Maria Cristina Tiemi Hamada Cogubum

Transmissão de Imagens Utilizando SVD

Trabalho de Conclusão de Curso apresentado à Universidade Federal de São Carlos - Campus Sorocaba ao Curso de Licenciatura em Matemática Departamento De Física/Química/Matemática

Orientadora: Profa. Dra. Silvia Maria Simões de Carvalho

Sorocaba, SP
2013

Maria Cristina Tiemi Hamada Cogubum

Transmissão de Imagens Utilizando SVD

Trabalho de Conclusão de Curso apresentado à Universidade Federal de São Carlos - Campus Sorocaba como parte dos requisitos para formação em Licenciatura em Matemática.

Aprovação em 20/09/2013

Banca Examinadora:

Profa. Dra. Silvia M. S. Carvalho (Orientadora) - UFSCar
Profa. Dra. Magda da Silva Peixoto - UFSCar
Prof. Dr. Mayk Vieira Coelho - UNIFAL

Sorocaba, SP
2013

À minha família

5

"Nunca deixe que lhe digam
Que não vale a pena acreditar no sonho que se tem
Quem acredita sempre alcança"

Flávio Venturini - Renato Russo

Agradecimentos

Gostaria de agradecer a todos que, de alguma forma, contribuíram para a realização desse trabalho. Em especial para:

- Mãe e Pai, pelo suporte, incentivo, dedicação e por tudo que fizeram e fazem por mim. Sem vocês, não teria chegado onde cheguei.

- Deus, pela fé e bênção.

- A "Obatian da Praça da Árvore" pelos cuidados comigo desde criança. Onde quer que ela esteja, está torcendo por mim e feliz por essa conquista.

- A "Obatian do Kikuyo", que sempre torceu por mim e me apoiou nos estudos.

- A minha professora, orientadora e amiga Silvia Maria Simões Carvalho, pelo brilhante trabalho de orientação e por acreditar em mim.

- Família, pelo apoio e torcida que sempre dedicaram a mim. Em especial ao Tio Massao, Tia Marisa, Maira, Cleber, Tia Satoe, Tio Natal, Vá, Nini, Vivi, Tia Marina, Tio Geraldo e Tio Tadao.

- Amigos, pelo incentivo, motivação, amor, companheirismo, e por tudo que fizeram por mim durante esses anos. Em especial a Mari, Tiago, Danilo, William, Miriam, Rick, Vitor, Victor, Camila, Fernanda e Rose.

- Aos colegas de faculdade, pela companhia e convivência.

- Ao Departamento de Matemática da UFSCar - Campus Sorocaba, pela base fornecida. Em especial aos professores, Magda, Wladimir, Silvia, Laércio, Paulo, Adilson e Geraldo.

- Membros da banca examinadora: Professora Magda da Silva Peixoto e Professor Mayk Coelho.

- Ao Professor Aurélio Oliveira, pelas considerações e auxílio que enriqueceram o trabalho.

Resumo

Nesse trabalho é apresentado o método de Decomposição em Valores Singulares. Ele está na base de alguns dos mais eficientes algoritmos de compressão de imagem. O armazenamento e transmissão de dados têm um papel extremamente importante em diversas áreas, e o estudo das técnicas que permitam a sua execução com o mínimo espaço de armazenamento e o menor tempo de processamento cresce cada vez mais. Desenvolvemos alguns conceitos importantes como reflexões de Householder, decomposição QR e decomposição em valores singulares, fundamentais para a realização deste trabalho.

Palavras-chave: Reflexões de Householder, Decomposição QR, Decomposição em Valores Singulares, Compressão de Imagens.

Abstract

In this work the Singular Values Decomposition method is presented. It is at the basis of some of the most efficient algorithms of image compression. The storage and transmission of data have an extremely important role in many areas, and the study of techniques that allow it's execution with the minimum of storage space and less processing time gets more and more important. We develop some important concepts as the Householder reflections, decomposition QR and Singular Value Decomposition, which are fundamental for the accomplishment of this work.

Keywords: Householder Reflection, QR decomposition, Singular Values Decomposition, Image Compression..

Lista de Figuras

3.1 Reflexão de Householder. 35

4.1 Gauss. 46
4.2 Teste 1. 48
4.3 Teste 2. 49

Lista de Símbolos e Siglas

$lin(A)$ - Espaço-linha da matriz A

$col(A)$ - Espaço-coluna da matriz A

$nul(A)$ - Espaço-nulo da matriz A

$posto(A)$ - Posto da matriz A

e - vetor de uns

$\|$ - Paralelismo

$\perp$ - Ortogonalidade

λ - Autovalor da matriz

x - Autovetor associado a λ

V_λ - Subespaço de autovetores associados a λ

A^T - Transposta da matriz A

A^{-1} - Inversa da matriz A

I_n - Matriz identidade

$||A||$ - Norma da matriz A

σ - Valor singular de A

Capítulo 1

Introdução

As imagens digitais são indispensáveis em várias áreas como, por exemplo, medicina, biologia e astronomia. Geralmente, são transmitidas na forma de matrizes enormes, que requerem muito espaço de armazenamento. Devido a estas circunstâncias, torna-se viável pesquisar por métodos eficazes que otimizem esses processos, acelerando a transmissão eletrônica de imagens e reduzindo seu espaço de armazenamento, sem muita perda de informação.

Um destes métodos é a Decomposição em Valores Singulares (SVD - Singular Values Decomposition), que pode ser aplicada a quaisquer tipos de matrizes: quadradas, retangulares, singulares e não singulares.

A Decomposição em Valores Singulares é um tema relativamente recente na Matemática e consiste em fatorar uma matriz qualquer $m \times n$ em um produto de matrizes $U\Sigma V^T$, onde $U_{m\times m}$, $V_{n\times n}$ e $\Sigma_{m\times n}$. Ela foi descoberta independentemente por Beltrami (1873) e por Jordan (1874), para matrizes quadradas reais. Em 1915, Autonne demonstrou a SVD para matrizes quadradas complexas. Em 1936, Carl Eckart e Gale Young demonstraram para matrizes retangulares gerais.

A SVD também é a base de muitos dos melhores algoritmos computacionais para resolução de sistemas lineares. Contudo, para compreender o método, foi essencial aprofundar conhecimentos de álgebra linear. Por isso, o trabalho está dividido da seguinte forma:

O segundo capítulo apresenta alguns conceitos básicos de álgebra linear, como espaço vetorial, subespaço, combinação linear, conjunto linearmente independente e dependente, base, vetor unitário, conjunto ortogonal e ortonormal. Além disso, traz definições e propriedades básicas para o estudo que segue, como Processo de Gram-Schmidt , espaço-linha, espaço-coluna, espaço-nulo, posto, nulidade, posto completo, projeção ortogonal, autovalores, autovetores, polinômio característico, autoespaço, multiplicidade algébrica e geométrica de um autovalor, matriz semelhante e matriz diagonalizável.

O terceiro capítulo trata de conceitos mais avançados, partindo de matriz ortogonal, passando por norma matricial, hiperplano, reflexões de Householder, decomposição QR e chegando à decomposição em valores singulares, tema deste trabalho.

Por fim, o quarto capítulo mostra uma aplicação da SVD: a compressão de imagens digitais.

1.1 Importância Histórica

A decomposição em valores singulares foi originalmente desenvolvida por geômetras estudando geometria diferencial. Eles desejavam determinar se uma forma bilinear real pode tornar-se igual a uma outra por transformações ortogonais independentes dos dois espaços no qual ela age. Eugenio Beltrami e Camille Jordan descobriram de forma independente, em 1873 e 1874, respectivamente, que os valores singulares das formas bilineares, representados por uma matriz, formam um conjunto completo de invariantes para formas bilineares sob substituições ortogonais. James Joseph Sylvester também chegou à decomposição em valores singulares para matrizes quadradas reais em 1889, independentemente de Beltrami e Jordan. Sylvester chamou os valores singulares de multiplicadores canônicos da matriz A. O quarto matemático a descobrir a decomposição em valores singulares de forma independente foi Autonne em 1915, que chegou a ela via a decomposição polar.

A primeira prova da decomposição singular para matrizes retangulares e complexas parece ter sido realizada por Carl Eckart e Gale Young em 1936: eles viam a SVD como uma generalização da transformação de eixo principal para matrizes hermitianas. Em 1907, Erhard Schmidt definiu um conceito análogo ao de valores singulares para operadores integrais (que são compactos, sob certas premissas técnicas); parece que ele não estava ciente do trabalho paralelo de valores singulares para matrizes finitas. Essa teoria foi levada adiante por Émile Picard em 1910, que é o primeiro a chamar os números de valores singulares.

De 1964 aos dias atuais, a área de processamento de imagens vem apresentando crescimento expressivo e suas aplicações permeiam quase todos os ramos da atividade humana. Em Medicina, o uso de imagens no diagnóstico médico tornou-se rotineiro e os avanços em processamento de imagens vem permitindo tanto o desenvolvimento de novos equipamentos quanto a maior facilidade de interpretação de imagens produzidas por equipamentos mais antigos, como por exemplo, o raio X. Em Biologia, a capacidade de processar automaticamente imagens obtidas de microscópios, por exemplo contando o número de células de um certo tipo presentes em uma imagem, facilita sobremaneira a execução de tarefas laboratoriais com alto grau de precisão e repetibilidade. O processamento e a interpretação automática de imagens captadas por satélites auxiliam os trabalhos nas áreas de Geografia, Sensoriamento Remoto, Geoprocessamento e Meteorologia, dentre outras. Técnicas de restauração de imagens auxiliam arqueologistas a recuperar fotos borradas de artefatos raros, já destruídos. O uso de robôs dotados de visão artificial em tarefas tais como controle de qualidade em linhas de produção aumenta a cada ano, num cenário de crescente automação industrial. Inúmeras outras áreas tão distintas como Astronomia, Segurança, Publicidade e Direito, vem sendo beneficiadas com os avanços nas áreas de processamento de imagens e visão por computador.

1.2 Comentários Adicionais

Neste trabalho é apresentado um algoritmo em Matlab que faz a compressão da imagem e o cálculo do SVD da matriz referida. O próximo capítulo apresenta uma revisão de álgebra linear em especial nos aspectos usados no decorrer do trabalho.

Capítulo 2

Conceitos Básicos de Álgebra Linear

Apresentam-se aqui algumas definições e teoremas básicos, necessários para a compreensão dos próximos capítulos. Não foram desenvolvidas maiores explicações, uma vez que estas noções já devem ser de conhecimento do leitor para entendimento desse trabalho.

A álgebra linear constitui um ramo da Matemática na qual o foco está em matrizes, espaços vetoriais e transformações lineares. Historicamente, originou-se a partir de sistemas lineares de equações. Além disso, ela ocupa importante papel em diversas áreas - desde as ciências sociais às ciências exatas - permitindo seu uso diário em áreas como economia, aviação, exploração petrolífera e circuitos eletrônicos.

Definição 2.0.1 (Espaço Vetorial) *Seja V um conjunto não vazio qualquer de objetos no qual estejam definidas duas operações, a adição e a multiplicação por escalares. Por adição entende-se uma regra que associa a cada par de objetos u e v em V um objeto $u+v$, denominado soma de u com v; por multiplicação por escalar entende-se uma regra que associa a cada escalar α e cada objeto u em V um objeto $\alpha.u$, denominado múltiplo escalar de u por α. Se todos os axiomas seguintes forem satisfeitos por todos os objetos u, v e w em V e quaisquer escalares α e β, diz-se que V é um espaço vetorial e que os objetos de V são vetores.*

1. *Se u e v são objetos em V, então $u+v$ é um objeto em V.*

2. $u+v=v+u$

3. $u+(v+w)=(u+v)+w$

4. *Existe um objeto 0 em V, denominado vetor nulo de V, ou vetor zero, tal que $0+u=u+0=u$, com qualquer u em V.*

5. *Dado qualquer u em V, existe algum objeto $-u$, denominado negativo de u, tal que $u + (-u) = (-u) + u = 0$.*

6. *Se α for qualquer escalar e u um objeto em V, então αu é um objeto em V.*

7. $\alpha(u + v) = \alpha u + \alpha v$

8. $(\alpha + \beta)u = \alpha u + \beta u$

9. $\alpha(\beta u) = (\alpha\beta)u$

10. 1u=u

Os escalares de um espaço vetorial podem ser números reais ou complexos. Os espaços vetoriais com escalares reais são denominados espaços vetoriais reais, e aqueles com escalares complexos são ditos espaços vetoriais complexos.

