Universidade Federal de Santa Catarina
Centro de Ciências Físicas e Matemáticas

Decomposição em Valores Singulares e Aplicações

Trabalho de Conclusão de Curso

Thiane Pereira Poncetta Coliboro

Florianópolis, fevereiro de 2009

Thiane Pereira Poncetta Coliboro

Decomposição em Valores Singulares e Aplicações

Trabalho de Conclusão de Curso apresentado ao Curso de Matemática - Habilitação Licenciatura - do Departamento de Matemática do Centro de Ciências Físicas e Matemáticas da Universidade Federal de Santa Catarina.

Orientador: Dr. Juliano de Bem Francisco

Florianópolis, fevereiro de 2009

Esta monografia foi julgada adequada como TRABALHO DE CONCLUSÃO DE CURSO no Curso de Matemática - Habilitação Licenciatura, e aprovada em sua forma final pela Banca Examinadora designada pela Portaria nº 04/CCM/2009.

Profª. Carmem Suzane Comitre Gimenez

Professora responsável pela disciplina

Banca Examinadora:

Prof. Juliano de Bem Francisco

Orientador

Prof. Fermín Sinforiano Viloche Bazán

Profª. Virgínia Silva Rodrigues

Florianópolis, 05 fevereiro de 2009.

Agradecimentos

Aos meus pais, Mª Bernadete e Milton, por fazerem sempre o melhor por mim.

Ao meu namorado Rodrigo, pela ajuda nas traduções e leituras preliminares além do amor incondicional, apoio e imensa paciência.

Aos amigos de faculdade, Mayna Volker, Marcelo Rosa e Leonardo Peixoto, pela cumplicidade e pelos bons momentos juntos.

Aos colegas do PET Matemática, pelo apoio durante esta fase final da graduação.

E ao meu orientador, professor Juliano, pela disposição para me orientar e sugestão do tema, pelas bibliografias recomendadas e pela dedicação durante a elaboração deste trabalho.

Apresentação

Este é o Trabalho de Conclusão de Curso previsto no currículo do Curso de Matemática habilitação licenciatura da Universidade Federal de Santa Catarina, desenvolvido sob a orientação do professor Dr. Juliano de Bem Francisco. Buscou-se elaborar um trabalho visando auxiliar os estudantes que desejam conhecer a Decomposição em Valores Singulares, explorando um pouco de suas aplicações teóricas e práticas. Sendo assim, trata-se de uma introdução ao assunto àqueles que não trabalham diretamente com este tema da Álgebra Linear.

A compreensão dos conteúdos dos cursos de Álgebra Linear I e II é pré-requisito para o início desta leitura. No primeiro capítulo é feita a apresentação de algumas definições e notações que serão utilizadas em todo o decorrer do trabalho. O Capítulo 2 trata de autovalores e autovetores e de algumas propriedades e teoremas que serão utilizados nos capítulos seguintes. Alguns resultados são menos relevantes para o estudo da Decomposição em Valores Singulares mas foram inclusos no intuito de se fazer um melhor detalhamento da teoria de autovalores e autovetores.

O terceiro capítulo contém as idéias centrais desta obra. Inicia-se com uma abordagem histórica desta decomposição, citando seus precursores e quais foram suas contribuições. Ainda neste capítulo prova-se a existência da Decomposição e apresenta-se algumas considerações a respeito das matrizes envolvidas neste tipo de fatoração.

O quarto e último capítulo traz aplicações da Decomposição em Valores Singulares, por exemplo, o uso desta decomposição na obtenção de fórmulas mais simples para normas de matrizes e compressão de imagens digitais

Sumário

Introdução

Uma das idéias mais importantes na teoria de matrizes com certeza é a decomposição de matrizes. Sua utilidade tem sido apreciada há muito tempo. Mais recentemente, elas tornaram-se um dos principais tópicos da Algebra Linear Numérica, servindo como anteparo computacional para uma grande variedade de problemas práticos.

Segundo G.W. Stewart, em *On The Early History Of The Singular Value Decomposition* (1993), a Decomposição em Valores Singulares foi descoberta por Eugenio Beltrami em 1873 e independentemente por Camile Jordan, no ano seguinte. Outros matemáticos como Sylvester, Schmidt, Autonne e Weyl também contribuiram para o desenvolvimento da teoria. Consiste de uma transformação por meio de matrizes unitárias que tem como produto final uma matriz diagonal. Essa transformação provém, em suma, da multiplicação de matrizes unitárias à direita e à esquerda de uma matriz A, sendo esta de ordem qualquer. O termo "valor singular" parece ter derivado, segundo Stewart, da literatura de equações integrais.

Das muitas decomposições úteis, a Decomposição em Valores Singulares tem assumido um papel especial em virtude de inúmeras razões. Em primeiro lugar, segundo G. Strang em *Linear Algebra and Its Applications* (1988), tem-se o fato desta decomposição ser realizada com matrizes unitárias e, com isso, não modificam a norma de vetores. Segundo, ela é estável, isto é, pequenas perturbações em A correspondem a pequenas perturbações em Σ. Por último, pois hoje existem algoritmos estáveis e eficientes para calcular a decomposição em valores singulares de uma matriz.

Uma observação interessante é que a maioria das decomposições de matrizes foram descobertas antes do uso de matrizes. Elas eram descobertas por meio de determinantes, sistemas de equações lineares e, especialmente, através de formas quadráticas e bilineares. Gauss foi o percursor de todo esse desenvolvimento, escrevendo em 1823, seu famoso algoritmos de eliminação, hoje conhecido como Eliminação Gaussiana. Ele também estava perto de descobrir a inversa de uma matriz por este processo de eliminação, no qual um sistema de equações $\boldsymbol{y} = A\boldsymbol{x}$ é transformado num sistema inverso $\boldsymbol{x} = B\boldsymbol{y}$. A habilidade de Gauss manipular formas quadráticas e sistemas de equações tornou possível o tratamento de mínimos quadrados, tanto do ponto de vista teórico quanto prático. Outros desenvolvimentos se seguiram.

> Cauchy estabeleceu as propriedade de autovalores e autovetores de um problema simétrico considerando o sistema homogêneo de equações correspondentes. Em 1846, Jacobi descobriu seu famoso algoritmo para diagonalização de uma matriz simétrica, e num artigo póstumo ele obteve a decomposição LU, decompondo uma forma bilinear ao mesmo estilo de Gauss. Weiertrass estabeleceu decomposições para pares de funções bilineares - o que hoje chamamos de Problema de Autovalores Generalizado. [8]

Dessa forma, o advento da Decomposição em Valores Singulares em 1873 é visto como resultado decorrente do estudo de decomposições.

No entanto, mesmo com essa linha de trabalhos desenvolvidos e que culminaram com Beltrami e Jordan, de acordo com L. N. Trefethen em *Numerical Linear Algebra* (1997), "a decomposição em valores singulares não se tornou conhecida na Matemática até a década de 60, quando Golub e outros matemáticos mostraram que poderia ser calculada eficazmente e usada como base para muitos algoritmos estáveis". O fato da decomposição em valores singulares poder ser aplicada para matrizes de qualquer ordem, inclusive complexas, faz essa decomposição muito especial, além de sua enorme aplicabilidade em problemas reais.

Capítulo 1

Noções Preliminares

Este capítulo introdutório apresenta algumas definições e resultados da Álgebra Linear que serão necessários no decorrer deste trabalho, estabelecendo também neste momento algumas notações que serão utilizadas. Será considerado que noções como vetores, matrizes, espaço vetorial e dimensão de um espaço são conhecidas pelo leitor.

Neste trabalho, os espaços vetoriais considerados serão o $\mathbb{C}^n$ e $\mathbb{C}^{m\times n}$, e em particular $\mathbb{R}^n$ e $\mathbb{R}^{m\times n}$.

1. Dado V espaço vetorial e $\boldsymbol{x} \in V$, denotaremos $\boldsymbol{x}$ por
$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}.$$
2. $\boldsymbol{x}^T = [x_1\ x_2\ \cdots\ x_n]$ é o vetor transposto de $\boldsymbol{x}$.
3. $\mathbb{R}^n = \{\boldsymbol{x}^T = [x_1\ x_2\ \cdots\ x_n]; x_j \in \mathbb{R}, j \in \{1, \ldots, n\}\}$.
4. $\mathbb{C}^n = \{\boldsymbol{x}^T = [x_1\ x_2\ \cdots\ x_n]; x_j \in \mathbb{C}, j \in \{1, \ldots, n\}\}$, onde cada x_j é da forma $x_j = a_j + b_j i$, em que a_j, $b_j \in \mathbb{R}$ e $i = \sqrt{-1}$.

Definição 1.1 (Produto interno) *Um produto interno ou produto escalar em um espaço vetorial complexo V é uma operação que associa a cada par de vetores $\boldsymbol{x}$ e $\boldsymbol{y}$ em V um número complexo $\langle \boldsymbol{x}, \boldsymbol{y} \rangle$ tal que as seguintes propriedades são válidas:*

(i) $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \overline{\langle \boldsymbol{y}, \boldsymbol{x} \rangle}$, $\boldsymbol{x}, \boldsymbol{y} \in V$.

(ii) $\langle \boldsymbol{x}+\boldsymbol{y}, \boldsymbol{z}\rangle = \langle \boldsymbol{x}, \boldsymbol{z}\rangle + \langle \boldsymbol{y}, \boldsymbol{z}\rangle$, $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in V$ e $c \in \mathbb{C}$.

(iii) $\langle c\boldsymbol{x}, \boldsymbol{y}\rangle = c\langle \boldsymbol{x}, \boldsymbol{y}\rangle$, $\boldsymbol{x}, \boldsymbol{y} \in V$.

(iv) $\langle \boldsymbol{x}, \boldsymbol{x}\rangle \geq 0$ e $\langle \boldsymbol{x}, \boldsymbol{x}\rangle = 0$ se, e somente se, $\boldsymbol{x} = 0$, $\forall \boldsymbol{x} \in V$.

Assim, o produto interno usual de $\mathbb{R}^n$ é definido como $\langle \boldsymbol{x}, \boldsymbol{y}\rangle = \boldsymbol{y}^T\boldsymbol{x} = \boldsymbol{x}^T\boldsymbol{y} = x_1y_1 + x_2y_2 + \ldots + x_ny_n$ e produto interno usual de $\mathbb{C}^n$ é definido como $\langle \boldsymbol{x}, \boldsymbol{y}\rangle = \overline{\boldsymbol{y}}^T\boldsymbol{x} = x_1\overline{y_1} + x_2\overline{y_2} + \ldots + x_n\overline{y_n}$.

Definição 1.2 (Norma) *Uma norma em um espaço vetorial V é uma aplicação que associa a cada vetor $\boldsymbol{x}$ um número real $\|\boldsymbol{x}\|$, de modo que as seguintes propriedades sejam satisfeitas para todos os vetores $\boldsymbol{x}$ e $\boldsymbol{y}$:*

(i) $\|\boldsymbol{x}\| \geq 0$ e $\|\boldsymbol{x}\| = 0$ se, e somente se, $\boldsymbol{x} = \boldsymbol{0}$.

(ii) $\|c\boldsymbol{x}\| = |c|\,\|\boldsymbol{x}\|$, $c \in \mathbb{C}$.

(iii) $\|\boldsymbol{x}+\boldsymbol{y}\| \leq \|\boldsymbol{x}\| + \|\boldsymbol{y}\|$.

Definição 1.3 (Norma euclidiana) *A norma euclidiana de um vetor $\boldsymbol{x} \in \mathbb{R}^n$ é o número não negativo $\|\boldsymbol{x}\|$ definido por $\|\boldsymbol{x}\| = \sqrt{\langle \boldsymbol{x}, \boldsymbol{x}\rangle} = \sqrt{\boldsymbol{x}^T\boldsymbol{x}}$. Assim $\|\boldsymbol{x}\| = \sqrt{x_1^2 + {x_2}^2 + \ldots + {x_n}^2}$. Se $\boldsymbol{x} \in \mathbb{C}^n$ então $\|\boldsymbol{x}\| = \sqrt{\overline{\boldsymbol{x}}^T\boldsymbol{x}}$.*

Definição 1.4 (Vetores ortogonais) *Seja V é um espaço vetorial munido de produto interno. Diz-se que os vetores $\boldsymbol{x}$ e $\boldsymbol{y}$ em V são ortogonais se $\langle \boldsymbol{x}, \boldsymbol{y}\rangle = 0$.*

Observação 1.1 *Se além de $\langle \boldsymbol{x}, \boldsymbol{y}\rangle = 0$ ocorrer que $\|\boldsymbol{x}\| = \|\boldsymbol{y}\| = 1$ então $\boldsymbol{x}$ e $\boldsymbol{y}$ são denominados vetores ortonormais.*

Definição 1.5 (Conjunto ortonormal) *Se V é um espaço vetorial então um conjunto de vetores $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n\} \in V$ é ortonormal se $\langle \boldsymbol{x}_i, \boldsymbol{x}_j\rangle = 0$ sempre que $i \neq j$ e $\|\boldsymbol{x}_i\| = 1, \forall i, j \in \{1, 2, \ldots, n\}$.*

Definição 1.6 (Combinação Linear) *Um vetor $\boldsymbol{x}$ é uma combinação linear de vetores $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k$ se existirem escalares $c_1, c_2, \ldots, c_k$, tais que $\boldsymbol{x} = c_1\boldsymbol{x}_1 + c_2\boldsymbol{x}_2 + \ldots + c_k\boldsymbol{x}_k$. Os escalares $c_1, c_2, \ldots, c_k$ são chamados coeficientes da combinação linear.*

Definição 1.7 (Conjunto Linearmente Dependente) *Um conjunto de vetores*

$$\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k\} \subseteq V$$

é linearmente dependente (LD) quando existem escalares $c_1, c_2, \ldots, c_k$, pelo menos um dos quais não-nulo, tais que $c_1\boldsymbol{x}_1 + c_2\boldsymbol{x}_2 + \ldots + c_k\boldsymbol{x}_k = \mathbf{0}$. Se $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k\}$ não é LD então diz-se que é linearmente independentes (LI).

Os vetores $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k\}$ são LD se, e somente se, pelo menos um dos vetores puder ser escrito como uma combinação linear dos demais.

Exemplo 1.1 *Os vetores* $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$, $\begin{bmatrix} 1 \\ 4 \end{bmatrix}$ *e* $\begin{bmatrix} 1 \\ 6 \end{bmatrix}$ *são LD pois* $\begin{bmatrix} 1 \\ 2 \end{bmatrix} = 2\begin{bmatrix} 1 \\ 4 \end{bmatrix} - \begin{bmatrix} 1 \\ 6 \end{bmatrix}$.

Observação 1.2 *Quando a única solução do sistema $c_1\boldsymbol{x}_1 + c_2\boldsymbol{x}_2 + \ldots + c_k\boldsymbol{x}_k = 0$ for a trivial, ou seja, $c_1 = c_2 = \ldots = c_k = 0$, então os vetores $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k$ são LI.*

Definição 1.8 (Matriz quadrada) *Diz-se que uma matriz $A \in \mathbb{C}^{m\times n}$ é quadrada quando $m = n$.*

Lembre-se que uma matriz $A \in \mathbb{C}^{m\times n}$ pode ser denotada como $A = [a_{ij}]_{m\times n}$, onde cada a_{ij} representa uma entrada da matriz, com $1 \leq i \leq m$ e $1 \leq j \leq n$

Definição 1.9 (Matriz triangular) *Diz-se que uma matriz $A \in \mathbb{C}^{n\times n}$ é triangular superior quando $a_{ij} = 0$, $\forall i > j$ e triangular inferior quando $a_{ij} = 0$, $\forall i < j$.*

Definição 1.10 (Matriz diagonal) *Diz-se que uma matriz $A \in \mathbb{C}^{n\times n}$ é diagonal quando $a_{ij} = 0$, $\forall i \neq j$.*

Definição 1.11 (Matriz identidade) *Diz-se que uma matriz $A \in \mathbb{C}^{n\times n}$ é a matriz identidade de ordem n se A é diagonal e $a_{ij} = 1$, $\forall i = j$. Denota-se a matriz identidade de ordem $n \times n$ por I_n ou simplesmente I, quando a ordem ficar subentendida.*

Definição 1.12 (Matriz simétrica) *Diz-se que uma matriz $A \in \mathbb{R}^{n\times n}$ é simétrica se $A = A^T$.*

Definição 1.13 (Matriz transposta conjugada) *Se* $A \in \mathbb{C}^{n\times n}$ *então a transposta conjugada de* A*, denotada por* A^H*, é definida por* $A^H = \overline{A}^T$*, onde* $\overline{A}$ *é a matriz cujas entradas são os conjugados complexos das correspondentes entradas de* A*.*

A notação A^H pode ser utilizada também em vetores. Assim, pode-se denotar $\overline{\boldsymbol{y}}^T\boldsymbol{x}$ por $\boldsymbol{y}^H\boldsymbol{x}$.

Observação 1.3 *Se* A*,* $B \in \mathbb{R}^{n\times n}$ *tem-se* $(AB)^T = B^TA^T$*. Da mesma maneira, se* A*,* $B \in \mathbb{C}^{n\times n}$ *segue que* $(AB)^H = B^HA^H$*.*

Definição 1.14 (Matriz hermitiana) *Uma matriz* $A \in \mathbb{C}^{n\times n}$ *é chamada hermitiana se* $A^H = A$*.*

Definição 1.15 (Matriz anti-simétrica) *Diz-se que uma matriz* $A \in \mathbb{R}^{n\times n}$ *é anti-simétrica se* $A = -A^T$*.*

Definição 1.16 (Matriz anti-hermitiana) *Uma matriz* $A \in \mathbb{C}^{n\times n}$ *é chamada anti-hermitiana se* $A^H = -A$*.*

Definição 1.17 (Matriz inversível) *Uma matriz* $A \in \mathbb{C}^{n\times n}$ *é dita inversível, ou não-singular, se existir* $B \in \mathbb{C}^{n\times n}$ *tal que* $AB = BA = I$*. A matriz* B *é chamada inversa de* A *e é denotada por* $B = A^{-1}$*.*

Proposição 1.1 *A inversa de uma matriz é única.*

Prova: Seja A uma matriz. Suponha, por absurdo, que A' e A'' são inversas de A. Então:

$$AA' = I = A'A \text{ e } AA'' = I = A''A.$$

Assim,

$$A' = A'I = A'(AA'') = (A'A)A'' = IA'' = A''.$$

Portanto, $A' = A''$, e a inversa é única. ■

Teorema 1.1 (Teorema das matrizes inversíveis) *Seja* $A \in \mathbb{R}^{n\times n}$*. As seguintes afirmações são equivalentes:*

(i) *A é inversível.*

(ii) $A\boldsymbol{x} = \boldsymbol{b}$ *tem uma única solução para todo* $\boldsymbol{b} \in \mathbb{R}^n$.

(iii) $A\boldsymbol{x} = \mathbf{0}$ *admite somente a solução trivial.*

(iv) $\det(A) \neq 0$

(v) *As colunas de* A *formam um conjunto linearmente independente.*

Prova: Serão demonstradas as seguintes implicações:

$$(i) \Rightarrow (ii) \Rightarrow (iii) \Rightarrow (iv) \Rightarrow (i) \text{ e } (v) \Leftrightarrow (iii)$$

$(i) \Rightarrow (ii)$ Tem-se que demonstrar que $A\boldsymbol{x} = \boldsymbol{b}$ tem uma solução e que ela é única. Seja $\boldsymbol{x}' = A^{-1}\boldsymbol{b}$. Então $\boldsymbol{x}'$ é uma solução para $A\boldsymbol{x} = \boldsymbol{b}$. De fato,

$$A\boldsymbol{x}' = A(A^{-1}\boldsymbol{b}) = (AA^{-1})\ \boldsymbol{b} = I\boldsymbol{b} = \boldsymbol{b}.$$

Suponha que $\boldsymbol{y}$ seja uma outra solução para o sistema. Então $A\boldsymbol{y} = \boldsymbol{b}$, e multiplicando-se A^{-1} à esquerda, em ambos os lados da equação tem-se

$$A^{-1}(A\boldsymbol{y}) = A^{-1}\boldsymbol{b} \Rightarrow (A^{-1}A)\boldsymbol{y} = A^{-1}\ \boldsymbol{b} \Rightarrow I\ \boldsymbol{y} = A^{-1}\ \boldsymbol{b} \Rightarrow \boldsymbol{y} = A^{-1}\boldsymbol{b}.$$

Assim $\boldsymbol{y} = \boldsymbol{x}'$ e, portanto, a solução é única.

$(ii) \Rightarrow (iii)$ Do item (ii) tem-se que, em particular, $A\boldsymbol{x} = \mathbf{0}$ tem uma única solução. No entanto $\boldsymbol{x} = \mathbf{0}$ é sempre uma solução do sistema homogêneo. Portanto $\boldsymbol{x} = \mathbf{0}$ é a única solução para $A\boldsymbol{x} = \mathbf{0}$.

$(iii) \Rightarrow (iv)$ Como $A\boldsymbol{x} = \mathbf{0}$ admite somente a solução trivial, então $\det(A) \neq 0$ pois, caso contrário, qualquer vetor $\boldsymbol{x}$ seria solução para $A\boldsymbol{x} = \mathbf{0}$.

$(iv) \Rightarrow (i)$ Seja R a forma escalonada reduzida (por linhas) de A. Então

$$E_r \cdots E_2 E_1 A = R,$$

onde E_1, $E_2, \ldots,$ E_r são as matrizes elementares correspondentes às operações elementares com linhas necessárias para transformar A em R. Como $\det(A) \neq 0$, por hipótese, e $\det(E_i) \neq 0, \forall i$, então $\det(R) \neq 0$. Assim, R não pode conter nenhuma linha nula. Além disso, como R é a forma escalonada reduzida então R é a matriz identidade I de ordem $n \times n$. Denotando por A^{-1} a matriz $E_r \cdots E_2 E_1$ tem-se

$$A^{-1}A = I.$$

Se as operações elementares forem efetuadas da seguinte forma

$$AE_r \cdots E_2E_1,$$

também obtém-se a matriz R. Assim $AA^{-1} = I$. Portanto, A é inversível.

$(v) \Rightarrow (iii)$ Denote $\{A_i\}_{i=1}^n \subseteq \mathbb{R}^n$ as colunas de A. Se $\{A_i\}_{i=1}^n$ formam um conjunto LI, então $c_1\boldsymbol{A}_1 + c_2\boldsymbol{A}_2 + \ldots + c_n\boldsymbol{A}_n = \boldsymbol{0}$ se, e somente se, $c_i = 0$, para todo $i \in \{1, 2, \ldots, n\}$.

Considere o sistema $A\boldsymbol{x} = \boldsymbol{0}$. Sabe-se que é possível escrevê-lo como combinação das colunas de A da seguinte forma:

$$x_1\boldsymbol{A}_1 + x_2\boldsymbol{A}_2 + \ldots + x_n\boldsymbol{A}_n = \boldsymbol{0}, \tag{1.1}$$

onde cada x_i é a i-ésima coordenada do vetor $\boldsymbol{x}$. Mas, a única forma da Equação (1.1) existir é $x_i = 0$, $\forall i$, já que as colunas da A são LI. Portanto, $\boldsymbol{x} = \boldsymbol{0}$ e, então, o sistema $A\boldsymbol{x} = \boldsymbol{0}$ admite somente a solução trivial.

