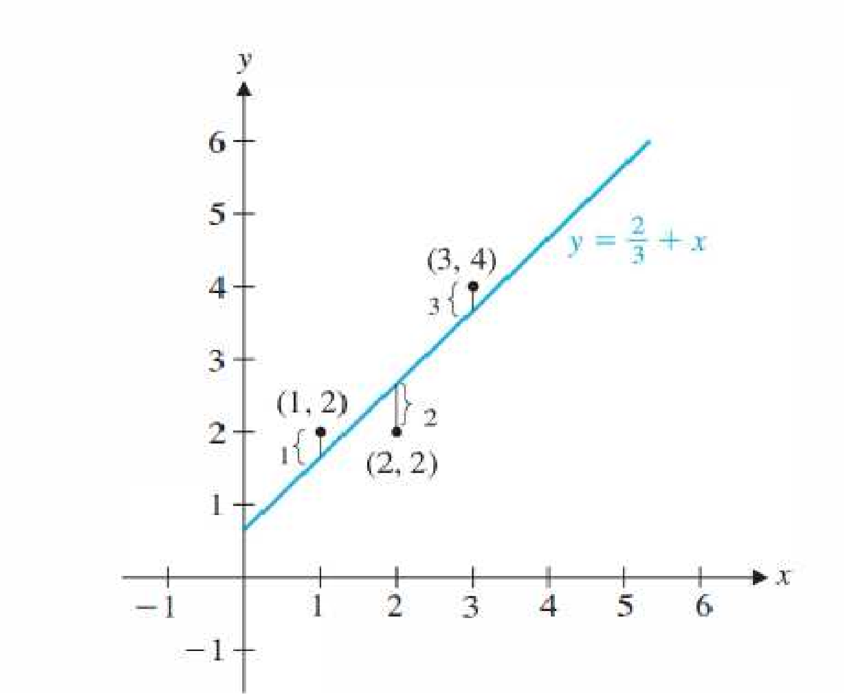

Figura 2 – Reta de Quadrados Mínimos (POOLE, 2011, 527)

3 Decomposição em Valores Singulares

No capítulo anterior, vimos que toda matriz simétrica A pode ser fatorada como

$$A = P \cdot D \cdot P^T,$$

em que P é uma matriz ortogonal e D é uma matriz diagonal formada pelos autovalores de A, associados aos vetores coluna de P, e podemos chamá-la de **Decomposição em Autovalores** de A.

Se A não for simétrica uma fatoração desse tipo não é possível, mas vimos também que podemos fatorar uma matriz quadrada A como $A = P \cdot D \cdot P^{-1}$, em que D é também uma matriz diagonal formada pelos autovalores de A, mas P agora é apenas uma matriz invertível. Entretanto, nem toda matriz é diagonalizável.

O resultado surpreendente que será apresentado nesse capítulo é o fato de que toda matriz, quadrada ou não, simétrica ou não, possui uma fatoração da forma $A = P \cdot D \cdot Q^T$, em que P e Q são ortogonais e D é uma matriz diagonal. Esse resultado notável é chamado de **Decomposição em Valores Singulares** cuja sigla em inglês é **SVD (Singular Value Decomposition)**, e é uma das mais importantes dentre todas as fatorações de matrizes. Assim nesse capítulo vamos mostrar algumas propriedades da SVD e como calcular uma SVD de uma matriz (AXLER, 1997; TREFETHEN; III, 1997; WATKINS, 2010; POOLE, 2011; ANTON; RORRES, 2012; OLVER; SHAKIBAN; SHAKIBAN, 2018).

3.1 Valores Singulares

Para qualquer matriz A de ordem $m \times n$, lembramos que a matriz A^TA, é simétrica. Com efeito,

$$\left(A^TA\right)^T = A^T\left(A^T\right)^T = A^TA.$$

Agora, pelo Teorema Espectral 2.9.1, A^TA pode ser diagonalizada ortogonalmente, pelo Teorema 2.9.2 sabemos que os autovalores são reais. Além disso, afirmamos que os autovalores de A^TA são não negativos. De fato, seja λ um autovalor de A^TA com o correspondente autovetor unitário v. Então:

$$0 \leq \|Av\|^2 = (Av) \cdot (Av) = v \cdot (A^TAv) = v \cdot \lambda v = \lambda\,(v \cdot v) = \lambda \cdot \|v\|^2 = \lambda.$$

Portanto $\lambda \geq 0$, e assim faz sentido extrair a raíz quadrada (positiva) desses autovalores.

Mostramos acima que $\lambda_i = \| Av_i \|^2$. Portanto $\sigma_i = \sqrt{\lambda_i} = \| Av_i \|$. Em outras palavras, os valores singulares de A são os comprimentos dos vetores Av_i com $i = 1, \ldots, n$.

Definição 3.1.1. *Seja A uma matriz de ordem $m \times n$ e $\lambda_1, \lambda_2, \ldots, \lambda_n$ os autovalores da matriz $A^T A$, os números $\sigma_i = \sqrt{\lambda_i} \geq 0$, com $i = 1, \ldots, n$, são denominados valores singulares da matriz A.*

Vamos supor que $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$, assim os valores singulares $\sigma_1 \geq \sigma_2 \geq, \ldots \geq \sigma_n$. Com os valores singulares nomeados desta forma, σ_1 irá sempre denotar o maior valor singular ou valor singular dominante.

Exemplo 3.1.2. *Seja A matriz de ordem 2×2 definida por*

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

Queremos encontrar os valores singulares da matriz A. Pela definição, precisamos encontrar os autovalores da matriz $A^T A$. Calculemos

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

Calculando o polinômio característico de $A^T A$ obtemos

$$P(\lambda) = det\left(A^T A - \lambda \cdot I\right) = \begin{vmatrix} 3-\lambda & 0 \\ 0 & 2-\lambda \end{vmatrix} = (3-\lambda)(2-\lambda) = 0.$$

Assim os autovalores de $A^T A$ são $\lambda_1 = 3$ e $\lambda_2 = 2$. Logo os valores singulares da matriz A são $\sigma_1 = \sqrt{\lambda_1} = \sqrt{3} = 1,732$ e $\sigma_2 = \sqrt{\lambda_2} = \sqrt{2} = 1,414$.

3.2 Decomposição em Valores Singulares

Uma intuição geométrica tem a vantagem de auxiliar na visualização da decomposição em valores singulares. Inicialmente vamos pensar no efeito de uma matriz A sobre um vetor v, por exemplo, digamos que $A = \begin{bmatrix} 2 & 1 \\ -1 & 1 \end{bmatrix}$ e $v = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$. Então

$$Av = \begin{bmatrix} 2 & 1 \\ -1 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 4 \\ 1 \end{bmatrix},$$

conforme ilustrado pela seguinte figura.

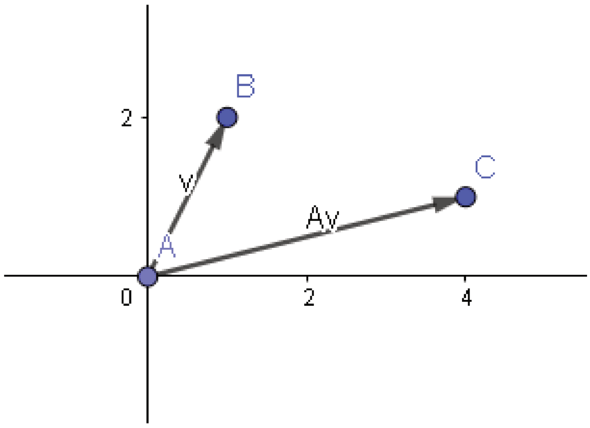

Figura 3 – Vetores v e Av, elaboração própria

A ideia é transformar o vetor v no vetor Av que basicamente consiste no ato de rotacionar e esticar o vetor v. Isso é o que acontece sempre que se manipula um vetor sob a multiplicação de uma matriz.

Procedendo de modo análogo, o que acontece com um círculo unitário S sob a multiplicação de uma determinada matriz A? Considere a seguinte figura:

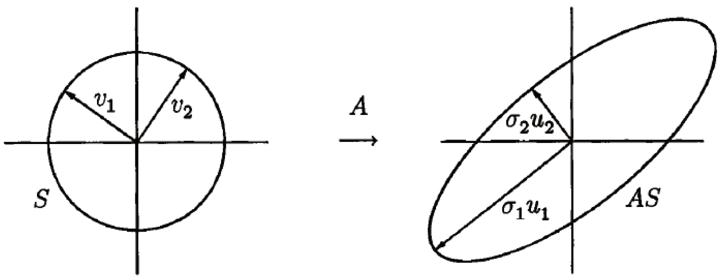

Figura 4 – SVD de uma matriz 2x2, Trefethen & III (1997, p. 26)

Observe que os vetores v_1 e v_2 ortogonais e unitários são transformados pela matriz A nos vetores $\sigma_1 u_1$ e $\sigma_2 u_2$ que produzem a rotação e o alongamento dos vetores v_1 e v_2. Assim a matriz A transforma o círculo unitário S em uma elipse cujos eixos principais são $\sigma_1 u_1$ e $\sigma_2 u_2$ em que σ_1 e σ_2 são os fatores de alongamento chamados de valores singulares da matriz A.

Considere agora uma esfera unitária S de $\mathbb{R}^n$ e A uma matriz de ordem $m \times n$ com $m \geq n$. Supondo que A tenha posto completo, ou seja, $posto(A) = n$, a imagem AS é uma hiperelipse de $\mathbb{R}^m$.