Teorema 2.0.1 *Sejam V um espaço vetorial, u um vetor em V e α um escalar. Então:*

1. *0.u=0*

2. *α.0=0*

3. *(-1).u=-u*

4. *Se α.u=0, então α=0 ou u=0*

É possível um espaço vetorial estar contido em outro espaço vetorial. Neste caso, tais espaços vetoriais denominam-se subespaços [9].

Definição 2.0.2 *Um subconjunto W de um espaço vetorial V é denominado subespaço de V se W for um espaço vetorial por si só com as operações de adição e multiplicação por escalar definidas em V.*

Teorema 2.0.2 *Se W for um conjunto de um ou mais vetores em um espaço vetorial V, então W é um subespaço de V se, e somente se, as condições seguintes forem satisfeitas.*

1. *Se u e v forem vetores em W, então u+v está em W.*

2. *Se alpha for um escalar qualquer e u algum vetor de W, então αu está em W.*

Em outras palavras, o Teorema acima diz que W é um subespaço de V se, e somente se, for fechado na adição e na multiplicação por escalar [2].

Teorema 2.0.3 *Se W_1, W_2, ..., W_r forem subespaços de um espaço vetorial V, então a interseção desses subespaços também será um subespaço de V.*

Definição 2.0.3 *Diz-se que um vetor w em um espaço vetorial V é uma combinação linear dos vetores $v_1, v_2, \ldots, v_r$ em V se w puder ser expresso na forma*

$$w = \alpha_1 v_1 + \alpha_2 v_2 + \ldots + \alpha_r v_r$$

em que $\alpha_1, \alpha_2, \ldots, \alpha_r$ são escalares e denominados coeficientes da combinação linear.

Teorema 2.0.4 *Seja $S = \{w_1, w_2, \ldots, w_r\}$ um conjunto não vazio de vetores em um espaço vetorial V.*

1. *O conjunto W de todas as combinações lineares possíveis de vetores de S é um subespaço de V.*
2. *O conjunto W do item acima é o "menor"subespaço de V que contém todos os vetores de S, no sentido de que qualquer outro subespaço de V que contenha todos aqueles vetores contém W.*

A definição seguinte fornece a notação e a terminologia relevantes relacionadas ao Teorema 2.0.4.

Definição 2.0.4 *Diz-se que o subespaço de um espaço vetorial V que é formado com todas as combinações lineares possíveis de vetores de um conjunto não vazio S é gerado por S, e diz-se que os vetores em S geram esse subespaço. Se $S = \{w_1, w_2, \ldots, w_r\}$, denota-se o gerado de S por $ger\{w_1, w_2, \ldots, w_r\}$ ou ger(S).*

Teorema 2.0.5 *As soluções de um sistema linear homogêneo $Ax = 0$, com $A_{m\times n}$, é um subespaço de $\mathbb{R}^n$.*

O conjunto das soluções de um sistema homogêneo em n incógnitas é um subespaço de $\mathbb{R}^n$. Diz-se que esse conjunto é o espaço solução do sistema.

As soluções de um sistema linear homogêneo $Ax = 0$ de m equações e n incógnitas s?o vetores em $\mathbb{R}^n$.

Teorema 2.0.6 *Se $S = \{v_1, v_2, \ldots, v_r\}$ e $S' = \{w_1, w_2, \ldots, w_k\}$ são conjuntos não vazios de vetores em um espaço vetorial V, então*

$$ger\{v_1, v_2, \ldots, v_r\} = ger\{w_1, w_2, \ldots, w_k\}$$

se, e somente se, cada vetor em S é uma combinação linear dos vetores em S', e cada vetor em S' é uma combinação linear dos vetores em S.

Definição 2.0.5 *Se $S = \{v_1, v_2, \ldots, v_r\}$ for um conjunto não vazio de vetores em um espaço vetorial V, então a equação vetorial*

$$k_1v_1 + k_2v_2 + \ldots + k_rv_r = 0$$

tem pelo menos uma solução, a saber,

$$k_1 = 0, k_2 = 0, \ldots, k_r = 0$$

Diz-se que essa é a solução trivial. Se essa for a única solução, diz-se que S é um conjunto linearmente independente. Se existem outras soluções além da trivial, diz-se que S é um conjunto linearmente dependente.

Teorema 2.0.7 *Um conjunto S de dois ou mais vetores é:*

1. *linearmente dependente se, e somente se, pelo menos um dos vetores de S pode ser expresso como uma combinação linear dos outros vetores em S.*

2. *linearmente independente se, e somente se, nenhum vetor em S pode ser expresso como uma combinação linear dos outros vetores em S.*

O teorema seguinte se refere à dependência e à independência linear de conjuntos de um ou dois elementos e conjuntos que contenham o vetor nulo.

Teorema 2.0.8 1. *Um conjunto finito que contenha $\vec{0}$ é linearmente dependente.*

2. *Um conjunto de exatamente um vetor é linearmente independente se, e somente se, esse vetor não é $\vec{0}$.*

3. *Um conjunto de exatamente dois vetores é linearmente independente se, e somente se, nenhum dos dois vetores é um múltiplo escalar do outro.*

O próximo teorema mostra que um conjunto linearmente independente em $\mathbb{R}^n$ pode conter, no máximo, n vetores.

Teorema 2.0.9 *Seja $S = \{v_1, v_2, \ldots, v_r\}$ um conjunto de vetores em $\mathbb{R}^n$. Se $r > n$, então S é linearmente dependente.*

Definição 2.0.6 *Se V for um espaço vetorial qualquer e $S = \{v_1, v_2, \ldots, v_n\}$ for um conjunto finito de vetores em V, diz-se que S é uma base de V se forem válidas as duas condições a seguir:*

1. *S é linearmente independente.*
2. *S gera V.*

Nem todo espaço vetorial tem uma base. Um exemplo é o espaço vetorial nulo, que não contém conjuntos linearmente independentes e, portanto, não tem base. Além disso, um espaço vetorial que não pode ser gerado por um número finito de vetores não tem base. Este espaço vetorial é de dimensão infinita, enquanto um que pode ser gerado é de dimensão finita.

Teorema 2.0.10 *Se $S = \{v_1, v_2, \ldots, v_n\}$ for uma base de um espaço vetorial V, então cada vetor em V pode ser expresso de maneira única como $v = c_1v_1 + c_2v_2 + \ldots + c_nv_n$.*

Definição 2.0.7 *Se $S = \{v_1, v_2, \ldots, v_n\}$ for uma base de um espaço vetorial V e se*

$$v = c_1v_1 + c_2v_2 + \ldots + c_nv_n$$

é a expressão de um vetor v em termos da base S, então os escalares $c_1, c_2, \ldots, c_n$ são chamados coordenadas de v em relação à base S. O vetor $(c_1, c_2, \ldots, c_n)$ em $\mathbb{R}^n$ construído com essas coordenadas é denominado vetor de coordenadas de v em relação à S e é denotado por

$$(v)_s = (c_1, c_2, \ldots, c_n)$$

Definição 2.0.8 *Seja $u = (x_1, x_2, \ldots, x_n)$ as coordenadas de u na base v. Diz-se que u é um vetor unitário se $||u|| = 1$, onde $||u|| = \sqrt{{x_1}^2 + {x_2}^2 + \ldots + {x_n}^2}$.*

Definição 2.0.9 *Diz-se que dois vetores não nulos u e v em $\mathbb{R}^n$ são ortogonais (ou perpendiculares) se $u.v = 0$. Também convenciona-se que o vetor nulo em $\mathbb{R}^n$ é ortogonal a cada vetor em $\mathbb{R}^n$. Um conjunto $S = \{u_1, u_2, \ldots, u_k\}$ é dito ortogonal se dois vetores distintos quaisquer em S são ortogonais, isto é, se $u_i.u_j = 0$ para $i \neq j$. Um conjunto ortonormal de vetores é um conjunto ortogonal de vetores unitários, ou seja, $S = \{u_1, u_2, \ldots, u_k\}$ é ortonormal se $u_i.u_j = 0$ para $i \neq j$ e $u_i.u_i = 1$ para $i = 1, 2, \ldots, k$.*

Um resultado importante sobre conjuntos ortogonais é dado no próximo teorema.

Teorema 2.0.11 *Seja $S = \{u_1, u_2, \ldots, u_k\}$ um conjunto ortogonal de vetores não nulos em $\mathbb{R}^n$. Então S é linearmente independente.*

Corolário 2.0.1 *Um conjunto ortonormal de vetores em $\mathbb{R}^n$ é linearmente independente.*

Definição 2.0.10 *Um conjunto ortonormal de n vetores em $\mathbb{R}^n$ é uma base para $\mathbb{R}^n$. Uma base ortogonal (respectivamente, ortonormal) para um espaço vetorial é uma base que é um conjunto ortogonal (respectivamente, ortonormal).*

Teorema 2.0.12 *Seja $S = \{u_1, u_2, \ldots, u_n\}$ uma base ortonormal para $\mathbb{R}^n$ e seja v um vetor qualquer em $\mathbb{R}^n$. Então*

$$v = r_1u_1 + r_2u_2 + \ldots + r_nu_n$$

onde $r_i = v.u_i \quad (1 \leq i \leq n)$.

Outro conceito básico e importante de Álgebra Linear é o de projeção ortogonal, pois permite obter um vetor que tem a menor distância de um vetor considerado, critério que serve de base para o método dos quadrados mínimos.

Definição 2.0.11 *Seja u um vetor não-nulo. Dado v qualquer, o vetor p é chamado projeção ortogonal de v sobre u, e indicado por $proj_uv$, se satisfaz as condições:*

1. *p paralelo a u (Notação: $p \parallel u$)*
2. *$(v - p)$ ortogonal a u (Notação: $(v - p) \perp u$)*

Proposição 2.0.1 *Seja u um vetor não-nulo. Qualquer que seja v, existe e é única a projeção ortogonal de v sobre u. Sua expressão em termos de u e v é*

$$proj_uv = \frac{v.u}{||u||^2}u$$

E a expressão de sua norma é

$$||proj_uv|| = \frac{|v.u|}{||u||}$$

Teorema 2.0.13 (Processo de Gram-Schmidt) *Seja W um subespaço não nulo de $\mathbb{R}^n$ com base $S = \{u_1, u_2, \ldots, u_m\}$. Então, existe uma base ortonormal $T = \{w_1, w_2, \ldots, w_m\}$ para W.*

O processo de Gram-Schmidt para construir uma base ortonormal $T = \{w_1, w_2, \ldots, w_m\}$ de um subespaço não nulo W de $\mathbb{R}^n$ com base $S = \{u_1, u_2, \ldots, u_m\}$ é o seguinte:

Passo 1: Seja $v_1 = u_1$.

Passo 2: Calcule os vetores $v_2, v_3, \ldots, v_m$ sucessivamente, um de cada vez, pela fórmula

$$v_i = u_i - \left(\frac{u_i.v_1}{v_1.v_1}\right).v_1 - \left(\frac{u_i.v_2}{v_2.v_2}\right).v_2 - \ldots - \left(\frac{u_i.v_{i-1}}{v_{i-1}.v_{i-1}}\right).v_{i-1}$$

O conjunto de vetores $T^* = \{v_1, v_2, \ldots, v_m\}$ é um conjunto ortogonal.

Passo 3: Seja $w_i = \frac{1}{||v_i||}.v_i$, com $1 \leq i \leq m$. Então, $T = \{w_1, w_2, \ldots, w_m\}$ é uma base ortonormal para W.

Exemplo 2.0.1 *Considere a base $S = \{u_1, u_2, u_3\}$ de $\mathbb{R}^3$, onde $u_1 = (1,1,1), u_2 = (-1,0,-1)$ e $u_3 = (-1,2,3)$. Use o processo de Gram-Schmidt para transformar S em uma base ortonormal para $\mathbb{R}^3$.*

Solução:

Passo 1: Seja $v_1 = u_1 = (1,1,1)$.