$(iii) \Rightarrow (v)$ Usando redução ao absurdo, supõe-se que as colunas de A formem um conjunto LD. Então pela Definição 1.7 tem-se que existem escalares $c_1, c_2, \ldots, c_k$, pelo menos um dos quais não-nulo, tais que $c_1\boldsymbol{A}_1 + c_2\boldsymbol{A}_2 + \ldots + c_n\boldsymbol{A}_n = \boldsymbol{0}$, onde $\boldsymbol{A}_i$ é a i-ésima coluna da matriz A. No entanto, $c_1\boldsymbol{A}_1 + c_2\boldsymbol{A}_2 + \ldots + c_n\boldsymbol{A}_n$ é um modo de representar a sistema de equações $A\boldsymbol{c}$, onde $\boldsymbol{c}^T = [c_1\ c_2\ \cdots\ c_n]$. Assim, $A\boldsymbol{x} = \boldsymbol{0}$ teria como solução o vetor $\boldsymbol{c} \neq \boldsymbol{0}$, contrariando a hipótese inicial. Logo, as colunas de A formam um conjunto linearmente independente. ■

Definição 1.18 (Matriz ortogonal) *Uma matriz $A \in \mathbb{R}^{n\times n}$ inversível é dita ortogonal se $A^{-1} = A^T$.*

Teorema 1.2 *Seja $A \in \mathbb{R}^{n\times n}$. As colunas de A formam um conjunto ortonormal se, e somente se, A é ortogonal.*

Prova: Suponha as colunas de A formam um conjunto ortonormal, então precisa-se provar que $A^TA = I$, ou seja:

$$(A^TA)_{ij} = \begin{cases} 1 \text{ se } i = j \\ 0 \text{ se } i \neq j \end{cases},$$

onde $(A^TA)_{ij}$ é o elemento que ocupa a posição (i, j) na matriz A^TA.

Seja a_i a i-ésima coluna de A (e a i-ésima linha de A^T), então $(A^TA)_{ij} = \langle a_i, a_j \rangle$. Assim, tem-se que mostrar que

$$\langle a_i, a_j \rangle = \begin{cases} 1 \text{ se } i = j \\ 0 \text{ se } i \neq j \end{cases}. \tag{1.2}$$

Mas, por hipótese, as colunas de A são ortonormais e, portanto, a Equação (1.2) de fato ocorrem.

Supondo A matriz ortogonal, a prova segue raciocínio semelhante. ■

Definição 1.19 (Matriz unitária) *Uma matriz* $U \in \mathbb{C}^{n \times n}$ *é chamada unitária se* $U^{-1} = U^H$.

O capítulo a seguir contém tópicos da teoria de autovalores e autovetores, e resultados importantes que auxiliarão na compreensão da Decomposição em Valores Singulares, apresentada no Capítulo 3.

Capítulo 2

Autovalores e Autovetores

Neste segundo capítulo será feito um estudo sobre autovalores e autovetores, também denominados valores próprios e vetores próprios, respectivamente. Os conceitos de autovalores e autovetores são úteis tanto em Matemática Pura como em Matemática Aplicada, sendo utilizados em diversas situações. Os autovalores são utilizados no estudo de equações diferenciais e em sistemas dinâmicos; eles fornecem informação crítica em projetos de engenharia e aparecem em áreas como a Física, Dinâmica Populacional e Genética.

Compreender o que são e conhecer algumas propriedades dos autovalores e autovetores de uma matriz é muito importante para o desenvolvimento deste trabalho. Como conseqüência da decomposição em valores singulares de uma matriz $A \in \mathbb{C}^{m\times n}$, obtém-se os chamados *valores singulares* que, como veremos mais adiante, são as raízes quadradas dos autovalores das matrizes AA^H e A^HA.

Definição 2.1 (Núcleo) *O núcleo de A, denotado por $N(A)$, é o conjunto formado pelos vetores solução do sistema $A\boldsymbol{x} = \boldsymbol{0}$, ou seja, $N(A) = \{\boldsymbol{x} \in \mathbb{C}^n : A\boldsymbol{x} = \boldsymbol{0}\}$. A dimensão de $N(A)$ é chamada nulidade de A, denotada por nulidade(A).*

Definição 2.2 (Autovalor e Autovetor) *Seja $A \in \mathbb{C}^{n\times n}$. Um vetor $\boldsymbol{x} \in \mathbb{C}^n$, $\boldsymbol{x} \neq \boldsymbol{0}$, é um autovetor de A se existe um escalar $\lambda \in \mathbb{C}$ tal que $A\boldsymbol{x} = \lambda\boldsymbol{x}$. Neste caso, o escalar λ é denominado autovalor.*

Pela Definição 2.2, para encontrar os autovalores de uma matriz $A \in \mathbb{C}^{n\times n}$ precisa-se resolver $(A-\lambda I)\boldsymbol{x} = \boldsymbol{0}$. Assim, para que λ seja um autovalor, a equação $(A-\lambda I)\boldsymbol{x} = \boldsymbol{0}$

deve apresentar uma solução não-nula, isto é, deve-se garantir a existência de $\lambda \in \mathbb{C}$ tal que $N(A - \lambda I) \neq \{\mathbf{0}\}$. Portanto, λ é autovalor de A se, e somente se,

$$\det(A - \lambda I) = 0. \tag{2.1}$$

A Equação (2.1) é denominada **equação característica** de A e os escalares que satisfazem essa equação são os autovalores de A. Além disso, considerando-se as definições e propriedades do determinante de uma matriz, verifica-se que essa equação define um polinômio de grau n com coeficientes complexos. Este polinômio é denominado **polinômio característico** de A e é denotado por $p(\lambda) = \det(A - \lambda I)$.

Assim, como consequência do Teorema Fundamental da Álgebra, segue que $p(\lambda)$ apresenta n raízes complexas, considerando-se suas respectivas multiplicidades. Conseqüentemente, toda matriz $A \in \mathbb{C}^{n\times n}$ admite n autovalores. Usando-se um argumento indutivo, prova-se que

$$p(\lambda) = \lambda^n + c_{n-1}\lambda^{n-1} + c_{n-2}\lambda^{n-2} + \ldots + c_1\lambda + c_0,$$

ou seja, o polinômio característico de A é um polinômio de grau n cujo coeficiente de λ^n tem módulo igual a 1.

A seguir, apresenta-se alguns resultados sobre autovalores e autovetores

Proposição 2.1 *Seja* $A \in \mathbb{C}^{n\times n}$.

(i) Se $A\boldsymbol{x} = \lambda_1\boldsymbol{x}$ e $A\boldsymbol{x} = \lambda_2\boldsymbol{x}$ então $\lambda_1 = \lambda_2$

Prova: Seja $\boldsymbol{x} \in \mathbb{C}^n$. Como $A\boldsymbol{x} = \lambda_1\boldsymbol{x}$ e $A\boldsymbol{x} = \lambda_2\boldsymbol{x}$ tem-se

$$\lambda_1\boldsymbol{x} = \lambda_2\ \boldsymbol{x} \Leftrightarrow \lambda_1\boldsymbol{x} - \lambda_2\ \boldsymbol{x} = 0 \Leftrightarrow (\lambda_1 - \lambda_2)\ \boldsymbol{x} = 0$$

Como $\boldsymbol{x}$ é não-nulo tem-se $\lambda_1 - \lambda_2 = 0 \Leftrightarrow \lambda_1 = \lambda_2$. ∎

(ii) Se $\boldsymbol{x}$ é um autovetor associado ao autovalor λ então qualquer múltiplo escalar não-nulo de $\boldsymbol{x}$ também é um autovetor associado a λ.

Prova: Como $\boldsymbol{x}$ é um autovetor associado ao autovalor λ tem-se $A\boldsymbol{x} = \lambda\boldsymbol{x}$. Seja $\alpha \in \mathbb{R}^*$ tal que $\alpha\boldsymbol{x}$ é múltiplo escalar de $\boldsymbol{x}$ então $A(\alpha\boldsymbol{x}) = (A\alpha)\ \boldsymbol{x} = \alpha(A\boldsymbol{x}) = \alpha(\lambda\boldsymbol{x}) = \lambda(\alpha\boldsymbol{x})$. Logo, $\alpha\boldsymbol{x}$ é autovetor associado a λ. ∎

(iii) Para qualquer número natural n, não-nulo, se $A\boldsymbol{x} = \lambda \boldsymbol{x}$ então $A^n \boldsymbol{x} = \lambda^n \boldsymbol{x}$.

Prova: Esta demonstração será feita usando indução sobre n. No caso $n = 1$ tem-se a hipótese $A\boldsymbol{x} = \lambda\boldsymbol{x}$. Tome como hipótese de indução $A^k\boldsymbol{x} = \lambda^k\boldsymbol{x}$. Para $k+1$ tem-se:

$$A^{k+1}\boldsymbol{x} = A(A^k\boldsymbol{x}) = A(\lambda^k\boldsymbol{x}) = \lambda^k(A\boldsymbol{x}) = \lambda^k(\lambda\boldsymbol{x}) = \lambda^{k+1}\boldsymbol{x}$$

Assim, $A^{k+1}\boldsymbol{x} = \lambda^{k+1}\boldsymbol{x}$, como se queria demonstrar. Por indução, o resultado é verdadeiro para número natural positivo. ■

(iv) Uma matriz é inversível se, e somente se, 0 não for um autovalor de A.

Prova: Pelo Teorema 1.1, como A é inversível, $\det(A) \neq 0$. Note que

$$\det(A) = \det(A - 0I) \neq 0,$$

o que significa que zero não é raiz do polinômio característico e, portanto, não é autovalor de A.

Se zero não é autovalor de A então $p(\lambda) = p(0) = \det(A - 0I) \neq 0$. Isso implica que $\det(A) = \det(A - 0I) \neq 0$. Logo, A é uma matriz inversível. ■

(v) Seja A uma matriz quadrada com autovalor λ e um correspondente autovetor $\boldsymbol{x}$. Se A é inversível então $\frac{1}{\lambda}$ é autovalor de A^{-1} com autovetor $\boldsymbol{x}$ associado.

Prova: Como $\boldsymbol{x}$ é autovetor associado ao autovalor λ tem-se $A\,\boldsymbol{x} = \lambda\boldsymbol{x}$. No entanto, note que:

$$A\,\boldsymbol{x} = \lambda\,\boldsymbol{x} \Rightarrow A^{-1}A\,\boldsymbol{x} = A^{-1}\lambda\,\boldsymbol{x} \Rightarrow \boldsymbol{x} = \lambda A^{-1}\,\boldsymbol{x} \Rightarrow A^{-1}\boldsymbol{x} = \frac{1}{\lambda}\boldsymbol{x}$$

Assim $\frac{1}{\lambda}$ é autovalor de A^{-1} com autovetor $\boldsymbol{x}$. ■

Exemplo 2.1 *Seja $A \in \mathbb{R}^{n\times n}$ tal que*

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

Calcular os autovalores de A e encontrar os autovetores associados a cada autovalor.

Primeiramente, calcula-se $p(\lambda) = \det(A - \lambda I)$.

$$p(\lambda) = \det \begin{bmatrix} 3-\lambda & 0 & 0 \\ 0 & 1-\lambda & -2 \\ 1 & 0 & 1-\lambda \end{bmatrix} = (3-\lambda)(1-\lambda)(1-\lambda) = -\lambda^3 + 5\lambda - 7\lambda + 3$$

Observando o determinante antes da expansão polinomial, nota-se que os autovalores são $\lambda_1 = 3$ e $\lambda_2 = \lambda_3 = 1$.

Logo após, é necessário encontrar $\boldsymbol{x} \in N(A - \lambda I)$, para $\lambda \in \{\lambda_1, \lambda_2\}$. Para tanto, deve-se que resolver o sistema $(A - \lambda I)\boldsymbol{x} = \mathbf{0}$.

Para $\lambda_1 = 3$ tem-se

$$(A - \lambda_1 I)\boldsymbol{x} = \mathbf{0} \Rightarrow (A - 3I)\boldsymbol{x} = \mathbf{0},$$

ou

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

Assim $x_1 = 2x_3$ e $x_2 = -x_3$ e portanto, $\boldsymbol{x} = \begin{bmatrix} 2x_3 & -x_3 & x_3 \end{bmatrix}^T$, $x_3 \in \mathbb{C}$ e $x_3 \neq 0$.

Tomando-se $x_3 = 1$ tem-se que $\boldsymbol{x} = \begin{bmatrix} 2 & -1 & 1 \end{bmatrix}^T$ é um autovetor associado ao autovalor $\lambda_1 = 3$.

Para $\lambda_2 = \lambda_3 = 1$ tem-se $(A - \lambda_2 I)\boldsymbol{x} = \mathbf{0} \Rightarrow (A - 1I)\boldsymbol{x} = \mathbf{0}$, ou seja:

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

Assim, $x_1 = x_3 = 0$ e x_2 qualquer. Portanto, $\boldsymbol{x} = \begin{bmatrix} 0 & x_2 & 0 \end{bmatrix}^T$, $x_2 \in \mathbb{C}$ e $x_2 \neq 0$. Tomando-se $x_2 = 1$ tem-se que $\boldsymbol{x} = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}^T$ é um autovetor associado ao autovalor $\lambda_2 = \lambda_3 = 1$.

Considere o seguinte lema:

Lema 2.1 *Seja $A \in \mathbb{C}^{n \times n}$ e sejam λ_1 e λ_2 autovalores de A, com autovetores x_1 e x_2, respectivamente. Se $\lambda_1 \neq \lambda_2$ então $\{\boldsymbol{x}_1, \boldsymbol{x}_2\}$ é um conjunto LI.*

Prova: Suponha que $\{\boldsymbol{x}_1, \boldsymbol{x}_2\}$ é um conjunto LD. Então existem escalares c_1 e c_2 em $\mathbb{C}$, pelo menos um dos quais não-nulo, tal que:

$$c_1\boldsymbol{x}_1 + c_2\boldsymbol{x}_2 = \mathbf{0} \tag{2.2}$$

Sem perda de generalidade, considere $c_1 \neq 0$ e multiplicando-se a Equação (2.2), à esquerda por A obtém-se:

$$\begin{aligned} c_1A\boldsymbol{x}_1 + c_2A\boldsymbol{x}_2 &= \mathbf{0} \\ c_1\lambda_1\boldsymbol{x}_1 + c_2\lambda_2\boldsymbol{x}_2 &= \mathbf{0}. \end{aligned} \tag{2.3}$$

Por outro lado, multiplicando-se a Equação (2.2) por λ_2:

$$c_1\lambda_2\boldsymbol{x}_1 + c_2\lambda_2\boldsymbol{x}_2 = \mathbf{0}. \tag{2.4}$$

Igualando-se as Equações (2.3) e (2.4) obtém-se,

$$c_1\lambda_1\boldsymbol{x}_1 + c_2\lambda_2\boldsymbol{x}_2 = c_1\lambda_2\boldsymbol{x}_1 + c_2\lambda_2\boldsymbol{x}_2 \Rightarrow c_1\lambda_1\boldsymbol{x}_1 - c_1\lambda_2\boldsymbol{x}_1 = \mathbf{0} \Rightarrow c_1(\lambda_1 - \lambda_2)\boldsymbol{x}_1 = \mathbf{0}.$$

Como $c_1 \neq 0$ e $\boldsymbol{x}_1$ é autovetor, e portanto não-nulo, resulta que $\lambda_1 - \lambda_2 = 0$, isto é, $\lambda_1 = \lambda_2$, contradizendo a hipótese de $\lambda_1 \neq \lambda_2$. Logo $\{\boldsymbol{x}_1, \boldsymbol{x}_2\}$ é um conjunto LI. ■

Teorema 2.1 *Seja $A \in \mathbb{C}^{n\times n}$ e sejam $\lambda_1, \lambda_2, \ldots, \lambda_m$ distintos autovalores de A, com $m \leq n$. Considere o autovetor $\boldsymbol{x}_i$ associado a cada λ_i, respectivamente. Então $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_m$ são linearmente independentes.*

Prova: Seja $A \in \mathbb{C}^{n\times n}$ e sejam $\lambda_1, \lambda_2, \ldots, \lambda_m$ distintos autovalores de A, com $m \leq n$. Considere o autovetor $\boldsymbol{x}_i$ associado a cada λ_i, respectivamente. Seja $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_m$ um conjunto de autovetores. Pelo Lema 2.1, se $m = 2$ então o conjunto $\{\boldsymbol{x}_1, \boldsymbol{x}_2\}$ é LI. Suponha que a tese seja válida para $m = k - 1$, $k \in \mathbb{N}$. Então deve-se demostrar que ela é válida para k.

Considere a combinação linear

$$c_1\boldsymbol{x}_1 + c_2\boldsymbol{x}_2 \ldots + c_k\boldsymbol{x}_k = 0. \tag{2.5}$$

Multiplicando a Equação (2.5) pela matriz A obtém-se:

$$c_1\lambda_1\boldsymbol{x}_1 + c_2\lambda_2\boldsymbol{x}_2 \ldots + c_k\lambda_k\boldsymbol{x}_k = 0. \tag{2.6}$$

Agora, multiplicando-se (2.5) por λ_k tem-se:

$$c_1\lambda_k\boldsymbol{x}_1 + c_2\lambda_k\boldsymbol{x}_2 \ldots + c_k\lambda_k\boldsymbol{x}_k = 0. \tag{2.7}$$

Subtraindo-se as Equações (2.6)e (2.7) chega-se que:

$$c_1(\lambda_1 - \lambda_k)\boldsymbol{x}_1 + c_2(\lambda_2 - \lambda_k)\boldsymbol{x}_2 \ldots + c_{k-1}(\lambda_{k-1} - \lambda_k)\boldsymbol{x}_{k-1} + c_k(\lambda_k - \lambda_k)\boldsymbol{x}_k = 0,$$

ou seja,

$$c_1(\lambda_1 - \lambda_k)\boldsymbol{x}_1 + c_2(\lambda_2 - \lambda_k)\boldsymbol{x}_2 \ldots + c_{k-1}(\lambda_{k-1} - \lambda_k)\boldsymbol{x}_{k-1} = 0.$$

No entanto, pela hipótese de indução, se

$$c_1(\lambda_1 - \lambda_k)\boldsymbol{x}_1 + c_2(\lambda_2 - \lambda_k)\boldsymbol{x}_2 \ldots + c_{k-1}(\lambda_{k-1} - \lambda_k)\boldsymbol{x}_{k-1} = 0$$

então $c_1(\lambda_1 - \lambda_k) = c_2(\lambda_2 - \lambda_k) = \ldots = c_{k-1}(\lambda_{k-1} - \lambda_k)$.

Como os $\lambda_i \neq \lambda_j$, se $i \neq j$, então $c_1 = c_2 = \ldots = c_{k-1} = 0$.

Assim, da Equação (2.5) resulta que

$$c_k\boldsymbol{x}_k = 0.$$

Como $\boldsymbol{x}_k$ é autovetor tem-se que $c_k = 0$. Portanto, o conjunto $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k\}$ é LI. ∎

Proposição 2.2 *Seja $A \in \mathbb{C}^{n\times n}$ triangular (superior, inferior ou diagonal) então os autovalores de A são as entradas da diagonal principal de A.*

Prova: Tendo em vista que se A é triangular então $A - \lambda I$ também o é, a demonstração segue do fato que o determinante de uma matriz triangular (superior, inferior ou diagonal) é o produto dos elementos da diagonal principal. ∎

O conjunto de todos os autovetores correspondentes a um autovalor λ de uma matriz $A \in \mathbb{C}^{n\times n}$ é exatamente o conjunto dos vetores não-nulos do núcleo de $A - \lambda I_n$. Assim, se a esse conjunto juntarmos o vetor nulo de $\mathbb{C}^n$, obteremos o $N(A - \lambda I)$.

Definição 2.3 (Espectro) *O conjunto formado por todos os autovalores de $A \in \mathbb{C}^{n\times n}$ é chamado de espectro de A e é denotado por $\lambda(A)$.*

Definição 2.4 (Auto-espaço) *Considere* $A \in \mathbb{C}^{n\times n}$ *e* λ *um autovalor de* A. *O conjunto formado por todos os autovetores correspondentes a* λ, *acrescido do vetor nulo, é chamado de auto-espaço de* λ *e é denotado por* E_λ. *Portanto,* $E_\lambda = N(A - \lambda I)$.

Assim, no Exemplo 2.1 tem-se E_{λ_1} como o subespaço gerado pelo vetor $\begin{bmatrix} 2 & -1 & 1 \end{bmatrix}^T$ e E_{λ_2} como o subespaço gerado pelo vetor $\begin{bmatrix} 0 & 1 & 0 \end{bmatrix}^T$.

Definição 2.5 (Multiplicidade algébrica e multiplicidade geométrica) *Seja* A *uma matriz* $n \times n$ *e* λ *um autovalor de* A. *Chama-se de multiplicidade algébrica, denotada por* $m_a(\lambda)$, *o maior elemento* $j \in \mathbb{N}$ *tal que* $p(t) = \det(A - tI) = (t-\lambda)^j q(t)$, *com* $q(t) \neq 0$, *ou seja, o maior* j *tal que* $(t-\lambda)^j$ *divide* $\det(A - tI)$. *Chama-se de multiplicidade geométrica de* λ *a dimensão do auto-espaço* E_λ, *ou seja,* $m_g(\lambda) = \dim(E_\lambda)$.

Teorema 2.2 *Se* A *é uma matriz hermitiana, então:*

(i) seus autovalores são reais.

(ii) autovetores correspondentes a autovalores distintos de A, são ortogonais.

Prova: Para o demonstrar o item (i) seja λ um autovalor de A associado a um autovetor $\boldsymbol{x}$, ou seja, $A\boldsymbol{x} = \lambda\boldsymbol{x}$. Por hipótese $A = A^H$ e, então, multiplicando-se $\boldsymbol{x}^H$ a esquerda de $A\boldsymbol{x} = \lambda\boldsymbol{x}$ tem-se:

$$\lambda = \frac{\boldsymbol{x}^H A\boldsymbol{x}}{\boldsymbol{x}^H\boldsymbol{x}} = \frac{\boldsymbol{x}^H A^H\boldsymbol{x}}{\boldsymbol{x}^H\boldsymbol{x}} = \left(\frac{\boldsymbol{x}^H A\boldsymbol{x}}{\boldsymbol{x}^H\boldsymbol{x}}\right)^H = \lambda^H.$$

Ou seja, $\lambda = \lambda^H$. Portanto $\lambda \in \mathbb{R}$.

Sejam $\boldsymbol{x}_1$ e $\boldsymbol{x}_2$ autovetores correspondentes aos autovalores distintos $\lambda_1 \neq \lambda_2$, respectivamente, isto é, $A\boldsymbol{x}_1 = \lambda_1\boldsymbol{x}_1$ e $A\boldsymbol{x}_2 = \lambda_2\boldsymbol{x}_2$. Então $\boldsymbol{x}_1^H(A\boldsymbol{x}_2) = \boldsymbol{x}_1^H(\lambda_2\boldsymbol{x}_2) = \lambda_2(\boldsymbol{x}_1^H\boldsymbol{x}_2)$.

Também tem-se que

$$\boldsymbol{x}_1^H(A\boldsymbol{x}_2) = (A\boldsymbol{x}_1)^H\boldsymbol{x}_2 = (\lambda_1\boldsymbol{x}_1)^H\boldsymbol{x}_2 = \lambda_1(\boldsymbol{x}_1^H\boldsymbol{x}_2)$$

pois A é hermitiana e, portanto, pelo Teorema 2.2, $\lambda_1^H = \lambda_1$.

