

UNIVERSIDADE FEDERAL DO AMAZONAS - UFAM
INSTITUTO DE CIÊNCIAS EXATAS - ICE
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
DEPARTAMENTO DE APOIO À PESQUISA
PROGRAMA INSTITUCIONAL DE INICIAÇÃO CIENTÍFICA

APLICAÇÃO DA DECOMPOSIÇÃO EM VALORES SINGULARES À
ANÁLISE DE DADOS MULTIVARIADOS

Bolsista: Stéfane Rêgo Gandra - CNPQ

Manaus
2013

UNIVERSIDADE FEDERAL DO AMAZONAS - UFAM
INSTITUTO DE CIÊNCIAS EXATAS - ICE
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
DEPARTAMENTO DE APOIO À PESQUISA
PROGRAMA INSTITUCIONAL DE INICIAÇÃO CIENTÍFICA

RELATÓRIO FINAL
PIB - E/0009/2012
APLICAÇÃO DA DECOMPOSIÇÃO EM VALORES SINGULARES À
ANÁLISE DE DADOS MULTIVARIADOS

Bolsista: Stéfane Rêgo Gandra - CNPq

Orientador: Prof. Dr. Sandro Dimy Barbosa Bitar

Manaus
2013

RESUMO

Neste trabalho investigamos a aplicação da teoria da Decomposição em Valores Singulares à Análise de Dados Multivariados, em especial à Análise de Componentes Principais e Compressão de Imagens. Os fundamentos teóricos discorrem sobre alguns elementos de Álgebra Linear e Algoritmos. Além disso, apresenta-se um resultado, ao final do relatório, relacionado a Compressão de Imagens via Decomposição em Valores Singulares.

Palavras-chave: Decomposição em Valores Singulares, Análise de Componentes Principais, Compressão de Imagens.

ABSTRACT

This paper presents the application of singular value decomposition the multivariate data analysis, in particular the Principal Component Analysis and Image Compression. The theoretical discourse on some Elements of Linear Algebra and Algorithms. Furthermore, presents a result, the end of the report, related to Image Compression via the Singular Value Decomposition.

Keywords: Singular Value Decomposition, Principal Component Analysis, Image Compression.

Sumário

Introdução	7
Objetivos	8
Capítulo 1 - Definições e Teoremas básicos	9
0.1 Matrizes	9
0.2 Produto de Matrizes	11
0.3 Matrizes blocadas (ou em bloco)	12
Capítulo 2 - Normas	13
0.4 Norma de um vetor	14
0.5 Norma de matrizes	14
0.6 Processo de Gram-Schmidt	16
Capítulo 3 - Decomposição em Valores Singulares	19
0.7 Decomposição em Valores Singulares	19
Capítulo 4 - Aplicação da Decomposição em Valores Singulares na Análise de Componentes Principais	21
0.8 Análise de Componentes Principais	22
0.8.1 Encontrando Subespaços Ótimos	23
Aplicação da Decomposição em Valores Singulares à Compressão de Imagens	25
Conclusão	30

Referências	31
Apêndice	32

Introdução

Em modelagem matemática quase sempre se depara com a necessidade de análise e tratamento de um grande volume de dados observados (Data Mining, Processamento de Imagens, Reconhecimento de Padrões, Bioestatística, Análise Econômica e outros). Uma grande quantidade de variáveis explicativas ou categóricas dificulta o entendimento do fenômeno a ser estudado. A dimensão do espaço sob o qual está sujeito o problema se torna um complicador e a busca por subespaços de dimensão menor que, de alguma forma, destaquem as variáveis latentes que sumarizam as informações relativas ao fenômeno ou problema em estudo, é desejado. Algumas técnicas em Análise de Dados Multivariados propõem a redução ao número de variáveis. São técnicas baseadas em modelos matemáticos que reduzem a dimensão do problema resguardando o máximo de variabilidade computada nos valores das variáveis originais com base nos dados observados. Neste relatório expõe-se como os fundamentos da ACP podem ser justificados com a teoria da SVD. O entendimento dessas técnicas necessita de uma boa formação matemática. Assim, este relatório consiste em expor toda base teórica das técnicas citadas anteriormente.

Como aplicação, utilizou-se os resultados da SVD num processo de compressão de Imagem. Sabemos que a transmissão e armazenamento de dados têm um papel de extrema importância para tecnologia moderna e é desejável que sua execução seja realizada com o mínimo espaço de armazenamento e o menor tempo de processamento. Muitas vezes, as informações processadas são muito extensas, ocupando muito espaço de armazenamento e levando muito tempo a serem transmitidas. Para resolver este problema, podemos condensar a informação, excluindo os dados redundantes. Tal efeito de eliminação de redundância é chamado de compressão. Registra-se que, o foco desse trabalho não é, exatamente o de examinar a eficiência da Decomposição em Valores Singulares, aplicada à compressão de imagens comparada a outros métodos, e sim difundir a idéia de que as decomposições que se estudam em Álgebra Linear servem para extrair das matrizes propriedades primordiais e, no caso da SVD, propriedades que expliquem quase fielmente uma imagem.

Foi utilizado o software FreeMat¹ para cálculos, simulações e implementação do modelo, assim como programação em linguagem baseadas em script.

¹FreeMat Version VERSION(Licensed under the GNU Public License. Web: <http://freemat.sf.net>)

Objetivos

O projeto tinha como principal objetivo um estudo em Álgebra Linear avançada relacionado à Decomposição em Valores Singulares (SVD) de uma matriz A , definida sobre o corpo dos números reais, e suas contribuições à Análise de Dados Multivariados. Neste relatório final será exposto os resultados que obtivemos, que foi caracterizado pelo fortalecimento, compreensão e aprendizado matemático de alguns conceitos essenciais a investigação dos principais Teoremas e Corolários relacionados com a Decomposição em Valores Singulares de uma matriz e, particularidades oriundas desses resultados, com ênfase na Análise de Componentes Principais(ACP) e na Compressão de Imagens.

Capítulo 1 - Definições e Teoremas básicos

0.1 Matrizes

Definição 1.1. *Uma matriz sobre o corpo dos números reais é um arranjo ordenado de $m \times n$ elementos ou escalares chamados de termos da matriz.*

Exemplo 1.1.

$$\begin{bmatrix} 2 & 0 \\ 1 & 3 \end{bmatrix}, \begin{bmatrix} 0 & 3 & 1 \\ 4 & 5 & 6 \\ 9 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 3 \\ 5 & 0 \\ 4 & 9 \end{bmatrix}, \begin{bmatrix} 9 & 0 & 7 \\ 5 & 1 & 5 \end{bmatrix} \text{ e } \begin{bmatrix} 1 & 8 & 9 & 0 \\ 4 & 3 & 7 & 9 \\ 8 & 6 & 4 & 9 \end{bmatrix}$$

A ordem de uma matriz descreve o número de linhas e colunas dessa matriz. Uma matriz de ordem $m \times n$ (pronuncia-se m por n) quando tem m linhas e n colunas. Consequentemente, os exemplos que acabamos de ver são matrizes 2×2 , 3×3 , 3×2 , 2×3 e 3×4 , respectivamente. Uma matriz $1 \times m$ é chamada de matriz linha (ou vetor linha), e uma matriz $n \times 1$ é chamada de matriz coluna (ou vetor coluna).