Passo 2: Calculando v_2 e v_3:

$$v_2 = u_2 - \left(\frac{u_2.v_1}{v_1.v_1}\right).v_1 = \left(-\frac{1}{3}, \frac{2}{3}, -\frac{1}{3}\right)$$

Multiplicando v_2 por 3 para eliminar as frações, tem-se $(-1,2,-1)$, que utiliza-se agora como v_2. Então:

$$v_3 = u_3 - \left(\frac{u_3.v_1}{v_1.v_1}\right).v_1 - \left(\frac{u_3.v_2}{v_2.v_2}\right).v_2 = (-2,0,2)$$

Logo, $T^* = \{v_1, v_2, v_3\} = \{(1,1,1), (-1,2,-1), (-2,0,2)\}$ é uma base ortogonal para $\mathbb{R}^3$.

Passo 3: Sejam

$$w_1 = \frac{1}{||v_1||}.v_1 = \left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right)$$

$$w_2 = \frac{1}{||v_2||}.v_2 = \left(-\frac{1}{\sqrt{6}}, \frac{2}{\sqrt{6}}, -\frac{1}{\sqrt{6}}\right)$$

$$w_3 = \frac{1}{||v_3||}.v_3 = \left(-\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}\right)$$

Então, $T = \{w_1, w_2, w_3\} = \{(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}), (-\frac{1}{\sqrt{6}}, \frac{2}{\sqrt{6}}, -\frac{1}{\sqrt{6}}), (-\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}})\}$ é uma base ortonormal para $\mathbb{R}^3$.

Definição 2.0.12 *Se A é uma matriz $m \times n$, então existem três espaços importantes associados com A:*

1. *O espaço-linha de A, denotado por lin (A), é o subespaço de $\mathbb{R}^n$ gerado pelos vetores-linha de A.*

2. *O espaço-coluna de A, denotado por $col\,(A)$, é o subespaço de $\mathbb{R}^m$ gerado pelos vetores-coluna de A.*

3. *O espaço-nulo de A, denotado por $nul\,(A)$, é o espaço-solução de $Ax = 0$. Esse é um subespaço de $\mathbb{R}^n$.*

Definição 2.0.13 *A dimensão do espaço-linha de uma matriz A é denominada o posto de A e é denotado por $posto\,(A)$; a dimensão do espaço nulo de A é denominada a nulidade de A.*

Teorema 2.0.14 (Teorema do Posto) *O espaço-linha e o espaço-coluna de uma matriz têm a mesma dimensão.*

Demonstração: Suponha que A seja uma matriz $m \times n$ e que $dim(lin(A)) = k$. Isso significa que a forma escalonada reduzida R de A tem k vetores-linha não-nulos $R_1, \ldots, R_k$. Como A e R tem o mesmo espaço-linha, segue que os vetores-linha $A_1, \ldots, A_m$ de A podem ser escritos como combinações lineares dos vetores-linha de R:

$$A_1 = c_{11}R_1 + c_{12}R_2 + \ldots + c_{1k}R_k,$$

$$A_2 = c_{21}R_1 + c_{22}R_2 + \ldots + c_{2k}R_k,$$

$$\vdots$$

$$A_m = c_{m1}R_1 + c_{m2}R_2 + \ldots + c_{mk}R_k.$$

Isso implica que o j-ésimo elemento da linha A_i é dado por

$$A_{ij} = c_{i1}R_{1j} + c_{i2}R_{2j} + \ldots + c_{ik}R_{kj}.$$

Logo, a j-ésima coluna da matriz A é dada por

$$\begin{bmatrix} A_{1j} \\ A_{2j} \\ \vdots \\ A_{mj} \end{bmatrix} = R_{1j} \begin{bmatrix} c_{11} \\ c_{21} \\ \vdots \\ c_{m1} \end{bmatrix} + R_{2j} \begin{bmatrix} c_{12} \\ c_{22} \\ \vdots \\ c_{m2} \end{bmatrix} + \ldots + R_{kj} \begin{bmatrix} c_{1k} \\ c_{2k} \\ \vdots \\ c_{mk} \end{bmatrix}.$$

Portanto, o espaço-coluna de A também é gerado por k vetores. Isto implica que

$$dim(col(A)) \leq dim(lin(A)), \tag{2.1}$$

já que não sabe-se em princípio se estes vetores são linearmente independentes. Mas isso vale para qualquer matriz, em particular também vale para transposta de A, logo também tem-se

$$dim(col(A^T)) \leq dim(lin(A^T)). \tag{2.2}$$

Mas

$$dim(col(A^T)) = dim(lin(A)),$$
$$dim(lin(A^T)) = dim(col(A)),$$

logo (2.2) é equivalente a

$$dim(lin(A)) \leq dim(col(A)). \tag{2.3}$$

A partir de (2.1) e (2.3), conclui-se que

$$dim(lin(A)) = dim(col(A)).$$

Como consequência do teorema anterior, pode-se pensar no posto de uma matriz como a dimensão do espaço-coluna. Além disso, como a matriz transposta converte linhas em colunas e vice-versa, uma matriz e sua transposta possuem o mesmo posto.

Teorema 2.0.15 *Se A é uma matriz $m \times n$, então posto(A)=posto(A^T).*

O próximo conceito é um instrumento importante no estudo de sistemas lineares nos quais o número de equações e o número de incógnitas não são necessariamente iguais.

Definição 2.0.14 *Diz-se que uma matriz A de tamanho $m \times n$ tem posto-coluna máximo se seus vetores-coluna são linearmente independentes e tem posto-linha máximo se seus vetores-linha são linearmente independentes.*

Uma maneira alternativa para os conceitos de posto-coluna e posto-linha máximos é a que segue:

Teorema 2.0.16 *Seja A uma matriz $m \times n$.*

1. *A tem posto-coluna máximo se, e somente se, posto $(A) = n$.*

2. *A tem posto-linha máximo se, e somente se, posto $(A) = m$.*

Sabe-se que o posto linha (coluna) de uma matriz $A \in \mathbb{R}^{m \times n}$ é o número de linhas (colunas) linearmente independentes. Ainda, pelo Teorema do Posto, o posto linha é igual ao posto coluna. Por fim, define-se posto completo.

Definição 2.0.15 *Uma matriz A de tamanho $m \times n$ tem posto completo se posto $(A) = min\{m, n\}$.*

Definição 2.0.16 *Se A for uma matriz $n \times n$, então um vetor não nulo x em $\mathbb{R}^n$ é denominado autovetor de A se Ax for um múltiplo escalar de x, isto é,*

$$Ax = \lambda x$$

com algum escalar λ. O escalar λ é denominado autovalor de A, e diz que x é um autovetor associado a λ.

O próximo teorema fornece um procedimento para encontrar os autovalores de A [3].

Teorema 2.0.17 *Se A for uma matriz $n \times n$, então λ é um autovalor de A se, e somente se, λ satisfaz a equação*

$$det(\lambda I - A) = 0 \tag{2.4}$$

Essa equação é a equação característica de A.

Quando o determinante $det(\lambda I - A)$ do lado esquerdo de (1.1) é expandido, resulta um polinômio $p(\lambda)$ de grau n denominado polinômio característico de A. Em geral, o polinômio característico de uma matriz $n \times n$ é da forma

$$p(\lambda) = \lambda^n + c_1 \lambda^{n-1} + \ldots + c_n$$

em que 1 é o coeficiente de λ^n. Como um polinômio de grau n tem, no máximo, n raízes distintas, tem-se que uma matriz $n \times n$ tem, no máximo, n autovalores distintos.

Teorema 2.0.18 *Se A for uma matriz $n \times n$ triangular (superior, inferior ou diagonal), então os autovalores de A são as entradas na diagonal principal de A.*

Teorema 2.0.19 *Se A for uma matriz $n \times n$, são equivalentes as afirmações seguintes.*

1. *λ é um autovalor de A.*
2. *O sistema $(\lambda I - A)x = 0$ de equações tem soluções não triviais.*
3. *Existe algum vetor não nulo x tal que $Ax = \lambda x$.*
4. *λ é uma solução da equação característica $det(\lambda I - A) = 0$.*

Encontrados os autovalores de uma matriz A, deve-se encontrar os autovetores. Os autovetores de A associados a um autovalor λ são os vetores não nulos x que satisfazem $Ax = \lambda x$. Equivalentemente, os autovetores associados a λ são os vetores não nulos que satisfazem a equação

$$(\lambda I - A)x = 0$$

Esse espaço é denominado autoespaço de A associado a λ.

Exemplo 2.0.2 *Encontre os autovalores e os autovetores de*

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

Solução: O objetivo é encontrar vetores $x \in \mathbb{R}^3$ e escalares $\lambda \in \mathbb{R}$ tais que $Ax = \lambda x$. O polinômio característico de A é

$$det(\lambda I - A) = det \begin{bmatrix} \lambda - 4 & -2 & 0 \\ 1 & \lambda - 1 & 0 \\ 0 & -1 & \lambda - 2 \end{bmatrix} = \lambda^3 - 7\lambda^2 + 16\lambda - 12$$

Logo, os autovalores de A satisfazem a equação

$$\lambda^3 - 7\lambda^2 + 16\lambda - 12 = 0 \tag{2.5}$$

Para resolver essa equação, deve-se procurar inicialmente soluções inteiras. Para isto, basta lembrar que todas as soluções inteiras (se houver) de uma equação polinomial

$$\lambda^n + c_1\lambda^{n-1} + \ldots + c_n = 0$$

de coeficientes inteiros são divisores do termo constante c_n. Assim, as únicas possíveis soluções inteiras de (1.3) são os divisores de -12, ou seja, ± 1, ± 2, ± 3, ± 4, ± 6, ± 12. Substituir cada um desses valores em (1.3) mostra que $\lambda = 2$ é uma raiz, e então, $\lambda - 2$ é uma solução inteira. Dividindo $\lambda^3 - 7\lambda^2 + 16\lambda - 12$ por $\lambda - 2$, tem-se que (1.3) pode ser reescrita como

$$(\lambda - 2)(\lambda^2 - 5\lambda + 6) = 0$$

Assim, as demais soluções de (1.3) satisfazem a equação quadrática

$$\lambda^2 - 5\lambda + 6 = 0$$

Assim, os autovalores de A são $\lambda = 2$ e $\lambda = 3$.
Conhecendo os autovalores, pode-se encontrar os autovetores correspondentes. Resolvendo $Ax = \lambda x$, para os casos:

- $\lambda = 2$

$$\begin{bmatrix} 4 & 2 & 0 \\ -1 & 1 & 0 \\ 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = 2 \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

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

A terceira equação implica que $y = 0$. Assim, pela segunda equação, tem-se que $x = 0$. Como nenhuma equação impõe uma restrição em z, os autovetores associados a $\lambda = 2$ são do tipo $(0, 0, z)$, ou seja, pertencem ao subespaço $[(0, 0, 1)]$.

- $\lambda = 3$

$$\begin{bmatrix} 4 & 2 & 0 \\ -1 & 1 & 0 \\ 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = 3 \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

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

Tanto a primeira equação quanto a segunda implica que $x = -2y$. E da terceira vem $z = y$. Os autovetores associados ao autovalor $\lambda = 3$ são do tipo $(-2y, y, y)$, ou seja, pertencem ao subespaço $[(-2, 1, 1)]$.

Definição 2.0.17 *Denomina-se multiplicidade algébrica de um autovalor a quantidade de vezes que ele aparece como raiz do polinômio característico.*

Definição 2.0.18 *Denomina-se multiplicidade geométrica de um autovalor λ a dimensão do subespaço V_λ de autovetores associados a λ.*

No exemplo acima, o autovalor $\lambda = 2$ tem multiplicidade algébrica igual a 2. Ou ainda, 2 é uma raiz dupla do polinômio característico. Ainda para o autovalor $\lambda = 2$, os autovetores são do tipo $(0, 0, z)$. Note que $\{(0, 0, z) \in \mathbb{R}^3; z \in \mathbb{R}\} = \{z(0, 0, 1) : z \in \mathbb{R}\} = [(0, 0, 1)]$ e portanto a dimensão deste subespaço é 1. Logo, a multiplicidade geométrica é 1.

