

IC - Álgebra Linear e Aplicações em Ciência de Dados (Apresentação 2)

Victor Moreli dos Santos

Universidade de São Paulo

victormoreli@usp.br

May 31, 2024

Overview

- 1 Autovetores e Autovalores
- 2 Diagonalização de uma Matriz
- 3 Matrizes Simétricas e Teorema Espectral
- 4 Matrizes Similares e Transformações de Similaridade
- 5 Matriz Definida Positiva
- 6 Quociente de Rayleigh
- 7 Singular Value Decomposition (SVD)
 - Construção
 - Teorema Fundamental da Álgebra Linear
 - Matrizes Particionadas
 - Interpretação Geométrica
 - Aplicações do SVD
 - Método dos mínimos quadrados
 - Matriz Inversa Generalizada
 - Compressão de dados

Autovetores e Autovalores

Os autovetores (x) e autovalores (λ) de uma matriz A quadrada ($n \times n$) vem da expressão:

$$Ax = \lambda x,$$

da qual concluí-se que: $(A - \lambda I)x = 0$. Ou seja,

$$\exists x \neq 0 \iff \det(A - \lambda I) = 0 \iff (A - \lambda I) \text{ é singular.}$$

O polinômio descrito por $\det(A - \lambda I) = 0$ é chamado de **polinômio característico de A** . Suas raízes são os autovalores de A e seu grau n revela que **A tem no máximo n autovalores diferentes**. Há a possibilidade de autovalores com multiplicidade, os quais podem ou não ter autovetores diferentes.

Propriedades

- Interpretação geométrica: autovetor é aquele que não muda sua direção (span) ao passar por uma transformação linear, sendo apenas multiplicado por um escalar (autovalor).
- Traço de uma matriz quadrada: a soma dos elementos da diagonal principal é igual a soma dos autovalores.
- O produto dos autovalores é igual ao determinante da matriz.
- Em matrizes triangulares, os autovalores são iguais aos elementos da diagonal principal.
- Uma matriz $n \times n$ tem n autovalores, podendo haver autovalores repetidos.
- Autovalores diferentes correspondem a autovetores diferentes, enquanto autovalores iguais podem ou não ter autovetores diferentes (exemplo da matriz identidade).

Diagonalização de uma Matriz

Sendo A uma matriz $n \times n$ com n autovetores linearmente independentes, e sendo S uma matriz cujas colunas são iguais a esses autovetores, então:

$$\begin{aligned} AS &= A \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix} = \begin{bmatrix} \lambda_1 x_1 & \dots & \lambda_n x_n \end{bmatrix} = \\ &\begin{bmatrix} x_1 & \dots & x_n \end{bmatrix} \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} = S\Lambda \\ S^{-1}AS &= \Lambda \end{aligned}$$

Λ : Matriz diagonal com os autovalores de A na diagonal principal

Existência

A existência de autovalores repetidos pode ou não ser acompanhada pela existência de n autovetores independentes tal que a matriz seja diagonalizável (*caso da matriz identidade).

Definições

- A é simétrica $\iff A = A^T$ (Requer que A seja quadrada)
- Q é ortonormal \iff seus vetores colunas são todos unitários e ortogonais entre si. $Q^T Q = I = Q Q^{-1} \therefore Q^T = Q^{-1}$

Sendo A $n \times n$ simétrica, seus autovalores são todos reais e podem ser escolhidos n autovetores ortogonais entre si e unitários tal que, da sua forma diagonalizada:

$$A = Q \Lambda Q^{-1} \Rightarrow \text{Teorema Espectral: } \boxed{A = Q \Lambda Q^T}$$

Definição - Similaridade entre Matrizes

Duas matrizes A e B são ditas **similares** se, e somente se:

$$B = M^{-1}AM$$

Propriedade: **Matrizes similares tem os mesmos autovalores. Um autovetor x de A corresponde a um autovetor $M^{-1}x$ de B .**

Dem: $Ax = \lambda x \Rightarrow MBM^{-1}x = \lambda x \Rightarrow B(M^{-1}x) = \lambda(M^{-1}x)$.

Transformações de Similaridade

Definição - Similaridade em transformações lineares

Pela definição anterior, é possível concluir que duas matrizes similares representam a mesma transformação linear T com respeito a duas bases diferentes:

$$\begin{aligned} [T]_{V \text{ to } V} &= [I]_{V \text{ to } V} [T]_{V \text{ to } V} [I]_{V \text{ to } V} \\ B &= M^{-1}AM \end{aligned}$$

Relação com teorema espectral: Se A é uma transformação linear e deseja-se representá-la por meio de uma matriz diagonal, basta escolher uma base que consista de autovetores, pois em

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

A e Λ são matrizes similares, ou seja, que representam a mesma transformação linear, com S sendo uma base composta dos autovetores de A .