Usamos a notação índice subscrito duplo para nos referirmos aos elementos de uma matriz A : o elemento de A na linha i e coluna j é denotado por a_{ij} . Com essa notação, uma matriz genérica $m \times n$ tem a forma :

$$A_{(i,j)_{m \times n}} = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} \end{bmatrix}$$

Os elementos da diagonal A são $\{a_{11}, a_{22}, a_{33}, \dots, a_{mm}\}$ quando $m = n$ e, se $m = n$ (isto é, se A tem o mesmo número de linhas e colunas), A é chamada de matriz quadrada. Uma matriz quadrada cujos elementos fora da diagonal são todos nulos é chamada de matriz diagonal. Uma matriz diagonal em que todos os elementos da diagonal são iguais é chamada de matriz escalar.

Definição 1.2. *Uma matriz quadrada com entradas 1 na diagonal principal e com entradas 0 fora da diagonal principal é chamada de matriz identidade e é denotada por I . Se for importante enfatizar seu tamanho, nós escreveremos I_n para a matriz*

identidade de ordem n .

Exemplo 1.2.

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, C = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \text{ e } D = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Se A é uma matriz $m \times n$, com $m = n$, então $AI_n = A$ e $I_n A = A$. Assim, uma matriz identidade desempenha na aritmética matricial o mesmo papel que o número 1 desempenha nas relações numéricas $a \times 1 = 1 \times a = a$.

Definição 1.3. Se A é uma matriz $m \times n$ qualquer, então a transposta de A denotada por A^T , é definida como a matriz $n \times m$ que resulta da permutação das linhas com as colunas de A , ou seja, a primeira coluna de A^T é a primeira linha de A , a segunda coluna de A^T é a segunda linha de A , e assim por diante. A seguir, alguns exemplos de matrizes e suas transpostas.

Exemplo 1.3.

$$A = \begin{bmatrix} 2 & 4 & 3 \\ 1 & 6 & 3 \\ 3 & 1 & 8 \end{bmatrix}, \text{ logo } A^T = \begin{bmatrix} 2 & 1 & 3 \\ 4 & 6 & 1 \\ 3 & 3 & 8 \end{bmatrix}$$

$$B = \begin{bmatrix} 1 & -4 & 0 & 2 \\ 3 & -9 & 1 & 3 \end{bmatrix}, \text{ logo } B^T = \begin{bmatrix} 1 & 3 \\ -4 & -9 \\ 0 & 1 \\ 2 & 3 \end{bmatrix}$$

Propriedades da Matriz Transposta.

Se os tamanhos das matrizes são tais que as operações indicadas podem ser efetuadas, então (RORRES ANTON (2001)) :

- $((A^T)^T = A$
- $(A + B)^T = A^T + B^T$ e $(A - B)^T = A^T - B^T$
- $(kA)^T$ onde k é um escalar qualquer
- $(AB)^T = B^T A^T$

Definição 1.4. Uma matriz ortogonal é uma matriz quadrada A com a propriedade $A^{-1} = A^T$. Decorre desta definição que uma matriz quadrada A é ortogonal se, e somente se, $AA^T = A^T A = I$.

Definição 1.5. Seja $T : V \rightarrow V$ um operador linear. Um vetor $v \in V$, $v \neq 0$, é dito autovetor, vetor próprio ou vetor característico do operador T , se existir $\lambda \in \mathbb{R}$ tal que $T(v) = \lambda v$. O escalar λ é denominado autovalor, valor próprio ou valor característico do operador linear T associado ao autovetor v .

Exemplo 1.4.

$$1) T : \mathbb{R}^2 \longrightarrow \mathbb{R}^2 \quad (x,y) \longrightarrow (3x, 8x - y).$$

(1,2) é autovetor de T associado ao autovalor $\lambda = 3$, pois $T(1,2) = (3,6) = 3(1,2)$.

Definição 1.6. Uma matriz quadrada A é chamada simétrica se $A = A^T$.

Toda matriz simétrica pode ser fatorada como $A = PDP^T$. Onde P é a matriz ortogonal e D uma matriz diagonal (formada pelos autovalores de A)(POOLE DAVID (2003)).

Teorema 1.1. Se A é ortogonalmente diagonalizável, então A é simétrica.

Demonstração:

Se A é diagonal então é trivialmente diagonalizável por uma matriz ortogonal, ou seja, existem uma matriz ortogonal Q e uma matriz diagonal D tais que $Q^T A Q = D$. Como $Q^{-1} = Q^T$ temos que $Q^T Q = I = Q Q^T$, e assim, $Q D Q^T = Q Q^T A Q Q^T = I A I = A$. Portanto, $A^T = (Q D Q^T)^T = (Q^T)^T D^T Q^T = Q D Q^T = A$.

0.2 Produto de Matrizes

A definição de produto usual de matrizes presta-se à descrição de sistemas lineares, como uma generalização de uma equação escalar do tipo

$$ax = b \tag{1.1.}$$

em que a, b são escalares, x é a solução escalar desejada. Ou seja, dado um sistema linear m x n

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

É desejável escrever o sistema em uma forma semelhante à equação (1.1.), isto é, como uma equação matricial

$$Ax = b \tag{1.2.}$$

onde $A_{(i,j)}$ é conhecida, x é uma matriz n x 1 de incógnitas e b é uma matriz m x 1 que representa os elementos à direita dos sinais de igualdade. Definimos, então,

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}, x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

e

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

O i -ésimo elemento de Ax é determinado pela i -ésima linha de A . Os elementos daquela linha são multiplicados pelos elementos correspondentes em x e os n produtos são, então, somados. Ou seja, por esta "regra", o i -ésimo elemento de Ax é determinado por :

$$a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n$$

Pela familiaridade com produtos internos reconhecemos essa fórmula, simplesmente, como o produto interno da n -upla correspondente a i -ésima linha de A com a n -upla correspondente ao vetor x

$$\begin{bmatrix} a_{i1} & a_{i2} & \cdots & a_{in} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n$$

Assim, é possível definir um produto de uma matriz A por uma outra matriz B se o número de colunas de A é igual ao número de linhas de B : cada coluna B dá origem a AB que é o resultado do produto da matriz A pela matriz coluna B . Ou melhor, dadas as matrizes $A = (a_{ij})_{m \times n}$ e $B = (b_{ij})_{n \times r}$, o produto $AB = C$ é definido do seguinte modo: cada coluna de C é o produto de A pela coluna de B correspondente $C = AB$. Desse modo, cada entrada (c_{ij}) da matriz C é definida por :

$$c_{ij} = \sum_{p=1}^n a_{ip}b_{pj}$$

0.3 Matrizes blocadas (ou em bloco)

Matrizes em bloco podem ser multiplicadas de modo análogo, como se os blocos fossem entradas numéricas, desde que, para um produto AB , a partição das colunas de A combine com a partição das linhas de B . Veja o exemplo que se segue.