Uma vez obtidos os autovalores e autovetores de uma matriz A, é simples obter os autovalores e autovetores de qualquer potência inteira positiva de A.

Teorema 2.0.20 *Se k for um inteiro positivo, λ um autovalor de uma matriz A e x um autovetor associado, então λ^k é um autovalor de A^k e x é um autovetor associado.*

O teorema seguinte estabelece uma relação entre os autovalores e a invertibilidade de uma matriz.

Teorema 2.0.21 *Uma matriz quadrada A é invertível se, e somente se, $\lambda = 0$ não é um autovalor de A.*

Demonstração: Suponha que A seja uma matriz $n \times n$ e observe primeiro que $\lambda = 0$ é uma solução da equação característica

$$\lambda^n + c_1\lambda^{n-1} + \ldots + c_n = 0$$

se, e somente se, o termo constante c_n for zero. Assim, é suficiente provar que A é invertível se, e somente se, $c_n \neq 0$. Mas

$$det(\lambda I - A) = \lambda^n + c_1\lambda^{n-1} + \ldots + c_n$$

ou, tomando $\lambda = 0$,

$$det(-A) = c_n$$

ou

$$(-1)^n det(A) = c_n$$

Segue da última equação que $det(A) = 0$ se, e somente se, $c_n = 0$ e isso, por sua vez, implica que A é invertível se, e somente se, $c_n \neq 0$.

Definição 2.0.19 *Se A e B forem matrizes quadradas, diz-se que B é semelhante a A se existir alguma matriz invertível P tal que $B = P^{-1}AP$.*

As matrizes semelhantes têm muitas propriedades em comum. Uma delas é que se $B = P^{-1}AP$, então A e B têm o mesmo determinante.

Definição 2.0.20 *Uma matriz quadrada A é dita diagonalizável se for semelhante a alguma matriz diagonal, isto é, se existir alguma matriz invertível P tal que $P^{-1}AP$ é diagonal. Neste caso, diz-se que a matriz P diagonaliza A.*

Teorema 2.0.22 *Se A for uma matriz $n \times n$, são equivalentes as afirmações seguintes:*

1. *A é diagonalizável.*
2. *A tem n autovetores linearmente independentes.*

O teorema anterior garante que uma matriz A de tamanho $n \times n$ com n autovetores linearmente independentes é diagonalizável. A seguir, apresenta-se um método para diagonalizar A.

Passo 1: Certifique-se que a matriz é realmente diagonalizável encontrando n autovetores linearmente independentes. Um modo de fazer isso é encontrar uma base de cada autoespaço e unir todos esses vetores em um único conjunto S. Se esse conjunto possuir menos do que n elementos, a matriz não é diagonalizável.

Passo 2: Forme a matriz $P = \begin{bmatrix} p_1 & p_2 & \dots & p_n \end{bmatrix}$ que tem os vetores de S como vetores coluna.

Passo 3: A matriz $P^{-1}AP$ será diagonal com os autovalores $\lambda_1, \lambda_2, \dots, \lambda_n$ correspondentes aos autovetores $p_1, p_2, \dots, p_n$ como entradas diagonais sucessivas.

Exemplo 2.0.3 *Encontre uma matriz P que diagonalize*

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

Solução: Verifica-se que a equação característica de A é

$$(\lambda - 1)(\lambda - 2)^2 = 0$$

e encontra-se as seguintes bases dos autoespaços:

- Para o autovalor $\lambda = 2 : p_1 = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}, p_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$
- Para o autovalor $\lambda = 1 : p_3 = \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}$

Tem-se um total de três vetores de base. Logo, a matriz

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

diagonaliza A. Para conferir, basta verificar que

$$P^{-1}AP = \begin{bmatrix} 1 & 0 & 2 \\ 1 & 1 & 1 \\ -1 & 0 & -1 \end{bmatrix} \begin{bmatrix} 0 & 0 & -2 \\ 1 & 2 & 1 \\ 1 & 0 & 3 \end{bmatrix} \begin{bmatrix} -1 & 0 & -2 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Vale ressaltar que, em geral, não existe uma ordem preferencial para as colunas de P.

Note que no Exemplo 2.0.2, a matriz A não é diagonalizável.

Teorema 2.0.23 *Se $v_1, v_2, \ldots, v_k$ forem autovetores de uma matriz A associados a autovalores distintos, então $\{v_1, v_2, \ldots, v_k\}$ é um conjunto linearmente independente.*

Teorema 2.0.24 *Se uma matriz A de tamanho $n \times n$ tem n autovalores distintos, então A é diagonalizável.*

Teorema 2.0.25 *Se A for uma matriz quadrada, valem as afirmações seguintes:*

1. *Dado qualquer autovalor de A, a multiplicidade geométrica é menor do que ou igual à multiplicidade algébrica.*

2. *A é diagonalizável se, e somente se, a multiplicidade geométrica de cada autovalor é igual à multiplicidade algébrica.*

Capítulo 3

Álgebra Linear Numérica

Neste capítulo, apresentam-se conceitos e teoremas mais avançados da Álgebra Linear, como matriz ortogonal, norma matricial, hiperplano, reflexões de Householder, decomposição QR e decomposição em valores singulares. Os últimos dois tópicos são de extrema importância para o próximo capítulo, pois permitem decompor matrizes de maneira eficiente e com baixo custo computacional.

3.1 Matriz Ortogonal

Definição 3.1.1 *Seja $A = [a_{ij}]_{n\times n}$ uma matriz real e A^t sua transposta. Se $A.A^t = A^t.A = I_n$ (ou seja, $A^t = A^{-1}$), diz-se que A é uma matriz ortogonal.*

Como exemplo de matrizes ortogonais, tem-se

$$A = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \qquad e \qquad B = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Para verificar isto, basta multiplicar cada uma pela sua transposta, obtendo assim a matriz identidade. Para a primeira matriz, tem-se:

$$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} . \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Com a proposição a seguir, pode-se verificar se uma matriz é ortogonal ou não.

Proposição 3.1.1 *Seja $A = [a_{ij}]_{n\times n}$ uma matriz. São equivalentes:*

1. *A é ortogonal;*
2. *As colunas de A formam um conjunto ortonormal em $\mathbb{R}^n$;*
3. *As linhas de A formam um conjunto ortonormal em $\mathbb{R}^n$.*

Considere agora três propriedades das matrizes ortogonais.

- *Propriedade I:* O produto de duas matrizes ortogonais é ortogonal.

  Demonstração: Para provar a afirmação, lembre-se que $(AB)^t = B^t A^t$. Sejam agora M e N ortogonais. Logo:

  $$(MN)(MN)^t = MNN^tM^t = M(NN^t)M^t = MM^t = I_d$$

  Portanto, MN é ortogonal.

- *Propriedade II:* O determinante de uma matriz ortogonal é ± 1.

  Demonstração: Seja A uma matriz ortogonal. Logo, $A.A^t = I$.

  Para provar esta afirmação, é preciso lembrar das propriedades de matrizes que A^t e A possuem o mesmo determinante e que o determinante da matriz identidade é igual a 1.

  Então, $det(A.A^t) = det(I)$ e $(detA).(detA^t) = 1$.

  Mas, $detA = detA^t$.

  Assim, $(detA)^2 = 1$, ou seja, $detA = \pm 1$.

- *Propriedade III:*

  1. Uma matriz é ortogonal se, e somente se, as colunas são vetores ortonormais.
  2. Uma matriz é ortogonal se, e somente se, as linhas são vetores ortonormais.

  Demonstração: A prova dos itens (1) e (2) é feita de maneira análoga. Seja

  $$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \dots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix}$$

  Na primeira parte da demonstração, o objetivo é mostrar que se A é ortogonal, isto implica que

  $$\begin{bmatrix} a_{11} \\ \vdots \\ a_{n1} \end{bmatrix}, \dots, \begin{bmatrix} a_{1n} \\ \vdots \\ a_{nn} \end{bmatrix}$$

  são ortonormais (o mesmo vale para as linhas). Para isto, calcula-se o produto de A pela sua transposta:

  $$A.A^t = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{nn} \end{bmatrix} . \begin{bmatrix} a_{11} & \dots & a_{n1} \\ \vdots & & \vdots \\ a_{1n} & \dots & a_{nn} \end{bmatrix}$$

  $$= \begin{bmatrix} {a^2}_{11} + \dots + {a^2}_{n1} & \dots & a_{11}a_{1n} + \dots + a_{n1}a_{nn} \\ \vdots & & \vdots \\ a_{1n}a_{11} + \dots + a_{nn}a_{n1} & \dots & {a^2}_{1n} + \dots + {a^2}_{nn} \end{bmatrix} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}$$

pois $A.A^t = I$.

Nota-se que $a^2{}_{11} + \ldots + a^2{}_{n1} = 1$. Mas isto quer dizer que

$$\begin{bmatrix} a_{11} \\ \vdots \\ a_{n1} \end{bmatrix}$$

é unitário. Da mesma forma, percorrendo a diagonal principal, observa-se que cada vetor coluna da matriz A é unitário. Saindo desta diagonal, o elemento na posição i,j $(i \neq j)$ é $a_{1i}a_{1j} + \ldots + a_{ni}a_{nj}$ e seu valor deve ser zero. Mas isto diz que o produto interno de

$$\begin{bmatrix} a_{1i} \\ \vdots \\ a_{ni} \end{bmatrix} \quad por \quad \begin{bmatrix} a_{1j} \\ \vdots \\ a_{nj} \end{bmatrix}$$

é nulo, ou seja, os vetores coluna são dois a dois ortogonais quando $i \neq j$.

A volta da demonstração é apenas uma adaptação da prova dada acima.

3.2 Norma Matricial

Definição 3.2.1 *Uma norma matricial sobre o conjunto de todas as matrizes $n \times n$ é uma função de valor real, $||.||$, definida nesse conjunto e que satisfaz, para todas as matrizes A e B de $n \times n$ e todos os números reais α:*

1. $||A|| \geq 0$.
2. $||A|| = 0$, *se e somente se A é a matriz com todas as entradas nulas.*
3. $||\alpha A|| = |\alpha|\ ||A||$.
4. $||A + B|| \leq ||A|| + ||B||$.
5. $||AB|| \leq ||A||\ ||B||$.

Uma norma de matriz em $n \times n$ é compatível com uma norma de vetor $||x||$ em $\mathbb{R}^n$ se, para toda matriz $A_{n\times n}$ e todo x em $\mathbb{R}^n$, tem-se $||Ax|| \leq ||A||\ ||x||$.

A distância entre as matrizes A e B de $n \times n$ com relação a essa norma matricial é $||A - B||$. Seja $A = (a_{ij})$ uma matriz $n \times n$. Alguns exemplos de normas matriciais são:

- $||A||_1 = \max_{1 \leq j \leq n}(\sum_{i=1}^{n} |a_{ij}|) \rightarrow$ A maior soma absoluta das colunas
- $||A||_2 = \sqrt{\lambda_{\max}} = \max\{\sqrt{|\lambda|} : \lambda$ é um autovalor de $A^T A\} \rightarrow$ Norma espectral
- $||A||_\infty = \max_{1 \leq i \leq n}(\sum_{j=1}^{n} |a_{ij}|) \rightarrow$ A maior soma absoluta das linhas

- $||A||_F = \sqrt{\sum_{i,j=1}^{n} |a_{ij}|^2} \rightarrow$ Norma de Frobenius

- $||A||_p = \sup_{x\neq 0} \frac{||Ax||_p}{||x||_p} \rightarrow$ p-norma

É importante entender que a Norma de Frobenius e a p-norma definem famílias de normas - a norma-2 em $\mathbb{R}^{3\times 2}$ é uma função diferente da norma-2 em $\mathbb{R}^{5\times 6}$. A desigualdade

$$||AB||_p \leq ||A||_p||B||_p \quad com \quad A \in \mathbb{R}^{m\times n}, B \in \mathbb{R}^{n\times q} \tag{3.1}$$

é uma observação sobre a relação entre três normas diferentes. Formalmente, diz-se que que as normas $||A||_1$, $||A||_2$ e $||A||_3$ em $\mathbb{R}^{m\times q}$, $\mathbb{R}^{m\times n}$, e $\mathbb{R}^{n\times q}$ são mutuamente consistentes se para todo $A \in \mathbb{R}^{m\times n}$ e $B \in \mathbb{R}^{n\times q}$ tem-se $||AB||_1 \leq ||A||_2||B||_3$

Nem toda norma matricial satisfaz a propriedade abaixo

$$||AB|| \leq ||A||\ ||B|| \tag{3.2}$$

As p-normas possuem a importante propriedade que para todo $A \in \mathbb{R}^{m\times n}$ e $x \in \mathbb{R}^n$ tem-se $||Ax||_p \leq ||A||_p||x||_p$. Genericamente, para qualquer norma vetorial $||.||_\alpha$ em $\mathbb{R}^n$ e $||.||_\beta$ em $\mathbb{R}^m$ tem-se $||Ax||_\beta \leq ||A||_{\alpha,\beta}||x||_\alpha$ onde $||A||_{\alpha,\beta}$ é uma norma matricial definida por

$$||A||_{\alpha,\beta} = \sup_{x\neq 0} \frac{||Ax||_\beta}{||x||_\alpha} \tag{3.3}$$

Diz-se que $||.||_{\alpha,\beta}$ é subordinada para as normas vetoriais $||.||_\alpha$ e $||.||_\beta$.