Assim $\lambda_2(\boldsymbol{x}_1^H\boldsymbol{x}_2) = \lambda_1(\boldsymbol{x}_1^H\boldsymbol{x}_2) \Rightarrow (\lambda_2 - \lambda_1)(\boldsymbol{x}_1^H\boldsymbol{x}_2) = 0$.

Como $\lambda_1 \neq \lambda_2$ então $\boldsymbol{x}_1^H\boldsymbol{x}_2 = 0$. ■

Proposição 2.3 *Seja $U \in \mathbb{C}^{n\times n}$ uma matriz unitária e $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{C}^n$ então:*

(i) $\|U\boldsymbol{x}\|^2 = \|\boldsymbol{x}\|^2$;

(ii) $\langle U\boldsymbol{x}, U\boldsymbol{y}\rangle = \langle \boldsymbol{x}, \boldsymbol{y}\rangle$;

(iii) Se λ é um autovalor de U então $|\lambda| = 1$.

Prova: De fato,

$$\|U\boldsymbol{x}\|^2 = (U\boldsymbol{x})^H U\boldsymbol{x} = \boldsymbol{x}^H U^H U\boldsymbol{x} = \boldsymbol{x}^H\boldsymbol{x} = \|\boldsymbol{x}\|^2 .$$

Além disso, $\langle U\boldsymbol{x}, U\boldsymbol{y}\rangle = (U\boldsymbol{x})^H U\boldsymbol{y} = \boldsymbol{x}^H U^H U\boldsymbol{y} = \boldsymbol{x}^H\boldsymbol{y} = \langle \boldsymbol{x}, \boldsymbol{y}\rangle$.

Seja λ autovalor de U então

$$U\,\boldsymbol{x} = \lambda\,\boldsymbol{x} \Leftrightarrow \|U\boldsymbol{x}\| = \|\lambda\boldsymbol{x}\| \Rightarrow \|\boldsymbol{x}\| = |\lambda|\,\|\boldsymbol{x}\| \Rightarrow |\lambda| = 1.$$

■

Observação 2.1 *Os vetores coluna de uma matriz unitária formam um conjunto ortonormal em $\mathbb{C}^n$.*

A verificação desta observação é análoga à prova da Teorema 1.2, para matrizes ortogonais.

2.1 Semelhança, Diagonalização e Decomposição de Matrizes

Conforme exposto na Proposição 2.2, encontrar os autovalores de matrizes triangulares (em particupar, diagonais) é uma tarefa bastante simples já que estes mostram-se na diagonal principal de tais matrizes. Seria bastante interessente, e também útil, se fosse possível relacionar qualquer matriz com um matriz triangular de forma que ambas possuíssem os mesmo autovalores.

A seguir, será apresentada a diagonalização de matriz, procedimento que permite fatorar determinadas matrizes num produto de matrizes, mantendo seus autovalores em uma matriz diagonal. Também serão apresentadas duas outras decomposições matriciais: a Decomposição de Schur e a Decomposição Espectral.

Definição 2.6 (Matrizes semelhantes) *Sejam $A, B \in \mathbb{C}^{n\times n}$. Diz-se que A é semelhante a B se existir uma matriz $P \in \mathbb{C}^{n\times n}$ inversível tal que $A = PBP^{-1}$. Denota-se A semelhante a B por $A \sim B$.*

Se a matriz P for também unitária, ou seja, $P^{-1} = P^H$, diz-se que A é *unitariamente semelhante* a B.

Teorema 2.3 *Para $A, B \in \mathbb{C}^{n\times n}$ tal que $A \sim B$ tem-se:*

(i) $\det A = \det B$

(ii) *A é inversível se, e somente se, B é inversível.*

(iii) *A e B têm o mesmo polinômio característico*

(iv) *A e B têm os mesmos autovalores.*

Prova: (i) Note que

$$\det A = \det PBP^{-1} = \det P \det B (\det P)^{-1} = \det P \det B \frac{1}{\det P} = \det B$$

(ii) Se A é inversível então existe A^{-1} tal que $AA^{-1} = A^{-1}A = I$. Assim

$$\begin{aligned} AA^{-1} &= I \qquad (2.8)\\ PBP^{-1}(PBP^{-1})^{-1} &= I \\ PBP^{-1}PB^{-1}P^{-1} &= I \\ PBB^{-1}P^{-1} &= I \\ BB^{-1}P^{-1} &= P^{-1} \\ BB^{-1} &= P^{-1}P \\ BB^{-1} &= I \end{aligned}$$

Para obter $B^{-1}B$ basta iniciar a Equação (2.8) com $A^{-1}A$. Portanto, B é inversível. Além disso, se por hipótese tem-se que B é inversível, a Equação (2.8) continua válida. Assim, A é inversível.

(iii) Como o polinômio característico da matriz A é dado por $\det(A - \lambda I)$ tem-se

$$\det(A - \lambda I) = \det(PBP^{-1} - \lambda PIP^{-1}) = \det(P(B - \lambda I)P^{-1}) = \det(B - \lambda I).$$

Ou seja, B tem o mesmo polinômio característico de A.

(iv) Seja λ autovalor de A então $A\boldsymbol{x} = \lambda\boldsymbol{x}$, para algum $\boldsymbol{x}$. Assim

$$A\boldsymbol{x} = \lambda\, \boldsymbol{x} \Rightarrow PBP^{-1}\boldsymbol{x} = \lambda\, \boldsymbol{x} \Rightarrow BP^{-1}\boldsymbol{x} = P^{-1}\lambda\, \boldsymbol{x} \Rightarrow B(P^{-1}\boldsymbol{x}) = \lambda(P^{-1}\boldsymbol{x})$$

Ou seja, λ é autovalor de B correspondente ao autovetor $P^{-1}\boldsymbol{x}$. Analogamente, mostra-se que todo autovalor de B é também autovalor de A. ∎

Definição 2.7 (Matriz diagonalizável) *Uma matriz $A \in \mathbb{C}^{n\times n}$ é diagonalizável se existem uma matriz diagonal $D \in \mathbb{C}^{n\times n}$ e uma matriz inversível $P \in \mathbb{C}^{n\times n}$ tal que $A = PDP^{-1}$, isto é, A é semelhante a uma matriz diagonal.*

Teorema 2.4 *Uma matriz $A \in \mathbb{C}^{n\times n}$ é diagonalizável se, e somente se, A tem n autovetores linearmente independentes.*

Prova: Suponha que A seja semelhante à matriz diagonal D onde $D = P^{-1}AP$, ou equivalentemente, $AP = PD$. Sejam $\boldsymbol{p}_1,\ \boldsymbol{p}_2, \ldots,\ \boldsymbol{p}_n$ os vetores coluna de P, e $\lambda_1, \lambda_2, \ldots, \lambda_n$ os elementos da diagonal de D. Então:

$$A\begin{bmatrix} \boldsymbol{p}_1 & \boldsymbol{p}_2 & \ldots & \boldsymbol{p}_n \end{bmatrix} = \begin{bmatrix} \boldsymbol{p}_1 & \boldsymbol{p}_2 & \cdots & \boldsymbol{p}_n \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} \tag{2.9}$$

ou

$$\begin{bmatrix} A\boldsymbol{p}_1 & A\boldsymbol{p}_2 & \ldots & A\boldsymbol{p}_n \end{bmatrix} = \begin{bmatrix} \lambda_1\boldsymbol{p}_1 & \lambda_2\boldsymbol{p}_2 & \ldots & \lambda_n\boldsymbol{p}_n \end{bmatrix}. \tag{2.10}$$

Tem-se, assim, n equações:

$$\begin{aligned} A\boldsymbol{p}_1 &= \lambda_1\boldsymbol{p}_1 \\ A\boldsymbol{p}_2 &= \lambda_2\boldsymbol{p}_2 \\ &\vdots \\ A\boldsymbol{p}_n &= \lambda_n\boldsymbol{p}_n \end{aligned}$$

Portanto, os vetores coluna da P são os autovetores de A cujos autovalores correspondentes são os elementos da diagonal de D , na mesma ordem. Como P é inversível, pelo Teorema 1.1, seus vetores coluna são linearmente independentes.

Reciprocamente, se A tem n autovetores $\boldsymbol{p}_1, \boldsymbol{p}_2, \ldots, \boldsymbol{p}_n$ linearmente independentes, com os autovalores correspondentes $\lambda_1, \lambda_2, \ldots, \lambda_n$, respectivamente, então:

$$\begin{aligned} A\boldsymbol{p}_1 &= \lambda_1\boldsymbol{p}_1 \\ A\boldsymbol{p}_2 &= \lambda_2\boldsymbol{p}_2 \\ &\vdots \\ A\boldsymbol{p}_n &= \lambda_n\boldsymbol{p}_n \end{aligned}$$

Esta constatação resulta da Equação (2.10), a qual é equivalente à Equação (2.9).

Conseqüentemente, chamando-se de P a matriz com colunas $\boldsymbol{p}_1, \boldsymbol{p}_2, \ldots, \boldsymbol{p}_n$ então a Equação (2.9) pode ser escrita como $AP = PD$. Como as colunas de P são linearmente independentes, P é inversível e, assim $D = P^{-1}AP$. Portanto, A é diagonalizável. ■

Como resultado do Teorema 2.4, tem-se que as colunas de P são os autovetores de A enquanto os elementos da diagonal de D são os autovalores de A, com cada autovetor $\boldsymbol{p}_i$ associado ao autovalor λ_i. Diz-se que A, quadrada de ondem n, tem um conjunto completo de autovetores se A apresenta n autovetores linearmente independentes.

Exemplo 2.2 *Seja* $A = \begin{bmatrix} 4 & -5 \\ 2 & -3 \end{bmatrix}$. *Calculando os autovalores e autovetores tem-se* $\lambda_1 = -1$ *com* $\boldsymbol{x}_1 = \begin{bmatrix} 1 & 1 \end{bmatrix}^T$ *e* $\lambda_2 = 2$ *com* $\boldsymbol{x}_2 = \begin{bmatrix} 5 & 2 \end{bmatrix}^T$.

Assim, tomando-se $P = \begin{bmatrix} 1 & 5 \\ 1 & 2 \end{bmatrix}$ *e* $D = \begin{bmatrix} -1 & 0 \\ 0 & 2 \end{bmatrix}$ *tem-se de* $A = PDP^{-1}$ *que* $AP = PD$. *De fato:*

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

e

$$PD = \begin{bmatrix} 1 & 5 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} -1 & 0 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} -1 & 10 \\ -1 & 4 \end{bmatrix}.$$

Observação 2.2 *Nem todas as matrizes são diagonalizáveis, neste caso, dizemos que* A *é uma matriz defectiva.*

Observação 2.3 *Se $A \in \mathbb{R}^{n\times n}$ é simétrica e A tem n autovalores distintos, então existe $Q \in \mathbb{R}^{n\times n}$ ortogonal e $D \in \mathbb{R}^{n\times n}$ matriz diagonal contendo os autovalores de A tal que $A = QDQ^T$. Neste caso, $Q = \begin{bmatrix} \boldsymbol{x}_1 & \cdots & \boldsymbol{x}_n \end{bmatrix}$, onde $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$ são autovalores de A normalizados.*

A verificação da Observação 2.3 segue diretamente dos Teoremas 2.1 e 2.2.

Teorema 2.5 (Decomposição de Schur) *Dada $A \in \mathbb{C}^{n\times n}$, existem $U \in \mathbb{C}^{n\times n}$ matriz unitária e $T \in \mathbb{C}^{n\times n}$ matriz triangular superior tal que $A = UTU^H$.*

Prova: Esta demonstração utiliza indução sobre n.

Para $n = 1$, o resultado é imediato, pois dada $A \in \mathbb{C}^{1\times 1}$ tem-se $A = [1][A][1]$.

Seja a tese do teorema válida para $k \in \mathbb{N}$. Tem-se que provar que é também válida para $k + 1$.

De fato, seja $A \in \mathbb{C}^{(k+1)\times(k+1)}$, λ_1 autovalor de A associado ao autovetor $\boldsymbol{x}_1$ e $\|\boldsymbol{x}_1\| = 1$.

Seja

$$U_1 = \begin{bmatrix} \boldsymbol{x}_1 & M_1 \end{bmatrix}_{(k+1)\times(k+1)},$$

com $\boldsymbol{x}_1 = \begin{bmatrix} x_1 & \cdots & x_{k+1} \end{bmatrix}^T \in \mathbb{C}^{k+1}$ e $M_1 \in \mathbb{C}^{(k+1)\times k}$ com colunas ortonormais de maneira que U_1 seja unitária, ou seja, $U_1^H U_1 = U_1 U_1^H = I$

Note que:

$$\begin{aligned} U_1^H A U_1 &= \begin{bmatrix} \boldsymbol{x}_1^H \\ M_1^H \end{bmatrix} A \begin{bmatrix} \boldsymbol{x}_1 & M_1 \end{bmatrix} = \begin{bmatrix} \boldsymbol{x}_1^H \\ M_1^H \end{bmatrix} \begin{bmatrix} A\boldsymbol{x}_1 & AM_1 \end{bmatrix} = \\ &= \begin{bmatrix} \boldsymbol{x}_1^H \\ M_1^H \end{bmatrix} \begin{bmatrix} \lambda_1 \boldsymbol{x}_1 & AM_1 \end{bmatrix} = \begin{bmatrix} \lambda_1 & \boldsymbol{x}_1^H A M_1 \\ 0 & M_1^H A M_1 \end{bmatrix}. \end{aligned}$$

No entanto, $B = M_1^H A M_1 \in \mathbb{C}^{k\times k}$, e pela hipótese de indução tem-se que existe $U_2 \in \mathbb{C}^{k\times k}$ unitária tal que $B = U_2 T_2 U_2^H$, onde $T_2 \in \mathbb{C}^{k\times k}$ é triangular superior.

Então:

$$\begin{aligned} U_1^H A U_1 &= \begin{bmatrix} \lambda_1 & \boldsymbol{x}_1^H A M_1 \\ 0 & M_1^H A M_1 \end{bmatrix} = \\ &= \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & U_2 \end{bmatrix}}_{U_3} \underbrace{\begin{bmatrix} \lambda_1 & \boldsymbol{x}_1^H A M_1 \\ 0 & T_2 \end{bmatrix}}_{T} \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & U_2^H \end{bmatrix}}_{U_3^H} = U_3 T U_3^H \end{aligned}$$

onde T é triangular superior de ordem $(k+1)\times(k+1)$ e $U_3 \in \mathbb{C}^{(k+1)\times(k+1)}$ unitária. Assim $U_1^H A U_1 = U_3 T U_3^H$, com U_1 e U_3 unitárias. Logo,

$$A = (U_1U_3)T(U_3^H U_1^H) = (U_1U_3)T(U_1U_3)^H.$$

Tomando-se $U = U_1U_3 \in \mathbb{C}^{(k+1)\times(k+1)}$, tem-se $UU^H = U_3^H U_1^H U_1 U_3 = U_3^H U_3 = I = UU^H$, ou seja, U é unitária. Portanto, $A = UTU^H$, com T triangular superior e U unitária. $\blacksquare$

O Teorema 2.5 mostra que toda matriz $A \in \mathbb{C}^{n\times n}$ é unitariamente semelhante a uma matriz triangular superior.

Observação 2.4 *A diagonal de T contém os autovalores de A.*

De fato, por construção, tem-se que A é semelhante a T, de $A = UTU^H$, e pelas propriedades de matrizes semelhantes, $\lambda(A) = \lambda(T)$.

Definição 2.8 (Traço) *Chama-se de traço de uma matriz $A \in \mathbb{C}^{n\times n}$ a soma das entradas a_{ij} de A, com $i = j$. Note que $Tr(A) \in \mathbb{C}$*

Teorema 2.6 *Seja $A \in \mathbb{C}^{n\times n}$ com autovalores $\lambda_1, \lambda_2, \ldots, \lambda_n$. Então*

$$Tr(A) = \lambda_1 + \lambda_2 + \ldots + \lambda_n$$

e

$$\det(A) = \lambda_1 \cdot \lambda_2 \cdot \ldots \cdot \lambda_n.$$

Prova: Pelo Teorema 2.5 tem-se que A pode ser fatorada como $A = UTU^H$, onde U é uma matriz unitária e T uma matriz triangular superior. Desta forma

$$Tr(A) = Tr(UTU^H) = Tr(TU^HU) = Tr(T),$$

pois dadas matrizes $B, C \in \mathbb{C}^{n\times n}$ tem-se $Tr(BC) = Tr(CB)$.

No entanto, a matriz T é tal que os autovalores de A formam a diagonal principal de T. Logo

$$Tr(A) = Tr(T) = \lambda_1 + \lambda_2 + \ldots + \lambda_n.$$

Utilizando-se a decomposição de A novamente tem-se que $det(A) = det(UTU^H)$. Mas

$$det(UTU^H) = det(U)det(T)det(U^H) = det(T)det(U)det(U^H) = det(T).$$

Assim $det(A) = det(T) = \lambda_1 \cdot \lambda_2 \cdot \ldots \cdot \lambda_n$. $\blacksquare$

Teorema 2.7 (Decomposição espectral para matrizes hermitianas) *Toda matriz $A \in \mathbb{C}^{n\times n}$ hermitiana pode ser diagonalizada por uma matriz unitária, ou seja, existe $U \in \mathbb{C}^{n\times n}$ matriz unitária e $D \in \mathbb{R}^{n\times n}$ matriz diagonal tal que $A = UDU^H$.*

Prova: É necessário mostrar que existe $U \in \mathbb{C}^{n\times n}$ unitária tal que $A = UDU^H$, com D matriz diagonal. Pelo Teorema 2.5, segue que existem U unitária e T triangular superior tal que $A = UTU^H$. Por hipótese, $A = A^H$ mas $A^H = UT^HU^H$. Assim:

$$UTU^H = UT^HU^H \Rightarrow T = T^H.$$

Como T é triangular superior e $T = T^H$ tem-se que T é diagonal com os elementos da diagonal todos reais. Denotando T por D tem-se $A = UDU^H$, com $D \in \mathbb{R}^{n\times n}$. ■

O Teorema 2.7 pode ser assim enunciado: Toda matriz $A \in \mathbb{C}^{n\times n}$ hermitiana é unitariamente semelhante a uma matriz diagonal.

Observação 2.5 *Se $A \in \mathbb{R}^{n\times n}$ então, como $D \in \mathbb{R}^{n\times n}$ e cada coluna de U pertence ao $N(A - \lambda_i I)$, segue que $U \in \mathbb{R}^{n\times n}$, isto é, U é ortogonal e $A = UDU^T$.*

Observação 2.6 *Se $A \in \mathbb{C}^{n\times n}$ é hermitiana, então os autovetores de A formam uma base ortonormal para $\mathbb{C}^n$.*

De fato, pois pelo Teorema 2.7 as colunas de U são os autovetores de A e, como U é unitária, formam uma base para $\mathbb{C}^n$.

Definição 2.9 (Matriz normal) *Uma matriz $N \in \mathbb{C}^{n\times n}$ é chamada normal se $NN^H = N^HN$.*

Note que se A é uma matriz hermitiana então $A^HA = AA = AA^H$, isto é, A é normal.

Considere o seguinte lema

Lema 2.2 *Se $T \in \mathbb{C}^{n\times n}$ é triangular superior e normal então T é diagonal*

Prova: Seja T_i a i-ésima coluna de T. Como T é normal então T_i^H é a i-ésima coluna de T. Assim,

$$\|T_i\|^2 = \left\|T_i^H\right\|^2.$$

Então, a norma da i-ésima coluna de T é igual à norma da i-ésima linha de T.

Considere:

$$T = \begin{bmatrix} t_{11} & t_{12} & \cdots & t_{1n} \\ 0 & t_{22} & \cdots & t_{2n} \\ \vdots & 0 & \ddots & \vdots \\ 0 & \cdots & 0 & t_{nn} \end{bmatrix}.$$

Assim, para $i = 1$ tem-se:

$$\begin{aligned} \|T_1\|^2 = \left\|T_1^H\right\|^2 \quad &\Rightarrow \quad |t_{11}|^2 = |t_{11}|^2 + |t_{12}|^2 + \ldots + |t_{1n}|^2 \Rightarrow \\ &\Rightarrow \quad |t_{12}| = \ldots = |t_{1n}| = 0 \end{aligned}$$

Analogamente, para $i = 2$ tem-se:

$$|t_{12}|^2 + |t_{22}|^2 = |t_{22}|^2 + |t_{23}|^2 + \ldots + |t_{2n}|^2 \Rightarrow |t_{23}| = |t_{24}| = \ldots = |t_{2n}| = 0$$

pois $|t_{12}| = 0$, do passo anterior.

Em geral, para $1 \leq i \leq n$ tem-se:

$$\begin{aligned} |t_{1i}|^2 + |t_{2i}|^2 + \ldots + |t_{ii}|^2 = |t_{ii}|^2 + |t_{i(i+1)}|^2 + \ldots + |t_{in}|^2 \Rightarrow \\ \Rightarrow \quad |t_{i(i+1)}|^2 + \ldots + |t_{in}|^2 = 0 \Rightarrow |t_{i(i+1)}| = \ldots = |t_{in}| = 0 \end{aligned}$$

Portanto, T é diagonal. ■

Teorema 2.8 (Decomposição espectral para matrizes normais) *Se $N \in \mathbb{C}^{n\times n}$ é uma matriz normal então existe uma matriz unitária $U \in \mathbb{C}^{n\times n}$ e $D \in \mathbb{C}^{n\times n}$ matriz diagonal tal que $N = UDU^H$.*

Prova: Como N é normal então $N^H N = N N^H$. Pelo Teorema 2.5, existe $U \in \mathbb{C}^{n\times n}$ unitária e $T \in \mathbb{C}^{n\times n}$ triangular superior tal que

$$N = UTU^H \Rightarrow N^H = UT^H U^H$$

Assim $NN^H = (UTU^H)(UT^H U^H) = UTT^H U^H$. Por outro lado,

$$N^H N = (UT^H U^H)(UTU^H) = UT^H T U^H.$$

Portanto, $UT^H TU^H = UTT^H U^H \Rightarrow T^H T = TT^H$ e assim, T é normal.

Assim, pelo Lema 2.2, a matriz T é diagonal. Denotando T por D tem-se

$$N = UDU^H,$$

com U unitária e D diagonal. ■

O Teorema 2.8 mostra que toda matriz normal é unitariamente semelhante a uma matriz diagonal.

Observação 2.7 *A recíproca da Teorema 2.8 também é válida, ou seja, se $N \in C^{n\times n}$ é unitariamente diagonalizável então N é normal.*

De fato, se $N = UDU^H$, tem-se:

$$\begin{aligned} N^H N &= (UD^H U^H)(UDU^H) = UD^H DU^H \\ NN^H &= (UDU^H)(UD^H U^H) = UDD^H U^H \end{aligned}$$

Mas, considerando d_i os elementos de D tem-se:

$$D^H D = \begin{bmatrix} \overline{d_1} & & \\ & \ddots & \\ & & \overline{d_n} \end{bmatrix} \begin{bmatrix} d_1 & & \\ & \ddots & \\ & & d_n \end{bmatrix} = \begin{bmatrix} \overline{d_1} d_1 & & \\ & \ddots & \\ & & \overline{d_n} d_n \end{bmatrix} = \begin{bmatrix} |d_1|^2 & & \\ & \ddots & \\ & & |d_n|^2 \end{bmatrix}.$$

Por outro lado,

$$DD^H = \begin{bmatrix} d_1 & & \\ & \ddots & \\ & & d_n \end{bmatrix} \begin{bmatrix} \overline{d_1} & & \\ & \ddots & \\ & & \overline{d_n} \end{bmatrix} = \begin{bmatrix} d_1 \overline{d_1} & & \\ & \ddots & \\ & & d_n \overline{d_n} \end{bmatrix} = \begin{bmatrix} |d_1|^2 & & \\ & \ddots & \\ & & |d_n|^2 \end{bmatrix}.$$

Então $D^H D = DD^H$ e, consequentemente $N^H N = NN^H$, isto é, N é normal.