Vimos no capítulo anterior que para algumas matrizes A é possível obter a Decomposição Espectral da forma $A = P \cdot D \cdot P^T$ em que P é uma matriz ortogonal $n \times n$ e D é uma matriz diagonal $n \times n$, na qual os elementos da diagonal são os autovalores da matriz A, ou seja, $a_{ii} = \lambda_i$ para $i = 1, 2, \ldots, n$. As colunas de P são os autovetores correspondentes, digamos p_i, $i = 1, 2, \ldots, n$. Desta forma,

$$A = P \cdot D \cdot P^T \Longleftrightarrow AP = PD \Longleftrightarrow Ap_i = \lambda_i p_i, i = 1, 2, \ldots, n.$$

A proposta para matrizes gerais A de ordem $m \times n$ é construir uma decomposição da forma

$$A = U \cdot \varSigma \cdot V^T,$$

em que U é uma matriz ortogonal de ordem $m \times m$, $\varSigma$ uma matriz diagonal de ordem $m \times n$ e V uma matriz ortogonal de ordem $n \times n$.

Para não causar estranheza, vamos definir a diagonal principal de uma matriz $m \times n$ como a fileira de entradas que começa no canto superior esquerdo e se estende diagonalmente até onde for possível. O teorema a seguir garante a existência dessa decomposição.

Teorema 3.2.1 (Decomposição por Valores Singulares - SVD)**.** *Seja A uma matriz de ordem $m \times n$ com valores singulares $\sigma_1 \geq \sigma_2 \geq \ldots, \geq \sigma_r > 0$ e $\sigma_{r+1} = \cdots = \sigma_n = 0$. Então existem uma matriz ortogonal $U_{m\times m}$, uma matriz ortogonal $V_{n\times n}$ e uma matriz não quadrada diagonal $\Sigma_{m\times n}$ de modo que $A = U \cdot \Sigma \cdot V^T$.*

Demonstração. Queremos mostrar que a matriz A de ordem $m \times n$ pode ser fatorada como $A = U \cdot \Sigma \cdot V^T$, em que $U_{m\times m}$ e $V_{n\times n}$ são ortogonais e $\Sigma_{m\times n}$ é diagonal. Se os valores singulares de A são $\sigma_1 \geq \sigma_2 \geq \ldots, \geq \sigma_r > 0$ e $\sigma_{r+1} = \cdots = \sigma_n = 0$, então Σ terá a forma em blocos

$$\Sigma_{m\times n} = \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix},$$

em que

$$D_{r\times r} = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_r \end{bmatrix}$$

e 0 é uma matriz nula com ordem adequada.

Seja $\{v_1, \ldots, v_n\}$ uma base de autovetores ortonomais de A^TA em $\mathbb{R}^n$. A matriz ortogonal V será dada por $V = [v_1 \cdots v_n]$. Para construir a matriz U, primeiro notamos que $\{Av_1, \ldots, Av_n\}$ é um conjunto ortogonal de $\mathbb{R}^m$. Para constatar isso, suponha que v_i seja o autovetor da matriz simétrica A^TA correspondente ao autovalor λ_i. Então, para $i \neq j$, temos

$$\begin{aligned} (Av_i) \cdot (Av_j) &= (Av_i)^T(Av_j) = (v_i^TA^T)(Av_j) = v_i^T(A^TAv_j) \\ &= v_i \cdot (A^TAv_j) = v_i \cdot (\lambda_j \cdot v_j) = \lambda_j\,(v_i \cdot v_j) \\ &= \lambda_j \cdot 0 = 0, \end{aligned}$$

visto que os vetores v_i são ortogonais.

Agora, lembre-se que para cada $i = 1, \ldots, n$ o valor singular é $\sigma_i = \parallel Av_i \parallel$, e que os primeiros r valores singulares são não nulos. Portanto, podemos normalizar os vetores do conjunto $\{Av_1, \ldots, Av_n\}$ considerando $u_i = \frac{Av_i}{\|Av_i\|} = \frac{Av_i}{\sigma_i}$ ou $Av_i = \sigma_i u_i$, com $i = 1, \ldots, r$. Isso garante que $\{u_1, \ldots, u_r\}$ seja um conjunto ortonormal de $\mathbb{R}^m$, mas se $r < m$ esse conjunto não será uma base para $\mathbb{R}^m$. Nesse caso, podemos estender o conjunto $\{u_1, \ldots, u_r\}$ para uma base ortonormal $\{u_1, \ldots, u_r, u_{r+1}, \ldots, u_m\}$ de $\mathbb{R}^m$.

Resta verificar que, com U, V e Σ já descritas anteriormente, temos $A = U \cdot \Sigma \cdot V^T$. Como a matriz V é ortogonal, isto é, $V^T = V^{-1}$, isso equivale a mostrar que $AV = U\Sigma$. Sabemos que $Av_i = \sigma_i u_i$ para $i = 1, \ldots, r$ e $\parallel Av_i \parallel = \sigma_i = 0$ para $i = r+1, \ldots, n$. Disso, por meio de manipulações matriciais vemos que

$$AV = A\left[v_1 \cdots v_n\right] = \left[Av_1 \cdots Av_n\right] = \left[Av_1 \cdots Av_r, 0 \cdots 0\right] = \left[\sigma_1 \cdot u_1 \cdots \sigma_r \cdot u_r, 0 \cdots 0\right].$$

Portanto,

$$AV = \left[u_1 \cdots u_m\right] \cdot \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 & \\ 0 & \sigma_2 & \cdots & 0 & \\ \vdots & \vdots & \ddots & \vdots & O \\ 0 & 0 & \cdots & \sigma_r & \\ & O & & & O \end{bmatrix} = U\Sigma.$$

□

Uma fatoração do tipo $A = U \cdot \Sigma \cdot V^T$ vista no teorema acima é chamada de **decomposição em valores singulares (SVD)** de A. As colunas de U são chamadas de **vetores singulares à esquerda** de A, e as colunas de V são chamadas de **vetores singulares à direita** de A.

Exemplo 3.2.2. *Seja a matriz A de ordem 3×2 dada por*

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

Já trabalhamos com esta matriz no exemplo anterior. A decomposição por valores singulares da matriz A é dada por

$$A_{3\times 2} = U_{3\times 3} \cdot \Sigma_{3\times 2} \cdot V^T_{2\times 2}$$

em que U e V são matrizes ortogonais e Σ uma matriz diagonal. Já sabemos que os valores singulares da matriz A são $\sigma_1 = \sqrt{3}$ e $\sigma_2 = \sqrt{2}$, então a matriz diagonal $\Sigma_{3\times 2}$ é dada por

$$\Sigma = \begin{bmatrix} \sqrt{3} & 0 \\ 0 & \sqrt{2} \\ 0 & 0 \end{bmatrix}.$$

Para calcular a matriz V precisamos obter os autovetores ortonormais da matriz A^TA. Inicialmente vamos calcular os autovetores associados ao autovalor $\lambda_1 = 3$. Por definição de autovalores e autovetores temos

$$A^TAv_1 = \lambda_1 v_1 \Longleftrightarrow \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} = 3 \cdot \begin{bmatrix} x \\ y \end{bmatrix} \Longleftrightarrow \begin{cases} 3x = 3x \\ 2y = 3y \end{cases}.$$

Assim x é um número real qualquer e $y = 0$, logo $V_{\lambda_1} = span\{(1,0)\}$. O que implica que $v_1 = (1,0)$.

Analogamente, é possível mostrar que $V_{\lambda_2} = span\{(0,1)\}$. Desta forma, a matriz V é dada por

$$V = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = V^T.$$

Para calcular a matriz U devemos lembrar que $u_i = \frac{Av_i}{\|Av_i\|} = \frac{Av_i}{\sigma_i}$ com $i = 1,2,3$. Assim

$$u_1 = \frac{1}{\sqrt{3}} \cdot \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ -1 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{3}} \end{bmatrix} \text{ e } u_2 = \frac{1}{\sqrt{2}} \cdot \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ -1 & 1 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}.$$

são dois dos três vetores coluna de U. Note que os vetores u_1 e u_2 são ortonormais como esperado. Assim podemos estender o conjunto $\{u_1, u_2\}$ a uma base de $\mathbb{R}^3$. Vamos simplificar os cálculos removendo convenientemente as raízes multiplicando u_1 e u_2 por escalares apropriados. Desta forma vamos encontrar um vetor u_3 ortogonal a

$$\sqrt{3}u_1 = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix} \text{ e } \sqrt{2}u_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}.$$

Para satisfazer essas duas condições de ortogonalidade, o vetor u_3 deve ser uma solução do sistema linear homogêneo

$$\begin{bmatrix} 1 & 1 & -1 \\ 0 & 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Longleftrightarrow \begin{cases} x + y - z = 0 \\ y + z = 0. \end{cases}$$

Resolvendo o sistema obtemos $y = -z$ e $x = 2z$. Tomando $z = 1$ obtemos $\tilde{u}_3 = (2, -1, 1)$. Normalizando o vetor $\tilde{u}_3$ obtemos $u_3 = \left(\frac{2}{\sqrt{6}}, -\frac{1}{\sqrt{6}}, \frac{1}{\sqrt{6}}\right)$. Disso, escrevemos a matriz U como

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

Portanto, a decomposição por valores singulares (SVD) de A é dada por

$$A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{3}} & 0 & \frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \end{bmatrix} \cdot \begin{bmatrix} \sqrt{3} & 0 \\ 0 & \sqrt{2} \\ 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = U \cdot \Sigma \cdot V^T.$$

Observe que os cálculos envolvidos na decomposição em valores singulares são bastante trabalhosos, e neste caso é mais conveniente utilizar softwares matemáticos para efetuar esses cálculos como, por exemplo, o software Mathematica, o MatLab ou o Python.