Para $A \in \mathbb{R}^{m\times n}$, tem-se as seguintes propriedades:

- $||A||_2 \leq ||A||_F \leq \sqrt{n}||A||_2$

- $\max_{i,j}(|a_{ij}|) \leq ||A||_2 \leq \sqrt{mn}\max_{i,j}|a_{ij}|$

- $\frac{1}{\sqrt{n}}||A||_\infty \leq ||A||_2 \leq \sqrt{m}||A||_\infty$

- $\frac{1}{\sqrt{m}}||A||_1 \leq ||A||_2 \leq \sqrt{n}||A||_1$

Teorema 3.2.1 *Se $||.||$ é uma norma vetorial de $\mathbb{R}^n$, então*

$$||A|| = \max_{||x||=1} ||Ax||$$

é uma norma matricial.

Ela denomina-se norma matricial natural, ou induzida, associada com a norma vetorial.

Teorema 3.2.2 *Se $||x||$ é uma norma de vetor em $\mathbb{R}^n$, então $||A|| = \max_{||x||=1}||Ax||$ define uma norma matricial em $M_{n\times n}$ que é compatível com a norma de vetor que a induz.*

Demonstração: (1) Certamente, $||Ax|| \geq 0$ para todos os vetores x, então, em particular, essa desigualdade é verdadeira se $||x|| = 1$. Assim, $||A|| = \max_{||x||=1} ||Ax|| \geq 0$. Se $||A|| = 0$, então deve-se ter $||Ax|| = 0$ e, portanto, $Ax = 0$, para todo x tal que $||x|| = 1$. Em particular, $Ae_i = 0$ para cada um dos vetores da base canônica e_i em $\mathbb{R}^n$. Mas Ae_i é a i-ésima coluna de A, logo, $A = 0$. Reciprocamente, se $A = 0$, tem-se que $||A|| = 0$.
(2) Seja α um escalar. Então:

$$||\alpha A|| = \max_{||x||=1} ||\alpha Ax|| = \max_{||x||=1} |\alpha|||Ax|| = |\alpha| \max_{||x||=1} ||Ax|| = |\alpha| \; ||A||$$

(3) Seja B uma matriz $n \times n$ e seja y um vetor unitário para o qual

$$||A + B|| = \max_{||x||=1} ||(A + B)x|| = ||(A + B)y||$$

Então:

$$||A + B|| = ||(A + B)y|| = ||Ay + By|| \leq ||Ay|| + ||By|| \leq ||A|| + ||B||$$

Agora, se $x = 0$, a desigualdade $||Ax|| \leq ||A|| \; ||x||$ é verdadeira, já que ambos os lados são nulos. Se $x \neq 0$, então

$$\frac{||Ax||}{||x||} \leq \max_{x \neq 0} \frac{||Ax||}{||x||} = ||A||$$

Portanto, $||Ax|| \leq ||A|| \; ||x||$.
(4) Seja z um vetor unitário tal que $||AB|| = \max_{||x||=1} ||(AB)x|| = ||ABz||$. Então:

$$||AB|| = ||ABz|| = ||A(Bz)|| \leq ||A|| \; ||Bz|| \leq ||A|| \; ||B|| \; ||z|| = ||A|| \; ||B||.$$

As desigualdades vêm da definição de uma norma matricial ser compatível a uma norma vetorial.

Isso completa a demonstração de que $||AB|| = \max_{||x||=1} ||(Ax)||$ define uma norma de matriz em $M_{n \times n}$ que é compatível com a norma de vetor $||x||$ que a induz.

A decomposição em valores singulares, que será estudada adiante, tem um papel importante na análise e resolução de sistemas lineares que são difíceis de resolver devido à sensibilidade a erros de arredondamento. No caso de um sistema linear consistente (quando existe pelo menos uma solução) $Ax = b$, isso ocorre quando a matriz de coeficientes é "quase singular", isto é, se um ou mais de seus valores singulares está próximo de zero. Esses sistemas lineares são ditos malcondicionados. Uma boa medida de quanto os erros de arredondamento afetam a precisão de uma solução calculada é dada pelo quociente do maior valor singular com o menor valor singular de A. Essa razão denomina-se número de condição de A e é denotada por

$$cond(A) = \frac{\sigma_1}{\sigma_k}$$

Quanto maior o número de condição, maior é a sensibilidade do sistema a pequenos erros de arredondamento.

3.3 Hiperplanos

Definição 3.3.1 *O conjunto de pontos* $(x_1, x_2, \ldots, x_n)$ *de* $\mathbb{R}^n$ *que satisfaz uma equação linear da forma*

$$a_1x_1 + a_2x_2 + \ldots + a_nx_n = b \qquad (a_1, a_2, \ldots, a_n \text{ nem todos nulos}) \tag{3.4}$$

é denominado um hiperplano de $\mathbb{R}^n$.

Assim, por exemplo, retas são hiperplanos em $\mathbb{R}^2$ e planos são hiperplanos em $\mathbb{R}^3$. Se $b = 0$ em (2.1), então a equação simplifica para

$$a_1x_1 + a_2x_2 + \ldots + a_nx_n = 0 \qquad (a_1, a_2, \ldots, a_n \text{ nem todos nulos}) \tag{3.5}$$

e diz-se que o hiperplano passa pela origem.

Uma equação linear $a_1x + a_2y = b$, na qual a_1 e a_2 não são ambos nulos, representa uma reta no plano xy e uma equação linear $a_1x + a_2y + a_3z = b$ na qual a_1, a_2 e a_3 não são todos nulos representa um plano no sistema xyz de coordenadas.

Quando for conveniente, pode-se expressar as equações (3.4) e (3.5) na notação de produto escalar como:

$$a.x = b \quad (a \neq 0) \tag{3.6}$$

e

$$a.x = 0 \quad (a \neq 0) \tag{3.7}$$

respectivamente, onde $a = (a_1, a_2, \ldots, a_n)$ e $x = (x_1, x_2, \ldots, x_n)$. A forma da Equação (3.7) revela que para um vetor não-nulo a fixado, o hiperplano $a.x = 0$ consiste de todos os vetores x de $\mathbb{R}^n$ que são ortogonais ao vetor a. Por isso, diz-se que (3.7) é o hiperplano pela origem com vetor normal a ou então o complemento ortogonal de a. Denota-se este hiperplano pelo símbolo $a^\perp$.

3.4 Reflexões de Householder

Nesta seção, apresenta-se um método planejado por Alton Householder que tem uma extensa aplicação, como por exemplo, a aproximação dos autovalores.

O método de Householder é utilizado para encontrar uma matriz simétrica tridiagonal B que seja similar a uma matriz simétrica dada A [3]. Sabe-se que A é similar a uma matriz diagonal D desde que exista uma matriz ortogonal Q com a propriedade que $D = Q^{-1}AQ = Q^tAQ$. Como a matriz Q (e consequentemente a matriz D) é geralmente difícil de ser calculada, o método de Householder oferece uma solução. Após este método ter sido implementado, métodos eficientes, como o algoritmo QR, podem ser usados para aproximações dos autovalores das matrizes simétricas tridiagonais resultantes.

Se v é um vetor não-nulo em $\mathbb{R}^2$ ou $\mathbb{R}^3$, então existe uma relação simples entre a projeção ortogonal sobre a reta $ger\{v\}$ e a reflexão no hiperplano $v^{\perp}$ que é ilustrada na Figura 3.1 em $\mathbb{R}^3$. Denotando a projeção ortogonal de um vetor x sobre a reta por $proj_v x$ e a reflexão de x no hiperplano por $refl_{v^{\perp}} x$, então a figura sugere que $x - refl_{v^{\perp}} x = 2proj_v x$ ou, equivalentemente, que $refl_{v^{\perp}} x = x - 2proj_v x$.

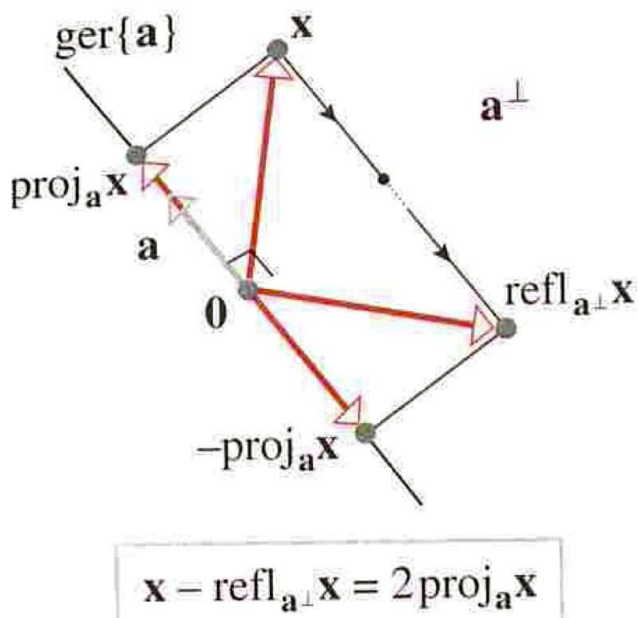

Fig. 3.1: Reflexão de Householder.

Motivados por esse resultado, apresenta-se a seguinte definição:

Definição 3.4.1 *Se v é um vetor não-nulo em $\mathbb{R}^n$ e se x é um vetor qualquer em $\mathbb{R}^n$, então a reflexão de x no hiperplano $v^{\perp}$ é denotada por $refl_{v^{\perp}} x$ e definida por*

$$refl_{v^{\perp}} x = x - 2proj_v x \tag{3.8}$$

O operador $T : \mathbb{R}^n \to \mathbb{R}^n$ definido por $T(x) = refl_{v^{\perp}} x$ é denominado a reflexão de $\mathbb{R}^n$ no hiperplano $v^{\perp}$.

Sabe-se dos conceitos de Projeção Ortogonal [9], [7] que:

$$refl_{a^{\perp}} x = x - 2\frac{x.a}{||a||^2} a \tag{3.9}$$

e que a matriz canônica $H_{a^{\perp}}$ da reflexão $refl_{a^{\perp}}$ é:

$$H_{a^{\perp}} = I - \frac{2}{a^T a} a a^T \tag{3.10}$$

No caso especial que o hiperplano é dado por um vetor unitário u temos $u^T u = ||u||^2 = 1$ de modo que as Equações (3.9) e (3.10) simplificam para:

$$refl_{u^{\perp}}x = x - 2(x.u)u \tag{3.11}$$

e

$$H_{u^{\perp}} = I - 2uu^T \tag{3.12}$$

Por causa do trabalho pioneiro do matemático norte-americano Alton Scott Householder na aplicação de reflexões em hiperplanos a algoritmos numéricos importantes, as reflexões em hiperplanos são, muitas vezes, denominadas **reflexões de Householder** ou **transformações de Householder** em sua homenagem.

Definição 3.4.2 *Uma matriz $n \times n$ da forma*

$$H = I - \frac{2}{a^T a} aa^T \tag{3.13}$$

na qual a é um vetor não-nulo em $\mathbb{R}^n$, é denominada uma matriz de Householder. Geometricamente, H é a matriz canônica da reflexão de Householder no hiperplano $a^{\perp}$.