Forma Jordan

Como visto na seção de diagonalização, matrizes que não tem um conjunto completo de autovetores independentes não podem ser diagonalizadas, mas com a teoria de similaridade é possível chegar numa matriz '**mais diagonal o possível**', que é chamada de Forma de Jordan (J).

Definição

A forma Jordan de uma família de matrizes similares é uma matriz que contém os autovalores na diagonal principal e um '1' acima da diagonal para cada autovetor faltante.

$$J = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$$

Essa teoria é importante na resolução de equações diferenciais com matrizes.

Matriz Definida Positiva

Definição - Definida Positiva

Matriz A (simétrica) é definida positiva: $A \succ 0 \iff x^T A x > 0 \forall x \neq 0$

- Todos os autovalores são positivos;
- Todos os pivôs (sem troca de linha) são positivos;
- Todas as submatrizes a esquerda tem determinantes positivos;
- $\exists R$ com colunas independentes tal que $A = R^T R$.

Definição - Semidefinida Positiva

Matriz A (simétrica) é semidefinida positiva: $A \succcurlyeq 0 \iff x^T A x \geq 0 \forall x \neq 0$.

- Todos os autovalores são positivos ou nulos;
- Todos os pivôs (sem troca de linha) são positivos ou nulos;
- Todas as submatrizes a esquerda tem determinantes não negativos;
- $\exists R$ possivelmente com colunas dependentes tal que $A = R^T R$.

Matriz Definida Positiva - Exemplos e propriedades

Algumas propriedades

- Interpretação geométrica: a matriz A é definida positiva se sua transformação linear correspondente mantiver um vetor na mesma 'direção';
- Toda matriz definida positiva é inversível. Dem: A é invertível $\iff \det(A) \neq 0 = \lambda_1 \dots \lambda_n$, para uma matriz positiva definida: $\lambda_1 \dots \lambda_n > 0 \neq 0$;
- D é diagonal com entradas positivas, então D é definida positiva (pois os autovalores de uma matriz diagonal estão em sua diagonal principal);
- A única matriz de projeção definida positiva é a Identidade. Dem: $Px = \lambda x \Rightarrow \lambda = 1$ ou $\lambda = 0$. A opção definida positiva é a de $\lambda = 1$ na qual a matriz de projeção é a identidade, que projeta um vetor sobre si mesmo.

Quociente de Rayleigh

Definição

O quociente de Rayleigh de uma matriz, em função de um vetor x , é dado pela seguinte equação:

$$R(x) = \frac{x^T A x}{x^T x}$$

Se x é um autovetor (v) de A , então:

$$R(v) = \frac{v^T A v}{v^T v} = \lambda \frac{v^T v}{v^T v} = \lambda$$

Esse quociente é limitado inferior e superiormente pelo menor e pelo maior autovalor, respectivamente, tal que:

$$\lambda_1 \leq R(x) \leq \lambda_n$$

Os autovetores intermediários representam "saddle points" = pontos de sela.

Quociente de Rayleigh

Definição

Ponto de sela é um ponto na superfície do gráfico de uma função que não corresponde a um máximo ou mínimo local, mas cujas derivadas em direções ortogonais são todas nulas (ponto crítico).

Propriedade

Dado um elemento da diagonal de A , por exemplo: a_{11}

$$\lambda_1 \leq a_{11} \leq \lambda_n$$
$$R(x) = a_{11} \iff x = (1, 0, \dots, 0)$$

Singular Value Decomposition (SVD)

As decomposições estudadas até aqui contêm restrições que reduzem significativamente suas aplicabilidades (Ex.: teorema espectral, matriz deve ser simétrica). Uma decomposição muito poderosa que funciona para qualquer matriz retangular é o SVD:

Definição

Sendo A $m \times n$ (de rank r), seu SVD é:

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

- u_1, \dots, u_r : base ortonormal para o espaço-coluna
- u_{r+1}, \dots, u_m : base ortonormal para o espaço-nulo à esquerda
- v_1, \dots, v_r : base ortonormal para o espaço-linha
- v_{r+1}, \dots, v_n : base ortonormal para o espaço-nulo
- Σ contém os valores singulares (σ) na diagonal em ordem decrescente

SVD - Construção

Sendo A $m \times n$: $A^T A$ e AA^T são simétricas e positiva definidas.

$$A^T A = (U\Sigma V^T)^T (U\Sigma V^T) = V\Sigma^T U^T U\Sigma V^T = V\Sigma^T \Sigma V^T \rightarrow$$

Diagonalização de $A^T A$:

- $\Sigma^T \Sigma$: matriz diagonal com os autovalores ao quadrado (σ^2)
- V é a matriz dos autovetores de $A^T A$

$$AA^T = (U\Sigma V^T)(U\Sigma V^T)^T = U\Sigma V^T V\Sigma^T U^T = U\Sigma^T \Sigma U^T \rightarrow$$