Existe outra forma de decompor uma matriz em valores singulares análoga a decomposição espectral para uma matriz simétrica. Tal decomposição é obtida a partir da SVD por uma expansão de um produto externo

Teorema 3.2.3 (Forma SVD por Produto Externo)**.** *Seja $A_{m\times n}$ uma matriz com valores singulares $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$ e $\sigma_{r+1} = \cdots = \sigma_n = 0$. Sejam $u_1, \ldots, u_r$ os vetores singulares à esquerda, e sejam $v_1, \ldots, v_r$ os vetores singulares à direita de A, correspondentes a esses valores singulares. Então,*

$$A = \sigma_1 \cdot u_1 \cdot v_1^T + \cdots + \sigma_r \cdot u_r \cdot v_r^T.$$

Demonstração. Podemos mostrar essa versão da SVD tomando como base o que já fizemos anteriormente para obter a decomposição espectral de uma matriz simétrica. Usando a multiplicação por blocos e a representação coluna - linha do produto temos

$$A = U \cdot \Sigma \cdot V^T = [u_1 \cdots u_m] \cdot \begin{bmatrix} \begin{matrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_r \end{matrix} & 0 \\ 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} v_1^T \\ \vdots \\ v_n^T \end{bmatrix}.$$

De outra forma, segue que

$$A = \left[u_1 \cdots u_r \vdots u_{r+1} \cdots u_m\right] \cdot \begin{bmatrix} \begin{matrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_r \end{matrix} & 0 \\ 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} v_1^T \\ \vdots \\ v_r^T \\ \cdots \\ v_{r+1}^T \\ \vdots \\ v_n^T \end{bmatrix}.$$

Agora, reescrevamos o produto de matriz da seguinte forma

$$A = [u_1 \cdots u_r] \cdot \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_r \end{bmatrix} \cdot \begin{bmatrix} v_1^T \\ \vdots \\ v_r^T \end{bmatrix} + [u_{r+1} \cdots u_m] \cdot [0] \cdot \begin{bmatrix} v_{r+1}^T \\ \vdots \\ v_n^T \end{bmatrix}.$$

Disso,

$$A = [u_1 \cdots u_r] \cdot \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_r \end{bmatrix} \cdot \begin{bmatrix} v_1^T \\ \vdots \\ v_r^T \end{bmatrix}.$$

Concluimos assim que

$$A = [\sigma_1 \cdot u_1 \cdots \sigma_r \cdot u_r] \cdot \begin{bmatrix} v_1^T \\ \vdots \\ v_r^T \end{bmatrix} = \sigma_1 \cdot u_1 \cdot v_1^T + \cdots + \sigma_r \cdot u_r \cdot v_r^T.$$

□

Teorema 3.2.4. *Sejam $A = U \cdot \Sigma \cdot V^T$ uma decomposição a valores singulares de uma matriz $A_{m \times n}$ não nula com posto r. Então, a esfera unitária de $\mathbb{R}^n$ por uma transformação matricial que associa o vetor x ao vetor Ax é:*

a) *a superfície de um elipsóide em $\mathbb{R}^m$, se $r = n$;*

b) *o elipsóide sólido em $\mathbb{R}^m$, se $r < n$.*

Demonstração. Sejam $u_1, \ldots, u_m$ os vetores singulares à esquerda e $v_1, \ldots, v_n$ os vetores singulares à direita de A, respectivamente. Já que o $posto(A) = r$, os valores singulares de A satisfazem $\sigma_1 \geq \sigma_2 \geq \ldots, \geq \sigma_r > 0$ e $\sigma_{r+1} = \cdots = \sigma_n = 0$. Seja

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

um vetor unitário em $\mathbb{R}^n$. Como a matriz V é ortogonal então V^T também é, e por isso,

$$V^T x = \begin{bmatrix} v_1^T \\ \vdots \\ v_n^T \end{bmatrix} x = \begin{bmatrix} v_1^T x \\ \vdots \\ v_n^T x \end{bmatrix}$$

é um vetor unitário. Logo

$$\left(v_1^T x\right)^2 + \cdots + \left(v_1^T x\right)^2 = 1.$$

Pela decomposição em valores singulares do produto externo, temos

$$A = \sigma_1 \cdot u_1 \cdot v_1^T + \cdots + \sigma_r \cdot u_r \cdot v_r^T.$$

Portanto

$$Ax = \sigma_1 \cdot u_1 \cdot v_1^T x + \cdots + \sigma_r \cdot u_r \cdot v_r^T x = \left(\sigma_1 v_1^T x\right) u_1 + \cdots + \left(\sigma_r v_r^T x\right) u_r.$$

Escrevendo os escalares $y_i = \sigma_i v_i^T x$, com $i = 1, \ldots, r$, ou, equivalentemente, $v_i^T x = \frac{y_i}{\sigma_i}$ temos

$$Ax = y_1 \cdot u_1 + \cdots + y_r \cdot u_r.$$

a) Se $r = n$, devemos ter $n \leq m$ e

$$Ax = y_1 \cdot u_1 + \cdots + y_n \cdot u_n = Uy$$

em que $y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$. Portanto $\| Ax \| = \| Uy \| = \| y \|$, já que U é ortogonal. Assim

$$\left(v_1^T x\right)^2 + \cdots + \left(v_1^T x\right)^2 = \left(\frac{y_1}{\sigma_1}\right)^2 + \cdots + \left(\frac{y_n}{\sigma_n}\right)^2 = 1.$$

O que mostra que os vetores Ax formam a superfície de um elipsóide em $\mathbb{R}^m$.

b) Se $r < n$, a equação do ítem (a) passa a ser a inequação

$$\left(\frac{y_1}{\sigma_1}\right)^2 + \cdots + \left(\frac{y_n}{\sigma_n}\right)^2 \leq 1,$$

já que estão faltando algumas parcelas. Dessa forma, essa inequação corresponde a um elipsóide sólido em $\mathbb{R}^m$. □

Exemplo 3.2.5. *Para descrever a imagem da esfera unitária de $\mathbb{R}^3$ sob a ação da matriz*

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

é necessário calcular inicialmente sua SVD. Por meio do software Mathematica, o comando SingularValueDecomposition nos mostra que

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

Como $r = posto\,(A) = 2 < 3 = n$ temos que a imagem da esfera unitária irá satisfazer a inequação

$$\left(\frac{y_1}{\sigma_1}\right)^2 + \left(\frac{y_2}{\sigma_2}\right)^2 \leq 1 \text{ ou } \frac{y_1^2}{2} + y_2^2 \leq 1.$$

em relação aos eixos coordenados $y_1 y_2$ em $\mathbb{R}^2$.

Em geral, podemos descrever o efeito de uma matriz A de ordem $m \times n$ sobre a esfera unitária de $\mathbb{R}^n$ em relação aos efeitos de cada fator na sua SVD, $A = U\varSigma V^T$, da direita para esquerda. Como V^T é uma matriz ortogonal, ela preserva a esfera, isto é, ela leva a esfera unitária nela mesma. Com respeito à matriz $\varSigma$ de ordem $m \times n$, os elementos da diagonal $\sigma_{r+1} = \cdots = \sigma_n = 0$ anulam $n-r$ das dimensões da esfera unitária, restando uma esfera unitária de dimensão r, a qual os elementos da diagonal $\sigma_1, \cdots, \sigma_r$ irão distorcer em um elipsóide. Por fim, a matriz ortogonal U alinha os eixos desse elipsóide com a base ortonormal de vetores $u_1 \ldots, u_r$ de $\mathbb{R}^m$.

3.3 SVD Reduzida

Algebricamente, as linhas e colunas nulas da matriz Σ são supérfulas e podem ser eliminadas multiplicando-se por extenso a expressão $U\Sigma V^T$ por meio de multiplicação em blocos e subdividindo-se as matrizes conforme indicado na fórmula da SVD. Os produtos que envolvem blocos nulos como fatores desaparecem, restando

$$A = [u_1 \cdots u_r] \cdot \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_r \end{bmatrix} \cdot \begin{bmatrix} v_1^T \\ \vdots \\ v_r^T \end{bmatrix}$$

que é denominada **SVD reduzida ou condensada** de A. Vamos denotar as matrizes do lado direito por $A = \widehat{U}\widehat{\Sigma}\widehat{V}^T$ cujas ordens são $m \times r$, $r \times r$ e $r \times n$, respectivamente. Observamos que a matriz $\widehat{\Sigma}$ é invertível já que suas entradas diagonais são positivas.

Multiplicando o lado direito de $A = \widehat{U}\widehat{\Sigma}\widehat{V}^T$ por colunas e linhas, obtemos

$$A = \sigma_1 \cdot u_1 \cdot v_1^T + \cdots + \sigma_r \cdot u_r \cdot v_r^T$$

que é denominada **expansão em valores singulares reduzida** de A. Esse resultado é aplicável a qualquer matriz.

3.3.1 Algumas Aplicações Básicas da SVD

A partir da SVD de uma matriz A é possível simplificar cálculos, como por exemplo, da norma Euclidiana, do número de condição e da norma de Frobenius de uma matriz conforme mostram os teoremas a seguir.