Teorema 3.4.1 *As matrizes de Householder são simétricas e ortogonais.*

O fato de matrizes de Householder serem ortogonais significa que as reflexões de Householder preservam comprimentos. O próximo teorema mostra que a recíproca é verdadeira, ou seja, que dois vetores de mesmo comprimento são relacionados por alguma reflexão de Householder.

Teorema 3.4.2 *Se v e w são vetores distintos em $\mathbb{R}^n$ de mesmo comprimento, então a reflexão de Householder no hiperplano $(v - w)^{\perp}$ leva v em w e reciprocamente.*

O Teorema 3.4.2 é importante porque fornece uma maneira de usar reflexões de Householder para transformar um vetor dado em um vetor no qual algum componente específico é zero. Por exemplo, os vetores

$$v = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$$

e

$$w = \begin{bmatrix} ||v|| \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

têm o mesmo comprimento, de modo que o Teorema 2.4.2 garante que existe alguma reflexão de Householder que leva v em w. Além disso, o escalar $||v||$ poderia ter sido colocado em qualquer lugar de w, de modo que existem reflexões de Householder que levam v em um vetor com zeros em quaisquer $n - 1$ posições selecionadas.

Pode-se usar as reflexões de Householder para construir decomposições QR. Suponha, por exemplo, que procura-se uma decomposição QR de uma matriz 4×4 invertível.

$$A = \begin{bmatrix} x_{11} & x_{12} & x_{13} & x_{14} \\ x_{21} & x_{22} & x_{23} & x_{24} \\ x_{31} & x_{32} & x_{33} & x_{34} \\ x_{41} & x_{42} & x_{43} & x_{44} \end{bmatrix}_{4\times 4} \tag{3.14}$$

Deve-se encontrar três matrizes ortogonais Q_1,Q_2 e Q_3 tais que $Q_3Q_2Q_1A = R$ seja uma matriz triangular superior. Então pode-se reescrever essa equação como

$$A = Q_1^{-1}Q_2^{-1}Q_3^{-1}R = Q_1^TQ_2^TQ_3^TR = QR \tag{3.15}$$

que será uma decomposição QR de A com $Q = Q_1^TQ_2^TQ_3^T$.

Seja Q_1 a matriz de Householder de uma reflexão de Householder que "zera" as segunda, terceira e quarta entradas da primeira coluna de A. Assim, o produto Q_1A é da forma

$$Q_1A = \begin{bmatrix} x_{11} & x_{12} & x_{13} & x_{14} \\ 0 & x_{22} & x_{23} & x_{24} \\ 0 & x_{32} & x_{33} & x_{34} \\ 0 & x_{42} & x_{43} & x_{44} \end{bmatrix} \tag{3.16}$$

onde as entradas representadas com um x não precisam ser iguais àquelas em (3.11).

Agora, deseja-se introduzir zeros abaixo da diagonal principal na segunda coluna sem destruir os zeros já criados em (3.13). Para obter isso, olha-se para a submatriz 3×3 no canto inferior direito de (3.13) e toma-se a matriz de Householder H_2 de tamanho 3×3 de uma reflexão de Householder que "zera" as segunda e terceira entradas da primeira coluna da submatriz. Formando a matriz

$$Q_2 = \left[\begin{array}{c:ccc} 1 & 0 & 0 & 0 \\ \hdashline 0 & & & \\ 0 & & H_2 & \\ 0 & & & \end{array}\right]$$

tem-se que Q_2 é ortogonal e Q_2Q_1A é da forma

$$Q_2Q_1A = \begin{bmatrix} x_{11} & x_{12} & x_{13} & x_{14} \\ 0 & x_{22} & x_{23} & x_{24} \\ 0 & 0 & x_{33} & x_{34} \\ 0 & 0 & x_{43} & x_{44} \end{bmatrix} \tag{3.17}$$

Em seguida, olha-se para a submatriz 2×2 no canto inferior direito de (3.14) e toma-se a matriz de Householder H_3 de tamanho 2×2 de uma reflexão de Householder que "zera" a segunda entrada da primeira coluna da submatriz. Formando a matriz

$$Q_3 = \left[\begin{array}{cc:cc} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \hdashline 0 & 0 & \multicolumn{2}{c}{\multirow{2}{*}{H_3}} \\ 0 & 0 & & \end{array}\right]$$

tem-se que Q_3 é ortogonal e $Q_3Q_2Q_1A$ é da forma

$$Q_3Q_2Q_1A = \begin{bmatrix} x_{11} & x_{12} & x_{13} & x_{14} \\ 0 & x_{22} & x_{23} & x_{24} \\ 0 & 0 & x_{33} & x_{34} \\ 0 & 0 & 0 & x_{44} \end{bmatrix} \tag{3.18}$$

A matriz R do lado direito dessa equação é triangular superior, de modo que (3.12) é uma decomposição QR de A.

3.5 Decomposição QR

Apresenta-se aqui um algoritmo que tem por base o processo de Gram-Schmidt, denominado decomposição QR[8]. Este método tem alcançado, nos últimos anos, importância crescente como o fundamento matemático de uma variedade de algoritmos numéricos, inclusive os de calcular autovalores de matrizes grandes.

Teorema 3.5.1 *Se A é uma matriz $m \times n$ com colunas linearmente independentes, então A pode ser fatorada como $A = QR$, onde Q é uma matriz $m \times n$ cujas colunas formam uma base ortonormal para o espaço coluna de A e R é uma matriz triangular superior invertível $n \times n$.*

O procedimento para encontrar a decomposição QR de uma matriz $A_{m\times n}$ com colunas linearmente independentes é o seguinte:

Passo 1: Represente as colunas de A por $u_1, u_2, \ldots, u_n$ e seja $W = \mathbb{R}^n$ o subespaço de $\mathbb{R}^n$ que tem esses vetores como base.

Passo 2: Use o processo de Gram-Schmidt para transformar a base $\{u_1, u_2, \ldots, u_n\}$ para W em uma base ortonormal $\{w_1, w_2, \ldots, w_n\}$. Seja $Q = \begin{bmatrix} w_1 & w_2 & \ldots & w_n \end{bmatrix}$ a matriz cujas colunas são $w_1, w_2, \ldots, w_n$.

Passo 3: Calcule $R = [r_{ij}]$, onde $r_{ji} = u_i.w_j$.

3.6 Decomposição em Valores Singulares

Nesta seção, apresenta-se uma extensão da teoria de diagonalização de matrizes $n \times n$ simétricas para matrizes $m \times n$ arbitrárias. Os resultados que aqui são desenvolvidos têm aplicações à compressão, ao armazenamento e à transmissão de informação digitalizada e formam a base dos melhores algoritmos computacionais que estão atualmente disponíveis para resolução de sistemas lineares.

Qualquer matriz simétrica A pode ser expressa como $A = PDP^T$, onde P é uma matriz ortogonal $n \times n$ de autovetores de A e D é a matriz diagonal cujas entradas diagonais são os autovalores associados aos vetores coluna de P. Diz-se que essa é uma decomposição em autovalores de A.

Teorema 3.6.1 (Teorema de Schur) *Se A é uma matriz $n \times n$ com entradas reais e autovalores reais, então existe uma matriz ortogonal P tal que $P^T AP$ é uma matriz triangular superior da forma*

$$P^T AP = \begin{bmatrix} \lambda_1 & x_{12} & x_{13} & \dots & x_{1n} \\ 0 & \lambda_2 & x_{23} & \dots & x_{2n} \\ 0 & 0 & \lambda_3 & \dots & x_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \lambda_n \end{bmatrix}$$

na qual $\lambda_1, \lambda_2, \dots, \lambda_n$ são os autovalores da matriz A repetidos de acordo com a multiplicidade. Pode-se denotar a matriz triangular superior acima por S (de Schur), e então a equação pode ser reescrita como $A = PSP^T$, que é denominada uma decomposição de Schur de A.

Teorema 3.6.2 (Teorema de Hessenberg) *Se A é uma matriz $n \times n$ não simétrica, com entradas reais, então existe uma matriz ortogonal P tal que $P^T AP$ é da forma*

$$P^T AP = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1(n-2)} & x_{1(n-1)} & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2(n-2)} & x_{2(n-1)} & x_{2n} \\ 0 & x_{32} & \ddots & x_{3(n-2)} & x_{3(n-1)} & x_{3n} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & x_{(n-1)(n-2)} & x_{(n-1)(n-1)} & x_{(n-1)n} \\ 0 & 0 & \dots & 0 & x_{n(n-1)} & x_{nn} \end{bmatrix}$$

Pode-se denotar esta matriz por H (de Hessenberg), e então a equação pode ser reescrita como $A = PHP^T$, que é denominada uma decomposição de Hessenberg superior de A.

Em outras palavras, o Teorema 3.6.1 afirma que qualquer matriz quadrada A com autovalores reais é ortogonalmente semelhante a uma matriz triangular superior que tem os autovalores de A na diagonal principal, enquanto o Teorema 3.6.2 afirma que qualquer matriz quadrada com entradas reais é ortogonalmente semelhante a uma matriz na qual cada entrada abaixo da primeira subdiagonal é zero. Os elementos da primeira subdiagonal estão destacados na Matriz 3.6. Essa matriz é denominada matriz de Hessenberg superior.

$$\begin{bmatrix} x_{11} & x_{12} & x_{13} & x_{14} & x_{15} \\ X & x_{22} & x_{23} & x_{24} & x_{25} \\ x_{31} & X & x_{33} & x_{34} & x_{35} \\ x_{41} & x_{42} & X & x_{44} & x_{45} \\ x_{51} & x_{52} & x_{53} & X & x_{55} \end{bmatrix}$$

Matriz 3.6

As decomposições em autovalores, de Hessenberg e de Schur são importantes em algoritmos numéricos porque, além das matrizes D,H e S possuírem formatos mais simples do que A, as matrizes ortogonais que aparecem nessas fatorações não aumentam os erros de arredondamento, já que assim a matriz n?o possui elementos próximos de zero.

Teorema 3.6.3 *Se A é uma matriz $m \times n$, então*

1. *A e A^TA têm o mesmo espaço nulo.*

2. *A e A^TA têm o mesmo espaço linha.*

3. *A^T e A^TA têm o mesmo espaço coluna.*

4. *A e A^TA têm o mesmo posto.*

Teorema 3.6.4 *Se A é uma matriz $n \times n$, então as afirmações seguintes são equivalentes:*

1. *A é ortogonalmente diagonalizável.*

2. *A tem um conjunto ortonormal de n autovetores.*

3. *A é simétrica.*

Teorema 3.6.5 *Se A é uma matriz $m \times n$, então*

1. *A^TA é ortogonalmente diagonalizável.*

2. *Os autovalores de A^TA são não negativos.*

Demonstração (1): A matriz A^TA, por ser simétrica, é ortogonalmente diagonalizável pelo Teorema 3.6.4.
Demonstração 2: Como A^TA é ortogonalmente diagonalizável, existe alguma base ortonormal de $\mathbb{R}^n$ consistindo em autovetores de A^TA, a saber $v_1, v_2, \ldots, v_n$. Se $\lambda_1, \lambda_2, \ldots, \lambda_n$ forem os autovalores correspondentes, então, dado qualquer $1 \leq i \leq n$, tem-se

$$||Av_i||^2 = Av_i.Av_i = v_i.A^TAv_i = v_i.\lambda_i v_i = \lambda_i(v_i.v_i) = \lambda_i||v_i||^2 = \lambda_i$$

Segue dessa relação que $\lambda_i \geq 0$.

Definição 3.6.1 *Se A for uma matriz $m \times n$ e $\lambda_1, \lambda_2, \ldots, \lambda_n$ os autovalores de $A^T A$, então os números*

$$\sigma_1 = \sqrt{\lambda_1}, \quad \sigma_2 = \sqrt{\lambda_2}, \ldots, \quad \sigma_n = \sqrt{\lambda_n}$$

são denominados valores singulares de A.

Vamos supor que os autovalores de $A^T A$ são tais que $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n \geq 0$ e, portanto, que $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_n \geq 0$.