Exemplo 1.5. Sejam

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

blocadas da seguinte maneira:

• blocos de A :

$$A_{11} = \begin{bmatrix} 2 & -3 & 1 \\ 1 & 5 & -2 \end{bmatrix}, A_{12} = \begin{bmatrix} 0 & -4 \\ 3 & -1 \end{bmatrix}, A_{21} = [0 \quad -4 \quad -2] \text{ e } A_{22} = [7 \quad -1];$$

• blocos de B : $B_1 = \begin{bmatrix} 6 & 4 \\ -2 & 1 \\ -3 & 7 \end{bmatrix}$ e $B_2 = \begin{bmatrix} -1 & 3 \\ 5 & 2 \end{bmatrix}$

O produto AB pode ser escrito como

$$AB = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} [B_1 \quad B_2] = \begin{bmatrix} A_{11}B_1 + A_{12}B_2 \\ A_{21}B_1 + A_{22}B_2 \end{bmatrix}$$

Agora,

$$A_{11}B_1 = \begin{bmatrix} 15 & 12 \\ 2 & -5 \end{bmatrix}, \quad A_{12}B_2 = \begin{bmatrix} -20 & -8 \\ -8 & 7 \end{bmatrix}$$

$$A_{21}B_1 = [14 \quad -18], \quad A_{22}B_2 = [-12 \quad 19]$$

Logo, o bloco superior de AB é $\begin{bmatrix} -5 & 4 \\ -6 & 2 \end{bmatrix}$ e o seu bloco inferior é $[2 \quad 1]$

$$\text{Portanto } AB = \begin{bmatrix} -5 & 4 \\ -6 & 2 \\ 2 & 1 \end{bmatrix}$$

O produto de matrizes em blocos é a forma mais genérica de considerar o produto de duas matrizes. Este produto terá importância na teoria relativa à Decomposição em Valores Singulares reduzida.

Capítulo 2 - Normas

0.4 Norma de um vetor

Definição 2.1. Uma norma no espaço vetorial \mathbb{R}^n é uma função $\|\cdot\| : \mathbb{R}^n \rightarrow \mathbb{R}$ que satisfaz as seguintes propriedades ou axiomas :

- $\|v\| \geq 0$, para todo $v \in \mathbb{R}^n$;
- $\|v\| = 0 \iff v = 0$; (se não for válida esta propriedade, diz-se então que é uma semi-norma);
- $\|\alpha v\| = |\alpha| \|v\|$, para todo $\alpha \in \mathbb{R}$, e para todo $v \in \mathbb{R}^n$
- $\|v + u\| \leq \|v\| + \|u\|$, (desigualdade triangular)

Dessa desigualdade, segue que $\|v - u\| \geq \|v\| - \|u\|$.

Exemplos de normas no \mathbb{R}^n são as seguintes:

- $\|v\|_2 = \sqrt{\langle v, v \rangle} = \sqrt{v^T v} = \sqrt{\sum_{i=1}^n v_i^2}$; (norma euclidiana)
- $\|v\|_\infty = \max_{i=1, \dots, n} |v_i|$; (norma do máximo)
- $\|v\|_1 = \sum_{i=1}^n |v_i|$

Exemplo 1.6. Se $v = [1 \ -1 \ 1]$, $\|v\|_2 = \sqrt{6}$, $\|v\|_\infty = 2$ e $\|v\|_1 = 4$

0.5 Norma de matrizes

Definição 2.2. Seja $M = (m_{ij})$ o espaço vetorial das matrizes $m \times n$ reais ou complexas. Uma norma $\|\cdot\|$ é uma função que associa a cada matriz um número real não negativo e satisfaz as propriedades:

- $\|A\| = 0 \iff A = 0$ (matriz cujas entradas são nulas)
- $\|\lambda A\| = |\lambda| \|A\| \forall \lambda \in \mathbb{R}$
- $\|A + B\| \leq \|A\| + \|B\|$

Particularmente, iremos trabalhar com a Norma-2 ou Norma operacional euclidiana que será definida a seguir.

Definição 2.3. Quando uma matriz $A \in M$ é vista como um operador entre os espaços euclidianos \mathbb{R}^n e \mathbb{R}^m , a norma natural é dada pela norma operacional:

$$\|A\|_2 = \sup \frac{\|Ax\|}{\|x\|} \text{ para } x \in \mathbb{R}^n, x \neq 0.$$

Definição 2.4. Se uma norma vetorial de \mathbb{R}^n é dada, então define-se a correspondente norma matricial induzida como segue:

$$\|A\| = \max\{\|Ax\|; x \in \mathbb{R}^n \text{ com } \|x\| = 1\}$$

A norma-2 de uma matriz também pode ser definida como sendo o maior valor singular de uma matriz A , ou seja, $\|A\|_2 = \sqrt{\rho(A^T A)}$ (POOLE DAVID (2003)).

Observação 1.1. Acima $\rho(A^T A)$ denota o raio espectral da matriz A cujo é $\rho(A) = \max_{1 \leq i \leq r} |\lambda_i(A)|$ que são os valores próprios da matriz A .

Teorema 1.2. Se $(Q_{ij})_{n \times n}$ é uma matriz ortogonal e J é uma matriz de mesma dimensão, então :

- (a) $Qx \cdot Qy = x \cdot y \forall x$ e y em \mathbb{R}^n
- (b) $\|Qx\|_2 = \|x\|_2 \forall x$ e y em \mathbb{R}^n
- (c) $\|JQ\|_2 = \|QJ\|_2 = \|J\|_2$

Demonstração:

(a) Para isso usaremos o fato de que, se x e y são vetores (coluna) em \mathbb{R}^n então o produto interno $x \cdot y = x^T y$.

$$x^T y = [x_1 \quad x_2 \quad \cdots \quad x_n] \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \sum_{i=1}^n x_i y_i$$

Se Q é ortogonal então $Q^T Q = I$ e temos:

$$Qx \cdot Qy = (Qx)^T Qy = x^T Q^T Qy = x^T Iy = x^T y = x \cdot y \forall x, y \text{ em } \mathbb{R}^n$$

(b) Se Q é ortogonal então $Q^T Q = I$ e temos:

$$Qx \cdot Qy = x \cdot y \text{ para todo } x, y \in \mathbb{R}^n, \text{ então tomando } x = y, \text{ temos :}$$

$$Qx \cdot Qy = (Qx)^T Qx = x^T Q^T Qx = x^T x = x \cdot x, \text{ logo } \|Qx\|_2 = \|x\|_2.$$

(c) Vamos demonstrar que $\|JQ\| = \|J\|_2$.

Sabemos que a norma de uma matriz é definida por :

$$\|A\|_2 = \|A\| = \sup \frac{\|Ax\|}{\|x\|} \text{ para } x \in \mathbb{R}^n, x \neq 0. \quad (1.3.)$$

Assim temos por (1.3.):

$$\| JQ \| = \max_{\|x\|=1} \| JQ \|_2.$$

Como Q é ortogonal, sabemos que $JQx = J(Qx)$ e $\| Qx \|_2 = \| x \|_2$.

$$\text{Logo, } \| JQ \|_2 = \max_{\|x\|_2=1} \| JQ \|_2$$

$$\| JQ \|_2 = \max_{\|Qx\|_2=1} \| J(Qx) \|_2$$

$$\| JQ \|_2 = \max_{\|y\|_2=1} \| J(y) \|_2 = \| J \|_2$$

$$\text{Então, } \| JQ \|_2 = \| J \|_2$$

2. Pode-se demonstrar que a norma-2 de uma matriz também pode ser definida como $\| A \|_2 = \sqrt{\rho(A^T A)}$ (POOLE DAVID (2003)). Sendo Q ortogonal temos que $Q^T Q = I$ logo temos:

$$\| QJ \|_2 = \sqrt{|(QJ)^T QJ|}$$

$$\| QJ \|_2 = \sqrt{\rho(J^T Q^T QJ)}$$

$$\| QJ \|_2 = \sqrt{\rho(J^T J)}$$

$$\| QJ \|_2 = \| J \|_2$$

$$\text{Então. } \| QJ \|_2 = \| J \|_2$$

0.6 Processo de Gram-Schmidt

A partir de uma base para um espaço vetorial qualquer (com produto interno \langle , \rangle), o processo de Gram-Schmidt é usado para gerar uma base ortogonal desse espaço (em relação a esse produto interno). Inicialmente descrevemos esses processo para uma base $\beta = \{v_1, v_2\}$.