2.2 Autovalores de Produtos e de Potências de Matrizes

Proposição 2.4 *Se $\lambda_1, \lambda_2, \dots, \lambda_n$ são autovalores de A então $\lambda_1^k, \lambda_2^k, \dots, \lambda_n^k$, com $k \in \mathbb{N}$, são os autovalores de A^k. Além disso, se $A = PDP^{-1}$ é a diagonalização de A então $A^k = PD^k P^{-1}$.*

Prova: Por hipótese $A\boldsymbol{x}_i = \lambda_i \boldsymbol{x}_i$ e portanto para $k = 1$, a proposição é válida.

Considere a proposição válida para algum $k \in \mathbb{N}$, ou seja, $(\lambda_i)^k$ são os autovalores de A^k com autovetores $\boldsymbol{x}_i$, respectivamente.

Conseqüentemente, para $k + 1$ tem-se:

$$A^{k+1}\boldsymbol{x}_i = A(A^k\boldsymbol{x}_i) = A(\lambda_i^k\boldsymbol{x}_i) = \lambda_i^k(A\boldsymbol{x}_i) = \lambda_i^k\lambda_i\boldsymbol{x}_i = \lambda_i^{k+1}\boldsymbol{x}_i.$$

Portanto temos que $(\lambda_i)^k$ é o conjunto de autovalores de A^k com autovetores $\boldsymbol{x}_i$, para $i \in 1, 2, \ldots, n$, respectivamente.

Além disso, se D contém os autovalores de A, então os elementos de D^k são os λ_i^k, com $i \in \{1, 2, \ldots, n\}$.

Logo, $A^k = PD^kP^{-1}$. ■

Proposição 2.5 *Se A e B são matrizes quadradas e $\boldsymbol{x}$ é um autovetor de A e B, simultaneamente, isto é, $A\ \boldsymbol{x} = \lambda\boldsymbol{x}$ e $B\ \boldsymbol{x} = \mu\boldsymbol{x}$, então $\lambda\mu$ é um autovalor de AB com autovetor $\boldsymbol{x}$.*

Prova: Sabe-se que $A\ \boldsymbol{x} = \lambda\boldsymbol{x}$ e $B\ \boldsymbol{x} = \mu\boldsymbol{x}$, então:

$$AB\ \boldsymbol{x} = A\mu\boldsymbol{x} = \mu A\boldsymbol{x} = \mu\lambda\boldsymbol{x}.$$

Assim $\mu\lambda$ é um autovalor de AB com autovetor $\boldsymbol{x}$. ■

Teorema 2.9 *Sejam A e B matrizes diagonalizáveis. Então A e B possuem a mesma matriz de autovetores se, e somente se, $AB = BA$.*

Prova: Sejam A e B matrizes com a mesma matriz de autovetores. Como A e B são diagonalizáveis com a mesma matriz de autovetores tem-se

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

e

$$B = PD'P^{-1}$$

onde D e D' são matrizes diagonais. Assim:

$$\begin{aligned} AB = (PDP^{-1})(PD'P^{-1}) = PDD'P^{-1} &= \\ = PD'DP^{-1} &= (PD'P^{-1})(PDP^{-1}) = BA. \end{aligned}$$

Para provar a recíproca, considera-se um caso particular, onde os autovalores são todos distintos. O caso geral será omitido.

Sejam $\lambda_1, \lambda_{2,} \ldots, \lambda_n$ os autovalores de A, todos distintos entre si. Então, para cada i, com $i \in 1, 2, \ldots, n$ tem-se $A\boldsymbol{x}_i = \lambda_i \boldsymbol{x}_i$, onde $\boldsymbol{x}_i$ é autovetor associadao à λ_i. Conseqüentemente,

$$BA\boldsymbol{x}_i = B(\lambda_i \boldsymbol{x}_i) = \lambda_i(B\boldsymbol{x}_i) \Rightarrow A(B\boldsymbol{x}_i) = \lambda_i(B\boldsymbol{x}_i).$$

Ou seja, $B\boldsymbol{x}_i$ é um autovetor de A associado ao autovalor λ_i. Portanto

$$(B\boldsymbol{x}_i) \in N(A - \lambda_i I).$$

Assim, $B\boldsymbol{x}_i$ pertence ao auto-espaço de A associado a λ_i.

Assumindo que $\lambda_i \neq \lambda_j$ se $i \neq j$, então os autovetores associados à λ_i são todos múltiplos de $\boldsymbol{x}_i$. Logo, existe μ_i tal que $B\boldsymbol{x}_i = \mu_i \boldsymbol{x}_i$, com $i \in 1, 2, \ldots, n$.

Portanto, $\boldsymbol{x}_i$ também é autovetor de B. ■

Capítulo 3

A Decomposição em Valores Singulares

No capítulo anterior, o Teorema 2.8 mostrou que toda matriz A normal pode ser fatorada como $A = UDU^H$, onde U é unitária - ortogonal - e D é diagonal. Se a matriz A não é normal, o Decomposição Espectral não é possível. É possível, ainda, diagonalizar A, ou seja, fatorá-la como $A = PDP^{-1}$, onde D é diagonal mas P é somente inversível. Entretando, nem todas as matrizes são diagonalizáveis. A Decomposição em Valores Singulares é uma fatoração que pode ser aplicada a qualquer tipo de matriz, real ou complexa, quadrada ou não, hermitiana ou não, normal ou não.

A Decomposição em Valores Singulares — também denominada *SVD*, do inglês, *Singular Value Decomposition* — consiste na fatoração de uma matriz A como o produto $U\Sigma V^H$, onde U e V são matrizes unitárias e Σ é uma "matriz diagonal". Assim, como veremos, a SVD de uma matriz A de ordem $m \times n$ é o produto $U\Sigma V^H$, onde U e V são unitárias de ordem $m \times m$ e $n \times n$, respectivamente, e Σ é diagonal da mesma ordem de A. A extensão do conceito de matriz diagonal para matrizes retangulares será apresentada no decorrer deste capítulo.

3.1 A Criação da SVD

Existem cinco matemáticos que contribuíram grandemente para a teoria da SVD. Já citados anteriormente, Beltrami, Jordan, Sylvester, Schmidt e Weyl foram os responsáveis por estabelecer a existência da decomposição em valores singulares e pelo desenvolvimento da teoria. Beltrami, Jordan e Sylvester chegaram a esta decomposição através do que hoje chama-se de Álgebra Linear enquanto Schmidt e Weyl trabalharam com equações integrais.

As informações históricas apresentadas a seguir têm como base o artigo *On The Early History Of The Singular Value Decomposition* escrito por G. W. Stewart.

Eugênio Beltrami (1835 - 1899) Juntos, Beltrami e Jordan são os progenitores da decomposição em valores singulares. Eugenio foi responsável pela primeira publicação e Jordan foi quem completou e deu elegância a teoria.

Beltrami começou com a forma bilinear $f(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{x}^T A \boldsymbol{y}$, sendo A uma matriz real de ordem n. Substituindo $\boldsymbol{x} = U\boldsymbol{\xi}$ e $\boldsymbol{y} = V\boldsymbol{\eta}$ tem-se $f(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{\xi}^T S \boldsymbol{\eta}$, onde $S = U^T AV$. Percebendo o que acontecia quando U e V eram matrizes ortogonais e trabalhando com as equações

$$\det(AA^T - \sigma^2 I) = 0 \text{ e } \det(A^T A - \sigma^2 I) = 0,$$

onde σ são elementos da diagonal de S, ele chegou ao que se conhece, hoje, como decomposição em valores singulares.

No entanto, a teoria desenvolvida por ele era pouco formal.

Camille Jordan (1838 - 1921) Pode realmente ser considerado como o co-autor da decomposição em valores singulares. Mesmo tendo publicado sua descoberta depois de Beltrami, é evidente que seu trabalho é independente. De fato, "Mémoire sur les formes bilinéaires" trata de três problemas dos quais a redução de uma forma bilinear para um forma diagonal usando basicamente substituições ortogonais. Ele começou com a forma $P = \boldsymbol{x}^T A \boldsymbol{y}$ e procurou o máximo e o mínimo de P sujeito a $\|\boldsymbol{x}\|^2 = \|\boldsymbol{y}\|^2 = 1$.

Ao contrário de Beltrami, Jordan desenvolveu a teoria com economia e elegância, evitando certos passos que tornavam a aborgadem adotada por Beltrami pouco acessível.

James Joseph Sylvester (1814 - 1897) Sylvester escreveu uma nota e dois artigos tratando da SVD. A nota aparece no final de um artigo publicado em *The Messenger of Mathematics* intutulado "A new proof that a general quadric may be reduced to its canonical form (that is, a linear function of squares) by means of a real orthogonal substitution"[1]. No artigo ele descreve um algoritmo iterativo para reduzir uma forma quadrática a uma forma diagonal. Na nota, Sylvester aponta que uma iteração análoga

[1]Uma nova prova de que uma quádrica geral pode ser reduzida a sua forma canônica através de

pode ser usada para diagonalizar uma forma bilinear e, segundo ele, esta nota é a regra efetiva desta redução. A regra, na verdade, tinha muitos pontos em comum com o algoritmo apresentado por Beltrami.

Erhard Schmidt[2] (1876 - 1959) A abordagem adotada agora parte da Álgebra Linear e entra no campo das equações integrais. No tratamento destas equações, com núcleos não simétricos, Schmidt introduziu uma SVD análoga em dimensão infinita. Mas ele foi além da mera existência dessa decomposição mostrando como ela poderia ser usada para obter aproximações ótimas de posto inferior para um operador. Ao fazê-lo, ele transformou a SVD de uma curiosidade matemática em uma importante ferramenta teórica e computacional.

Hermann Weyl (1885 - 1955) A contribuição de Weyl foi desenvolver uma teoria geral de perturbações e usá-la para dar uma prova elegante do teorema das aproximações. Além disso, seu tratamento de equações integrais levou em conta o caso em que os núcleos eram simétricos.

Outros matemáticos também contribuíram para a SVD como, por exemplo, Autonne (1913) estendendo essa decomposição para matrizes complexas; e Eckart e Young (1936), desenvolvendo a teoria para matrizes retangulares.

3.2 Subespaços Fundamentais e Propriedades de A^HA e AA^H

Na Definição 2.1 do capítulo anterior considerava-se como núcleo de A o conjunto formado pelos vetores solução do sistema $A\boldsymbol{x} = \mathbf{0}$, ou seja, $N(A) = \{\boldsymbol{x} \in \mathbb{C}^n : A\boldsymbol{x} = \mathbf{0}\}$. O núcleo de uma matriz é um subespaço bastante importante do espaço vetorial das matrizes. A seguir serão apresentados mais alguns subespaços. Para tanto, considere $A \in \mathbb{C}^{m\times n}$.

Definição 3.1 (Espaço coluna) *O espaço coluna de A é o subespaço de $\mathbb{C}^m$ gerado*

uma substituição ortogonal real.

[2]Um dos criadores do Processo de Ortogonalização de Gram-Schmidt e que foi aluno de David Hilbert

pelas colunas de A. Assim

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

Definição 3.2 (Posto) *A dimensão de $R(A)$ chama-se posto da matriz A e é denotado por $posto(A)$.*

O posto de uma matriz também pode ser visto como o número de colunas linearmente independentes da matriz.

Além disso, se $posto(A) = min(m, n)$ diz-se que A tem *posto completo*.

Teorema 3.1 (do Posto) *Se $A \in \mathbb{R}^{m\times n}$ então $posto(A) + nulidade(A) = n$. Além disso, $posto(A) + nulidade(A^T) = m$.*

Prova: *A demonstração desse teorema pode ser encontrada em [6], página 437.* ■

O Teorema 3.1 também é válido quando $A \in \mathbb{C}^{m\times n}$. Neste caso, substitui-se A^T por A^H .

Maiores considerações acerca de *posto* e *nulidade* de matrizes com entradas complexas podem ser encontradas em *Matrix Analysis and Applied Linear Algebra*, de Carl D. Meyer, páginas 199 e 218.

Definição 3.3 (Núcleo à esquerda) *Dada matriz A, o núcleo à esquerda de A, ou $N(A^H)$, é o conjunto dos vetores solução do sistema $A^H\boldsymbol{x} = \boldsymbol{0}$, ou seja, $N(A^H) = \{\boldsymbol{x} \in \mathbb{C}^m : A^H\boldsymbol{x} = \boldsymbol{0}\}$.*

Definição 3.4 (Espaço linha) *O espaço linha de A é o conjunto $R(A^H) = \{\boldsymbol{y} \in \mathbb{C}^n : \boldsymbol{y} = A^H\boldsymbol{x}, \boldsymbol{x} \in \mathbb{C}^m\}$.*

Os subespaços $N(A)$, $R(A)$, $N(A^H)$ e $R(A^H)$ são denominados *subespaços fundamentais.*

Definição 3.5 (Complemento Ortogonal) *Seja M um subespaço de V, onde V é um espaço vetorial com produto interno. O complemento ortogonal de M, denotado por $M^\perp$, é o conjunto de todos os vetores de V que são ortogonais aos vetores de M, ou seja,*

$$M^\perp = \{\boldsymbol{x} \in V \quad ; \quad \langle \boldsymbol{m}, \boldsymbol{x}\rangle = 0 \quad \forall \boldsymbol{m} \in M\}.$$

Considere o teorema abaixo.

Teorema 3.2 *Se M é um subespaço de dimensão finita tal que $M \subseteq V$, onde V é um espaço vetorial com produto interno então*

$$V = M \cup M^{\perp} \text{ e } M \cap M^{\perp} = \{\mathbf{0}\}$$

Prova: *A demonstração deste teorema pode ser encontrada em [5], página 404.* ■

Quanto a estes subespaços fundamentais pode-se fazer as seguintes considerações:

(i) $N(A^H) \cup R(A) = \mathbb{C}^m$ e $N(A^H) \cap R(A) = \{\mathbf{0}\}$

(ii) $N(A) \cup R(A^H) = \mathbb{C}^n$ e $N(A) \cap R(A^H) = \{\mathbf{0}\}$

Para verificar o item (i), tem-se que encontrar os vetores $\boldsymbol{x}$ tais que $\boldsymbol{x} \in R(A)^{\perp}$.

Seja $\boldsymbol{x} \in R(A)^{\perp}$. Pela Definição 3.5 tem-se que $\langle A\widehat{\boldsymbol{x}}, \boldsymbol{x}\rangle = \mathbf{0}$, com $\widehat{\boldsymbol{x}} \in \mathbb{C}^n$ e $\boldsymbol{x} \in \mathbb{C}^m$. Assim

$$\langle A\widehat{\boldsymbol{x}}, \boldsymbol{x}\rangle = 0 \Rightarrow \boldsymbol{x}^H A\widehat{\boldsymbol{x}} = 0 \Rightarrow (A^H\boldsymbol{x})^H\widehat{\boldsymbol{x}} = 0 \Rightarrow \left\langle \widehat{\boldsymbol{x}}, A^H\boldsymbol{x}\right\rangle = 0 \tag{3.1}$$

Como $\widehat{\boldsymbol{x}}$ é um vetor qualquer então $\left\langle \widehat{\boldsymbol{x}}, A^H\boldsymbol{x}\right\rangle = 0$ se, e somente se, $A^H\boldsymbol{x} = \mathbf{0}$. Portanto, $\boldsymbol{x} \in N(A^H)$.

Assim, $R(A)^{\perp} \subset N(A^H)$. De forma análoga verifica-se que $R(A)^{\perp} \supset N(A^H)$. Logo, $R(A)^{\perp} = N(A^H)$.

Considere o caso em que $V = \mathbb{C}^m$ e $M = N(A^H)$. Então, do Teorema 3.2 e de $N(A^H) \subseteq \mathbb{C}^n$, tem-se que

$$N(A^H) \cup R(A) = \mathbb{C}^m \text{ e } N(A^H) \cap R(A) = \{\mathbf{0}\}$$

Para o item (ii), considere $V = \mathbb{C}^n$ e $M = N(A)$, já que de forma análoga ao desenvolvido na Equação (3.1) mostra-se que $R(A^H)^{\perp} = N(A)$.

Proposição 3.1 *Seja $A \in \mathbb{C}^{n\times n}$ hermitiana com posto$(A) = r$ e $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n\}$ os autovetores A associados aos autovalores $\lambda_1, \lambda_2, \ldots, \lambda_n$, respectivamente. Então $\lambda_1, \lambda_2, \ldots, \lambda_r$ são todos não-nulos e $\lambda_{r+1} = \lambda_{r+2} = \ldots = \lambda_n = 0$.*

Prova: Se $posto(A) = r$ então $dimN(A) = n-r$. Seja $p = n-r$ e considere $y_1, y_2, ..., y_p$ base de $N(A)$. Assim,

$$Ay_i = 0, i = 1, ..., p.$$

Portanto $y_i, i = 1, ..., p$ são autovetores de A associados a $\lambda = 0$.

Suponha que exista outro autovetor y_j associado a $\lambda = 0$ tal que $\{y_1, y_2, ..., y_p, y_j\}$ é LI. Então $dimN(A) = p + 1$, contradição. Então A tem $p = n - r$ autovetores LI associados a $\lambda = 0$. ∎

Lema 3.1 *Se $A = A^H$ então, para todo vetor $\boldsymbol{x} \in \mathbb{C}^n$, tem-se que $\boldsymbol{x}^H A\boldsymbol{x}$ é real.*

Prova: Apenas considere o transposto-conjugado de $\boldsymbol{x}^H A\boldsymbol{x}$. Mas como, a princípio, é um número complexo, $(\boldsymbol{x}^H A\boldsymbol{x})^H = \overline{\boldsymbol{x}^H A\boldsymbol{x}}$. Então tem- se a seguinte expressão:

$$\overline{\boldsymbol{x}^H A\boldsymbol{x}} = (\boldsymbol{x}^H A\boldsymbol{x})^H = \boldsymbol{x}^H A^H \boldsymbol{x} = \boldsymbol{x}^H A\boldsymbol{x}$$

pois A é hermitiana. Portanto, se um número complexo é igual ao seu conjugado, segue que ele é real ∎

Definição 3.6 (Matriz definida positiva) *Uma matriz hermitiana $A \in \mathbb{C}^{n\times n}$ é definida positiva se $\boldsymbol{x}^H A\boldsymbol{x} > 0$, $\forall \boldsymbol{x} \neq \mathbf{0}$ com $\boldsymbol{x} \in \mathbb{C}^n$.*

Definição 3.7 (Matriz semidefinida positiva) *Uma matriz hermitiana $A \in \mathbb{C}^{n\times n}$ é semidefinida positiva se $\boldsymbol{x}^H A\boldsymbol{x} \geq 0$, $\forall \boldsymbol{x} \in \mathbb{C}^n$.*

Definição 3.8 (Matriz definida negativa) *Uma matriz hermitiana $A \in \mathbb{C}^{n\times n}$ é definida negativa se $\boldsymbol{x}^H A\boldsymbol{x} < 0$, $\forall \boldsymbol{x} \in \mathbb{C}^n$ e $\boldsymbol{x} \neq \mathbf{0}$.*

Teorema 3.3 *Seja $A \in \mathbb{C}^{n\times n}$ hermitiana. Se A é definida positiva então todos os seus autovalores são positivos.*

Prova: Seja λ_i um autovalor da A com autovetor $\boldsymbol{x}_i$, então $A\boldsymbol{x}_i = \lambda_i \boldsymbol{x}_i$. Como A é definida positiva então $\boldsymbol{x}_i^H A\boldsymbol{x}_i > 0$. Note que

$$0 < \boldsymbol{x}_i^H A\boldsymbol{x}_i = \boldsymbol{x}_i^H \lambda_i \boldsymbol{x}_i = \boldsymbol{x}_i^H \boldsymbol{x}_i \lambda_i = \|\boldsymbol{x}_i\|^2 \lambda_i.$$

Como $\boldsymbol{x}_i$ é autovetor então é diferente do vetor nulo e, consequentemente $\|\boldsymbol{x}_i\|^2 > 0$. Assim, como $\|\boldsymbol{x}_i\|^2 \lambda_i > 0$ tem-se que $\lambda_i > 0$. ∎

Teorema 3.4 *Seja $A \in \mathbb{C}^{n\times n}$ hermitiana e semidefinida positiva então todos os seus autovalores são não negativos.*

Prova: A prova deste Teorema segue o mesmo raciocínio da do Teorema 3.3. ■

Quando se trabalha com matrizes retangulares, a chave está em considerar as matrizes A^HA e AA^H. A seguir, apresenta-se algumas propriedades dessas matrizes, relacionadas com as definições apresentadas anteriormente e que serão necessárias para a verificação da existência da Decomposição em Valores Singulares para qualquer matriz $A \in \mathbb{C}^{m\times n}$.

Observação 3.1 *Seja $A \in \mathbb{C}^{m\times n}$. Então são válidas cada uma das afirmações abaixo*

(i) $N(A^HA) = N(A)$ e $N(AA^H) = N(A^H)$.

De fato, tome $\boldsymbol{x} \in N(A)$. Então $A\boldsymbol{x} = \mathbf{0} \Rightarrow A^HA\boldsymbol{x} = A^H\mathbf{0} = \mathbf{0} \Rightarrow \boldsymbol{x} \in N(A^HA)$. Considere $\boldsymbol{x} \in N(A^HA)$. Então $A^HA\boldsymbol{x} = \mathbf{0} \Rightarrow \boldsymbol{x}^HA^HA\boldsymbol{x} = \boldsymbol{x}^H\mathbf{0} = 0 \Rightarrow (A\boldsymbol{x})^HA\boldsymbol{x} = 0 \Rightarrow \|A\boldsymbol{x}\|^2 = 0 \Rightarrow A\boldsymbol{x} = \mathbf{0}$. Ou seja, $\boldsymbol{x} \in N(A)$

Analogamente, mostra-se que $N(AA^H) = N(A^H)$.

(ii) $posto(A^HA) = posto(A) = posto(A^H) = posto(AA^H)$

Pelo Teorema 3.1 tem-se

$$posto(A^HA) = n - nulidade(A^HA) = n - nulidade(A) = posto(A)$$

já que $N(A^HA) = N(A)$.

Da mesma forma

$$posto(AA^H) = m - nulidade(AA^H) = n - nulidade(A^H) = posto(A^H).$$

Além disso

$$posto(A) + nulidade(A^H) = m \text{ e } posto(A^H) + nulidade(A^H) = m,$$

assim, $posto(A) = posto(A^H)$.

(iii) A^HA e AA^H são matrizes hermitianas e semidefinidas positivas.