Teorema 3.3.1. *Seja A uma matriz de $m \times n$ com valores singulares*

$$\sigma_1 \geq \sigma_2 \geq \ldots, \geq \sigma_r > 0. \text{ Então } \parallel A \parallel_2 = \sigma_1.$$

Demonstração. Da decomposição em valores singulares de A temos

$$A = U\Sigma V^T \iff AV = U\Sigma$$

e $Av_j = \sigma_j u_j$, $j = 1, 2, \ldots, r$. Agora, considerando que $\parallel v_1 \parallel_2 = 1$ e que $\parallel u_1 \parallel_2 = 1$ temos $\frac{\|Av_1\|_2}{\|v_1\|_2} = \parallel \sigma_1 u_1 \parallel_2 = \sigma_1 \parallel u_1 \parallel_2 = \sigma_1$. Portanto

$$\parallel A \parallel_2 = \underset{x \neq 0}{max} \frac{\parallel Ax \parallel_2}{\parallel x \parallel_2} \geq \frac{\parallel Av_1 \parallel_2}{\parallel v_1 \parallel_2} = \sigma_1. \tag{3.1}$$

Seja $x \in \mathbb{R}^n$. Então x pode ser escrito como combinação linear dos vetores singulares à direita de A, $x = c_1v_1 + \cdots + c_nv_n$. Como os vetores $v_1, \ldots, v_n$ são ortonormais,

$$\| x \|_2^2 = \| c_1v_1 + \cdots + c_nv_n \|_2^2 = \| c_1v_1 \|_2^2 + \cdots + \| c_nv_n \|_2^2 = | c_1 |^2 + \cdots + | c_n |^2 .$$

Calculando Ax obtemos

$$Ax = c_1Av_1 + \cdots + c_rAv_r + c_{r+1}Av_{r+1} + \cdots + c_nAv_n = c_1\sigma_1u_1 + \cdots + c_r\sigma_ru_r + 0 + \cdots + 0,$$

em que r é o posto da matriz A. Como $u_1, \ldots, u_r$ são ortonormais então

$$\| Ax \|_2^2 = | \sigma_1c_1 |^2 + \cdots + | \sigma_rc_r |^2 .$$

Assim

$$\| Ax \|_2^2 \leq \sigma_1^2 \left(| c_1 |^2 + \cdots + | c_n |^2 \right) \leq \sigma_1^2 \| x \|_2^2,$$

o que implica que

$$\| A \|_2 = \max_{x \neq 0} \frac{\| Ax \|_2}{\| x \|_2} \leq \sigma_1. \tag{3.2}$$

Da Equação 3.1 e Equação 3.2 temos $\| A \|_2 = \sigma_1$. □

Teorema 3.3.2. *Seja A uma matriz de ordem $n \times n$ não singular e $\sigma_1 \geq \sigma_2 \geq \ldots, \geq \sigma_n > 0$. Então $cond_2(A) = \frac{\sigma_1}{\sigma_n}$.*

Demonstração. Da decomposição em valores singulares de A e do fato de A ser não singular temos que os valores singulares de A são não nulos, isto é, $\sigma_i > 0$, $i = 1, 2 \ldots, n$, pois caso contrário A teria posto incompleto e não seria invertível.

Se $A = U\Sigma V^T$ então $A^{-1} = \left(U\Sigma V^T\right)^{-1} = V\Sigma^{-1}U^T$. Assim os valores singulares de A^{-1} são $\frac{1}{\sigma_n} \geq \cdots \geq \frac{1}{\sigma_1}$, logo $\| A^{-1} \|_2 = \frac{1}{\sigma_n}$. Portanto

$$cond_2(A) = \| A \|_2 \| A^{-1} \|_2 = \sigma_1 \cdot \frac{1}{\sigma_n} = \frac{\sigma_1}{\sigma_n}.$$

□

Teorema 3.3.3. *Seja A uma matriz $m \times n$ e sejam $\sigma_1, \sigma_2, \ldots, \sigma_r$ todos os valores singulares não nulos de A. Então $\| A \|_F^2 = \sigma_1^2 + \cdots + \sigma_r^2$.*

Demonstração. Lembre-se que se A é uma matriz de ordem $m \times n$ e Q é uma matriz ortogonal de ordem $n \times n$, então $\| QA \|_F = \| A \|_F$. Dada a decomposição em valores singulares de A, $A = U\Sigma V^T$, temos

$$\| A \|_F^2 = \| U\Sigma V^T \|_F^2 = \| \Sigma V^T \|_F^2 = \| \left(\Sigma V^T\right)^T \|_F^2 = \| V\Sigma^T \|_F^2 = \| \Sigma^T \|_F^2 = \sigma_1^2 + \cdots + \sigma_r^2.$$

□

Geralmente calculamos o posto de uma matriz por redução de linhas para a forma escalonada e contando o número de linhas não nulas. Entretanto, erros de arredondamento podem afetar esse processo, especialmente se a matriz for malcondicionada. Elementos que deveriam ser nulos podem acabar sendo números muito pequenos, interferindo em nossa capacidade de determinar com precisão o posto de uma matriz.

Na prática a decomposição por valores singulares é frequentemente utilizada para encontrar um posto de uma matriz, já que ela é muito mais confiável quando erros de arredondamento estão presentes. A ideia básica por trás dessa abordagem é que as matrizes ortogonais U e V da SVD preservam o comprimento e, portanto, não introduzem nenhum erro adicional, qualquer erro que ocorrer tenderá a aparecer na matriz Σ.

Teorema 3.3.4. *Seja $A = U\Sigma V^T$ uma decomposição em valores singulares da matriz A de ordem $m \times n$. Sejam $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$ todos os valores singulares não nulos de A. Então o posto de A é igual ao número de valores singulares não nulos, isto é, $posto(A) = r$.*

Demonstração. Como $A = U\Sigma V^T$ temos $posto(A) = posto\left(U\Sigma V^T\right)$. Afirmamos que se U é invertível, então $posto(UA) = posto(A)$ e $posto(AU) = posto(A)$. Com efeito, para provar que $posto(UA) = posto(A)$ devemos mostrar que $N(UA) = N(A)$, em que A é uma matriz $m \times n$ e U é uma matriz $n \times n$ invertível, ou seja, devemos mostrar que

$$N(A) \subset N(UA) \text{ e } N(UA) \subset N(A).$$

Mostremos inicialmente que $N(A) \subset N(UA)$. Considerando $x \in N(A)$, segue que $Ax = 0$. Logo $U(Ax) = 0$, o que implica que $(UA)x = 0$. Portanto $x \in N(UA)$. Mostremos agora que $N(UA) \subset N(A)$. Para isso considere $x \in N(UA)$, então $(UA)x = 0$, logo $U(Ax) = 0$ o que implica que $Ax \in N(U)$. Como U é invertível $N(U) = 0$, então $Ax = 0$, logo $x \in N(A)$. Portanto $N(UA) = N(A)$.

Pelo Teorema do Posto,

$$dim(N(UA)) = n - posto(UA) \text{ e } dim(N(A)) = n - posto(A).$$

Assim $n - posto(UA) = n - posto(A)$, portanto $posto(UA) = posto(A)$. Analogamente mostra-se que $posto(AU) = posto(A)$. Usando a afirmação temos,

$$posto(A) = posto\left(U\Sigma V^T\right) = posto\left(\Sigma V^T\right) = posto(\Sigma) = r.$$

$\square$

Exemplo 3.3.5. *Sejam as matrizes*

$$A = \begin{bmatrix} 8,1650 & -0,0041 & -0,0041 \\ 4,0825 & -3,9960 & 4,0042 \\ 4,0825 & 4,0042 & -3,9960 \end{bmatrix} \quad e \quad B = \begin{bmatrix} 8,17 & 0 & 0 \\ 4,08 & -4 & 4 \\ 4,08 & 4 & -4 \end{bmatrix}.$$

Note que a matriz B foi obtida ao arredondarmos os valores dos elementos de A para a segunda casa decimal. Calculando os postos dessas matrizes aparentemente iguais, encontramos posto(A) = 3 e posto(B) = 2. De fato, se calcularmos a SVD de A e de B pelo software Mathematica, por exemplo, obtemos

$$\Sigma_A = \begin{bmatrix} 10 & 0 & 0 \\ 0 & 8 & 0 \\ 0 & 0 & 0,01 \end{bmatrix} \quad e \quad \Sigma_B = \begin{bmatrix} 10 & 0 & 0 \\ 0 & 8 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Desta forma os valores singulares de A são 10, 8 e 0,01, que implica no posto(A) = 3. Os valores singulares de B são 10, 8 e 0, portanto B tem posto 2. Isso implica que a matriz A é invertível, mas B não. A explicação para essas diferenças críticas entre duas matrizes aproximadamente iguais são retratadas pelas suas SVDs.

4 Aplicações

Nesse capítulo vamos tratar de uma das mais impressionantes aplicações da decomposição em valores singulares, que é seu uso em comprimir imagens digitais para que possam ser transmitidas eletronicamente de forma eficiente. A transmissão e o armazenamento de grandes quantidades de dados digitalizados têm se tornado um dos maiores problemas de nosso mundo tecnológico, daí vem a importância de discutir o papel da decomposição em valores singulares na compressão de dados digitalizados que possam ser transmitidos mais rapidamente ocupando menos espaço de armazenamento. Além disso, vamos mostrar como utilizar essa decomposição para encontrar a solução ótima de um sistema linear através da pseudo-inversa (TREFETHEN; III, 1997; MULLER; MAGAIA; HERBST, 2004; WATKINS, 2010; POOLE, 2011; ANTON; RORRES, 2012; OLVER; SHAKIBAN; SHAKIBAN, 2018; PEREIRA; MENEGUETTEJUNIOR; BOTTA, 2020).

4.1 Pseudo-Inversa

A pseudo-inversa foi definida pela primeira vez pelo matemático americano Eliakim Moore na década de 1920 e redescoberto pelo físico matemático britânico Sir Roger Penrose na década de 1950. Por conta disso a pseudo-inversa também é conhecida como a inversa de Moore-Penrose (ANTON; RORRES, 2012; OLVER; SHAKIBAN; SHAKIBAN, 2018; PELLEGRINI, 2020).