Seja $u_1 = v_1$. Teremos de encontrar a partir de v_2 um novo vetor u_2 ortogonal a u_1 , isto é $\langle u_1, u_2 \rangle = 0$. Para tal tomamos $u_2 = v_2 - \alpha_{12}u_1$, onde α_{12} é um número escolhido de modo que $\langle u_1, u_2 \rangle = 0$, ou melhor, $\langle u_1, v_2 - \alpha_{12}u_1 \rangle = 0$ o que resulta em $\alpha_{12} = \frac{\langle v_2, u_1 \rangle}{\langle u_1, u_1 \rangle}$. Ficamos então com :

$$u_1 = v_1$$

$$u_2 = v_2 - \alpha_{12}u_1 = v_2 - \frac{\langle u_1, v_2 \rangle}{\langle u_1, u_1 \rangle} u_1$$

Observemos que u_2 foi obtido de v_2 , subtraindo-se deste a projeção ortogonal de v_2 na direção de u_1 , $u_2 = v_2 - \alpha_{12}u_1 = v_2 - \frac{\langle u_1, v_2 \rangle}{\langle u_1, u_1 \rangle} u_1$

Como definido para uma base β com dois vetores, podemos estender o processo para uma base com um número arbitrário de vetores. Vamos, agora, transpor esse processo para os casos em que as bases são formada por vetore do \mathbb{R}^n e o produto interno é o usual.

Sejam $\{v_1, \dots, v_n\}$ vetores colunas linearmente independentes de uma matriz $A =$

$a_{(m,n)}$. Então,

$$u_1 = v_1$$

$$u_2 = v_2 - \alpha_{12}u_1$$

$$u_3 = v_3 - \alpha_{13}u_1 - \alpha_{23}u_2$$

\vdots

$$u_n = v_n - \alpha_{1n}u_1 - \alpha_{2n}u_2 - \cdots - \alpha_{(n-1)n}u_{(n-1)}$$

Genericamente, definindo $\alpha_{ij} = \frac{u_i^T v_j}{u_i^T u_i}$, obtemos:

$$u_i = v_j - \sum_{i=1}^{(j-1)} \alpha_{ij}u_i$$

Processo de Gram-Schmidt modificado

O processo de Gram-Schmidt não produz, em geral, resultados precisos quando trabalhamos com aritmética de precisão finita. Na prática, os vetores resultantes não são perpendiculares devido a erros de aproximação no cálculo de $u_1, u_2, u_3, \dots, u_n$. Podemos obter melhor precisão numérica usando uma versão modificada do método de Gram-Schmidt, em que:

- Na primeira etapa, chamando $u_1 = v_1$, subtraímos de cada vetor v_i sua projeção ortogonal sobre u_1 . Temos, então, que:

$$v_2^{(1)} = v_2 - \alpha_{12}u_1$$

$$v_3^{(1)} = v_3 - \alpha_{13}u_1$$

$$v_4^{(1)} = v_4 - \alpha_{14}u_1$$

\vdots

$$v_n^{(1)} = v_n - \alpha_{1n}u_1$$

em que $\alpha_{1j} = \langle u_1, v_j \rangle$. Definimos $u_2 = v_2^{(1)}$

- Na segunda etapa, modificamos os $(n - 2)$ vetores restantes de modo a torná-los ortogonais a u_2 (observe que pela primeira etapa, eles já são ortogonais a u_1).

$$v_3^{(2)} = v_3^{(1)} - \alpha_{23}^{(1)}u_2$$

$$v_4^{(2)} = v_4^{(1)} - \alpha_{24}^{(1)}u_2$$

$$v_5^{(2)} = v_5^{(1)} - \alpha_{25}^{(1)}u_2$$

\vdots

$$v_n^{(2)} = v_n^{(1)} - \alpha_{2n}^{(1)}u_2$$

em que $\alpha_{2j} = \langle u_2, v_j \rangle$, $j \geq 3$. Definimos $u_3 = v_3^{(2)}$. Analogamente, determinamos u_4, u_5, \dots, u_n

O processo de Gram-Schmidt modificado pode, ser descrito por meio de matrizes de projeção:

$$v_1^{(0)} = u_1$$

$$v_2^{(1)} = v_2^{(0)} - \frac{u_1^T v_2}{u_1^T u_1} u_1 = v_2^{(0)} - \frac{(v_1^{(0)})^T v_2^{(0)}}{((v_1^{(0)})^T v_1^{(0)})} v_1^{(0)} = v_2^{(0)} - P_1 v_2^{(0)} = (I - P_1) v_2^{(0)}$$

onde $P = v_1^{(0)}(v_1^{(0)})^T$.

Desta forma, para um k qualquer, temos:

$$v_k^{(0)} = v_k$$

$$v_k^{(1)} = (I - P_1) v_k^{(0)}$$

$$v_k^{(2)} = (I - P_2) v_k^{(1)} = (I - P_2)(I - P_1) v_k^{(0)}$$

⋮

$$v_k^{(n)} = (I - P_n) v_k^{(n-1)} = (I - P_n)(I - P_{(n-1)}) \cdots (I - P_2)(I - P_1) v_k^{(0)}$$

E u_k será tomado da mesma forma que anteriormente, ou seja, $u_k = v_k^{(k-1)}$

Capítulo 3 - Decomposição em Valores Singulares

0.7 Decomposição em Valores Singulares

Para qualquer matriz $(A_{ij})_{m \times n}$, a matriz $(A^T A_{ij})_{n \times n}$ é simétrica e, portanto pode ser diagonalizada ortogonalmente (POOLE DAVID (2003)). Os autovalores de $A^T A$ são todos reais e não negativos. Para mostrar isso, suponha que v_i seja o autovetor de $A^T A$ correspondente ao autovalor λ_i , então se $1 \leq i \leq n$, temos:

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

Portanto, faz sentido extrair a raiz quadrada desses autovalores.

Definição 3.1. (Valores Singulares.) Se A é uma matriz $m \times n$, os valores singulares de A são as raízes quadradas dos autovalores de $A^T A$, e são denotados por $\sigma_1, \sigma_2, \sigma_3, \dots, \sigma_n$. É uma convenção ordenar os valores singulares de modo que $\sigma_1 \geq \sigma_2 \geq \sigma_3 \geq \dots \geq \sigma_n$.

Significado dos Valores Singulares de uma matriz.