Antes de enunciar o resultado principal desta seção, que diz respeito a uma maneira específica de fatorar uma matriz A arbitrária de tamanho $m \times n$, convém definir diagonal principal para matrizes que não são quadradas.

Definição 3.6.2 *Define-se a diagonal principal de uma matriz $m \times n$ como a fileira de entradas que começa no canto superior esquerdo e estende-se diagonalmente até onde for possível. Diz-se que as entradas nessa diagonal principal são as entradas diagonais da matriz.*

Teorema 3.6.6 *Se A é uma matriz $m \times n$, então A pode ser expressa como*

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

em que U e V são matrizes ortogonais e Σ é uma matriz $m \times n$ cujas entradas diagonais são os valores singulares de A e as demais entradas são nulas.
Essa fatoração é denominada **decomposição em valores singulares** *de A e abrevia-se com as iniciais em inglês Singular Value Decomposition (SVD).*

O próximo teorema mostra que a decomposição em valores singulares de uma matriz A está relacionada com os quatro espaços fundamentais de A.

Teorema 3.6.7 *Seja $A = U \Sigma V^T$ uma decomposição por valores singulares de uma matriz $A_{m \times n}$. Sejam $\sigma_1, \ldots, \sigma_r$ todos os valores singulares não nulos de A e u_i a coluna de U. Então:*

1. *O posto de A é r.*
2. *$\{u_1, u_2, \ldots, u_r\}$ é uma base ortonormal para $col(A)$.*
3. *$\{u_{r+1}, \ldots, u_m\}$ é uma base ortonormal para $nul(A^T)$.*
4. *$\{v_1, v_2, \ldots, v_r\}$ é uma base ortonormal para $lin(A)$.*
5. *$\{v_{r+1}, \ldots, v_n\}$ é uma base ortonormal para $nul(A)$.*

3.7 Compressão de dados usando Decomposição em Valores Singulares

A transmissão e o armazenamento eficientes de grandes quantidades de dados digitalizados têm se tornado um dos maiores problemas do mundo tecnológico. Discute-se aqui o papel que a decomposição em valores singulares tem na compressão de dados digitalizados de modo que possam ser transmitidos mais rapidamente e que ocupem menos espaço de armazenamento [5].

O Teorema 3.6.5 tem sua versão expandida e reduzida. Nesta última versão, tem-se

$$A = \begin{bmatrix} u_1 & u_2 & \dots & u_k \end{bmatrix} \begin{bmatrix} \sigma_1 & 0 & \dots & 0 \\ 0 & \sigma_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_k \end{bmatrix} \begin{bmatrix} v_1^T \\ v_2^T \\ \vdots \\ v_k^T \end{bmatrix}$$

que é denominada uma decomposição em valores singulares reduzida de A. Multiplicando o lado direito da igualdade, tem-se

$$A = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \dots + \sigma_k u_k v_k^T$$

que é denominada expansão em valores singulares reduzida de A. Esse resultado aplica-se a qualquer matriz.

Detalhes desta aplicação serão explicados no próximo capítulo.

Capítulo 4

Aplicação de SVD à Imagens

A decomposição por valores singulares é uma ferramenta extremamente importante, tanto prática quanto teoricamente. Neste capítulo, apresentam-se algumas de suas aplicações.

4.1 Posto

Até agora, não há preocupação em calcular o posto de uma matriz sob um ponto de vista computacional. O cálculo é feito por redução nas linhas para a forma escalonada e conta-se o número de linhas não nulas [4]. No entanto, erros de arredondamento podem afetar este processo, principalmente se a matriz for mal condicionada. Elementos que deveriam ser zero podem ser números muito pequenos, interferindo na capacidade de determinar com precisão o posto e outras quantidades relacionadas à matriz.

Na prática, o método SVD é frequentemente utilizado para encontrar o posto de uma matriz, pois é muito mais confiável quando ocorrem erros por arredondamento. A ideia é que as matrizes ortogonais U e V de SVD preservam o comprimento e, portanto, não introduzem nenhum erro adicional, e qualquer erro que ocorrer tenderá a aparecer na matriz Σ.

4.2 Norma de Matrizes

O método SVD pode fornecer fórmulas simples para expressões que envolvem normas de matrizes. Considere, por exemplo, a norma de Frobenius para uma matriz. O teorema a seguir mostra que ela é determinada pelos valores singulares da matriz [6].

Teorema 4.2.1 *Seja A uma matriz $m \times n$ e sejam $\sigma_1, \ldots, \sigma_r$ todos os valores singulares não nulos de A. Então:*

$$||A||_F = \sqrt{\sigma_1^2 + \ldots + \sigma_r^2}$$

A demonstração desse resultado depende do teorema abaixo:

Teorema 4.2.2 *Se A é uma matriz $m \times n$ e Q é uma matriz ortogonal $m \times m$, então*

$$||QA||_F = ||A||_F \tag{4.1}$$

Demonstração do Teorema 4.2.1: Seja $A = U\Sigma V^T$ uma decomposição por valores singulares de A. Então, utilizando a equação (4.1) duas vezes, temos

$$\begin{aligned} ||A||_F^2 &= ||U\Sigma V^T||_F^2 \\ &= ||\Sigma V^T||_F^2 = ||(\Sigma V^T)^T||_F^2 \\ &= ||V\Sigma^T||_F^2 = ||\Sigma^T||_F^2 = \sigma_1^2 + \ldots + \sigma_r^2 \end{aligned}$$

que prova o resultado.

4.3 Compressão de Imagem Digital

As decomposições em valores singulares também podem ser utilizadas para "comprimir" informação visual com o objetivo de reduzir seu espaço de armazenamento e acelerar sua transmissão eletrônica por satélite, fax, Internet ou semelhante. O primeiro passo na compressão de uma imagem visual é representá-la como uma matriz numérica, a partir da qual a imagem possa ser recuperada quando for necessário [1].

Por exemplo, uma fotografia em preto e branco pode ser escaneada como um arranjo retangular de pixels (pontos) e armazenada como uma matriz A, associando a cada pixel um valor numérico de acordo com seu tom de cinza. Se utilizar 256 níveis de cinza, sendo $0 =$ branco e $255 =$ preto, então as entradas na matriz serão números inteiros entre 0 e 255. A imagem pode ser recuperada a partir da matriz A imprimindo ou exibindo os pixels com seus níveis de cinza associados.

Se a matriz A for de tamanho $m \times n$, então pode-se armazenar cada uma de suas mn entradas individualmente. Um procedimento alternativo é calcular a decomposição em valores singulares reduzida

$$A = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \ldots + \sigma_k u_k v_k^T$$

na qual $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_k$ e armazenar os números σ, e os vetores u_i e v_i. Quando for preciso, a matriz A pode ser reconstruída a partir da equação acima. Como cada vetor u tem m entradas e cada vetor v tem n entradas, este método requer espaço de armazenamento para

$$km + kn + k = k(m + n + 1)$$

números. Supõe-se, entretanto, que os valores singulares $\sigma_{r+1}, \ldots, \sigma_k$ sejam suficientemente pequenos, a ponto de serem ignorados, e produzam, assim, uma aproximação aceitável

$$A_r = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \ldots + \sigma_r u_r v_r^T$$

de A e da imagem que A representa. Diz-se que esta é a aproximação de posto r de A. Essa matriz requer espaço de armazenamento para apenas

$$rm + rn + r = r(m + n + 1)$$

números, ao contrário dos mn números requeridos para um armazenamento "entrada à entrada" de A. Por exemplo, uma aproximação de posto 100 de uma matriz A de tamanho 1000×1000 requer espaço de armazenamento para apenas

$$100(1000 + 1000 + 1) = 200100$$

números, ao contrário de 1 milhão de números requeridos no armazenamento "entrada à entrada", dando uma compressão de quase 80%.

Considere agora uma figura em preto e branco, de tamanho 340×280 pixels. Cada pixel é um dos 256 tons de cinza, que pode-se representar por números entre 0 e 255. Pode-se armazenar estas informações em uma matriz A 340×280, mas transmitir e manipular esses 95200 números é muito caro. A ideia por trás da compressão de imagens é que algumas partes da figura são menos importantes que outras. Por exemplo, em uma fotografia de alguém em pé ao ar livre, pode haver muito céu de plano de fundo, enquanto o rosto da pessoa contém muitos detalhes. Provavelmente, pode-se evitar uma transmissão de um pixel a cada dois ou três pixels no plano de fundo, mas desejando manter todos os pixels na região do rosto.

O menor valor singular na SVD da matriz A provém das partes "menos interessantes" da imagem, e pode-se ignorar a maioria delas. Suponha, então, a SVD de A na forma de produto externo

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

Seja $k \leq r$ e defina

$$A_k = \sigma_1 u_1 v_1^T + \ldots + \sigma_k u_k v_k^T$$

Então, A_k é uma aproximação de A que corresponde a manter somente os k primeiros valores singulares e os correspondentes vetores singulares.

Vale ressaltar que este procedimento reduz o número de colunas das matrizes U e V necessárias para reproduzir uma aproximação de A, e consequentemente, reduz a quantidade de dados para sua representação. Porém, há uma diretriz, mas não uma metodologia para selecionar os valores de σ a serem ignorados: isto está mais relacionado à prática do usuário.

Os valores singulares tornam-se rapidamente pequenos da primeira para a última coluna nas matrizes diagonais. O primeiro valor singular de cada matriz da imagem possui um valor muito alto, e consideravelmente maior que todos os demais. Por este motivo, ele é o valor singular que "guarda" mais informações da imagem e, visivelmente, é o principal formador desta. Por outro lado, a maior parte dos valores singulares são bem pequenos e, em consequência disto, a(s) matriz(es) da imagem pode(m) ser apromimada(s) por outra(s) de posto muito menor. Isto ocorre porque geralmente os valores de um dado elemento desta(s) matriz(es) estão próximos dos valores dos elementos vizinhos. Em outras palavras, uma "linha de píxeis" qualquer na imagem é, normalmente, muito parecida à "linha de píxeis" vizinha.

Para o exemplo de 340×280, é suficiente transmitir somente os dados correspondentes aos 20 primeiros valores singulares. Em vez de transmitir 95200 números, precisa-se enviar somente 20 valores singulares mais os 20 vetores $u_1, \ldots, u_{20}$ de $\mathbb{R}^{340}$ e os 20 vetores $v_1, \ldots, v_{20}$ de $\mathbb{R}^{280}$, o que dá um total de

$$20 + 20.340 + 20.280 = 12420$$

números. Isso representa uma economia substancial de quase 99%!.

Exemplo 4.3.1 *A foto do matemático Gauss é uma imagem em* 340×280 *pixels. Ela tem* 256 *tons de cinza, portanto, a matriz* A *correspondente é* 340×280*, com elementos entre* 0 *e* 255*.*

A matriz A*, portanto, tem posto* 280*. Aproximando* A *por* A_k*, obtém-se uma imagem que corresponde aos* k *primeiros valores singulares de* A*. A figura a seguir mostra várias dessas imagens para valores* k *de* 2 *a* 256*. De início, a imagem está muito embaçada, mas ela toma forma muito rapidamente. Note que* A_{32} *já deu uma aproximação bem razoável para a imagem real (a qual vem de* $A = A_{280}$*).*

Alguns dos valores singulares de A *são* $\sigma_1 = 49,096, \sigma_{16} = 22,589, \sigma_{32} = 10,187, \sigma_{64} = 484, \sigma_{128} = 182, \sigma_{256} = 5$ *e* $\sigma_{280} = 0,5$*. O menor valor singular contribui muito pouco para a imagem, o que é a razão de a aproximação rapidamente parecer tão próxima da original.*

Fig. 4.1: Gauss.

O processamento digital da imagem foi realizado em Matlab. Para iniciar é feita a leitura da foto, onde uma variável recebe uma "matriz"$m \times n \times 3$, onde m e n são a dimensão da figura em pixels. Cada elemento (i, j, k) representa a quantidade do k-ésimo tom no pixel da posição i, j (um inteiro entre 0 e 255). Para $k = 1$ tem-se a quantidade de vermelho, $k = 2$ de verde e$k = 3$ de azul.