A noção de pseudo-inversa e o estudo de suas propriedades nos auxiliam a responder um questionamento bastante natural com respeito ao sistema linear $Ax = b$. Qual é ou quais são os valores de x tais que o erro $r = \|b - Ax\|_2$ é o menor possível e qual entre esses vetores x é a solução ótima, quer dizer, tem norma mínima ? Pelo que foi descrito, esse, na realidade, é um problema de quadrados mínimos aplicado à resolução de sistemas lineares. No caso em que a matriz A quadrada possui colunas linearmente independentes, A tem posto completo e assim A é invertível, logo a única solução do sistema $Ax = b$ é dada por $x = A^{-1}b$. Agora se a matriz A não se encaixa nesse perfíl, ou seja, se $m > n$ e A é uma matriz $m \times n$ com colunas linearmente independentes, então o sistema $Ax = b$ não possui nenhuma solução exata. Isto é, podemos nos deparar com um sistema indeterminado ou com um sistema inconsistente. Nestas situações, temos duas alternativas; se o sistema é indeterminado, temos infinitas soluções e devemos encontrar aquela com a menor norma possível; se o sistema é inconsistente, não há solução, e neste caso devemos encontrar o vetor x para o qual a distância $\|b - Ax\|_2$ seja a menor possível.

Há diversos métodos para obter a pseudoinversa de uma matriz, por exemplo: a decomposição em posto completo, a decomposição por blocos, entre outros que não serão tratados aqui porquê a decomposição em valores singulares resolve esse problema para qualquer matriz.

Sabemos que nem toda matriz quadrada tem inversa. Ao generalizar o conceito de matriz inversa veremos que toda matriz, mesmo que seja singular ou que não seja quadrada possui o que vamos definir a seguir de **pseudo-inversa** em termos da decomposição em valores singulares.

Definição 4.1.1. *A pseudo-inversa de uma matriz A de ordem $m \times n$ com decomposição em valores singulares $A = U\Sigma V^T$ é a matriz A^+ de ordem $n \times m$ dada por $A^+ = V\Sigma^{-1}U^T$.*

Observe que a última equação da definição corresponde à SVD de A^+ e, portanto, seus valores singulares não nulos são os recíprocos dos valores singulares não nulos de A. Além disso, se A for uma matriz quadrada não singular, então sua pseudo-inversa A^+ coincide com sua inversa ordinária A^{-1}. De fato, lembrando que a inversa de uma matriz ortogonal é igual a sua transposta temos

$$A^{-1} = \left(U\Sigma V^T\right)^{-1} = \left(V^T\right)^{-1}\Sigma^{-1}U^{-1} = V\Sigma^{-1}U^T = A^+.$$

De forma mais geral, se a matriz A tiver colunas linearmente independentes, ou seja, tiver posto completo n, não é necessário usar a SVD para calcular a pseudo-inversa como mostra o lema a seguir.

Lema 4.1.2. *Se A é uma matriz de ordem $m \times n$ de posto n, então $A^+ = \left(A^TA\right)^{-1}A^T$.*

Demonstração. Inciamos por substituir A por sua SVD

$$A^TA = \left(U\Sigma V^T\right)^T\left(U\Sigma V^T\right) = V\Sigma^TU^TU\Sigma V^T = V\Sigma^T\Sigma V^T = V\Sigma^2V^T,$$

já que $\Sigma^T = \Sigma$ e $U^TU = I$. Agora se A tem posto completo n, então V é uma matriz ortogonal de ordem $n \times n$, e assim $V^{-1} = V^T$. Portanto,

$$\begin{aligned}\left(A^TA\right)^{-1}A^T &= \left(V\Sigma^2V^T\right)^{-1}\left(U\Sigma V^T\right)^T\\ &= V\Sigma^{-2}V^{-1}V\Sigma^TU^T\\ &= V\Sigma^{-2}\Sigma U^T = V\Sigma^{-1}U^T = A^+,\end{aligned}$$

o que demonstra o Lema. □

Exemplo 4.1.3. *Considerando a matriz $A = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}$ note que suas colunas são linearmente independentes, então podemos usar o lema anterior para calcular sua pseudo-inversa. Assim, vamos calcular A^+ iniciando com o cálculo da inversa de A^TA:*

$$A^TA = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{bmatrix} \cdot \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} = \begin{bmatrix} 3 & 6 \\ 6 & 14 \end{bmatrix}.$$

Logo,

$$\left(A^TA\right)^{-1} = \frac{1}{6}\begin{bmatrix} 14 & -6 \\ -6 & 3 \end{bmatrix} = \begin{bmatrix} \frac{7}{3} & -1 \\ -1 & \frac{1}{2} \end{bmatrix}.$$

Portanto, a pseudo-inversa de A é

$$A^+ = \left(A^TA\right)^{-1}A^T = \begin{bmatrix} \frac{7}{3} & -1 \\ -1 & \frac{1}{2} \end{bmatrix}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{bmatrix} = \begin{bmatrix} \frac{4}{3} & \frac{1}{3} & -\frac{2}{3} \\ -\frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix}.$$

Teorema 4.1.4. *O problema de quadrados mínimos $Ax = b$ possui uma única solução por quadrados mínimos $\bar{x}$ de comprimento mínimo, que é dada por $x^* = A^+b$*

Demonstração. Seja uma matriz A uma matriz de ordem $m \times n$ com posto r e decomposição em valores singulares $A = U\Sigma V^T$ de modo que $A^+ = V\Sigma^{-1}U^T$. Escrevamos $y = V^Tx$ e $c = U^Tb$, os quais podem ser escritos em forma matricial como $y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$ e $c = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix}$, em que y_1 e c_1 estão em $\mathbb{R}^r$.

Agora nosso objetivo é minimizar $\| b - Ax \|$ ou equivalentemente $\| b - Ax \|^2$. Como U^T é ortogonal temos

$$\| b - Ax \|^2 = \| U^T(b - Ax) \|^2 = \| U^Tb - U^TU\Sigma V^Tx \|^2 = \| c - \Sigma y \|^2 .$$

Assim,

$$\| c - \Sigma y \|^2 = \left\| \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} - \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \right\|^2 = \left\| \begin{bmatrix} c_1 - Dy_1 \\ c_2 \end{bmatrix} \right\|^2.$$

O valor mínimo ocorre quando $c_1 - Dy_1 = 0$ ou, equivalentemente, $y_1 = D^{-1}c_1$. Com isso todas as soluções por quadrados mínimos são da forma $x = Vy$. Agora vamos definir

$$x^* = Vy^* = V\begin{bmatrix} D^{-1}c_1 \\ 0 \end{bmatrix}.$$

Afirmamos que x^* é a solução por quadrados mínimos de menor comprimento. De fato, supondo que

$$\bar{x} = V\bar{y} = V\begin{bmatrix} D^{-1}c_1 \\ y_2 \end{bmatrix}$$

seja uma solução por quadrados mínimos diferente, ou seja, $y_2 \neq 0$. Então,

$$\| x^* \| = \| Vy^* \| = \| y^* \| < \| \bar{y} \| = \| V\bar{y} \| = \| \bar{x} \|$$

Desta forma,

$$x^* = Vy^* = V\begin{bmatrix} D^{-1}c_1 \\ 0 \end{bmatrix} = V\begin{bmatrix} D^{-1} & 0 \\ 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = V\Sigma^+U^Tb = A^+b,$$

o que demonstra o nosso teorema. □

Exemplo 4.1.5. *Vamos encontrar a solução por quadrados mínimos de comprimento mínimo do sistema linear* $Ax = b$ *, em que*

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

cujas equações correspondentes são

$$\begin{cases} x + 2y = 3 \\ 2x + 4y = 5. \end{cases}$$

Observe que elas são claramente inconsistentes. Então não é possível usar o Lema 4.1.2 para encontrar a pseudo-inversa já que a matriz A *não é invertível. Assim, também não é possível encontrar a solução do sistema com esse mesmo artifício. Neste caso, a solução por quadrados mínimos é a que resolve o problema. Além disso, as colunas de* A *são linearmente denpendentes, por isso existirão infinitas soluções por quadrados mínimos, dentre as quais queremos aquela de comprimento mínimo. A decomposição em valores singulares de* A *resolverá esse problema.*

Inicialmente, o polinômio característico de A *é dado por*

$$p(\lambda) = \begin{vmatrix} 1-\lambda & 2 \\ 2 & 4-\lambda \end{vmatrix} = \lambda^2 - 5\lambda,$$

cujas raízes são $\lambda_1 = 5$ *e* $\lambda_2 = 0$*. Como a matriz* A *é simétrica, os valores singulares de* A *são os valores absolutos dos autovalores de* A*, logo*

$$\Sigma = \begin{bmatrix} 5 & 0 \\ 0 & 0 \end{bmatrix}.$$

Agora, para construirmos a matriz V *vamos calcular os autovetores unitários correspondentes.*

Os autovetores associados ao autovalor $\lambda_1 = 5$ *são encontrados por*

$$\begin{aligned} Av = \lambda v &\iff \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 5 \begin{bmatrix} x \\ y \end{bmatrix} \\ &\iff \begin{cases} x + 2y = 5x \\ 2x + 4y = 5y \end{cases} \\ &\iff 2x - y = 0 \\ &\iff y = 2x. \end{aligned}$$

Assim, o autoespaço associado a $\lambda_1 = 5$ *é formado por vetores do tipo* $v = (x, 2x) = x(1, 2)$*. Então, neste caso, o autovetor unitário é* $v_1 = \left(\frac{1}{\sqrt{5}}, \frac{2}{\sqrt{5}}\right)$*.*

Para encontrarmos os autovetores associados ao autovalor $\lambda_2 = 0$ procedemos da seguinte forma:

$$\begin{aligned} Av = \lambda v &\iff \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 0 \begin{bmatrix} x \\ y \end{bmatrix} \\ &\iff \begin{cases} x + 2y = 0 \\ 2x + 4y = 0 \end{cases} \\ &\iff x + 2y = 0 \\ &\iff x = -2y. \end{aligned}$$

Assim, o autoespaço associado a $\lambda_1 = 0$ é formado por vetores do tipo

$$v = (2y, -y) = y(2, -1).$$

Então, neste caso, o autovetor unitário é $v_2 = \left(\frac{2}{\sqrt{5}}, -\frac{1}{\sqrt{5}}\right)$. Portanto a matriz V é dada por

$$V = \begin{bmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}} \end{bmatrix}$$

que é igual a sua transposta já que V é simétrica.