Considere os autovetores de $A^T A$. Como $A^T A$ é simétrica, sabemos que existe uma base ortonormal para \mathbb{R}^n formada por autovetores de $A^T A$. Seja $\{v_1, v_2, \dots, v_n\}$ a base correspondente aos autovalores de $A^T A$, ordenados de modo que $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \dots \geq \lambda_n$.

$$\lambda_i = \|Av_i\|^2$$

$$\text{Portanto, } \sigma_i = \sqrt{\lambda_i} = \|Av_i\|$$

Em outras palavras, os valores singulares de A são os comprimentos dos vetores Av_1, Av_2, \dots, Av_n .

Teorema 1.3. Suponha que $v_1, v_2, v_3, \dots, v_n$ é uma base ortonormal consistindo em autovetores de $A^T A$ ordenados de tal forma que os autovalores associados satisfazam $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$. Suponha que A tem r valores singulares não nulos. Então $\{Av_1, Av_2, Av_3, \dots, Av_r\}$ é uma base ortogonal para o espaço gerado pelas colunas de A ($\text{Col}A$) e $\text{posto}(A) = r$.

Demonstração:

Como v_i e $\lambda_j v_j$ são ortogonais para $i \neq j$, temos: $\langle Av_i, Av_j \rangle = \langle v_i, A^T Av_j \rangle = \langle v_i, \lambda_j v_j \rangle =$

$$\lambda_j \langle v_i, v_j \rangle = 0$$

Logo, $Av_1, Av_2, Av_3, \dots, Av_n$ é um conjunto ortogonal. Além disso, como os comprimentos dos vetores $Av_1, Av_2, Av_3, \dots, Av_n$ são os valores singulares de A e como existem r valores singulares não-nulos, $Av_i \neq 0$ se, e somente se $1 \leq i \leq r$. Portanto são vetores linearmente independentes e pertencem ao espaço gerado pelas colunas de A. Finalmente qualquer que seja y em ColA

$$\begin{aligned} y &= Ax \\ y &= Ac_1v_1 + Ac_2v_2 + \dots + Ac_rv_r + Ac_{r+1}v_{r+1} + \dots + Ac_nv_n \\ y &= c_1Av_1 + c_2Av_2 + \dots + c_rAv_r + c_{r+1}Av_{r+1} + \dots + c_nAv_n \\ y &= c_1Av_1 + c_2Av_2 + \dots + c_rAv_r + 0 + \dots + 0 \end{aligned}$$

Enfim, y pertence a $\{Av_1, Av_2, Av_3, \dots, Av_r\}$, o que mostra que é uma base (ortogonal) para o espaço gerado pelas colunas de A e que $\text{Posto}(A) = r$.

Teorema 1.4.(Decomposição em Valores Singulares): *Seja $(A_{ij})_{m \times n}$ com valores singulares $\sigma_1 > \sigma_2 > \dots > \sigma_r > 0$ e $\sigma_{r+1} = \sigma_{r+2} = \dots = \sigma_n = 0$. Então existem duas matrizes ortogonais, $(U_{ij})_{m \times m}$ e $(V_{ij})_{n \times n}$, e uma matriz $(S_{ij})_{m \times n}$ tais que $A = USV^T$.*

$$D = \begin{bmatrix} \sigma_1 & 0 & 0 & 0 \\ 0 & \sigma_2 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \sigma_r \end{bmatrix} \quad S = \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix}$$

Demonstração:

Para construir a matriz ortogonal V, vamos primeiro achar uma base ortonormal $v_1, v_2, v_3, \dots, v_n$ de \mathbb{R}^n formada por autovetores da matriz simétrica. Então, $A^T A \in M(n \times n)$. Então, $V = [v_1 \ v_2 \ v_3 \ \dots \ v_n]$ é uma matriz ortogonal $n \times n$.

Para a matriz ortogonal U, primeiro notamos que $Av_1, Av_2, Av_3, \dots, Av_r$ é um conjunto ortogonal de vetores de \mathbb{R}^m . Para ver isso, suponha que v_i seja o autovetor de $A^T A$ correspondente ao autovalor λ_i . Então, para $i \neq j$, temos: $\langle Av_i, Av_j \rangle = \langle v_i, A^T Av_j \rangle = \langle v_i, \lambda_j v_j \rangle = \lambda_j \langle v_i, v_j \rangle = 0$

Visto que os autovetores v_i são ortogonais. Agora lembre-se de que o valor singular satisfaz $\sigma_i = \|Av_i\|$, e que os primeiros r desses são não nulos. Portanto, podemos normalizar $Av_1, Av_2, Av_3, \dots, Av_r$ considerando $u_i = 1/\sigma_i \times Av_i$ para $1 \leq i \leq r$.

Isto garante que $u_1, u_2, u_3, \dots, u_r$ seja um conjunto ortonormal de \mathbb{R}^m mas, se $r < m$, ele não será uma base para \mathbb{R}^m . Neste caso, estendemos o conjunto (Processor de Gram-Schmidt) $u_1, u_2, u_3, \dots, u_r$ para uma base ortonormal $u_1, u_2, u_3, \dots, u_m$ de \mathbb{R}^m . Então, consideremos $U = [u_1 \ u_2 \ u_3 \ \dots \ u_m]$. Possuímos as seguintes informações:

$$\begin{aligned} V^T &= V^{-1} \\ Av_i &> 0, & \forall 1 \leq i \leq r. \\ Av_i &= \sigma_i u_i, & \forall 1 \leq i \leq r. \\ \|Av_i\| &= \sigma_i = 0, & \forall r + 1 \leq i \leq n. \end{aligned}$$

$$\text{Assim, } Av = A[v_1 v_2 \dots v_n]$$

$$Av = [Av_1 Av_2 \dots Av_n]$$

$$Av = [\sigma_1 u_1 \sigma_2 u_2 \dots \sigma_r u_r]$$

$$Av = [u_1 u_2 \dots u_m] \begin{bmatrix} \sigma_1 & 0 & 0 & 0 \\ 0 & \sigma_2 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \sigma_n \end{bmatrix}$$

$$Av = US$$

$$A = USV^T$$

Onde a SVD de A também pode ser escrita como o somatório de r matrizes de posto 1 (SVD reduzida):

$$Av = [Av_1 Av_2 \dots Av_r Av_{r+1} \dots Av_n]$$

$$Av = [Av_1 Av_2 \dots Av_r 0 \dots 0]$$

$$Av = [\sigma_1 u_1 \sigma_2 u_2 \dots \sigma_r u_r 0 \dots 0]$$

$$Av = [u_1 u_2 \dots u_r] \begin{bmatrix} \sigma_1 & 0 & 0 & 0 \\ 0 & \sigma_2 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \sigma_r \end{bmatrix}$$

$$Av = US$$

$$A = USV^T$$

$$A = \sum_{i=1}^r \sigma_i u_i v_i^T$$

Capítulo 4 - Aplicação da Decomposição em Valores Singulares na Análise de Componentes Principais

0.8 Análise de Componentes Principais

A Análise de Componentes Principais é um método que tem por finalidade básica, a análise dos dados usados visando sua redução, eliminação de sobreposições e a escolha das formas mais representativas de dados a partir de combinações lineares das variáveis originais.

Considere uma matriz de dados A:

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

É frequente pretender descrever um conjunto de dados constituído por m indivíduos caracterizados por n variáveis do tipo medidas quantitativas.