De fato, A^HA e AA^H são hermitianas pois $(A^HA)^H = A^HA$ e $(AA^H)^H = AA^H$. Além disso, $\boldsymbol{x}^H A^H A\boldsymbol{x} = \|A\boldsymbol{x}\|^2 \geq 0$, $\forall \boldsymbol{x} \in \mathbb{C}^n$, e

$$\boldsymbol{x}^H AA^H \boldsymbol{x} = (A^H\boldsymbol{x})^H A^H \boldsymbol{x} = \|A\boldsymbol{x}\|^2 \geq 0, \forall \boldsymbol{x} \in \mathbb{C}^n.$$

Em particular, se $posto(A) = n$ então A^HA é hermitiana definida positiva e se $posto(A) = m$ então AA^H é hermitiana definida positiva.

(iv) Se λ é um autovalor de A^HA com autovetor $\boldsymbol{x}$ então λ é autovalor de AA^H com autovetor $A\boldsymbol{x}$.

De fato, $A^HA\boldsymbol{x} = \lambda\boldsymbol{x} \Rightarrow A(A^HA)\boldsymbol{x} = A\lambda\boldsymbol{x} \Rightarrow (AA^H)A\boldsymbol{x} = \lambda A\boldsymbol{x}$.

(v) A^HA e AA^H têm o mesmos autovalores não-nulos.

Seja λ autovalor não-nulo de A^HA, então de (iv) tem-se $AA^H(A\boldsymbol{x}) = \lambda(A\boldsymbol{x})$, ou seja, λ também é autovalor de AA^H.

Da mesma forma, sendo δ autovalor não-nulo de AA^H tem-se

$$AA^H\boldsymbol{x} = \delta\, \boldsymbol{x} \Rightarrow A^HA(A^H\boldsymbol{x}) = \delta(A^H\boldsymbol{x})$$

e, então, δ também é autovalor de A^HA.

Portanto, todos os autovalores não-nulos de A^HA são também de AA^H, e vice-versa.

(vi) Se $\boldsymbol{x}_1$ e $\boldsymbol{x}_2$ são autovetores da matriz A^HA, ortonormais entre si, então $A\boldsymbol{x}_1$ e $A\boldsymbol{x}_2$ também são ortogonais entre si.

Note que $A^HA\boldsymbol{x}_1 = \lambda_1\boldsymbol{x}_1$ e $A^HA\boldsymbol{x}_2 = \lambda_2\boldsymbol{x}_2$. Como $\boldsymbol{x}_1^H\boldsymbol{x}_2 = 0$ tem-se $(A\boldsymbol{x}_1)^H(A\boldsymbol{x}_2) = \boldsymbol{x}_1^H A^H A\boldsymbol{x}_2 = \boldsymbol{x}_1^H\lambda_2\boldsymbol{x}_2 = \lambda_2\boldsymbol{x}_1^H\boldsymbol{x}_2 = \lambda_2 0 = 0$

(vii) Se $\boldsymbol{x}$ é um autovetor de A^HA associado a um autovalor não-nulo λ então $A\boldsymbol{x}$ é um autovetor de AA^H associado ao mesmo autovalor.

De fato, $AA^H(A\boldsymbol{x}) = A(A^HA\boldsymbol{x}) = A(\lambda\boldsymbol{x}) = \lambda(A\boldsymbol{x})$.

3.3 A Decomposição em Valores Singulares de uma Matriz

A seguir, apresenta-se o teorema que mostra a existência da decomposição em valores singulares de uma matriz.

Definição 3.9 (Matriz diagonal generalizada) *Diz-se que* $A \in \mathbb{C}^{m \times n}$ *é diagonal generalizada se* $a_{ij} = 0$ *sempre que* $i \neq j$. *O conjunto* a_{ii}, *com* $i \in \{1, 2, \ldots, k\}$ *e* $k = min\{m, n\}$, *é denominado diagonal de* A.

Teorema 3.5 (Existência da decomposição em valores singulares) *Dada uma matriz* $A \in \mathbb{C}^{m \times n}$ *existem matrizes unitárias* $U \in \mathbb{C}^{m \times m}$ *e* $V \in \mathbb{C}^{n \times n}$, *e* $\Sigma \in \mathbb{R}^{m \times n}$ *matriz diagonal tal que* $A = U \Sigma V^H$.

Note que Σ não é uma matriz quadrada assim, Σ é uma matriz diagonal generalizada.

Prova: Para provar o Teorema 3.5 tem-se que demonstrar que U e V são matrizes unitárias, ou seja, que suas colunas formam um conjunto ortonormal (Observação 2.1), que Σ é diagonal generalizada e que $U^H A V = \Sigma$. Sem perda de generalidade, considere $posto(A) = r \leq n$ e $m \geq n$.

Seja $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\}$ autovetores ortonormais de $A^H A$ associados aos autovalores $\lambda_1, \lambda_2, \ldots, \lambda_n$, respectivamente.

Da Proposição 3.1 resulta que, como $posto(A^H A) = posto(A)$ então $\lambda_1, \lambda_2, \ldots, \lambda_r$ são todos não-nulos e $\lambda_{r+1} = \ldots = \lambda_n = 0$. Além disso, todos os autovalores de $A^H A$ são não negativos já que $A^H A$ é semidefinida positiva (Observação 3.1 item (iii)).

Assim, tem-se

$$\begin{aligned} A^H A \boldsymbol{v}_i &= \lambda_i \boldsymbol{v}_i \text{ para } i \in \{1, 2, \ldots, r\} \\ A^H A \boldsymbol{v}_i &= \lambda_i \boldsymbol{v}_i = 0 \boldsymbol{v}_i = \mathbf{0} \text{ para } i \in \{r+1, \ldots, n\}, \end{aligned}$$

com $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r$ ordenados de modo que $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_r$.

Seja V a matriz formada por $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\}$, autovetores ortonormais de $A^H A$, onde cada $\boldsymbol{v}_i$ é a i-ésima coluna de V. Então V é uma matriz unitária de ordem $n \times n$.

Define-se σ_i e $\boldsymbol{u}_i$ como

$$\sigma_i = \|A\boldsymbol{v}_i\|, \text{ para } i \in \{1, 2, \ldots, r\} \text{ e} \tag{3.2}$$

$$\boldsymbol{u}_i = \frac{A\boldsymbol{v}_i}{\sigma_i}, \text{ para } i \in \{1, 2, \ldots, r\}. \tag{3.3}$$

Note que se $i \in \{r+1, \ldots, n\}$ tem-se $\sigma_i = 0$ pois $A\boldsymbol{v}_i = \mathbf{0}$ já que $\boldsymbol{v}_i \in N(A^H A) = N(A)$. Neste caso, defini-se $\boldsymbol{u}_i = \mathbf{0}$.

Da Equação (3.2) retira-se

$$\sigma_i^2 = \|A\boldsymbol{v}_i\|^2 = (A\boldsymbol{v}_i)^H A\boldsymbol{v}_i = \boldsymbol{v}_i^H A^H A\boldsymbol{v}_i = \boldsymbol{v}_i^H \lambda_i \boldsymbol{v}_i = \lambda_i \boldsymbol{v}_i^H \boldsymbol{v}_i = \lambda_i. \tag{3.4}$$

Para encontrar a matriz U, seja $U_1 = \{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_r\}$. Tomando-se $\boldsymbol{u}_i$ e $\boldsymbol{u}_j$ do conjunto U_1 tem-se

$$\begin{aligned}\langle \boldsymbol{u}_i, \boldsymbol{u}_j \rangle &= \left\langle \frac{A\boldsymbol{v}_i}{\sigma_i}, \frac{A\boldsymbol{v}_j}{\sigma_j} \right\rangle = \left(\frac{A\boldsymbol{v}_i}{\sigma_i}\right)^H \frac{A\boldsymbol{v}_j}{\sigma_j} = \frac{1}{\sigma_i\sigma_j} \boldsymbol{v}_i^H A^H A\boldsymbol{v}_j = \\ &= \frac{1}{\sigma_i\sigma_j} \boldsymbol{v}_i^H \lambda_i \boldsymbol{v}_i = \frac{\lambda_i}{\sigma_i\sigma_j} \boldsymbol{v}_i^H \boldsymbol{v}_i = \begin{cases} \frac{\lambda_i}{\sigma_i\sigma_j} \text{se } i = j \\ 0 \text{ se } i \neq j \end{cases}.\end{aligned}$$

Pela Equação (3.4), se $i = j$ então $\dfrac{\lambda_i}{\sigma_i\sigma_j} = \dfrac{\lambda_i}{\lambda_i} = 1$. Ou seja,

$$\langle \boldsymbol{u}_i, \boldsymbol{u}_j \rangle = \begin{cases} 1 \text{ se } i = j \\ 0 \text{ se } i \neq j \end{cases}.$$

O que mostra que U_1 é um conjunto ortonormal.

Note que para $i \in \{1, 2, \ldots, r\}$, pela Equação (3.3), tem-se

$$A^H \boldsymbol{u}_i = A^H \left(\frac{A\boldsymbol{v}_i}{\sigma_i}\right) = \frac{1}{\sigma_i} A^H A\boldsymbol{v}_i = \frac{1}{\sigma_i} \lambda_i \boldsymbol{v}_i = \frac{1}{\sigma_i} \sigma_i^2 \boldsymbol{v}_i = \sigma_i \boldsymbol{v}_i,$$

ou seja, $A^H \boldsymbol{u}_i = \sigma_i \boldsymbol{v}_i$.

Além disso, $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_r\}$ são autovetores de AA^H associados aos autovalores não-nulos $\lambda_1, \lambda_2, \ldots, \lambda_r$ pois

$$AA^H \boldsymbol{u}_i = A\sigma_i \boldsymbol{v}_i = \sigma_i A\boldsymbol{v}_i = \sigma_i \sigma_i \boldsymbol{u}_i = \lambda_i \boldsymbol{u}_i.$$

Agora, como $posto(A^H A) = posto(AA^H) = r$ então $\dim(N(AA^H)) = m - r$.

Seja $U_2 = \{\boldsymbol{u}_{r+1} \ldots, \boldsymbol{u}_m\}$ uma base ortonormal de $N(AA^H)$.

Como estes $\boldsymbol{u}_i$ formam uma base $N(AA^H)$, são também elementos de $N(AA^H)$. Então $AA^H\boldsymbol{u}_i = \boldsymbol{0}$ e, assim, $\boldsymbol{u}_{r+1}\ldots,\boldsymbol{u}_m$ são autovetores de AA^H associados ao autovalor zero.

De fato, sendo $U_1 = \{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_r\}$ e $U_2 = \{\boldsymbol{u}_{r+1}\ldots,\boldsymbol{u}_m\}$, sabe-se que U_1 e U_2 são, separadamente, conjuntos ortonormais.

Tome $\boldsymbol{u}_i \in U_1$ e $\boldsymbol{u}_j \in U_2$ então

$$\langle \boldsymbol{u}_i, \boldsymbol{u}_j\rangle = \boldsymbol{u}_i^T\boldsymbol{u}_j = \left(\frac{A\boldsymbol{v}_i}{\sigma_i}\right)^H \boldsymbol{u}_j = \frac{1}{\sigma_i}\boldsymbol{v}_i^T A^T \boldsymbol{u}_j = 0,$$

pois, como $\boldsymbol{u}_j \in N(AA^H) = N(A^H)$ então $A^H\boldsymbol{u}_j = \boldsymbol{0}$.

Portanto, $\langle \boldsymbol{u}_i, \boldsymbol{u}_j\rangle = 0$ e assim, $U_1\cup U_2$ é um conjunto ortonormal.

Seja U a matriz formada por $U_1\cup U_2 = \{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_m\}$, autovetores ortonormais de AA^H, onde cada $\boldsymbol{u}_i$ é a i-ésima coluna de U. Então U é uma matriz unitária de ordem $m \times m$.

Considere, agora, o produto $\boldsymbol{u}_i^H A\boldsymbol{v}_i$. Então,

- se $i, j \leq r$: $\boldsymbol{u}_i^H A\boldsymbol{v}_j = \boldsymbol{u}_i^H \sigma_i \boldsymbol{u}_j = \sigma_i \boldsymbol{u}_i^H \boldsymbol{u}_j = \begin{cases} \sigma_i \text{ se } i = j \\ 0 \text{ se } i \neq j \end{cases}.$

- se $i > r$: $\boldsymbol{u}_i^H A\boldsymbol{v}_j = (A^H\boldsymbol{u}_i)^H\boldsymbol{v}_j = \boldsymbol{0}^H\boldsymbol{v}_j = 0.$

- se $j > r$: $\boldsymbol{u}_i^H A\boldsymbol{v}_j = \boldsymbol{u}_i^H\boldsymbol{0} = 0.$

Assim,

$$\boldsymbol{u}_i^H A\boldsymbol{v}_j = \begin{cases} \sigma_i & \text{se } i = j \text{ e } i, j \leq r \\ 0 & \text{caso contrário} \end{cases}.$$

Portanto, U^HAV é uma matriz diagonal de ordem $m \times n$. Ao denotar-se tal matriz diagonal por Σ tem-se $A = U\Sigma V^H$, já que U e V, definidas anteriormente, são unitárias.

■

Quando A é uma matriz real, trabalha-se com AA^T e A^TA, já que $\overline{A} = A$. Assim, a decomposição em valores singulares de A é o produto $U\Sigma V^T$, onde U e V são matrizes ortogonais.

Considerando as matrizes U, V e Σ que compõem a decomposição em valores singulares de uma matriz, pode-se fazer as seguintes observações:

Observação 3.2 *Seja $U\Sigma V^H$ a SVD de uma matriz A.*

(i) Os elementos na posição $i = j$ da matriz Σ, denotados por σ_i, são denominados valores singulares da matriz A e a matriz Σ pode ser representada como

$$\Sigma = \begin{bmatrix} D & O \\ O & O \end{bmatrix}, \tag{3.5}$$

onde

$$D = \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_r \end{bmatrix}.$$

De agora em diante será assumido que $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r > 0$ e O matriz nula com a devida ordem. Além disso, $r = \min\{m, n\}$ se, e somente se, A tem posto completo.

(ii) As colunas de U são denominadas vetores singulares à esquerda e são os autovetores ortonormais de AA^H.

(iii) As colunas de V são denominadas vetores singulares à direita e são os autovetores ortonormais de A^HA.

(iv) Os valores singulares não-nulos da matriz A são as raízes quadradas dos autovalores não-nulos de A^HA, que são os mesmos de AA^H.

(Equação (3.4)).

Observação 3.3

(v) A^H tem os mesmo valores singulares que A.

De fato, seja $A = U\Sigma V^T$ a decomposição em valores singulares de A. Então $A^H = V\Sigma^H U^H$. Considerando-se a matriz Σ apresentada em (3.5), tem-se que em Σ^H contém os mesmos valores positivos que Σ, e portanto A^H tem os mesmos valores singulares que A.

Observação 3.4

(vi) Os vetores singulares à direita (à esquerda) de A^H são os vetores singulares à esquerda (à direita) de A.

(vii) $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_r\}$ *é uma base ortonormal para* $R(A^H)$*;*

$\{\boldsymbol{v}_{r+1}, \ldots, \boldsymbol{v}_n\}$ *é uma base ortonormal para* $N(A)$*;*

$\{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_r\}$ *é uma base ortonormal para* $R(A)$*;*

$\{\boldsymbol{u}_{r+1}, \ldots, \boldsymbol{u}_m\}$ *é uma base ortonormal para* $N(A^H)$*.*

(viii) Os valores singulares $\sigma_1, \sigma_2, \ldots, \sigma_n$ *de* A *são únicos, no entanto, a SVD de uma matriz* A *não é única pois as bases anteriormente citadas não são únicas.*

Corolário 3.1 *Dada uma matriz* $A \in \mathbb{C}^{m \times n}$*,* $m \geq n$ *e* $posto(A) = r$*, existe* $\widetilde{U} \in \mathbb{C}^{m \times r}$ *e* $V \in \mathbb{C}^{r \times n}$ *matrizes ortogonais e* $\widetilde{\Sigma} \in \mathbb{R}^{r \times r}$ *matriz diagonal, com entradas positivas, tal que* $A = \widetilde{U}\widetilde{\Sigma}V^H$*.*

Este corolário defini a chamada *Decomposição em Valores Singulares Reduzida*, onde são considerados apenas os elementos positivos da diagonal de Σ e o conjunto ortonormal $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n\}$.

Observação 3.5 *Seja* $A \in \mathbb{C}^{m \times n}$*. Se* $m \geq n$ *tem-se que os valores singulares de* A *são as raízes quadradas dos autovalores de* $A^H A$ *e se* $m < n$*, os valores singulares de* A *são as raízes quadradas dos autovalores de* AA^H*.*

Revendo a prova do Teorema 3.5, segue que para encontrar a decomposição em valores singulares de uma matriz A, com $posto(A) = r$, é necessário seguir os passos seguintes:

Se $m \geq n$

1. Calcular a matriz $A^H A$, seus autovalores e encontrar os autovetores associados a cada autovalor. Assim, obtém-se a matriz V, onde cada coluna de V é um autovetor normalizado, já que $A^H A$ é hermitiana;

2. Calcular os valores singulares σ_i e montar a matriz Σ;

3. Obter os vetores $\boldsymbol{u}_i$ da seguinte forma: $\boldsymbol{u}_i = \dfrac{A\boldsymbol{v}_i}{\sigma_i}$, para $i \in \{1, 2, \ldots, r\}$ e, para $i \in \{r+1, \ldots, m\}$ calcular uma base ortonormal para $N(AA^H)$, ou seja, encontrar os autovetores de AA^H associados ao autovalor nulo. Então, construir a matriz U, onde sua i-ésima coluna é o vetor $\boldsymbol{u}_i$.

No caso em que $m < n$, calcula-se inicialmente os autovalores e autovetores de AA^H e, caso haja algum autovalor nulo, acha-se uma base ortonormal para $N(A^H A)$.

A seguir, apresenta-se um exemplo da decomposição em valores singulares de uma matriz real.

Exemplo 3.1 *Seja* $A = \begin{bmatrix} 2 & 0 & 1 \\ 0 & 2 & 0 \end{bmatrix}$. *Como* A *é real então* $AA^H = AA^T$. *Para encontrar a SVD de* A *serão seguidos os passos apresentados anteriormente:*

1. Encontrando a matriz V.

   Calculando AA^Tobtém-se

$$AA^T = \begin{bmatrix} 2 & 0 & 1 \\ 0 & 2 & 0 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 2 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 5 & 0 \\ 0 & 4 \end{bmatrix}.$$

   Assim, os autovalores de AA^T são dados por

$$p(\lambda) = (5 - \lambda)(4 - \lambda).$$

   Buscando-se as raízes do polinômio característico tem-se $\lambda_1 = 5$ e $\lambda_2 = 4$.

   Resolvendo $(A - \lambda_i I)\boldsymbol{x} = \mathbf{0}$ para cada λ_i, chega-se aos seguintes autovetores associados:

$$\boldsymbol{v}_1 = \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix} \text{ e } \boldsymbol{v}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}.$$

   Note que $\boldsymbol{v}_1$ e $\boldsymbol{v}_2$ formam um conjunto de vetores ortogonais. Vale ressaltar que, como AA^T é simétrica então seus autovetores são sempre ortogonais (Observação 2.6).

   Normalizando os autovetores encontrados tem-se:

$$\boldsymbol{V}_1 = \begin{bmatrix} \frac{2}{\sqrt{5}} \\ 0 \\ \frac{1}{\sqrt{5}} \end{bmatrix} \text{ e } \boldsymbol{V}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}.$$

   Assim, $V = \begin{bmatrix} \boldsymbol{V}_1 & \boldsymbol{V}_2 \end{bmatrix} = \begin{bmatrix} \frac{2}{\sqrt{5}} & 0 \\ 0 & 1 \\ \frac{1}{\sqrt{5}} & 0 \end{bmatrix}.$

2. Calculando os vetores singulares à esquerda e encontrando Σ.

   Como os autovalores de AA^T são 5 e 4 então os valores singulares de A são $\sigma_1 = \sqrt{5}$ e $\sigma_2 = \sqrt{4} = 2$.

   Assim, $\Sigma = \begin{bmatrix} \sqrt{5} & 0 & 0 \\ 0 & 2 & 0 \end{bmatrix}$ pois Σ tem a mesma ordem de A.

3. Encontrando a matriz U.

   Da Equação (3.3) tem-se que cada coluna de U é dada por $\boldsymbol{u}_i = \dfrac{A\boldsymbol{V}_i}{\sigma_i}$. Assim:

$$\begin{aligned} \boldsymbol{u}_1 &= \frac{A\boldsymbol{V}_1}{\sigma_1} = \frac{1}{\sqrt{5}} \begin{bmatrix} 2 & 0 & 1 \\ 0 & 2 & 0 \end{bmatrix} \begin{bmatrix} \frac{2}{\sqrt{5}} \\ 0 \\ \frac{1}{\sqrt{5}} \end{bmatrix} = \begin{bmatrix} 2 & 0 & 1 \\ 0 & 2 & 0 \end{bmatrix} \begin{bmatrix} \frac{2}{5} \\ 0 \\ \frac{1}{5} \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \\ \boldsymbol{u}_2 &= \frac{A\boldsymbol{V}_2}{\sigma_2} = \frac{1}{2} \begin{bmatrix} 2 & 0 & 1 \\ 0 & 2 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & \frac{1}{2} \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}. \end{aligned}$$

   Como $\boldsymbol{u}_1$ e $\boldsymbol{u}_2$ são ortonormais (vetores canônicos) então

$$U = \begin{bmatrix} \boldsymbol{u}_1 & \boldsymbol{u}_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Assim, a decomposição em valores singulares (reduzida) de A é

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \sqrt{5} & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} \frac{2}{\sqrt{5}} & 0 & \frac{1}{\sqrt{5}} \\ 0 & 1 & 0 \end{bmatrix}.$$

Note que, no Exemplo 3.1 não foi necessário calcular os autovetores de A^TA pois os autovalores de AA^T são todos diferentes de zero.

No exemplo a seguir, trabalha-se com autovalores nulos.

Exemplo 3.2 *Seja* $A = \begin{bmatrix} 1 & 0 & 1 \\ 0 & -3 & 0 \\ 1 & 0 & 1 \end{bmatrix}$

Calculando AA^Tobtém-se

$$A^TA = \begin{bmatrix} 2 & 0 & 2 \\ 0 & 9 & 0 \\ 2 & 0 & 2 \end{bmatrix}.$$

:

Assim, os autovalores e autovetores normalizados de A^TA são respectivamente

$$\begin{aligned} \boldsymbol{v}_1 &= \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}^T \text{ associado ao autovalor 9,} \\ \boldsymbol{v}_2 &= \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \end{bmatrix}^T \text{ associado ao autovalor 4 e} \\ \boldsymbol{v}_3 &= \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \end{bmatrix}^T \text{ associado ao autovalor 0.} \end{aligned}$$

Note que $\boldsymbol{v}_1$, $\boldsymbol{v}_2$ e $\boldsymbol{v}_3$ formam um conjunto de vetores ortogonais.

Assim, $V = \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \boldsymbol{v}_3 \end{bmatrix} = \begin{bmatrix} 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ 1 & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}$.

Como os autovalores de A^TA são 9, 4 e 0 então os valores singulares de A são $\sigma_1 = 3$ e $\sigma_2 = 2$.