Para encontrar a matriz U calculamos

$$u_1 = \frac{1}{\sigma_1} \cdot Av_1 = \frac{1}{5} \cdot \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \cdot \begin{bmatrix} \frac{1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \end{bmatrix}$$

Precisamos estender $\{u_1\}$ para uma base ortonormal $\{u_1, u_2\}$ do $\mathbb{R}^2$. Para simplificar os cálculos vamos nos livrar da inconveniente raíz mutiplicando o vetor u_1 por um escalar adequado, digamos $\sqrt{5}u_1 = (1, 2)$. Agora, satisfazendo a ortogonalidade de u_1 e u_2 temos

$$\begin{bmatrix} 1 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 0 \iff x + 2y = 0 \iff x = -2y.$$

Então, o vetor u_2 é um vetor do tipo $(2x, -y) = y\,(2, -1)$, assim $u_2 = \left(\frac{2}{\sqrt{2}}, -\frac{1}{\sqrt{5}}\right)$. Portanto, a matriz U é dada por

$$U = \begin{bmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}} \end{bmatrix},$$

que pela mesma razão de V é igual a sua transposta.

Em posse das matrizes U, Σ e V, podemos escrever a pseudo-inversa de A:

$$A^{+} = V\Sigma^{-1}U^{T} = \begin{bmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}} \end{bmatrix} \begin{bmatrix} \frac{1}{5} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}} \end{bmatrix} = \begin{bmatrix} \frac{1}{25} & \frac{2}{25} \\ \frac{2}{25} & \frac{4}{25} \end{bmatrix}.$$

Portanto,

$$\bar{x} = A^{+}b = \begin{bmatrix} \frac{1}{25} & \frac{2}{25} \\ \frac{2}{25} & \frac{4}{25} \end{bmatrix} \begin{bmatrix} 3 \\ 5 \end{bmatrix} = \begin{bmatrix} \frac{13}{25} \\ \frac{26}{25} \end{bmatrix}.$$

Seguindo as equivalências (Equação 2.5), para verificar se a solução por quadrados mínimos está correta, basta verificar se ela satisfaz as equações normais $A^T A\bar{x} = A^T b$. De fato,

$$A^T A\bar{x} = \begin{bmatrix} 5 & 10 \\ 10 & 20 \end{bmatrix} \begin{bmatrix} \frac{13}{25} \\ \frac{26}{25} \end{bmatrix} = \begin{bmatrix} 13 \\ 26 \end{bmatrix} \quad e \quad A^T b = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} 3 \\ 5 \end{bmatrix} = \begin{bmatrix} 13 \\ 26 \end{bmatrix}.$$

Isto implica que a solução encontrada por quadrados mínimos é a solução de menor comprimento, ou seja, é a solução mais próxima da origem.

Algoritmo 1: *Solução por mínimos quadrados de comprimento mínimo*

Entrada: matrix A mxn, matrix B mx1

Saída: matrix solução nx1

início

Atribuir valores para a matriz A mxn;

Atribuir valores para a matriz B mx1;

Atribuir x e y;

Calcular a pseudo-inversa A^+ de A;

Calcular a matriz solução X: n x 1 por meio do produto entre as matrizes A^+ e B;

Calcular a equivalência com a Equação 2.5;

fim

O Algoritmo 1 descreve o exemplo 4.1.5.

4.2 Compressão de Imagem Digital

Dentre muitas aplicações, a decomposição em valores singulares (SVD) é uma das mais impressionantes. Ela pode ser utilizada para comprimir imagens digitais com o objetivo de reduzir seu espaço de armazenamento e acelerar sua transmissão eletrônica com eficiência, por satélite, internet ou semelhante, uma vez que a transmissão e o armazenamento eficientes de grandes quantidades de dados digitalizados têm se tornado um dos maiores problemas do nosso mundo tecnológico.

Como vimos no capítulo anterior, para cada matriz $A \in \mathbb{R}^{m\times n}$, existem matrizes ortogonais $U = \begin{bmatrix} u_1 & u_2 & \cdots & u_m \end{bmatrix}$, com $u_i \in \mathbb{R}^m$, $i = 1, \ldots, m$, $V^T = \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix}$, com

$v_j \in \mathbb{R}^n$, $j = 1, \ldots, m$, e a matriz diagonal $\Sigma = diag\{\sigma_1, \sigma_2, \ldots, \sigma_r\} \in \mathbb{R}^{r \times r}$, em que σ_i são os valores singulares de A satisfazendo $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$ e, além disso, $r = min\{m, n\}$.

O problema que agora vamos considerar está relacionado com reduzir o total de informações a serem transmitidas, obviamente com perda considerável de informações a respeito de uma determinada figura, mas sem perder nenhuma informação essencial. O primeiro passo para isso é representá-la como uma matriz numérica, a partir da qual a imagem possa ser restaurada quando necessário. Por exemplo, uma fotografia em preto e branco pode ser escaneada como um arranjo retangular de pixels (abreviação da expressão em inglês picture elements) e armazendada como uma matriz A. A fim de medir a perda de dados durante a implementação da SVD para restauração de imagens vamos considerar o seguinte teorema.

Teorema 4.2.1 (Melhor Aproximação de Posto Baixo). *Seja a matriz* $A \in \mathbb{R}^{n \times m}$ *com* $posto\,(A) = r > 0$ *e sua SVD,* $A = U\Sigma V^T$ *com valores singulares* $\sigma_1 \geq \sigma_2 \geq \ldots, \geq \sigma_r > 0$. *Para* $k = 1, \ldots, r-1$ *defina* $A_k = U\Sigma_k V^T$, *em que* $\Sigma_k = diag\,(\sigma_1, \ldots, , \sigma_k, 0, \ldots, 0)$. *Então,* $posto\,(A_k) = k$ *e* $\| A - A_k \|_2 = min\,\{\| A - B \|_2;\ posto\,(B) \leq k\} = \sigma_{k+1}$. *Ou seja, de todas as matrizes de posto menor ou igual a k,* A_k *é a mais próxima de A.*

Demonstração. Pela SVD da matriz $A_k = U\Sigma_k V^T$, concluímos que $posto\,(A_k) = k$. Como

$$A - A_k = U\Sigma V^T - U\Sigma_k V^T = U\,(\Sigma - \Sigma_k)\,V^T$$

é evidente que o maior valor singular de $A - A_k$ é σ_{k+1}. Portanto,

$$\| A - A_k \|_2 = \| U\,(\Sigma - \Sigma_k)\,V^T \|_2 = \| \Sigma - \Sigma_k \|_2 = \sigma_{k+1}.$$

Mostremos agora que, para qualquer outra matriz B com $posto\,(B) \leq k$, temos $\| A - B \|_2 \geq \sigma_{k+1}$. De fato, considerando uma tal matriz B, temos que $dim\,(Im(B)) \leq k$, o que implica, pelo Teorema do Núcleo e da Imagem, que a dimensão do núcleo de B é dada por $dim\,(N\,(B)) \geq m - k$. O subespaço $S = span\,(v_1, \ldots, v_{k+1})$ tem dimensão $k+1$, lembrando que os vetores $v_1, \ldots, v_m$ representam as colunas da matriz V. Como $N\,(B)$ e S são subespaços do $\mathbb{R}^m$, a soma de suas dimenções excede m. De fato,

$$dim\,(N\,(B)) + dim\,(S) \geq (m - k) + (k + 1) = m + 1.$$

Então $N\,(B) \cap S \neq \{0\}$. Seja $w \in N\,(B) \cap S$ e suponha que $\| w \|_2 = 1$. Como $w \in N\,(B)$ então $Bw = 0$, e pelo fato de w pertencer a S, o vetor w pode ser expresso como combinação linear dos vetores de S, isto é,

$$w = c_1 v_1 + \ldots + c_{k+1} v_{k+1}.$$

Pela ortonormalidade de S temos,

$$| c_1 |^2 + \cdots + | c_{k+1} |^2 = \| w \|_2^2 = 1.$$

Assim,

$$(A - B)\, w = Aw - Bw = Aw = \sum_{i=1}^{k+1} c_i A v_i = \sum_{i=1}^{k+1} c_i \sigma_i u_i.$$

Como $u_1, \ldots, u_{k+1}$ também são ortonormais,

$$\| (A - B)\, w \|_2^2 = \sum_{i=1}^{k+1} \mid \sigma_i c_i \mid^2 \geq \sigma_{k+1}^2 \sum_{i=1}^{k+1} \mid c_i \mid^2 = \sigma_{k+1}^2.$$

Portanto,

$$\|A - B \|_2 \geq \frac{\| (A - B)\, w \|_2}{\|w \|_2} \geq \sigma_{k+1}.$$

Concluímos assim que qualquer matriz de posto menor ou igual a k está a uma distância maior ou igual à σ_{k+1} de A, e, também, que A_k é uma matriz de posto k que está o mais próximo possível de A. □

Desta forma, para qualquer matriz A_k, $\|A - A_k \|_2$ ou $\|A - A_k \|_F$ é a medida para a perda de dados no processamento de compressão de imagens. Segue do Teorema 4.2.1 que a qualidade da restauração de imagens pode ser estimada pela equação $\|A - A_k \|_2 = \sigma_{k+1}$ se a imagem restaurada tiver posto k.