Definição 4.1. *Sejam x , y e z vetores de \mathbb{R}^m . Uma distância em \mathbb{R}^m é uma aplicação $d : \mathbb{R}^m \times \mathbb{R}^m \rightarrow \mathbb{R}$ satisfazendo as seguintes condições:*

- i. $d(x, y) \geq 0$, $d(x, y) = 0 \iff x = y$
- ii. $d(x, y) = d(y, x)$
- iii. $d(x, y) \leq d(x, z) + d(z, y)$ (des. triangular)

Para estabelecermos distâncias entre os indivíduos da matriz A, não é razoável trabalharmos com a distância usual, pois as variáveis podem apresentar unidades de medida diferentes ou até unidades de medidas não comparáveis.

Necessitamos então de uma distância que elimine este efeito das unidades diferentes. Uma solução é dividirmos a coluna j da matriz X por $\sqrt{S_{jj}}$.

Note que, por exemplo, a distância entre a primeira e segunda coluna é dada por :

$$(x_{11} - x_{12})^2/S_{11} + \dots + (x_{m1} - x_{m2})^2/S_{mm} \quad (1.4.)$$

Assim se tomarmos a matriz :

$$D_s = \begin{bmatrix} 1/S_{11} & \dots & 0 \\ \vdots & \vdots & \vdots \\ 0 & 0 & 1/S_{nn} \end{bmatrix}$$

Assim, (1.4.) pode ser escrito como sendo : $(x_1 - x_2)^T D_s (x_1 - x_2)$

0.8.1 Encontrando Subespaços Ótimos

Dada as observações dos m indivíduos (m vetores em \mathbb{R}^n), queremos encontrar um subespaço vetorial de \mathbb{R}^n com dimensão menor que n , que melhor represente nossos dados.

Isso é equivalente a resolver ao seguinte problema:

Seja x_j um indivíduo e $p(x_j)$ a projeção sobre o subespaço que estamos procurando. Temos que a distância entre x_j e $p(x_j)$ é dada por :

$$d_j^2 = (x_j - p(x_j))^T D_s (x_j - p(x_j))$$

Tentaremos então minimizar :

$$\sum_j w_j d_j^2 \quad (1.5.)$$

Ou seja, encontrar um subespaço vetorial de \mathbb{R}^n que minimiza (1.5.). Note que estamos atribuindo "pesos" para as j observações, e conseqüentemente para as j distâncias.

Definição 4.2. *Um subespaço p -dimensional em \mathbb{R}^n ($k \leq n$) é um subconjunto de vetores da forma*

$$u + \lambda_1 v_1 + \dots + \lambda_k v_p$$

onde $u \in \mathbb{R}^n$ é um vetor fixo,

Assim :

$$\sum_j w_j d_j^2 =$$

$$\sum_j w_j (x_j - b_j)^T D_s (x_j - b_j) =$$

$$\sum_j w_j (a_j - y_j)^T D_s (a_j - y_j) , \text{ onde } a_j = x_j - \tilde{x} \text{ e } y_j = f_{j1}u_1 + \dots + f_{jp^*}u_{p^*}.$$

Seja $A = [a_1 \dots a_n]$ e $Y = [y_1 \dots y_n]$. Estamos então tentando encontrar uma matriz Y que minimize :

$$\|A - Y\|^2 = \sum_j w_j (a_j - y_j)^T D_s (a_j - y_j) \quad (1.6.)$$

entre todas as matrizes Y de posto p^*

Teorema 1.5. Dada uma matriz A de posto r , existe uma matriz A_p de posto p onde

$$\|A - A_p\|_2^2 = \sigma_{p+1}^2$$

Seja $A = \sum_{i=1}^r \sigma_i u_i v_i^T$ a SVD de A . Então $A_p = \sum_{i=1}^p \sigma_i u_i v_i^T$ e $A - A_p = \sum_{i=p+1}^r \sigma_i u_i v_i^T$. Seja v o vetor singular $A - A_p$. Expressamos v como combinação linear de $v_1, v_2, v_3, \dots, v_r$. Assim, escrevemos $v = \sum_{i=1}^r \alpha_i v_i$. Logo temos:

$$\begin{aligned} |(A - A_p)v| &= \left| \sum_{i=p+1}^r \sigma_i u_i v_i^T \sum_{i=1}^r \alpha_i v_i \right| \\ &= \left| \sum_{i=p+1}^r \alpha_i \sigma_i u_i v_i^T v_i \right| \\ &= \left| \sum_{i=p+1}^r \alpha_i \sigma_i u_i \right| \\ &= \sqrt{\sum_{i=p+1}^r \alpha_i^2 \sigma_i^2} \end{aligned}$$

Onde v atinge seu máximo quando esta sujeito a restrição de que $|v|^2 = \sum_{i=1}^r \alpha_i^2 = 1$, que ocorre quando $\alpha_{p+1} = 1$ e o restante dos α_i são iguais a 0. Assim,

$$\|A - A_p\|_2^2 = \sigma_{p+1}^2$$

Finalmente, provamos que $Y = A_p$ é a melhor aproximação para a matriz A pela norma-2.

Teorema 1.6. Seja A uma matriz $m \times n$. Para todas matrizes B de posto no máximo p é possível provar que

$$\|A - A_p\|_2 \leq \|A - B\|_2$$

Se A é uma matriz de posto menor ou igual a p , o teorema é obviamente verdadeiro pois $\|A - A_p\|_2 = 0$.

Vamos supor que posto de A é maior que p . Sabemos que

$$\|A - A_p\|_2^2 = \sigma_{p+1}^2$$

Suponha agora que existe alguma matriz B de posto no máximo p , de tal modo que B é uma melhor aproximação para A pela norma-2 do que A_p .

Isto é, $\|A - B\|_2 < \sigma_{p+1}$. Sabemos que espaço nulo de B , $Null(B)$ (conjunto de vetores v tais que $Bv = 0$), tem dimensão de pelo menos $n-p$.

Seja v_1, v_2, \dots, v_{p+1} , onde $p+1$ é o primeiro dos vetores singulares de A . Por um argumento de dimensão, segue-se que existe um $z = 0$ em $Null(B) \cap Span\{v_1, v_2, \dots, v_{p+1}\}$. De forma que $|z| = 1$. Agora nós vamos mostrar que o vetor z , que se encontra no espaço do primeiro vetor singular (v_{p+1}) de A , faz com que $(A - B)z \geq \sigma_{p+1}$. Assim a norma-2 de $A - B$ é pelo menos σ_{p+1} contradizendo o que assumimos, que $\|A - B\|_2 < \sigma_{p+1}$.
Primeiro

$$\|A - B\|_2^2 \geq |(A - B)z|^2$$

Com $Bz = 0$

$$\|A - B\|_2^2 \geq |Az|^2$$

Como z esta em $\text{Span}\{v_1, v_2, \dots, v_{p+1}\}$

$$|Az|^2 = \left| \sum_{i=1}^n \sigma_i u_i v_i^T z \right|^2 = \sum_{i=1}^n \sigma_i^2 (v_i^T z)^2 = \sum_{i=1}^{p+1} \sigma_i^2 (v_i^T z)^2 \geq \sigma_{p+1}^2 = \sum_{i=1}^{p+1} (v_i^T z)^2 = \sigma_{p+1}^2$$

Segue-se que $\|A - B\|_2^2 \geq \sigma_{p+1}^2$ contradizendo a suposição que $\|A - B\|_2 < \sigma_{p+1}$. Isso prova o teorema.