Então é realizado o cálculo da SVD para cada uma das matrizes de diferentes tonalidades, e seu posto é calculado. Quanto maior o posto da matriz, melhor será a nitidez da imagem e maior o custo computacional de sua transferência.

A foto original é uma imagem em 893×1230 pixels. A matriz A, portanto, tem posto 893. Aproximando A por A_k, obtém-se uma imagem que corresponde aos k primeiros valores singulares de A. De início, a imagem está pouco nítida, mas note que quanto maior o valor de k, mais próxima da imagem original a imagem reproduzida está.

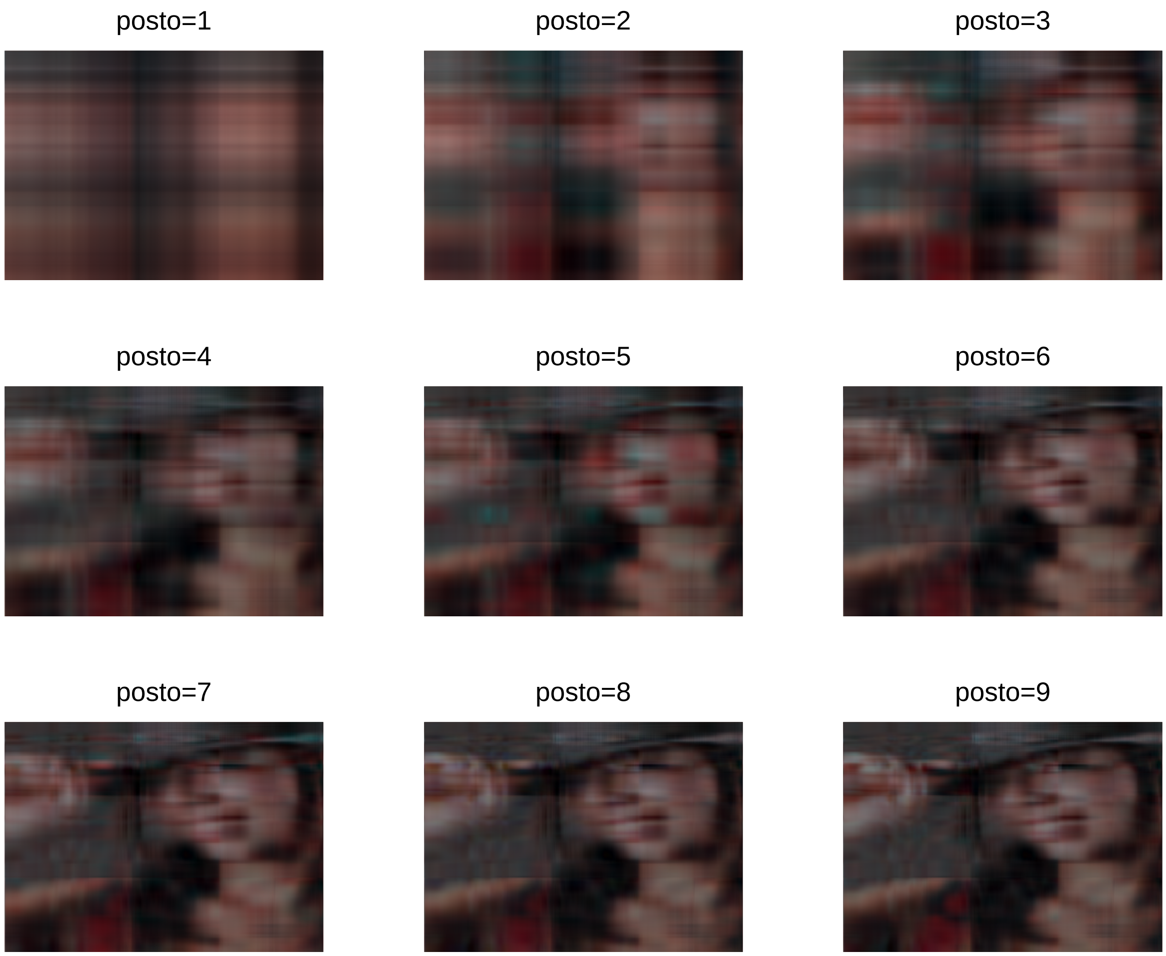

Fig. 4.2: Teste 1.

Fig. 4.3: Teste 2.

Conclusão

As imagens digitais são muito importantes na aplicação em diversas áreas. Elas são transmitidas na forma de matrizes extensas, que exigem muito espaço de armazenamento. Por isso, é necessário o estudo por métodos que acelerem a transmissão eletrônica de imagens e utilizem menos espaço de armazenamento, através da compressão de imagem. Um destes métodos é a Decomposição em Valores Singulares (SVD - Singular Values Decomposition), aplicada a quaisquer tipos de matrizes.

A Decomposição em Valores Singulares têm aplicações à compressão, armazenamento e transmissão de imagens digitais. Ela está relacionada ao problema matemático de aproximar uma matriz por outra de posto menor.

Os objetivos deste trabalho foram revisar conceitos de álgebra linear, discutir a teoria da Decomposição em Valores Singulares, estudar uma aplicação da SVD e verificar a eficiência da SVD através de um algoritmo no software MATLAB. Considera-se que todos os objetivos estabelecidos foram alcançados devido aos resultados obtidos mediante a aplicação do método proposto.

Para a compressão de uma imagem digital qualquer, faz-se necessária sua representação em forma matricial, onde cada elemento da matriz equivale ao pixel de posição correspondente na imagem, e o valor do elemento representa a intensidade luminosa do pixel. Após o processo de representação da imagem digital em forma matricial, pode-se iniciar o processo da compressão por SVD propriamente dita. Foi desenvolvido um programa no software MATLAB que realiza a decomposição em valores singulares de uma matriz e armazena apenas as informações de maior relevância, ou seja, os valores singulares de maior módulo. A operação da SVD é expressa da seguinte forma: $A = U\Sigma V^T$ onde U e V são matrizes ortogonais e Σ é uma matriz diagonal de valores singulares. Essa expressão também pode ser escrita na seguinte forma: $A = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \ldots + \sigma_n u_n v_n^T$ onde σ_i, u_i e v_i são a i-ésima coluna das matrizes σ, U e V, respectivamente.

A compressão por SVD foi testada com imagens coloridas, mas é possível realizá-la utilizando somente tons de cinza.

Neste trabalho, observamos a eficiência da SVD como ferramenta de compressão de imagens e algumas relações entre quantidade de dados armazenados e fidelidade da imagem comprimida.

Vale citar que o trabalho que aqui se conclui exigiu uma série de conhecimentos, de caráter diverso, para que os objetivos fossem alcançados.

Este trabalho também exigiu conceitos mais avançados de Álgebra Linear. Por isso, ressalta-se o aprendizado obtido por meio deste, já que foi possível estudar e compreender assuntos que não foram abordados no curso de licenciatura em matemática.

No entanto, não foi possível aprofundar o tema de maneira computacional. Contudo trabalhos futuros podem abordar este problema, com o desenvolvimento de algoritmos mais eficientes.

Por fim, acrescento que a experiência de pesquisa, produção de relatórios, orientação e apresentação, ao lado de todos os conteúdos aprendidos no decorrer do trabalho, forneceu base necessária para um possível mestrado em Matemática Aplicada.

Referências Bibliográficas

[1] H. Anton and R. C. Busby. *Álgebra Linear Contemporânea*. Bookman, Porto Alegre, 1 edição edition, 2011.

[2] J. L. Boldrini. *Álgebra Linear*. Harper Row do Brasil, São Paulo, 3 edicão edition, 1980.

[3] R. L. Burden. *Análise Numérica*. Pioneira Thomson Learnig, São Paulo, 1 edição edition, 2003.

[4] I.. Camargo and P. Boulos. *Geometria Analítica*. Prentice Hall, São Paulo, 3 edicão edition, 2005.

[5] Liliane Ribeiro da Silva. *Aplicação da Decomposição em Valores Singulares e Análise de Componentes Independentes em dados fMRI*. PhD thesis, Março 2011. Dissertação de Mestrado.

[6] G. H. Golub and F. Charles Van Loan. *Matrix Computations 2nd Edition*. The Johns Hopkins University Press, Baltimore, Maryland, 1989.

[7] A. Howard and C. Rorres. *Álgebra Linear com Aplicações*. Bookman, Porto Alegre, 10 edicão edition, 2012.

[8] D. Poole. *Álgebra Linear*. Cengage Learning, São Paulo, 1 edição edition, 2004.

[9] G. Strang. *Algebra Linear e suas Aplicações*. Cengage Learning, São Paulo, 2 edição edition, 2012.

Apêndice A

Programa em Matlab

O MATLAB (abreviação de "laboratório de matrizes- Matrix Laboratory) é um sistema para cálculos matemáticos e matriciais, o qual pode ser imaginado como uma espécie de linguagem de programação. Todas as variáveis são tratadas como matrizes pelo MATLAB, com uma característica especial: são dimensionadas automaticamente, fato que facilita sobremaneira a implementação de algoritmos matriciais. Outra vantagem do uso do MATLAB é o seu extenso conjunto de rotinas de representação gráfica.

Uma característica bastante útil e prática do MATLAB é a de que as suas variáveis não precisam ser dimensionadas antes de serem usadas. As variáveis são geradas e dimensionadas automaticamente ao serem referenciadas pela primeira vez em uma atribuição de valores, permanecendo na memória de trabalho até que esta seja limpa.

A toolbox do MATLAB permite trabalhar com 4 tipos de imagens:

- imagens indexadas;
- imagens de intensidade;
- imagens binarizadas e
- imagens RGB.

As imagens indexadas requerem duas matrizes: uma delas tem as dimensões da imagem e cada ponto desta matriz especifica um índice que serve para pesquisar em uma segunda matriz, que contém o mapa de cores, quais são os componentes R (Vermelho - Red), G (Verde - Green) e B (Azul - Blue) de cada pixel.

A.1 Programa

O programa utilizado nesse trabalho é apresentado a seguir:

A=imread('arquivo.jpg');
A=double(A); Converte a matriz para double
M=max(max(A)); M é um vetor de 3 componentes que contém a tonalidade máxima de cada cor.
Z=zeros(size(A));
$[U1, S1, V1] = svd(A(:, :, 1))$;
$[U2, S2, V2] = svd(A(:, :, 2))$; calcula a decomposião SVD de cada uma das
$[U3, S3, V3] = svd(A(:, :, 3))$; matrizes de diferentes tonalidades.
r=min(size(S1));
while $(S1(r, r) < tol)|(S2(r, r) < tol)|(S3(r, r) < tol)$
$r = r - 1$; calcula o posto da matriz de menor posto, dentre as três.
end
$cols = 3$;
$s = floor((r - 2 * cols^2)/cols^2 - 1)$; são impressas 3 páginas de figuras:
de posto 1 até $cols^2$, de posto $cols^2 + 1$ até $2 * cols^2$ e
de posto maiores que $2 * cols^2$ com um passo de tamanho s.
$k = 0$; $cont = 0$;
figure(1)
for $j = 1 : r$
$Z(:, :, 1) = Z(:, :, 1) + S1(j, j) * U1(:, j) * V1(:, j)'$;
$Z(:, :, 2) = Z(:, :, 2) + S2(j, j) * U2(:, j) * V2(:, j)'$;
$Z(:, :, 3) = Z(:, :, 3) + S3(j, j) * U3(:, j) * V3(:, j)'$;
para cada$j$, Z é a matriz de posto j mais próxima de A
if $((j <= 2 * cols^2)|(mod(j, s) == 1))\&(cont < 3 * cols^2 - 1)$
$k = mod(k, cols^2) + 1$;
subplot(cols,cols,k)
for$i = 1 : 3$
$maxZ = max(max(Z))$;
$W(:, :, i) = (M(i)/255) * abs(Z(:, :, i))/(maxZ(i))$;
end
image(W);
$cont = cont + 1$;
title(['posto=' num2str(j)])
if j==$cols^2$
elseif j==$2 * cols^2$
figure(3)
subplot(cols,cols,$cols^2$)
image(A/255); a última posição contém a figura original.
for i=length(num2str(r)):5
aux=[aux ' '];
title(['matriz original'; ' posto=' num2str(r) aux]);