O princípio fundamental do esquema de **compressão de imagem** baseado em SVD é selecionar um número apropriado de valores singulares para representar a imagem original. O primeiro passo é representar uma determinada imagem como uma matriz numérica de acordo com seus tons de cinza. Neste caso, dispomos de 256 tons de cinza, em que 0 corresponde ao branco e 255 ao preto. Assim as entradas da matriz são números inteiros entre 0 e 255, conforme ilustrado na seguinte figura.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1		15	30	45	60	75	90	105	120	135	150	165	180	195	210	225	240	255
2		15	30	45	60	75	90	105	120	135	150	165	180	195	210	225	240	255
3		15	30	45	60	75	90	105	120	135	150	165	180	195	210	225	240	255
4		15	30	45	60	75	90	105	120	135	150	165	180	195	210	225	240	255
5		15	30	45	60	75	90	105	120	135	150	165	180	195	210	225	240	255
6		15	30	45	60	75	90	105	120	135	150	165	180	195	210	225	240	255
7		15	30	45	60	75	90	105	120	135	150	165	180	195	210	225	240	255
8		15	30	45	60	75	90	105	120	135	150	165	180	195	210	225	240	255
9		15	30	45	60	75	90	105	120	135	150	165	180	195	210	225	240	255
10		15	30	45	60	75	90	105	120	135	150	165	180	195	210	225	240	255

Figura 5 – Matriz numérica em tons de cinza, Antonello (2020, p. 7)

Assim a imagem pode ser recuperada a partir da matriz A exibindo os pixels com seus tons de cinza associados.

Considerando a matriz A de ordem $m \times n$ poderíamos armazenar suas $m \cdot n$ entradas individualmente, o que corresponde a imagem original. Como procedimento alternativo, suponha que a imagem restaurada B é contruída a partir de sua SVD reduzida, isto é,

$$B = A_k = U\Sigma_k V^T = \sigma_1 u_1 v_1^T + \sigma_{2,} u_2 v_2^T + \ldots + \sigma_k u_k v_k^T. \tag{4.1}$$

Suponha agora que $\sigma_{r+1}, \ldots, \sigma_k$ sejam suficientemente pequenos, a ponto de poderem ser ignorados no produto externo (Equação 4.1), e dessa forma produzam uma aproximação aceitável de A e da imagem que A representa:

$$A_r = \sigma_1 u_1 v_1^T + \sigma_{2,} u_2 v_2^T + \ldots + \sigma_r u_r v_r^T \tag{4.2}$$

Pelo Teorema 4.2.1, dizemos que a Equação 4.2 é a aproximação de posto r da matriz A. Como cada vetor de U tem m entradas e cada vetor de V tem n entradas, o total necessário para o armazenamento A_k são as k colunas de U, as k linhas de V, bem como os primeiros k valores singulares, ou seja,

$$k + k \cdot m + k \cdot n = k\,(1 + m + n)\,.$$

Por exemplo, supondo que temos uma figura de tamanho 340×280 pixels, lembrando que cada pixel corresponde a um dos 256 tons de cinza, que podemos representar por números entre 0 e 255. Podemos armazenar essa informação em uma matriz A de ordem 340×280, porém transmitir e manipular esses $340 \times 280 = 95.200$ números é muito caro do ponto de vista computacional. A ideia por trás da compressão de imagens é que algumas partes da figura são menos importantes que outras. Em vez de transmitir 95.200 números, precisamos enviar, por exemplo, somente 20 valores singulares mais os 20 vetores singulares correspondentes à esquerda $u_1, \ldots, u_{20}$ de $\mathbb{R}^{340}$ e os vetores singulares correspondentes à direita $v_1, \ldots, v_{20}$ de $\mathbb{R}^{280}$, o que da um total de

$$k\,(1 + m + n) = 20\,(340 + 280 + 1) = 12.420$$

números.

Comparando os números $k\,(1 + m + n)$ com as entradas individuais $m \cdot n$ obtemos a taxa

$$\frac{k \cdot (1 + m + n)}{m \cdot n} = \frac{12420}{95200} = 0,13,$$

dando uma compressão de aproximadamente 87%.

Para exemplificar o processo de aproximação de A por matrizes de posto menor, vamos usar como exemplo uma imagem do Mercado Modelo, ponto turístico famoso de Salvador, em formato `.jpg` ou `.png`, que será convertida para uma matriz.

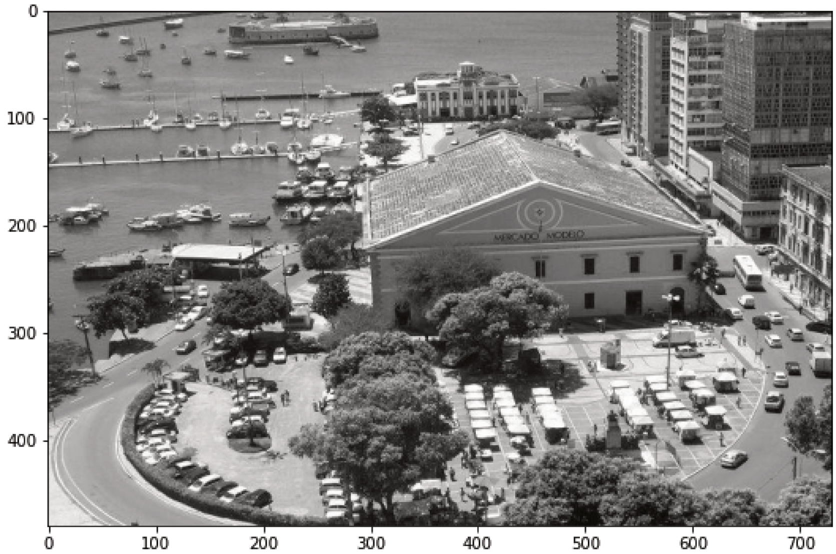

Figura 6 – Vista geral do Mercado Modelo, Google

A foto original do Mercado Modelo representada pela Figura 6 é uma imagem em 480×730 pixels. A matriz A correspondente tem dimensão $m \times n$, com elementos entre 0 e 255. A matriz A, portanto tem posto 480. Aproximando A por A_k, como descrito anteriormente, obtemos uma imagem que corresponde aos k primeiros valores singulares de A. As Figuras 7 e 8 a seguir mostram várias dessas imagens para valores de k de 5 a 60. De início a imagem está muito embassada, mas ela toma forma muito rapidamente. Ou seja, quanto maior o posto da matriz melhor será a nitidez da imagem e maior o custo computacional de sua transferência. Note que, para $n = 30$, na 8, já temos uma aproximação bem razoável da imagem original Algoritmo 2.

Algoritmo 2: *Compressão de Imagem*

Entrada: imagem, k valor singular

Saída: imagem processada

início

Carrega a imagem;

Inicializa a imagem carregada em tons de cinza;

Atribuí k valor singular;

Comprime a image;

fim

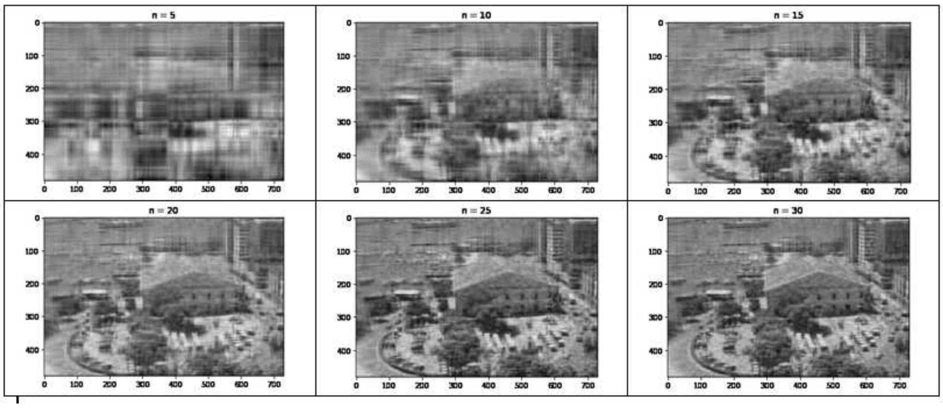

Figura 7 – Valores entre 5 e 30 para k

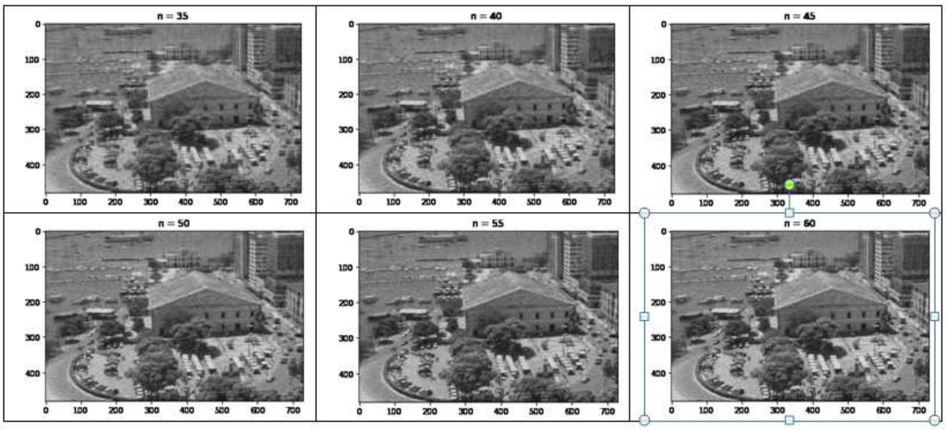

Figura 8 – Valores entre 35 e 60 para k

Assim, para $n = 30$ por exemplo, temos uma aproximação bem razoável da imagem original, porém foram perdidas várias informações sobre a imagem original, mas as informações essenciais foram conservadas permitindo fazer o processo inverso, ou seja, a partir dessa imagem representada por uma matriz de posto menor, podemos retomar a imagem original.