Por (Teorema 1.5. e 1.6.) provamos que o minimo é atingido quando:

$Y = A_{[p^*]} = \sum_{p=1}^{p^*} \lambda_p u_p v_p^T$, assim $A_{[p^*]}$ é a matriz de ordem $m \times n$ com posto p^* que melhor aproxima a matriz A no sentido de (1.6.).

Da SVD da matriz A , temos que $A = U D_\lambda v^T$, ou seja $A = U F^T$, onde $F = V D_\lambda$. Seja:

$$F^T = \begin{bmatrix} F_{11} & \dots & F_{n1} \\ \vdots & \vdots & \vdots \\ F_{1p} & \dots & F_{np} \end{bmatrix}$$

Da relação $[a_1 \ a_2 \ \dots \ a_n] = [u_1 \ u_2 \ \dots \ u_p] F^T$ vem que :

$$a_1 = f_{11}u_1 + \dots + f_{1p}u_p$$

$$a_n = f_{n1}u_1 + \dots + f_{np}u_p$$

De modo que os números f_{ii} fornecem as coordenadas do vetor a_j no espaço gerado pelas colunas de U . Temos que $u_1 \ u_2 \ \dots \ u_{p^*}$ formam uma base para o subespaço ótimo.

Desta forma, os f_{ii} , $1 \leq i \leq p^*$ são as coordenadas da projeção de a_j no subespaço ótimo. Agora observe que a projeção de a_1 sobre o espaço gerado por u_1, \dots, u_p é dada por :

$$p(a_1) = (a_{11})u_1 + \dots + (a_{1p^*})u_{p^*}. \text{Então temos que :}$$

$$f_{ji} = a_j \bullet u_i. \tag{1.7.}$$

Como $a_j \bullet m_i = a_j^T D_\lambda u_i$. Temos:

$$f(i) = \begin{bmatrix} f_{1i} \\ f_{ni} \end{bmatrix}$$

é denominado a i -ésima componente principal. E, assim a relação (1.7.) é uma fórmula para obtermos o vetor $f(i)$.

Compressão de Imagens utilizando a SVD

A transmissão e armazenamento de dados têm um papel de extrema importância na tecnologia atual e é desejável que sua execução seja realizada com o mínimo espaço de armazenamento e o menor tempo de processamento. Muitas vezes, as informações processadas são muito extensas, ocupando muito espaço de armazenamento e levando muito tempo durante o processo de transmissão. Para resolver esse problema, utilizamos a Decomposição em Valores Singulares (conhecida como SVD, acrônimo de Singular Value Decomposition) no processo de compressão de imagens com intuito que preservasse uma quantidade mínima de dados para recuperá-las.

Uma das aplicações da SVD é a compressão de imagens digitais. Uma imagem digital é uma matriz $m \times n$ em que as entradas são traduções numéricas da luminosidade e da cor dos "patches" (a tela fica dividida em $m \times n$ patches, como uma colcha de retalhos, que é a tradução literal de patches). No caso de imagem colorida podemos ter uma matriz $m \times n \times 3$, em que a primeira camada é o tom de vermelho, a segunda o tom de verde e, a terceira o tom de azul.

Se $A = USV^T$ é a SVD de A , $A = \sum_{i=1}^r \sigma_i u_i v_i^T$, em que $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$ são os valores singulares de A e, para todo i , u_i, v_i denota a i -ésima coluna de U e V , respectivamente, queremos encontrar uma aproximação para A que melhor represente a imagem, no caso, acuidade visual e relativa perfeição da imagem.

Então com a SVD podemos expressar os dados eficientemente com uma quantidade mínima de dados em vez de uma listagem de todas as entradas da matriz. Como um exemplo, suponhamos que um scanner é usado para digitalizar uma fotografica substituindo a imagem por uma matriz $m \times n$ pixels, na qual cada nível de cinza é representado numa escala de 0 a 1. Em qualquer escala, as características de uma imagem são traduzida redundantemente nas linhas ou colunas da matriz. Nós podemos rever estas características de forma menos redundante, através da aproximação da matriz original por uma matriz de posto menor do que o mínimo entre o m e n .

Temos do (Teorema 1.5.) que $Y = A_p = \sum_{i=1}^p \sigma_i u_i v_i^T$ é a melhor aproximação de A com posto p , no sentido de minimizar $\|A - Y\| = \|A - A_p\| = \sigma_{p+1}$. Lembrando somente que com $mp + np = (m + n)p$ elementos, podemos armazenar $\{u_1, \dots, u_p\}$ e $\sigma_1 v_1, \dots, \sigma_n v_n$, pois estamos guardando as matrizes $U_{m \times p}$ e $(SV_{p \times n}^T)$. Normalmente utilizaríamos mn elementos para armazenar A , que é um número que pode se tornar muito grande comparado a $(m + n)p$, se p é pequeno em relação a m ou a n .

Consequentemente, o erro relativo da aproximação de A por A_p é medido por $\frac{\sigma_{p+1}}{\sigma_p}$ e razão de compressão é medida por $\frac{(m+n)p}{mn}$.

Consideremos alguns aspectos de suma importância ao entendimento dos exemplos:

- Na obtenção da SVD utilizamos o FreeMat, que também foi utilizado na criação e compressão das imagens, vista a sua polivalência.

- O FreeMat apresenta diversos mapas de cores predefinidos, entre eles as opções Mapa default (Figura 1) e Mapa Gray (Figura 2), e outros podem ser criados. A matriz que armazena um mapa de cores é uma matriz $C_{m \times 3}$, em que C terá um número qualquer de linhas (até 64) e necessariamente três colunas. Cada linha da matriz é interpretada como uma cor, em que o primeiro elemento determina a intensidade de vermelho, o segundo de verde e o terceiro de azul.

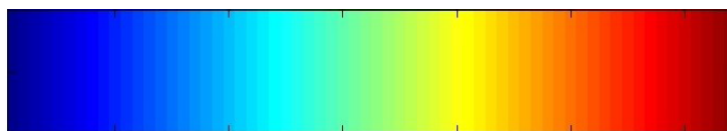


Figura 1: *Mapa Default. Fonte: Lorigiola Walisson (2004)*



Figura 2: *Mapa Gray. Fonte: Lorigiola Walisson (2004)*

- As imagens neste trabalho foram geradas com o comando *image*.
- Quando A é uma matriz $m \times n \times 3$ os elementos em $a(:, :, 1)$ são interpretados como as intensidades de vermelho, em $a(:, :, 2)$ como as intensidades de verde e em $a(:, :, 3)$ como as intensidades de azul, independente do mapa de cores utilizado.
- Quando importamos uma imagem ao FreeMat ele armazena em uma matriz (*nomedoarquivo*) $_{m \times n \times 3}$ no formato *uint8*; porém, para realizarmos cálculos devemos converter a matriz para o formato *double*; e, quando desejarmos visualizar a imagem devemos reconverter a matriz ao formato *uint8*; além disso, o formato *uint8* ocupa menos bytes do que *double*.
- Por fins computacionais, o FreeMat considera como sendo zero todos os valores singulares muito pequenos, e os anota com quatro casas decimais 0.0000; e, para aqueles valores singulares que são realmente zero, anota simplesmente 0.