Assim,

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

pois Σ tem a mesma ordem de A. Vale lembrar que 0 não é valor singular, mas que ele aparece na diagonal de Σ para que esta tenha a mesma ordem de A. Neste caso, poderia se considerar a SVD reduzida de A, obtendo-se, então

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

Cada coluna de U é dada por $\boldsymbol{u}_i = \dfrac{A\boldsymbol{v}_i}{\sigma_i}$, $i \in \{1, 2\}$. Assim:

$$\begin{aligned}
\boldsymbol{u}_1 &= \frac{A\boldsymbol{v}_1}{\sigma_1} = \frac{1}{3}\begin{bmatrix} 1 & 0 & 1 \\ 0 & -3 & 0 \\ 1 & 0 & 1 \end{bmatrix}\begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix} = \frac{1}{3}\begin{bmatrix} 0 \\ 3 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \quad \text{e} \\
\boldsymbol{u}_2 &= \frac{A\boldsymbol{v}_2}{\sigma_2} = \frac{1}{2}\begin{bmatrix} 1 & 0 & 1 \\ 0 & -3 & 0 \\ 1 & 0 & 1 \end{bmatrix}\begin{bmatrix} \frac{1}{\sqrt{2}} \\ 0 \\ \frac{1}{\sqrt{2}} \end{bmatrix} = \frac{1}{2}\begin{bmatrix} \frac{2}{\sqrt{2}} \\ 0 \\ \frac{2}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ 0 \\ \frac{1}{\sqrt{2}} \end{bmatrix}.
\end{aligned}$$

Note que $\boldsymbol{u}_1$ e $\boldsymbol{u}_2$ são ortonormais.

Agora é necessário encontrar o vetor $\boldsymbol{u}_2$, que não pode ser construído da mesma forma que $\boldsymbol{u}_1$ e $\boldsymbol{u}_2$ pois $\boldsymbol{v}_3$ está associado ao autovalor 0.

Então calcula-se o autovetor de AA^T que esteja associado ao autovalor nulo e que seja ortogonal com $\boldsymbol{u}_1$ e $\boldsymbol{u}_2$.

No entanto, neste caso como a matriz A é simétrica então $AA^T = A^TA$ e, portanto, o autovetor procurado é exatamente $\boldsymbol{v}_3 = \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \end{bmatrix}$.

Como $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{v}_3\}$ formam um conjunto ortonormal, estes formarão a matriz U.

Assim,

$$U = \begin{bmatrix} \boldsymbol{u}_1 & \boldsymbol{u}_2 & \boldsymbol{v}_3 \end{bmatrix} = \begin{bmatrix} 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -1 & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}.$$

Assim, a decomposição em valores singulares de A é

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

Seja A é uma matriz diagonal tal que

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

Exemplo 3.3 *A SVD de A é*

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

Exemplo 3.4 *Seja A é uma matriz linha tal que*

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

A SVD de A é

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

Proposição 3.2 *Se $A \in \mathbb{C}^{n\times n}$ é hermitiana com autovalores $\lambda_1, \lambda_2, \ldots, \lambda_n$ então seus valores singulares são $|\lambda_1|, |\lambda_2|, \ldots, |\lambda_n|$.*

Prova: Como λ_i, $i \in \{1, 2, \ldots, n\}$, é autovalor de A então. Como A é hermitiana então $A^H A = A^2$. Assim, os autovalores de $A^H A$ são os mesmos de A^2. Pela Proposição 2.4 tem-se que os autovalores de A^2 são $\lambda_1^2, \lambda_2^2, \ldots, \lambda_n^2$, ou seja, as autovalores de $A^H A$ são $\lambda_1^2, \lambda_2^2, \ldots, \lambda_n^2$. Como os valores singulares σ_i de A são as raízes quadradas dos autovalores de $A^T A$ então

$$\sigma_i = \sqrt{\lambda_i^2} = |\lambda_i|.$$

■

Proposição 3.3 *Se A é hermitiana definida positiva então a SVD de A é a decomposição espectral de A.*

Prova: Seja $\lambda_1, \lambda_2, \ldots, \lambda_n$ os autovalores de A. Como A é hermitiana, tem-se que:

(i) os valores singulares de A são $|\lambda_1|, |\lambda_2|, \ldots, |\lambda_n|$, em particular, pelo Teorema 3.4, os valores singulares de A são $\sigma_i = \lambda_i$, já que os autovalores de A são todos não negativos. Assim, se $A = U\Sigma V^T$, tem-se que $\Sigma = D$, onde D é a matriz diagonal da decomposição espectral de A, ou seja, a matriz que contém os autovalores de A.

(ii) $A = A^T$, e assim, $A^T A = A^2$. Assim, se $A = QDQ^T$ tem-se que $Q = U = V$, pois $A^2 = QD^2Q^T$ e, pela simetria de A, $A^2 = AA^T = A^T A$.

Assim, $A = U\Sigma V^T = QDQ^T$. ■

Observação 3.6 (Produto Externo) *A decomposição em valores singulares de uma matriz* $A \in \mathbb{C}^{m \times n}$ *pode ser descrita da seguinte forma:*

$$A = \sigma_1 \boldsymbol{u}_1 \boldsymbol{v}_1^H + \sigma_2 \boldsymbol{u}_2 \boldsymbol{v}_2^H + \ldots + \sigma_r \boldsymbol{u}_r \boldsymbol{v}_r^H,$$

onde $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_r$ *são os vetores singulares à esquerda,* $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r$ *são os vetores singulares à direita e* $\sigma_1 \geq \ldots \geq \sigma_r$ *e* $\sigma_{r+1} = \ldots = \sigma_n$.

Mesmo tendo sido descoberto no fim do século XIX, a Decomposição em Valores Singulares não era amplamente conhecida na Matemática Aplicada até a década de 60, quando mostrou-se que ela poderia ser computada eficazmente. O que a faz tão importante é a sua vasta aplicabilidade. No capítulo 4 são apresentadas algumas de suas aplicações, tanto de caráter teórico quanto prático.

Capítulo 4

Aplicações da SVD

A SVD é excelente para cálculos que exigem certa estabilidade. Uma das razões é o fato das matrizes U e V serem matrizes ortogonais, ou seja, preservam a norma de um vetor.

O fato da Decomposição em Valores Singulares não ser única não traz nenhuma consequência na maioria de suas aplicações.

4.1 Normas de Matrizes

A Decomposição em Valores Singulares pode fornecer fórmulas simples para certas expressões que envolvem normas de matrizes. As normas apresentadas neste trabalho são a Norma Espectral (ou Norma-2), e a Norma de Frobenius.

Definição 4.1 (Norma Espectral ou Norma-2) *Seja $A \in \mathbb{R}^{m\times n}$. Chama-se de Norma Espectral ou Norma-2 de uma matriz a norma induzida pela norma euclidiana, denotada por $\|A\|_2$. Neste caso,*

$$\|A\|_2 = \max_{\boldsymbol{x}\neq\mathbf{0}} \frac{\|A\boldsymbol{x}\|_2}{\|\boldsymbol{x}\|_2},$$

onde $\|\boldsymbol{x}\|_2 = \sqrt{x_1^2 + x_2{}^2 + \ldots + x_n{}^2}$.

Observação 4.1 *Utilizando a Norma Espectral*

(i) Se Q é ortogonal então $\|Q\|_2 = 1$.

De fato, da Proposição 2.3, $\|Q\boldsymbol{x}\|^2 = \|\boldsymbol{x}\|^2$.

(ii) $\|A\|_2 = \max_{\|\boldsymbol{x}\|=1} \|A\boldsymbol{x}\|_2$

Definição 4.2 (Norma de Frobenius) *Seja $A \in \mathbb{R}^{m\times n}$. A Norma de Frobenius da matriz A, denotada por $\|A\|_F$, é definida por*

$$\|A\|_F = \sqrt{\sum_{j=1}^{n}\sum_{i=1}^{m} |a_{ij}|^2} = \sqrt{tr(A^TA)} = \sqrt{tr(AA^T)}.$$

A Norma de Frobenius é obtida quando colocamos as elementos da matriz em um vetor e então calculamos a norma euclidiana. Em outras palavras $\|A\|_F$ é a raiz quadrada da soma dos quadrados dos elementos de A. [6]

Teorema 4.1 *Seja $A \in \mathbb{R}^{m\times n}$ com valores singulares $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0$. Então $\|A\|_2 = \sigma_1$ e $\|A\|_F^2 = \sigma_1^2 + \sigma_2^2 + \ldots + \sigma_n^2$.*

Prova: Considere a norma espectral $\|A\|_2$ e a seguinte igualdade:

$$\|A\|_2 = \max_{\boldsymbol{x}\neq \boldsymbol{0}} \frac{\|A\boldsymbol{x}\|_2}{\|\boldsymbol{x}\|_2} = \max_{\boldsymbol{x}\neq \boldsymbol{0}} \sqrt{\frac{\boldsymbol{x}^TA^TA\boldsymbol{x}}{\boldsymbol{x}^T\boldsymbol{x}}}.$$

Seja decomposição em valores singulares de A dada por $A = U\Sigma V^T$.

Note que

$$\begin{aligned} \frac{\boldsymbol{x}^TA^TA\boldsymbol{x}}{\boldsymbol{x}^T\boldsymbol{x}} &= \frac{\boldsymbol{x}^T(U\Sigma V^T)^TU\Sigma V^T\boldsymbol{x}}{\boldsymbol{x}^T\boldsymbol{x}} = \\ &= \frac{\boldsymbol{x}^TV\Sigma^TU^TU\Sigma V^T\boldsymbol{x}}{\boldsymbol{x}^T\boldsymbol{x}} = \\ &= \frac{\boldsymbol{x}^TV\Sigma^T\Sigma V^T\boldsymbol{x}}{\boldsymbol{x}^TVV^T\boldsymbol{x}} = \frac{(V^T\boldsymbol{x})^T\Sigma^T\Sigma(V^T\boldsymbol{x})}{(V^T\boldsymbol{x})^T(V^T\boldsymbol{x})}. \end{aligned}$$

Denotando $V^T\boldsymbol{x}$ por $\boldsymbol{y}$, vetor com n entradas, e $\Sigma^T\Sigma$ por B, matriz diagonal contendo $\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2$, tem-se

$$\frac{\boldsymbol{x}^TA^TA\boldsymbol{x}}{\boldsymbol{x}^T\boldsymbol{x}} = \frac{\boldsymbol{y}^TB\boldsymbol{y}}{\boldsymbol{y}^T\boldsymbol{y}}. \tag{4.1}$$

Sejam $\lambda_1, \lambda_2, \ldots, \lambda_n$ os autovalores de B então $\lambda_i = \sigma_i^2$.

Considerando a Equação (4.1) e denotando por $y_{1,}y_{2,}\ldots, y_n$ as i-ésimas coordenadas do vetor $\boldsymbol{y}$ tem-se

$$\frac{\boldsymbol{y}^TB\boldsymbol{y}}{\boldsymbol{y}^T\boldsymbol{y}} = \frac{\lambda_1y_1^2 + \lambda_2y_2^2 + \ldots + \lambda_ny_n^2}{y_1^2 + y_2^2 + \ldots + y_n^2}.$$

Como $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ obtém-se a seguinte desigualdade:

$$\lambda_1 = \frac{\lambda_1y_1^2 + \lambda_1y_2^2 + \ldots + \lambda_1y_n^2}{y_1^2 + y_2^2 + \ldots + y_n^2} \geq \frac{\lambda_1y_1^2 + \lambda_2y_2^2 + \ldots + \lambda_ny_n^2}{y_1^2 + y_2^2 + \ldots + y_n^2}.$$

Assim

$$\lambda_1 \geq \frac{\lambda_1 y_1^2 + \lambda_2 y_2^2 + \ldots + \lambda_n y_n^2}{y_1^2 + y_2^2 + \ldots + y_n^2} = \frac{\boldsymbol{y}^T B \boldsymbol{y}}{\boldsymbol{y}^T \boldsymbol{y}} = \frac{\boldsymbol{x}^T A^T A \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}}.$$

Ou seja,

$$\left(\frac{\|A\boldsymbol{x}\|_2}{\|\boldsymbol{x}\|_2} \right)^2 = \frac{\boldsymbol{x}^T A^T A \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}} \leq \lambda_1.$$

Então λ_1 é um limitante superior para a norma espectral.

Agora, é preciso encontrar um vetor $\widetilde{\boldsymbol{x}}$ tal que

$$\frac{\|A\widetilde{\boldsymbol{x}}\|_2}{\|\widetilde{\boldsymbol{x}}\|_2} = \sqrt{\lambda_1} = \sigma_1.$$

No entanto, do Teorema da Decomposição em Valores Singulares, sabe-se que $\sigma_i = \|A\boldsymbol{v}_i\|$. Além disso, $\boldsymbol{v}_i$ é um vetor unitário. Assim, tomando $\widetilde{\boldsymbol{x}} = \boldsymbol{v}_i$,

$$\frac{\|A\widetilde{\boldsymbol{x}}\|_2}{\|\widetilde{\boldsymbol{x}}\|_2} = \frac{\|A\boldsymbol{v}_1\|_2}{\|\boldsymbol{v}_1\|_2} = \|A\boldsymbol{v}_1\|_2 = \|A\boldsymbol{v}_1\| = \sigma_1.$$

Portanto, encontrou-se um vetor $\boldsymbol{x}$ tal que

$$\|A\|_2 = \max_{\boldsymbol{x} \neq \boldsymbol{0}} \frac{\|A\boldsymbol{x}\|_2}{\|\boldsymbol{x}\|_2} = \max_{\boldsymbol{x} \neq \boldsymbol{0}} \sqrt{\frac{\boldsymbol{x}^T A^T A \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}}} = \sqrt{\lambda_1} = \sigma_1.$$

Agora será mostrado que $\|A\|_F^2 = \sigma_1^2 + \sigma_2^2 + \ldots + \sigma_n^2$. Seja $\|A\|_F^2 = tr(A^T A)$ e $A = U\Sigma V^T$.

Sabe-se que se $B, C \in \mathbb{R}^{n \times n}$ então $tr(BC) = tr(CB)$. Assim

$$tr(A^T A) = tr[(U\Sigma V^T)^T U\Sigma V^T] = tr(V\Sigma^T \Sigma V^T) = tr(\Sigma^T \Sigma V^T V) = tr(\Sigma^T \Sigma).$$

Ou seja, $\|A\|_F^2 = \sigma_1^2 + \sigma_2^2 + \ldots + \sigma_n^2$. ■

Corolário 4.1 *Seja $A \in \mathbb{R}^{m \times n}$ então $\|A\|_2 \leq \|A\|_F \leq \sqrt{n}\,\|A\|_2$.*

Prova: Do Teorema 4.1 tem-se $\|A\|_2 = \sigma_1$ e $\|A\|_F^2 = \sigma_1^2 + \sigma_2^2 + \ldots + \sigma_n^2$.

Então

$$\begin{aligned} \|A\|_2 = \sigma_1 \leq \sqrt{\sigma_1^2} \quad &\leq \quad \sqrt{\sigma_1^2 + \sigma_2^2 + \ldots + \sigma_n^2} = \|A\|_F \\ \|A\|_2 \quad &\leq \quad \|A\|_F \end{aligned} \tag{4.2}$$

e

$$\begin{aligned} \|A\|_F = \sqrt{\sigma_1^2 + \sigma_2^2 + \ldots + \sigma_n^2} \quad &\leq \quad \sqrt{\sigma_1^2 + \sigma_1^2 + \ldots + \sigma_1^2} = \sqrt{n\sigma_1^2} = \sigma_1\sqrt{n} = \sqrt{n}\,\|A\|_2 \\ \|A\|_F \quad &\leq \quad \sqrt{n}\,\|A\|_2 \end{aligned} \tag{4.3}$$

Das Equações (4.2) e (4.3) obtém-se

$$\|A\|_2 \leq \|A\|_F \leq \sqrt{n}\,\|A\|_2 .$$

■

A Decomposição em Valores Singulares de uma matriz A pode também ser usada para encontrar uma matriz B, com $posto(B) = r$, tal que $\|A - B\|_F$ seja a menor possível. Ou seja, dado o posto de uma matriz, achar a matriz B com esse posto que esteja mais "próxima" da matriz A em relação a norma de Frobenius.

Considera-se $A \in \mathbb{R}^{n\times n}$ e $A = U\Sigma V^T$ sua decomposição em valores singulares, onde Σ é a matriz que contém os valores singulares $(\sigma_1, \sigma_2, \ldots, \sigma_r, \sigma_{r+1}, \ldots, \sigma_n)$.

Agora, seja Σ_r matriz diagonal com elementos $(\sigma_1, \sigma_2, \ldots, \sigma_r, 0, \ldots, 0)$ e defina $A_r = U\Sigma_r V^T$.

Assim,

$$\|A - A_r\|_F = \sqrt{\sigma_{r+1}^2 + \sigma_{r+2}^2 + \ldots + \sigma_n^2}$$

Assumindo $\{\sigma_1, \sigma_2, \ldots, \sigma_r\}$ não-nulos, o teorema a seguir mostra que a matriz B de posto r mais próxima de A é exatamente a matriz A_r. Mas antes considere o seguinte lema.

Lema 4.1 *Seja A uma matriz $m \times n$ ´e Q uma matriz de ordem m e ortogonal. Então $\|QA\|_F^2 = \|A\|_F^2$.*

Prova: *Usando a definição da norma de Frobenius tem-se*

$$\|QA\|_F^2 = tr((QA)^T QA) = tr(A^T Q^T QA) = tr(A^T A) = \|A\|_F^2$$

■

Teorema 4.2 *Sejam A e A_r as matrizes definidas acima. Então*

$$\|A - A_r\|_F = \min_{posto(B)=r} \|A - B\|_F .$$

Prova: Seja $U\Sigma V^T$ a SVD da matriz A e seja B uma matriz de posto r.

Note que $posto(B) = posto(U^T BV)$. De fato, seja B uma matriz tal que $posto(B) = r$. Como $nulidade(B) = nulidade(U^T BV)$, pois V é ortogonal e, consequentemente, inversível, tem-se que $posto(B) = posto(U^T BV)$.

Então, defina $C = U^T B V$. Assim, tem-se a seguinte equação:

$$\|A - B\|_F = \left\|U\Sigma V^T - UCV^T\right\|_F = \left\|U(\Sigma - C)V^T\right\|_F = \|\Sigma - C\|_F \,.$$

Dessa forma, minimizar $||A - B||_F$ sujeito a $posto(B) = r$ é equivalente a minimizar $||\Sigma - C||_F$ sujeito a $posto(C) = r$. Portanto, o problema consiste em minimizar $\|\Sigma - C\|_F$, sujeito a $posto(C) = r$. Como Σ é diagonal, $\|\Sigma - C\|_F$ é mínimo quando C também for diagonal e mais, $C = \Sigma_r$, já que os σ_i estão organizados em ordem não crescente ao longo da diagonal de Σ.

Logo, minimo $\|\Sigma - C\|_F$ sujeito a $posto(C) = r$ é igual a $\|\Sigma - \Sigma_r\|_F$.

Consequentemente, voltando para a variável B, temos que a solução de minimo$||A - B||$ sujeito a $posto(B) = r$ é

$$B = U\Sigma_r V^T = A_r.$$

■

O teorema anterior mostra que a matriz A_r é a matriz de posto r que está mais próxima de A em relação a norma de Frobenius e que esta distância é medida para valores singulares da matriz A.

4.2 O Problema de Mínimos Quadrados e a Pseudo-Inversa

4.2.1 O Problema de Mínimos Quadrados

Para efeito ilustrativo, pode-se pensar o métodos de mínimos quadrados como um método para encontrar um curva - função - que melhor se ajuste a um conjunto de pontos dados. Esse método era conhecido no século XVIII por Cotes[1] e, em 1806, Legrende[2]

[1]Roger Cotes (1682 - 1716) foi um matemático inglês que, enquanto esteve em Cambridge, editou a segunda edição do *Principia*, de Newton. Apesar de ter publicado pouco, fez importantes descobertas na teoria do logaritmo, cálculo integral e métodos numéricos.

[2]Adrien Marie Legendre (1752 - 1833) foi um matemático francês que trabalhou com astronomia, teoria de números e funções eliptícas.

publicou um artigo sobre ele em um livro sobre órbitas de cometas. Entretanto, em 1809, Gauss apresentou um artigo descrevendo o método, que alegou já conhecer desde 1795. Em 1801, Gauss calculou o reaparecimento de um novo asteróide, que havia sido descoberto no dia de ano novo daquele ano, atráves do método de mínimos quadrados e por isso, geralmente, recebe os créditos por esse método de aproximação. Esse método de ajuste de dados tem sido um ferramenta indispensável desde sua invenção, extendendo seu uso em diversas áreas da Matemática.

"Melhor se ajuste" no contexto do método de mínimos quadrados significa obter o menor erro possível. Cada ponto dado contém um erro em relação a curva escolhida para se ajustar ao conjunto total de dados, podendo este erro ser nulo no caso em que o ponto pertence a curva. Com o erro de cada um dos pontos do conjunto de dados, podemos formar um **vetor-erro**.

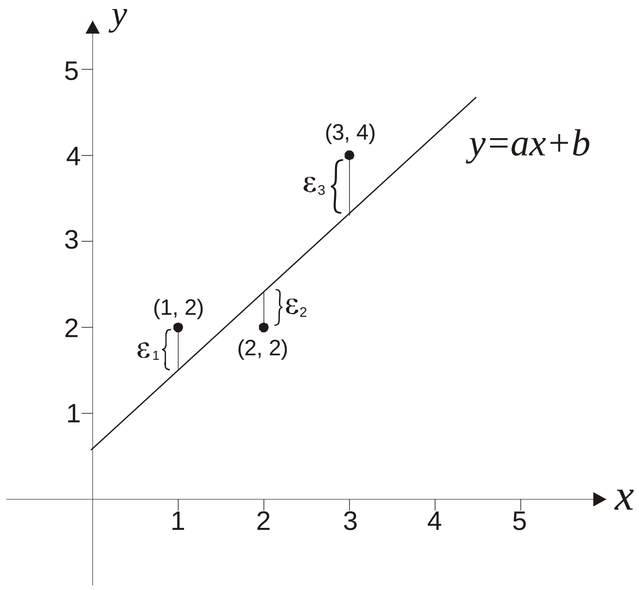

Figura 4.1: Encontrando a reta que minimiza o vetor-erro.

Considere a conjunto de pontos $P = \{(1,2),(2,2),(3,4)\}$ e uma reta $y = ax + b$ para ser a função que melhor de ajuste ao conjunto P, conforme a Figura 4.1.

Como se busca o menor erro possível, então $\|\boldsymbol{e}\|$ deve estar o mais próximo de zero. Utilizando a norma euclidiana e sendo $\boldsymbol{e} = \begin{bmatrix} \varepsilon_1 & \varepsilon_2 & \varepsilon_3 \end{bmatrix}^T$, tem-se que minimizar

$$\|\boldsymbol{e}\| = \sqrt{\varepsilon_1 + \varepsilon_2 + \varepsilon_3}. \tag{4.4}$$

É da Equação (4.4) que provém o termo "mínimos quadrados". O número $\|\boldsymbol{e}\|$ é chamado de **erro quadrático mínimo** da aproximação.

4.2.2 A solução do Problema

Seja P_n um conjunto de n pontos dados e r uma reta cuja equação é $y = ax + b$. Então o vetor-erro é

$$\boldsymbol{e} = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix},$$

onde $\varepsilon_i = y_i - (ax + b)$. Obtém-se com todas as equações

$$y_i = ax_i + b, 1 \leq i \leq n,$$

um sistema de duas incógnitas e n equações, o qual se pode escrever na notação matricial $A\boldsymbol{x} = \boldsymbol{b}$ onde

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

Logo, o vetor-erro é o vetor $\boldsymbol{b} - A\boldsymbol{x}$ e minimizá-lo significa minimizar $\boldsymbol{b} - A\boldsymbol{x}$.