5 Considerações Finais

Este trabalho ressalta a relevância das fatorações de matrizes apresentadas, como por exemplo, na simplificação do cálculo de potências de matrizes quadradas diagonalizáveis. O fato de matrizes simétricas poderem ser ortogonalmente diagonalizáveis é utilizado em muitas aplicações. Em particular, a decomposição em valores singulares que é classicamente utilizada na compactação de imagens digitais, que por sua vez, reduz de modo significativo seu espaço de armazenamento a fim de acelerar sua transmissão eletrônica com eficiência. Desta forma, abre-se espaço para o estudo de outras fatorações como a LU, QR, Schur e Cholesky, por exemplo. Isso deixa claro a continuidade desse trabalho acompanhado de suas aplicações computacionais. Isso mostra também que a Álgebra Linear não deve ser tratada como um componente curricular isolado já que é instrumento de aplicações em diferentes áreas.

Este mestrado juntamente com este trabalho contribuíram de modo significativo para o meu crescimento profissional uma vez que todo o curso foi dado prestigiando o casamento perfeito entre o tradicional quadro negro e giz, e os modernos laboratórios de informática onde conhecemos vários softwares matemáticos, como o Mathematica, o Matlab e o Python entre outros. Essa junção causou um impacto determinante no meu fazer pedagógico e refletiu diretamente na qualidade das minhas aulas. Nesta caso, optamos por apresentar a compressão de imagens por meio do Python uma vez que esse software é livre e está ao alcance de todos, além de poder ser utilizado online através do Google Colab.

Atuando como professor da rede federal de ensino nos cursos integrados de Tecnologia da Informação e Eletrotécnica, e nos cursos superiores de Licenciatura em Matemática e Ciência da Computação observo que as matrizes são objetos de estudo tanto do Ensino Médio quanto do Ensino Superior, o que permite uma maior interação entre as diversas áreas do conhecimento. Nesse contexto, a evidente vocação tecnológica do nosso *Campus* torna todas as consequências desse estudo algo muito natural. Por fim, cabe mencionar que este trabalho não esgotou as possibilidades de análise referente ao tema estudado, uma vez que as aplicações envolvendo a decomposição em valores singulares são inúmeras.

Pessoalmente, o processo de elaboração deste trabalho proporcionou uma satisfação e alegria indescritíveis, porque, além de permitir um estudo aprofundado e dinâmico dos conteúdos da Álgebra Linear me permitiu também associar os mesmos às aplicações desse conteúdo utilizando a tecnologia fornecida pelos softwares matemáticos citados neste trabalho.

Esperamos que a reflexão proposta a respeito das aplicações das fatorações de matrizes contagie outros colegas incentivando-os a embarcar no fantástico mundo da matemática aplicada, mundo esse que encheu minha vida de conhecimento e alegria.

5.1 Sugestões para Trabalhos Futuros

Como sugestão para trabalhos futuros poderíamos nos aprofundar nos estudos de outras fatorações de matrizes, dentre elas, a fatoração LU, a fatoração de Cholesky e a fatoração QR. entre outras. Estudo esse obviamente acompanhado de suas aplicações, inclusive computacionais. As diversas fatorações de matrizes são temas de extrema relevância para a área da Matemática Aplicada em todos os seus níveis.

Referências

ANTON, Howard; RORRES, Chris. *Álgebra linear com aplicações*. 12. ed. [S.l.]: Bookman Porto Alegre, 2012. Tradução de Claus Ivo Doering.

ANTONELLO, Ricardo. *Introdução a Visão Computacional com Python e OpenCV*. 2020. Livros. Disponível em: <https://cv.antonello.com.br/curso-python/>. Acesso em: 22 jun. 2020.

AXLER, Sheldon. *Linear Algebra Done Right*. New York: Springer Science & Business Media, 1997.

BOLDRINI, José Luiz *et al.* *Álgebra linear*. 3. ed. [S.l.]: Editora Harbra Ltda, 1980.

CALLIOLI, Carlos Alberto; DOMINGUES, Hygino Hugueros; COSTA, Roberto Celso Fabrício. *Álgebra linear e aplicações*. 6. ed. São Paulo: Atual, 2003.

COELHO, Flávio Ulhôa; LOURENçO, Mary Lilian. *Um Curso de Álgebra Linear*. 2. ed. São Paulo: Editora da USP, 2013.

GOLUB, Gene H; LOAN, Charles F Van. *Matrix computations*. 4. ed. Baltimore: Maryland, 2013.

LEON, Steven J. *Álgebra Linear com Aplicações*. Rio de Janeiro: LTC, 2011. Tradução de Valéria de Magalhães Iorio.

LIMA, Elon Lages. *Álgebra Linear*. 8. ed. Rio de Janeiro: IMPA, 2011.

LIPSCHUTZ, Seymour; LIPSON, Marc. *Algebra Linear*: Coleção schaum. 4. ed. Porto Alegre: Bookman Editora, 2011.

MULLER, Neil; MAGAIA, Lourenço; HERBST, Ben M. Singular value decomposition, eigenfaces, and 3d reconstructions. *SIAM review*, SIAM, v. 46, n. 3, p. 518–545, 2004.

OLVER, Peter J; SHAKIBAN, Chehrzad; SHAKIBAN, Chehrzad. *Applied linear algebra*. 2. ed. [S.l.]: Springer, 2018.

PELLEGRINI, Jerônimo C. *Álgebra linear com aplicações*. 2020. Curso de Álgebra Linear. Disponível em: <https://aleph0.info/cursos/al/notas/al.pdf>. Acesso em: 10 out. 2020.

PEREIRA, Clicia G. A.; MENEGUETTEJUNIOR, Messias; BOTTA, Vanessa. *Decomposição em Valores Singulares Aplicada à Compressão de Imagens*. 2020. Artigo. Disponível em: <https://impa.br/wp-content/uploads/2017/07/31CBM-P_CGAPereira.pdf>. Acesso em: 22 jun. 2020.

POOLE, David. *Álgebra linear*. São Paulo: Cengage Learning, 2011. Trad. Martha Salerno Monteiro.

SANTOS, Nathan Moreira. *Vetores e matrizes*: Uma introdução à Álgebra linear. Rio de Janeiro: Livros Técnicos e Científicos Editora, 2007.

STRANG, Gilbert. *Álgebra Linear e suas Aplicações*. 4. ed. São Paulo: Cengage Learning, 2009.

TREFETHEN, Lloyd N; III, David Bau. *Numerical linear algebra*. Philadelphia: Siam, 1997. v. 50.

WATKINS, David S. *Fundamentals of matrix computations*. 2. ed. [S.l.]: John Wiley & Sons, 2002.

______. ______. 3. ed. [S.l.]: John Wiley & Sons, 2010.

APÊNDICE A – Código para Compressão de Imagem

O código está escrito em Python para ilustrar a aplicação da compressão de imagens digitais com a SVD.

```
import matplotlib.pyplot as plt
import numpy as np
import cv2

img = cv2.imread('mercadomodelo.jpg')
img\_gray = cv2.cvtColor(img,cv2.COLOR\_BGR2GRAY)

plt.figure(figsize=(9,6))
plt.imshow(img\_gray,cmap='gray')
h,w=img\_gray.shape
print(h , w)
U,sigma,V = np.linalg.svd(img\_gray)
reconstimg = np.matrix(U[:,:1])*np.diag(sigma[:1])*np.matrix(V[:1,:])
plt.imshow(reconstimg, cmap='gray');
for i in range(0, 65,5):
reconstimg = np.matrix(U[:,:i])*np.diag(sigma[:i])*np.matrix(V[:i,:])
plt.imshow(reconstimg, cmap='gray')
title = "n = %s" % i
plt.title(title)
plt.show()
```

O matplotlib.pyplot é o pacote para fazer gráficos.

O pacote numpy é utilizado para fazer operações matemáticas.

O pacote cv2 lida com imagens.

Carrega a imagem em tons de cinza

Mostra a imagem

Configura o tamanho da imagem

Calcula a dimensão da matriz numérica onde h é altura (linhas) e w largura (colunas)

Calcula-se a SVD da matriz numérica pelo comando np.linalg.svd.

O comando for faz um Loop para i=0,5,...,61 de 5 em 5

Fatia-se a matriz com o parâmetro i.

Adiciona o título ao gráfico e mostam-se as imagens representadas pelas matrizes de posto menor.

APÊNDICE B – Código para cálculo da Pseudo-Inversa

O código está escrito em Python para ilustrar a resolução do sistema linear do exemplo 4.1.5 por meio do cálculo da pseudo-inversa.

```
import numpy as np
A = np.array([[1,0,1],[0,1,0],[1,0,1]])
B = np.array([[3],[5]])
Amais = np.linalg.pinv(A)
X = np.matmul(Amais,B)
print(X)
```

O pacote numpy lida com as operações matemáticas.

Declara-se A uma matriz qualquer de ordem m x n, neste caso m=n=2.

Declara-se B uma matriz qualquer m x 1,neste caso m=2 e n=1.

Calcula-se a pseudo-inversa de A, chamada A mais pelo comando np.linalg.pinv.

Calcula-se a matriz solução X tal que X é uma matriz nx1 obtida pelo produto entre A mais e B por meio do comando np.matmul ou np.dot.