Exemplos de imagens e suas respectivas compressões:

Considere a Figura 3. Esta imagem tem 499 x 425 pixels, o que corresponde a uma matriz $A_{499 \times 425}$. Tal imagem quando é importada do formato *jpg* para o ambiente FreeMat fica armazenada como uma matriz $A_{499 \times 425 \times 3}$, ou seja, como três matrizes $A_{499 \times 425}$, em que a primeira armazena as intensidades de vermelho, a segunda de verde

e a terceira de azul. As três matrizes que formam a foto $A1_{499 \times 425}$, $A2_{499 \times 425}$ e $A3_{499 \times 425}$ tem o mesmo posto $r = 425$. Podemos ver a imagem original em jpg(esquerda) e sua representação feita no FreeMat(direita).



Figura 3: *Landscapes*. Fonte: <http://wcriarte.com.br/>

Procederemos a aproximação de A por uma matriz de posto menor, consequentemente diminuiremos o número de bytes necessários para armazená-la. Para isso, faremos a compressão desta imagem e visualizaremos os resultados com as seguintes linhas de comando (Apêndice):

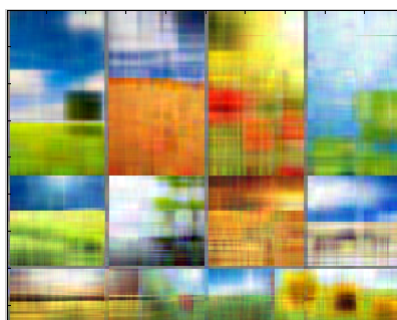


Figura 4: *Aproximação da matriz por uma matriz de posto $p = 10$. Comandos(Apêndice A)*



Figura 5: *Aproximação da matriz por uma matriz de posto $p = 50$. Comandos(Apêndice B)*



Figura 6: *Aproximação da matriz por uma matriz de posto $p = 150$. Comandos(Apêndice C)*



Figura 7: *Aproximação da matriz por uma matriz de posto $p = 200$. Comandos(Apêndice D)*

E finalmente, temos:

p	Erro relativo $\frac{\sigma_{p+1}}{\sigma_p}$	Razão de compressão
10	4,3569	924 x 10 /212075
50	0,2178	924 x 50 /212075
150	0,6535	924 x 150/212075
200	0,8714	924 x 200/212075

Figura 8: *Tabela de resultados*

Conclusão

Sobre a Análise de Componentes Principais, que se refere de modo geral a redução de dimensão, verificamos a importância da SVD na busca do subespaço que era essencial para encontrar as Componentes Principais, onde as definições e teoremas básicos eram imprescindíveis para compreensão da SVD e, conseqüentemente, para a teoria da ACP.

No que concerne à Compressão de Imagens, foi preciso relacionar o entendimento que tínhamos de Álgebra Linear com a Computação para geração de um algoritmo que encontrasse a matriz A_p de posto p mais próxima da minha matriz original A de posto r . Os experimentos realizados nos permitiram dizer que a compressão de imagens digitais via a SVD é eficiente, pois como verificamos o uso da SVD resulta em taxas satisfatórias de compressão.

A utilização de um software eficiente, como o FreeMat, contribuiu amplamente na geração, visualização e compressão das imagens.

Enfim, o projeto contribuiu, de um modo geral, para um melhor entendimento de como os conhecimentos em Álgebra Linear são importantes às ciências aplicadas, em especial, Análise de Dados Multivariados e Computação.

Referências

1. POOLE, David. Álgebra Linear com Aplicações, 3ª edição. Editora Saraiva - São Paulo, 2003.
2. LORIGIOLA, Walisson Pereira. Compressão de Imagens via DVS. Trabalho de conclusão de curso - Florianópolis-SC, 2004.
3. STRANG, G. Linear Algebra and its Applications. 4a ed. Belmont, California: Brooks/Cole Pub., 2005.
4. LIMA, E. L. Álgebra Linear. 7a ed. Rio de Janeiro: IMPA, 2008.
5. NOBLE B. and DANIEL, J.W. Applied Linear Algebra. 3a ed. Englewood Cliffs, NJ: Prentice-Hall Inc., 1988.
6. SANTOS, REGINALDO J. Matrizes, Vetores e Geometria Analítica. Belo Horizonte: Imprensa Universitaria da UFMG, 2010.
7. RORRES C. e HOWARD A., Algebra Linear com Aplicação, oitava edição. Editora Bookman, 2001.
8. LIPSCHUZ S., Algebra Linear Coleção Schaum. Makron Books.
9. FRANKLIN J. N., Matrix Theory, Dover publications, Inc., 1993.
10. TREFETHEN L.N. and BAU D. III, Numerical Linear Algebra. SIAM, Society for Industrial and Applied Mathematics, 1997.

Apêndice

Apêndice A

```
A=imread('teste.jpg');
[n,m,t]=size(A)
mtm=uint8(A)
image(mtm)
a=double(A)
x1=a(:,:,1)
x2=a(:,:,2)
x3=a(:,:,3)
[n,m,t]=size(x1)
[n,m,t]=size(x2)
[n,m,t]=size(x3)
[u1,s1,v1]=svd(x1)
[u2,s2,v2]=svd(x2)
[u3,s3,v3]=svd(x3)
p=10
a(:,:,1)=u1(:,1:p)*s1(1:p,1:p)*v1(:,1:p)';
a(:,:,2)=u2(:,1:p)*s2(1:p,1:p)*v2(:,1:p)';
a(:,:,3)=u3(:,1:p)*s3(1:p,1:p)*v3(:,1:p)';
mtm=(uint8(a));
image(mtm)
```

Figura 9: *Aproximação da matriz por uma matriz de posto $p = 10$*

Apêndice B

```
p=50
a(:,:,1)=u1(:,1:p)*s1(1:p,1:p)*v1(:,1:p)';
a(:,:,2)=u2(:,1:p)*s2(1:p,1:p)*v2(:,1:p)';
a(:,:,3)=u3(:,1:p)*s3(1:p,1:p)*v3(:,1:p)';
mtm=(uint8(a));
image(mtm)
```

Figura 10: *Aproximação da matriz por uma matriz de posto $p = 50$*

Apêndice C

```

p=150
a(:,:,1)=u1(:,1:p)*s1(1:p,1:p)*v1(:,1:p)';
a(:,:,2)=u2(:,1:p)*s2(1:p,1:p)*v2(:,1:p)';
a(:,:,3)=u3(:,1:p)*s3(1:p,1:p)*v3(:,1:p)';
mtm=(uint8(a));
image(mtm)

```

Figura 11: *Aproximação da matriz por uma matriz de posto $p = 150$*

Apêndice D

```

p=200
a(:,:,1)=u1(:,1:p)*s1(1:p,1:p)*v1(:,1:p)';
a(:,:,2)=u2(:,1:p)*s2(1:p,1:p)*v2(:,1:p)';
a(:,:,3)=u3(:,1:p)*s3(1:p,1:p)*v3(:,1:p)';
mtm=(uint8(a));
image(mtm)

```

Figura 12: *Aproximação da matriz por uma matriz de posto $p = 200$*