Assim, a solução para problema de mínimos quadrados pode ser escrita segundo a definição a seguir.

Definição 4.3 *Seja $A \in \mathbb{R}^{m \times n}$ e $\boldsymbol{b} \in \mathbb{R}^m$. Uma solução do problema de mínimos quadrados é o vetor $\boldsymbol{x} \in \mathbb{R}^n$ tal que $\|\boldsymbol{b} - A\boldsymbol{x}\|$ seja mínima.*

Qualquer vetor da forma $A\boldsymbol{x}$ está no espaço coluna de A e, quando $\boldsymbol{x}$ varia sobre todos os vetores de $\mathbb{R}^n$, significa que $A\boldsymbol{x}$ varia sobre todos os vetores de $R(A)$. Assim, uma solução por mínimos quadrados é equivalente a um vetor $\bar{\boldsymbol{y}} \in R(A)$ tal que

$$\|\boldsymbol{b} - \bar{\boldsymbol{y}}\| \leq \|\boldsymbol{b} - \boldsymbol{y}\|,$$

para todo $\boldsymbol{y} \in R(A)$. No entanto, sabe-se que o vetor no espaço coluna de A mais próximo de $\boldsymbol{b}$ é exatamente a projeção de $\boldsymbol{b}$ neste subespaço, ou seja, se $\bar{\boldsymbol{x}}$ é uma solução por mínimos quadrados tem-se $A\bar{\boldsymbol{x}} = proj_{R(A)}(\boldsymbol{b})$.

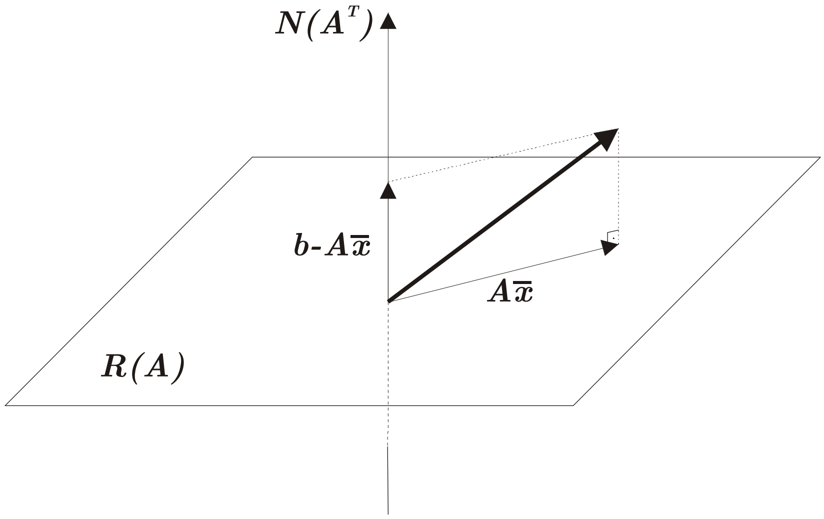

Figura 4.2: Representação geométrica do problema de mínimos quadrados.

Considere a Figura 4.2. Se $A\bar{\boldsymbol{x}}$ é a projeção de $\boldsymbol{b}$ em $R(A)$ então $\boldsymbol{b} - A\bar{\boldsymbol{x}}$ é ortogonal a $R(A)$, isto é,

$$(\boldsymbol{b} - A\bar{\boldsymbol{x}}) \in N(A^T).$$

Assim

$$\begin{aligned}
(\boldsymbol{b} - A\bar{\boldsymbol{x}}) &\in N(A^T) \\
A^T(\boldsymbol{b} - A\bar{\boldsymbol{x}}) &= \mathbf{0} \\
A^T\boldsymbol{b} - A^TA\bar{\boldsymbol{x}} &= \mathbf{0} \\
A^TA\bar{\boldsymbol{x}} &= A^T\boldsymbol{b}. \qquad (4.5)
\end{aligned}$$

A Equação (4.5) representa um sistema de equações conhecido como **equações normais** para $\bar{\boldsymbol{x}}$. Então $\bar{\boldsymbol{x}}$ é uma solução por mínimos quadrado de $A\boldsymbol{x} = \boldsymbol{b}$ se, e somente se, é solução da equação normal $A^TA\bar{\boldsymbol{x}} = A^T\boldsymbol{b}$.

Tratando-se de uma matriz genérica $A \in \mathbb{R}^{m\times n}$ não está se considerando o posto desta matriz. Observando-se o posto de uma matriz tem-se que:

- Se $A \in \mathbb{R}^{n\times n}$ e $posto(A) = n$, então A é inversível e a solução de $A\boldsymbol{x} = \boldsymbol{b}$ é $\boldsymbol{x} = A^{-1}\boldsymbol{b}$;

- Se $A \in \mathbb{R}^{m\times n}$, com $m > n$ e $posto(A) = n$, então A^TA é inversível e a solução de $A^TA\boldsymbol{x} = A^T\boldsymbol{b}$ é $\boldsymbol{x} = (A^TA)^{-1}A^T\boldsymbol{b}$;

- Se $A \in \mathbb{R}^{m\times n}$, com $m > n$ e $posto(A) = r < n$, então o sistema $A^T A\boldsymbol{x} = A^T\boldsymbol{b}$ tem infinitas soluções.

Geralmente, tratando-se de aproximações, a melhor solução para este problema é a solução de menor norma e é neste sentido que a decomposição em valores singulares pode auxiliar na solução do problema de mínimos quadrados.

4.2.3 A Matriz Pseudo-inversa e a SVD

Definição 4.4 (Solução de mínima norma) *Seja $A \in \mathbb{R}^{m\times n}$ e $\boldsymbol{b} \in \mathbb{R}^m$. A solução de mínima norma é o vetor $\bar{\boldsymbol{x}}$ em $\mathbb{R}^n$ tal que $\|\boldsymbol{b} - A\bar{\boldsymbol{x}}\| \leq \|\boldsymbol{b} - A\boldsymbol{x}\|$ para todo $\boldsymbol{x} \in \mathbb{R}^n$ e se $\boldsymbol{x}' \neq \bar{\boldsymbol{x}}$ é tal que $\|\boldsymbol{b} - A\boldsymbol{x}'\| \leq \|\boldsymbol{b} - A\boldsymbol{x}\|$, para todo $\boldsymbol{x} \in \mathbb{R}^n$, então $\|\bar{\boldsymbol{x}}\| \leq \|\boldsymbol{x}'\|$.*

Definição 4.5 (Matriz Pseudo-inversa) *Seja $A \in \mathbb{R}^{m\times n}$ e $A = U\Sigma V^T$. Chama-se de pseudo inversa (ou inversa generalizada) a matriz $A^+ = V\Sigma^+U^T$, onde*

$$\Sigma^+ = \begin{bmatrix} \frac{1}{\sigma_1} & 0 & & \cdots & & 0 \\ 0 & \ddots & 0 & & \vdots & \\ & 0 & \frac{1}{\sigma_r} & 0 & & \vdots \\ \vdots & & 0 & 0 & 0 & \\ & \vdots & & 0 & \ddots & 0 \\ 0 & & \cdots & & 0 & 0 \end{bmatrix} = \begin{bmatrix} D^{-1} & O \\ O & O \end{bmatrix},$$

com D dada na Observação 3.2 (i).

Além disso, se A tem dimensão $m \times n$, então $A^+ \in \mathbb{R}^{n\times m}$ e, consequentemente, $\Sigma^+ \in \mathbb{R}^{n\times m}$.

Exemplo 4.1 *Considere a matriz A apresentada no Exemplo 3.1 então*

$$A = \begin{bmatrix} 2 & 0 & 1 \\ 0 & 2 & 0 \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}}_{U} \underbrace{\begin{bmatrix} \sqrt{5} & 0 \\ 0 & 2 \end{bmatrix}}_{\Sigma} \underbrace{\begin{bmatrix} \frac{2}{\sqrt{5}} & 0 & \frac{1}{\sqrt{5}} \\ 0 & 1 & 0 \end{bmatrix}}_{V^T}.$$

Calculando a matriz pseudo-inversa de A, tem-se que $A^+ = V\Sigma^+U^T$. Então

$$A^+ = V\Sigma^+U^T = \begin{bmatrix} \frac{2}{\sqrt{5}} & 0 \\ 0 & 1 \\ \frac{1}{\sqrt{5}} & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{5}} & 0 \\ 0 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} \frac{2}{5} & 0 \\ 0 & \frac{1}{2} \\ \frac{1}{5} & 0 \end{bmatrix}.$$

De fato,

$$AA^+ = \begin{bmatrix} 2 & 0 & 1 \\ 0 & 2 & 0 \end{bmatrix} \begin{bmatrix} \frac{2}{5} & 0 \\ 0 & \frac{1}{2} \\ \frac{1}{5} & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Teorema 4.3 *O problema de mínimos quadrados $A\boldsymbol{x} = \boldsymbol{b}$ possui uma única solução $\bar{\boldsymbol{x}}$ de mínima norma por mínimos quadrados, que é dada por $\bar{\boldsymbol{x}} = A^+\boldsymbol{b}$.*

Prova: Seja $A \in \mathbb{R}^{m\times n}$ com SVD dada por $A = U\Sigma V^T$ e valores singulares $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$, onde $posto(A) = r$.

Seja $\boldsymbol{v} = V^T\boldsymbol{x}$ e $\boldsymbol{c} = U^T\boldsymbol{b}$. Considere $\boldsymbol{v}$ e $\boldsymbol{c}$ tais que

$$\boldsymbol{v} = \begin{bmatrix} \boldsymbol{v}_1 \\ \boldsymbol{v}_2 \end{bmatrix} \text{ e } \boldsymbol{c} = \begin{bmatrix} \boldsymbol{c}_1 \\ \boldsymbol{c}_2 \end{bmatrix},$$

onde $\boldsymbol{v}_1, \boldsymbol{c}_1 \in \mathbb{R}^r$. Então

$$\begin{gathered} \|\boldsymbol{b} - A\boldsymbol{x}\|^2 \\ \left\|U^T(\boldsymbol{b} - AVV^T\boldsymbol{x})\right\|^2 \\ \left\| \begin{bmatrix} \boldsymbol{c}_1 \\ \boldsymbol{c}_2 \end{bmatrix} - \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{v}_1 \\ \boldsymbol{v}_2 \end{bmatrix} \right\|^2 \\ \left\| \begin{matrix} \boldsymbol{c}_1 - D\boldsymbol{v}_1 \\ \boldsymbol{c}_2 \end{matrix} \right\|^2, \end{gathered} \tag{4.6}$$

utilizando o fato que a norma euclidiana é invariante por matrizes unitárias, apresentado na Proposição 2.3.

Observando a Equação (4.6), percebe-se que o vetor que minimiza $\|\boldsymbol{b} - A\boldsymbol{x}\|^2$ tem $\boldsymbol{v}_1 = \Sigma^{-1}\boldsymbol{c}_1$. Como o vetor $\boldsymbol{v}_2$ é arbitrário, pode-se tomar $\boldsymbol{v} = \begin{bmatrix} D^{-1}\boldsymbol{c}_1 \\ \boldsymbol{0} \end{bmatrix}$ e $\boldsymbol{v}' = \begin{bmatrix} D^{-1}\boldsymbol{c}_1 \\ \boldsymbol{v}_2 \end{bmatrix}$, com $\boldsymbol{v}_2 \neq \boldsymbol{0}$, vetores distintos que minimizam $\|\boldsymbol{b} - A\boldsymbol{x}\|^2$. E ainda,

$$\|\boldsymbol{v}\|^2 = \left\|D^{-1}\boldsymbol{c}_1\right\|^2 < \left\|D^{-1}\boldsymbol{c}_1\right\|^2 + \|\boldsymbol{v}_2\|^2 = \|\boldsymbol{v}'\|^2 .$$

Como $\boldsymbol{x} = V\boldsymbol{v}$ e V é ortogonal então $\|\boldsymbol{v}\|^2 = \|V\boldsymbol{v}\|^2 = \|\boldsymbol{x}\|^2$ e assim, $\|\boldsymbol{x}\|^2 < \|\boldsymbol{x}'\|^2$. Portanto,

$$\boldsymbol{x} = V\boldsymbol{v} = V\begin{bmatrix} D^{-1}\boldsymbol{c}_1 \\ \boldsymbol{0} \end{bmatrix} = V\begin{bmatrix} D^{-1} & 0 \\ 0 & 0 \end{bmatrix}\boldsymbol{c} = V\begin{bmatrix} D^{-1} & 0 \\ 0 & 0 \end{bmatrix}U^T\boldsymbol{b} = A^+\boldsymbol{b}. \tag{4.7}$$

■

> O Teorema 4.3 mostra como a decomposição em valores singulares de uma matriz A pode ser usada para resolver o problema de mínimos quadrados, mesmo quando a A não tem posto completo. Na verdade, a prova do Teorema mostra como encontrar o conjunto de todas as soluções . Elas são dadas por $V\boldsymbol{v}'$ onde $\boldsymbol{v}'$ é qualquer vetor da forma $\begin{bmatrix} D^{-1}\boldsymbol{c}_1 \\ \boldsymbol{v}_2 \end{bmatrix}$. No entanto, se $\boldsymbol{v}_2 = \boldsymbol{0}$ obtém-se a única solução de mínima norma. (...) Assim, a fórmula (4.7) representa uma maneira prática de resolver o problema de mínimos quadrados. ([8], pág 324)

Proposição 4.1 *Seja $A \in \mathbb{R}^{m\times n}$. Se $m < n$ e posto$(A) = m$ então*

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

Por outro lado, se $m > n$ e posto$(A) = n$ então

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

Prova: Supondo *posto*$(A) = m$ tem-se

$$A^T(AA^T)^{-1} = V\Sigma^TU^T(U\Sigma\Sigma^TU^T)^{-1} = V\Sigma^TU^TU(\Sigma\Sigma^T)^{-1}U^T = V\Sigma^T(\Sigma\Sigma^T)^{-1}U^T.$$

Note que

$$\Sigma^T(\Sigma\Sigma^T)^{-1} = \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_m \\ \vdots & \vdots & 0 \\ 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} \frac{1}{\sigma_1^2} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \frac{1}{\sigma_m^2} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sigma_1} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \frac{1}{\sigma_m} \\ \vdots & \vdots & 0 \\ 0 & 0 & 0 \end{bmatrix} = \Sigma^+.$$

Então $A^T(AA^T)^{-1} = A^+$.

Supondo $posto(A) = n$ tem-se

$$(A^TA)^{-1}A^T = (V\Sigma^T\Sigma V^T)^{-1}V\Sigma^TU^T = V(\Sigma^T\Sigma)^{-1}V^TV\Sigma^TU^T = V(\Sigma^T\Sigma)^{-1}\Sigma^TU^T.$$

Da mesma maneira, mostra-se que $(\Sigma^T\Sigma)^{-1}\Sigma^T = \Sigma^+$. E , portanto, $(A^TA)^{-1}A^T = A^+$. ■

Observação 4.2 *A solução por mínimos quadrados no caso em que $A \in \mathbb{R}^{m\times n}$, com $m > n$ e $posto(A) = n$, é $\boldsymbol{x} = A^+\boldsymbol{b}$.*

Observação 4.3 *Se $A \in \mathbb{R}^{n\times n}$ é inversível então $A^+ = A^{-1}$*

Seja $A = U\Sigma V^T$ a decomposição em valores singulares de A com

$$\Sigma = \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_n \end{bmatrix}$$

e $\sigma_i \neq 0$ para todo $1 \leq i \leq n$.

De fato, se A é inversível então

$$A^{-1} = (U\Sigma V^T)^{-1} = V\Sigma^{-1}U^T.$$

No entanto,

$$\Sigma^{-1} = \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_n \end{bmatrix}^{-1} = \begin{bmatrix} \frac{1}{\sigma_1} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \frac{1}{\sigma_n} \end{bmatrix} = \Sigma^+.$$

Portanto $A^{-1} = V\Sigma^+U^T = A^+$.

A observação anterior mostra um resultado esperado no caso em que A é inversível: o de que a pseudo-inversa da matriz A é exatamente a matriz inversa já conhecida.

Observação 4.4 *Seja $A \in \mathbb{R}^{m\times n}$ e A^+ a pseudo-inversa de A. Então $(A^+)^+ = A$, com as mesmas matrizes U e V.*

De fato, seja $A = U\Sigma V^T$ a SVD de A. Considere $m > n$. Então

$$(A^+)^+ = (V\Sigma^+U^T)^+ = V(\Sigma^+)^+U^T,$$

pois U e V são ortogonais (e, portanto, inversíveis). Assim

$$(\Sigma^+)^+ = \begin{bmatrix} \frac{1}{\sigma_1} & 0 & 0 & \cdots & 0 \\ 0 & \ddots & 0 & \cdots & 0 \\ 0 & 0 & \frac{1}{\sigma_n} & 0 & 0 \end{bmatrix}^+ = \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_n \\ \vdots & \vdots & 0 \\ 0 & 0 & 0 \end{bmatrix} = \Sigma$$

Definição 4.6 (Pseudo-inversa de Moore-Penrose) *Seja $A \in \mathbb{R}^{m\times n}$. A matriz pseudo-inversa de Moore-Penrose de A é a matriz X que satisfaz as seguintes condições:*

(i) $AXA = A$

(ii) $XAX = X$

(iii) $(AX)^T = AX$

(iv) $(XA)^T = XA$

Proposição 4.2 *A matriz pseudo-inversa de Moore-Penrose é única.*

Prova: Sejam X e Y, ambas não-nulas, matrizes pseudo-inversas de Moore-Penrose de uma matriz A então ambas satisfazem as condições (i) até (iv) da Definição 4.6. Considere agora as seguintes equações:

$$\begin{aligned} AX &= AX \\ AX &= AYAX \\ (AX)^T &= (AYAX)^T \\ AX &= AXAY \\ AX &= AY \\ YAX &= YAY \\ YAX &= Y \end{aligned} \tag{4.8}$$

utilizando as condições (i) e (iii); e

$$
\begin{aligned}
XA &= XA \\
XA &= XAYA \\
(XA)^T &= (XAYA)^T \\
XA &= YAXA \\
XA &= Y\ A \\
XAX &= YAX \\
X &= YAX \qquad (4.9)
\end{aligned}
$$

que utiliza as condições (i) e (iv).

Assim, das Equações (4.8) e (4.9) tem-se $X = YAX = Y$.

Portanto, a pseudo-inversa de Moore-Penrose é única. ■

Proposição 4.3 *Seja $A \in \mathbb{R}^{m\times n}$ então a matriz pseudo-inversa A^+ de A satisfaz condições de Moore-Penrose.*

Prova: Supondo $posto(A) = r$, $r \leq \max\{m, n\}$ e $A = U\Sigma V^T$, será verificado que A^+, dada por $V\Sigma^+U^T$, satisfaz as condições de (i) a (iv).

(i) $AA^+A = A(V\Sigma^+U^T)A = U\Sigma V^T(V\Sigma^+U^T)U\Sigma V^T = U\Sigma\Sigma^+\Sigma V^T = U\Sigma V^T = A$ se $\Sigma\Sigma^+\Sigma = \Sigma$. De fato:

$$
\begin{aligned}
\Sigma\Sigma^+\Sigma &= \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix}_{m\times n} \begin{bmatrix} D^{-1} & O \\ O & O \end{bmatrix}_{n\times m} \begin{bmatrix} D & O \\ O & O \end{bmatrix}_{m\times n} = \\
&= \begin{bmatrix} I_r & O \\ O & O \end{bmatrix}_{m\times m} \begin{bmatrix} D & O \\ O & O \end{bmatrix}_{m\times n} = \qquad (4.10) \\
&= \begin{bmatrix} D & O \\ O & O \end{bmatrix}_{m\times n} = \Sigma,
\end{aligned}
$$

com $D \in \mathbb{R}^{r\times r}$ e cada matriz nula O com a devida ordem.

(ii) $A^+AA^+ = (V\Sigma^+U^T)A(V\Sigma^+U^T) = (V\Sigma^+U^T)U\Sigma V^T(V\Sigma^+U^T) = V\Sigma^+\Sigma\Sigma^+U^T = A^+$ se $\Sigma^+\Sigma\Sigma^+ = \Sigma^+$. A exibição deste resultado segue o mesmo raciocínio da Equação (4.10).

(iii) $(AA^+)^T = [U\Sigma V^T(V\Sigma^+U^T)]^T = (U\Sigma\Sigma^+U^T)^T = U(\Sigma^+)^T\Sigma^TU^T$. Note que se $(\Sigma^+)^T\Sigma^T = \Sigma\Sigma^+$ tem-se

$$(AA^+)^T = U(\Sigma^+)^T\Sigma^TU^T = U\Sigma\Sigma^+U^T = U\Sigma V^TV\Sigma^+U^T = AA^+.$$

De fato

$$\begin{aligned}(\Sigma^+)^T\Sigma^T &= \begin{bmatrix} D^{-1} & O \\ O & O \end{bmatrix}_{m\times n} \begin{bmatrix} D & O \\ O & O \end{bmatrix}_{n\times m} = \\ &= \begin{bmatrix} D & O \\ O & O \end{bmatrix}_{m\times n} \begin{bmatrix} D^{-1} & O \\ O & O \end{bmatrix}_{n\times m} = \Sigma\Sigma^+,\end{aligned}$$

adequando-se as ordens das matrizes nulas O.

(iv) Análogo ao item anterior.

$\blacksquare$

Note que quando $posto(A) = n$ (ou $posto(A) = m$) pode-se usar a Proposição 4.1 e $A^+ = (A^TA)^{-1}A^T$ (ou $A^+ = A^T(AA^T)^{-1}$). Neste caso, as condições de Moore - Penrose são provadas facilmente, apenas substituindo A^+ pela devida representação. Por exemplo, no item (i) tem-se

$$AA^+A = A[(A^TA)^{-1}A^T]A = A[(A^TA)^{-1}(A^TA)] = AI = A,$$

no caso em que $posto(A) = n$ e

$$AA^+A = A[A^T(AA^T)^{-1}]A = [AA^T(AA^T)^{-1}]A = A,$$

no caso em que $posto(A) = m$.

Corolário 4.2 *A^+ é a Pseudo-Inversa de Moore-Penrose.*

Este fato decorre das Proposições 4.2 e 4.3.

Pode-se também utilizar as condições de Moore Penrose para caracterizar a solução do problema de mínimos quadrados, conforme sugere o enunciado abaixo. Vale ressaltar que, como o solução de mínima norma é única, a solução obtida utilizando as condições é a mesma apresentada anteriormente, ou seja, a matriz A^+.

Outra maneira de verificar que a solução de mínima norma é dada por $\boldsymbol{x} = A^+\boldsymbol{b}$ é considerando o seguinte problema.

Problema 4.1 *Dada matriz $A \in \mathbb{R}^{m\times n}$ e $\mathbf{b} \in \mathbb{R}^m$, encontrar $\boldsymbol{x} \in \mathbb{R}^n$ de menor norma que minimize $\|A\boldsymbol{x} - \boldsymbol{b}\|^2$.*

Resolução Considere o problema de min $\|A\boldsymbol{x} - \boldsymbol{b}\|$, sabendo que se $\boldsymbol{x}$ é solução deste problema então satistaz a equação normal $A^TA\boldsymbol{x} = A^T\boldsymbol{b}$. Então, utilizando as condições (i) e (iii) de Moore-Pensore:

$$\begin{aligned} A^TA\boldsymbol{x} &= A^T\boldsymbol{b} \\ A^TA\boldsymbol{x} &= A^T(A^+)^TA^T\boldsymbol{b} \\ A^TA\boldsymbol{x} &= A^T(AA^+)^T\boldsymbol{b} \\ A^TA\boldsymbol{x} &= A^TAA^+\boldsymbol{b}. \end{aligned}$$

Assim, de $A^TA(\boldsymbol{x} - A^+\boldsymbol{b}) = \mathbf{0}$, tem-se que $\boldsymbol{x} - A^+\boldsymbol{b} \in N(A^TA)$. Ou seja,

$$\boldsymbol{x} - A^+\boldsymbol{b} \in N(A). \tag{4.11}$$

Agora, tem-se que encontrar a solução de mínima norma, ou seja, min $\|\boldsymbol{x}\|^2$ sujeito a $A^TA\boldsymbol{x} - A^T\boldsymbol{b} = 0$. Para tanto, considere as funções f e g definidas da seguinte forma:

$$\begin{aligned} f \; : \; & \mathbb{R}^n \longrightarrow \mathbb{R} \\ & \boldsymbol{x} \longmapsto \|\boldsymbol{x}\|^2 \end{aligned}$$

e

$$\begin{aligned} g \; : \; & \mathbb{R}^n \longrightarrow \mathbb{R}^n \\ & \boldsymbol{x} \longmapsto A^TA\boldsymbol{x} - A\boldsymbol{b} \end{aligned}$$

Considerando as condições de otimalidade de Lagrange[3], f tem um mínimo em x_0 quando existe $\boldsymbol{\lambda} \in \mathbb{R}^n$ tal que

$$\nabla f(\boldsymbol{x}_0) = \nabla g(\boldsymbol{x}_0)\boldsymbol{\lambda}.$$

Sabendo que $\nabla f(\boldsymbol{x}) = 2\boldsymbol{x}$ e que $\nabla g(\boldsymbol{x}) = A^TA$ tem-se que f tem um mínimo em $\boldsymbol{x}_0$ quando $2\boldsymbol{x}_0 = A^TA\boldsymbol{\lambda}$.

[3]Mais detalhes em Vector Calculus, de J. E. Marsden e A. J. Tromba. O símbolo $\nabla g(x_0)$ denota o Jacobiano de g avaliado em x_0.

Utilizando as condições citadas, obtém-se que a solução de mínima norma, de maneira mais geral, deve ser da forma

$$\begin{cases} \boldsymbol{x} = A^T A\boldsymbol{\lambda}, \text{ para algum } \boldsymbol{\lambda} \in \mathbb{R}^n, \end{cases}$$

ou equivalentemente, $\boldsymbol{x} = A^T\boldsymbol{y}$ com $\boldsymbol{y} \in \mathbb{R}^m$ onde $\boldsymbol{y} = A\boldsymbol{\lambda}$.

$$\begin{aligned} \boldsymbol{x} - A^+\boldsymbol{b} &= \boldsymbol{x} - A^+AA^+\boldsymbol{b} \Rightarrow \\ \Rightarrow \boldsymbol{x} - A^T(A^+)^T A^+\boldsymbol{b} &= A^T\boldsymbol{y} - A^T(A^+)^T A^+\boldsymbol{b} \Rightarrow \\ \Rightarrow A^T(\boldsymbol{y} - (A^+)^T A^+\boldsymbol{b}) &= A^T\boldsymbol{y}', \end{aligned}$$

ou seja, $\boldsymbol{x} - A^+\boldsymbol{b} = A^T\boldsymbol{y}'$, utilizando as condições de Moore-Penrose (ii) e (iv) e $\boldsymbol{y}' = \boldsymbol{y} - (A^+)^T A^+\boldsymbol{b}$. Assim,

$$\boldsymbol{x} - A^+\boldsymbol{b} \in R(A^T). \tag{4.12}$$

Portanto, das Equações (4.11) e (4.12), $\boldsymbol{x} - A^+\boldsymbol{b} = \boldsymbol{0}$, ou seja, $\boldsymbol{x} = A^+\boldsymbol{b}$.

4.3 Decomposição Polar

Todo número complexo é o produto de um número não negativo r e um número $e^{i\theta}$ no círculo unitário: $z = re^{i\theta}$. Esta é a representação de z em coordenada polares. Entendendo um número complexo como uma matriz 1×1, r corresponderia a uma matriz hermitiana semidefinida positiva e $e^{i\theta}$ seria uma matriz unitária. Ou seja, $e^{i\theta}$ sendo complexo e $e^{-i\theta}\ e^{i\theta} = 1$.

Teorema 4.4 (Decomposição Polar) *Toda matriz $A \in \mathbb{C}^{n\times n}$ pode ser fatorada como $A = QS$ onde $Q \in \mathbb{C}^{n\times n}$ é unitária e $S \in \mathbb{C}^{n\times n}$ é hermitiana semidefinida positiva. Se A é inversível então S é hermitiana definida positiva.*

Também é possível fatorar um matriz A como $A = S'Q$ onde S' é uma matriz é hermitiana semidefinida positiva e Q é unitária. Esse tipo de fatoração é conhecida como Decomposição Polar Reversa.

Exemplo 4.2 *Seja* $A = \begin{bmatrix} 1 & -2 \\ 3 & -1 \end{bmatrix}$. *A decomposição polar e a decomposição polar reversa de* A *são, respectivamente:*

$$A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 3 & -1 \\ -1 & 2 \end{bmatrix} = QS \text{ e } A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} = S'Q.$$

Note que S *e* S' *são hermitianas definidas positivas e* Q *é unitária.*

Dada a decomposição em valores singulares de uma matriz A, obtém-se facilmente sua decomposição polar.

Neste caso considere $A \in \mathbb{R}^{n\times n}$ e $posto(A) = r$. Seja $A = U\Sigma V^T$ a SVD da matriz A. Então pode-se escrever A como

$$A = UV^TV\Sigma V^T = QS,$$

visto que V é ortogonal.

Note que $Q = UV^T$ também é uma matriz ortogonal pois

$$(UV^T)^TUV^T = VU^TUV^T = VV^T = I = UV^TVU^T = UV^T(UV^T)^T,$$

e $S = V\Sigma V^T$ é simétrica semidefinida positiva.

De fato, denotando-se $V^T\boldsymbol{x}$ por $\boldsymbol{y}$ tem-se

$$\boldsymbol{x}^TV\Sigma V^T\boldsymbol{x} = (V^T\boldsymbol{x})^T\Sigma(V^T\boldsymbol{x}) = \boldsymbol{y}^T\Sigma\boldsymbol{y} = y_1^2\sigma_1 + \ldots + y_r^2\sigma_r, \tag{4.13}$$

onde cada y_i é a i-ésima coordenada do vetor $\boldsymbol{y}$ e $\sigma_1, \ldots, \sigma_r$ são os valores singulares não-nulos de A.

Como cada σ_i é raiz quadrada de um autovalor não-nulo de A^TA, então $\sigma_i > 0$, $\forall i$. Além disso $y_1^2, \ldots, y_r^2 \geq 0$.

Portanto, da Equação (4.13)

$$\boldsymbol{x}^TV\Sigma V^T\, \boldsymbol{x} \geq 0.$$

Quando A é inversível tem-se que S é simétrica definida positiva pois A tem posto completo e, consequentemente, tem $p = min\{m, n\}$ valores singulares não-nulos.

Observação 4.5 *Se* $A \in \mathbb{R}^{n\times n}$ *for escrita como* $A = U\Sigma U^T U V^T$ *obtém-se a Decomposição Polar Reversa* $A = S'Q$*, onde* $S' = U\Sigma U^T$ *é simétrica semidefinida positiva e* $Q = UV^T$ *é ortogonal.*

Exemplo 4.3 *Considere a matriz* A *dada no Exemplo 3.1. Como* A *é real então* $A^H = A^T$*. Então*

$$A = \begin{bmatrix} 2 & 0 & 1 \\ 0 & 2 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \sqrt{5} & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} \frac{2}{\sqrt{5}} & 0 & \frac{1}{\sqrt{5}} \\ 0 & 1 & 0 \end{bmatrix} = U\Sigma V^T.$$

Assim

$$\begin{aligned} UV^T &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{2}{\sqrt{5}} & 0 & \frac{1}{\sqrt{5}} \\ 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} \frac{2}{\sqrt{5}} & 0 & \frac{1}{\sqrt{5}} \\ 0 & 1 & 0 \end{bmatrix} \quad e \\ V\Sigma V^T &= \begin{bmatrix} \frac{2}{5}\sqrt{5} & 0 \\ 0 & 1 \\ \frac{1}{5}\sqrt{5} & 0 \end{bmatrix} \begin{bmatrix} \sqrt{5} & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} \frac{2}{\sqrt{5}} & 0 & \frac{1}{\sqrt{5}} \\ 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} \frac{4}{\sqrt{5}} & 0 & \frac{2}{\sqrt{5}} \\ 0 & 2 & 0 \\ \frac{2}{\sqrt{5}} & 0 & \frac{1}{\sqrt{5}} \end{bmatrix}. \end{aligned}$$

Observe que UV^T *é ortogonal pois tem colunas ortonormais pois*

$$\left\langle \begin{bmatrix} \frac{2}{\sqrt{5}} \\ 0 \\ \frac{1}{\sqrt{5}} \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right\rangle = 0.$$

E ainda, $V\Sigma V^T$ *é simétrica semidefinida positiva.*

De fato, dado $\boldsymbol{x} \in \mathbb{R}^3$*,* $\boldsymbol{x} = \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix}$*, tem-se*

$$\begin{aligned} \boldsymbol{x} V\Sigma V^T \boldsymbol{x}^T &= \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \begin{bmatrix} \frac{4}{\sqrt{5}} & 0 & \frac{2}{\sqrt{5}} \\ 0 & 2 & 0 \\ \frac{2}{\sqrt{5}} & 0 & \frac{1}{\sqrt{5}} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \\ &= \frac{4}{\sqrt{5}} x_1^2 + \frac{4}{\sqrt{5}} x_1\ x_3 + 2x_2^2 + \frac{1}{\sqrt{5}} x_3^2 = \\ &= \frac{1}{\sqrt{5}} (4x_1^2 + 4x_1\ x_3 + x_3^2) + 2x_2^2 = \\ &= \frac{1}{\sqrt{5}} (2x_1 + x_3)^2 + 2x_2^2 \geq 0. \end{aligned}$$

Pode-se também encontrar a decomposição em valores singulares de uma matriz A por meio da decomposição polar de A. No caso em que A é real, sendo S simétrica e igual a $V\Sigma V^T$, tem-se

$$A = QS = QV\Sigma V^T = U\Sigma V^T,$$

onde $U = QV$.

A Decomposição Polar é utilizada na construção de *block reflectors, que* são uma generalização da transformação de Householder[4]. Essa decomposição é também usada em Mecânica Contínua (*Continuum Mechanics*) e, mais recentemente, em Robótica.

4.4 Compressão de Imagens Digitais

Das muitas aplicações da SVD, uma das que mais se destaca é seu uso para comprimir imagens digitais de modo que elas possam ser transmitidas eletronicamente com eficiência, por satélite, fax, Internet e outros meios.

Detectar e corrigir erros de transmissão são problemas comuns quando se trabalha com transmissão de dados em geral. No entanto, agora, o objetivo é reduzir o total de informações a serem transmitidas, sem perder nenhuma informação essencial.

Imagine um satélite levando recebendo um imagem de tamanho 340×280 *pixels*[5] e enviando-a para a Terra. Trabalhando-se com uma imagem em preto e branco, sabe-se que cada *pixel* é um dos 256 tons de cinza, que são representados por números entre 0 e 255. As informação em uma imagem podem ser entendidas como uma matriz A de ordem 340×280. Assim, transmitir a imagem significa manipular 95.200 números, o que é bastante trabalhoso. O ideal seria transmitir somente as informações relevantes na imagem.

A idéia por trás da compreensão de imagens é que algumas partes da figura são menos importantes que outras. Por exemplo, em uma fotografia de um barco no mar pode haver muito céu e mar de plano de fundo, enquanto o desenho da embarcação contém muitos detalhes. Provavelmente, se forem enviados todos os *pixels* que compõem o barco e for reduzido o números de *pixels* do plano de fundo, pouca diferença será notada na imagem recebida.

Com a Decomposição em Valores Singulares é possível encontrar essas informações relevantes e diminuir consideravelmente o número de *pixels* a serem enviados. A res-

[4]Mais detalhes sobre esta transformação em [3].

[5]Um *pixel* é o menor elemento num dispositivo de exibição (como por exemplo um monitor), ao qual é possivel atribuir-se uma cor. Ou seja, um *pixel* é o menor ponto que forma uma imagem digital, sendo que o conjunto de milhares de pixels formam a imagem inteira.

posta para esse problema está nos valores singulares. O menor valor singular na SVD da matriz A provém das partes menos relevantes da imagem e se pode ignorar a maioria delas.

Suponha a decomposição em valores singulares da matriz A na forma de produto externo, citado na Observação 3.6:

$$A = \sigma_1 \boldsymbol{u}_1 \boldsymbol{v}_1^T + \sigma_2 \boldsymbol{u}_2 \boldsymbol{v}_2^T + \ldots + \sigma_r \boldsymbol{u}_r \boldsymbol{v}_r^T.$$

Note que A é uma matriz real, já que seus elementos são o código de um tom de cinza, variando de 0 e 255.

Seja $k \leq r$ e $posto(A) = r$. Defina

$$A_k = \sigma_1 \boldsymbol{u}_1 \boldsymbol{v}_1^T + \sigma_2 \boldsymbol{u}_2 \boldsymbol{v}_2^T + \ldots + \sigma_k \boldsymbol{u}_k \boldsymbol{v}_k^T,$$

onde A_k é uma aproximação de A que corresponde a manter somente os k primeiro valores singulares e os correspondentes vetores singulares. Para o caso da imagem 340×280, supondo que seja suficiente transmitir somente os dados correspondentes aos 20 primeiros valores singulares, dentre os 280 no total, serão enviados somente

$$\begin{gathered} 20 \text{ vetores } \boldsymbol{u}_i \in \mathbb{R}^{340} \Rightarrow 20 \times 340 \textit{ pixels} \\ 20 \text{ valores singulares } \sigma_i \Rightarrow 20 \textit{ pixels} \\ 20 \text{ vetores } \boldsymbol{v}_i \in \mathbb{R}^{280} \Rightarrow 20 \times 280 \textit{ pixels} \end{gathered}$$

totalizando, assim 12.420 *pixels*.

Quando, através deste processo, obtém-se uma imagem ruim, basta aumentar o valor de k até obter a melhor qualidade possível dentro do objetivo de redução do números de dados transmitidos.

Para a visualização desta aplicação, será utilizado o software Matlab, já anteriormente citado.

Considere a imagem do matemático Beltrami.

A Figura 4.3 tem dimensão 265×326, ou seja, se fosse preciso enviar esta imagem, o número de *pixels* enviados seria igual a 86.390. Com o auxílio do programa Matlab, aproximamos a imagem de Beltrami por A_k, onde k representa o número de valores singulares que se deseja manter na matriz Σ da decomposição de A. Abaixo, seguem-se os comandos utilizados para realizar esta aproximação.

Figura 4.3: Eugenio Beltrami

```
>> A=imread('Beltrami.jpg');
>> figure(1);
>> imshow(A);
>> A1=double(A);
>> [U , S , V]=svd(A1);
>> k=20;
>> Ak=U( : , 1 : k ) * S( 1 : k , 1 : k ) * V( : , 1 : k )';
>> A2=uint8(Ak);
>> figure(2);
>> imshow(A2);
>> imwrite(A2,'Beltrami k=20.jpg');
```

Conforme varia-se o valor de k a qualidade da imagem se modifica, sendo melhor quanto mais valores singulares forem mantidos na matriz diagonal. A Figura 4.4 mostra várias aproximações da imagem original, com valores de k muito pequenos se comparados ao total de 265 valores singulares.

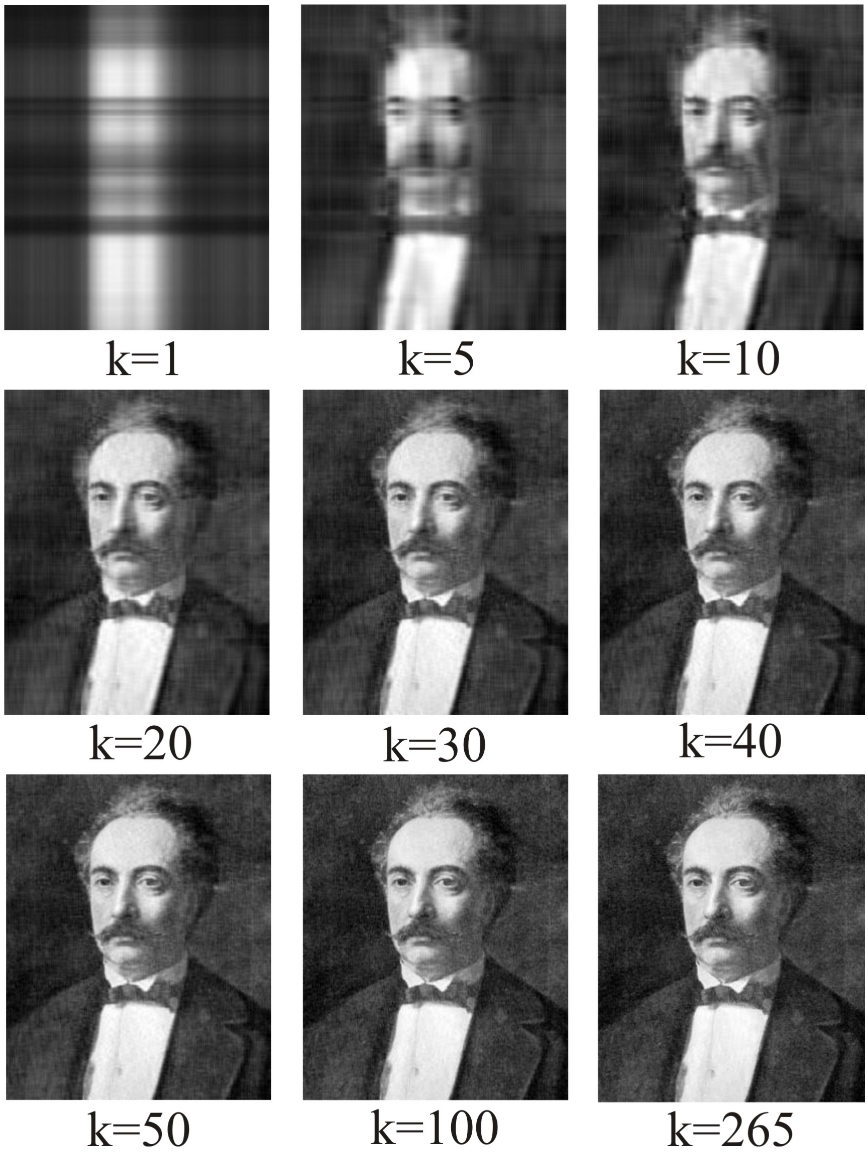

Figura 4.4: Eliminando valores singulares da imagem de Beltrami

Alguns dos valores singulares da matriz obtida são

$$\begin{aligned}
\sigma_1 &= 30.349 \\
\sigma_{10} &= 1.147 \\
\sigma_{50} &= 292 \\
\sigma_{100} &= 161 \\
\sigma_{200} &= 30 \\
\sigma_{265} &= 2
\end{aligned}$$

sendo que a partir de $\sigma_{146} = 100$ os valores singulares são menores que uma centena.

Conforme a Figura 4.4, pode-se notar que para $k = 50$ a imagem obtida fica muito próxima a original, e para esta quantidade de valores singulares, se fosse preciso enviá-la, seriam considerados 29.600 valores, ao contrário dos 86.390 iniciais.

Algumas informações foram perdidas mas os elementos essencias da imagem foram mantidos. Esse tipo de manipulação pode ser bastante útil em aplicações onde não são necessárias imagens de alta resolução.

Considerações Finais

A Decomposição em Valores Singulares tem outras aplicações que vão além das tratadas neste trabalho. Dentre elas, pode-se citar: o cálculo do posto numérico de uma matriz, o que via SVD torna-se mais preciso devido às matrizes unitárias que compõem essa decomposição; a determinação do número de condição de uma matriz, que quantifica o quanto pequenas alterações nos valores de seus elementos podem causar grandes mudanças na solução de um sistema linear onde a matriz representa a matriz dos coeficientes; e na solução do Problema de Procrustes Ortogonal[6], problema que procura uma matriz ortogonal que transforme, de maneira mais aproximada, uma matriz dada em uma segunda matriz, também dada.

Juntamente com esta decomposição, apresentam-se muitas outras que estão intimamente relacionadas com a SVD como, por exemplo, a decomposição UTV, que fatora uma matriz no produto de matrizes ortogonais e uma triangular. Tal decomposição foi apresentada por G. W. Stewart e pode ser encontrada com maiores detalhes em sua obra *Introduction to Matrix Computations.* Além disso, a SVD tem sua forma generalizada (GSVD), que é utilizada em problemas de mínimos quadrados.

Estes e outros temas relacionados à Decomposição em Valores Singulares são convidativas fontes de estudo para os que desejam conhecimentos mais aprofundados sobre o assunto. As aplicações existem, são muitas e permeiam de problemas teóricos a campos da Matemática Computacional. Espera-se que a abordagem introdutória que se faz neste trabalho desperte o interesse de novos estudantes, servindo como um impulso inicial às pesquisas futuras.

[6]Ver [3] página 601.

Referências Bibliográficas

[1] ANTON, H.; RORRES, C.. **Álgebra Linear com Aplicações**. 8. ed. Porto Alegre: Bookman, 2001.

[2] BOLDRINI, J. L. et al. **Álgebra Linear**. 3. ed. amp. e ver. São Paulo: Harbra Ltda, 1986.

[3] GOLUB, G. H.; VAN LOAN, C. F.. **Matrix Computations**. 3. ed. Maryland: Johns Hopkins University Press, 1996.

[4] MARSDEN, J. E.; TROMBA, A . J.. **Vector Calculus**, 4 ed., W.H. Fremann and Company, New York, 1996.

[5] MEYER, CARL D.. **Matrix Analysis and Applied Linear Algebra** . Philadelphia: SIAM, 2000.

[6] POOLE, D.. **Álgebra Linear**, 1 reimp. da 1. ed. Tradutoras técnicas Martha Salerno Monteiro. et. al. São Paulo: Thomson Learning, 2006.

[7] SOUTO, G.. **Decomposição em Valores Singulares.** 2000. v. 1. Trabalho de Conclusão de Curso (Graduação) - Curso de Matemática, Departamento de Matemática, Universidade Federal de Santa Catarina, Florianópolis, 2000.

[8] STEWART, G. W.. **Introduction to Matrix Computations**. New York: Academic Press, 1993.

[9] STEWART, G. W.. On The Early History Of The Singular Value Decomposition. **SIAM Review** v. 35, nº 4, p 551-565, December 1993.

[10] STRANG, G.. **Linear Algebra and its Aplications.** 3. ed. Massachusetts: Harcourt Brace, 1988.

[11] TREFETHEN, L. N.; BAU, D., III. **Numerical Linear Algebra**. Philadelphia: Society For Industrial And Applied Mathematics, 1997.