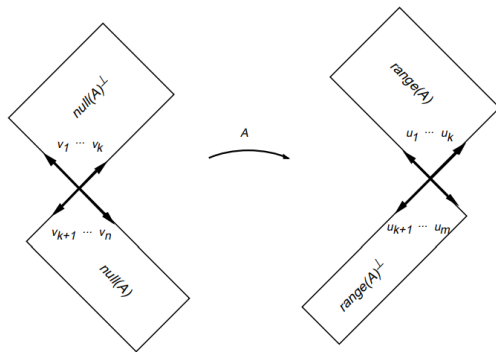
Diagonalização de AA^T :

- $\Sigma^T \Sigma$: matriz diagonal com os autovalores ao quadrado (σ^2)
- U é a matriz dos autovetores de AA^T

*Completa-se as matrizes U e V com os vetores das bases dos espaços nulos. Logo, U se torna uma base para \mathbb{R}^m e V para \mathbb{R}^n

SVD e o Teorema Fundamental da Álgebra Linear

O SVD se relaciona diretamente com o teorema fundamental da álgebra linear. Pensando em $A(m \times n)$ como uma transformação linear que leva vetores de \mathbb{R}^n para \mathbb{R}^m , o SVD fornece bases para subespaços em ambos esses espaços vetoriais: para o $\text{range}(A)$ (= Espaço Coluna de A), espaço nulo de A e seus complementos ortogonais.



SVD - Matrizes Particionadas

Sendo o rank k menor que m e n , o SVD pode ser representado da seguinte forma, usando matrizes particionadas:

$$A = \left[\begin{array}{ccc|ccc} u_1 & \cdots & u_k & u_{k+1} & \cdots & u_m \end{array} \right] \left[\begin{array}{ccc|ccc} \sigma_1 & & & & & \\ & \ddots & & & & \\ & & \sigma_k & & & 0 \\ \hline & & & 0 & & 0 \end{array} \right] \left[\begin{array}{c} v_1^T \\ \vdots \\ v_k^T \\ \hline v_{k+1}^T \\ \vdots \\ v_n^T \end{array} \right]$$

$$A = \left[\begin{array}{ccc} u_1 & \cdots & u_k \end{array} \right] \left[\begin{array}{ccc} \sigma_1 & & \\ & \ddots & \\ & & \sigma_k \end{array} \right] \left[\begin{array}{c} v_1^T \\ \vdots \\ v_k^T \end{array} \right] +$$

$$\left[\begin{array}{ccc} u_{k+1} & \cdots & u_m \end{array} \right] \left[\begin{array}{c} \mathbf{0} \\ v_{k+1}^T \\ \vdots \\ v_n^T \end{array} \right]$$

É notório que só contribuem para A efetivamente os primeiros k u 's e v 's, logo:

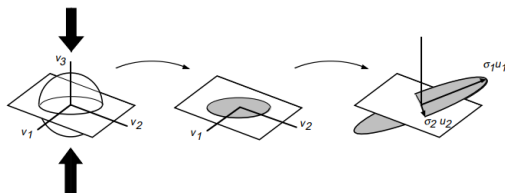
$$A = \begin{bmatrix} u_1 & \cdots & u_k \end{bmatrix} \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_k \end{bmatrix} \begin{bmatrix} v_1^T \\ \vdots \\ v_k^T \end{bmatrix}$$

Outra forma de escrever o SVD é utilizando a forma de 'outer product expansion':

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

SVD - Interpretação Geométrica

Como o SVD deforma o espaço:



A comprime $n - k$ dimensões do domínio e estica e contrai as outras na magnitude dos valores singulares.

SVD - Interpretação geométrica 2

O SVD fornece uma base para \mathbb{R}^m tal que os eixos coordenados não sejam distorcidos após a transformação A , ou seja, de forma que eles continuem ortogonais.

$A = U\Sigma V^T \Rightarrow AV = U\Sigma \rightarrow$ Aplicando a transformação linear A nos vetores colunas de V (base ortonormal para \mathbb{R}^n), são obtidos os vetores coluna de U (base ortonormal para \mathbb{R}^m , ainda ortogonais, mesmo após a transformação) multiplicados por escalares (valores singulares).

- Otimização do método dos quadrados mínimos
- Compressão de dados
- Inverso generalizado de uma matriz

Método dos mínimos quadrados

Contexto geral: encontrar a combinação linear de um conjunto de vetores (se tornam vetores coluna de uma matriz A) que é a melhor aproximação de um outro vetor (b).

Definição

O método dos mínimos quadrados buscar minimizar o erro dado por:

$$\|b - Ax\|$$

Pela interpretação geométrica, busca-se encontrar um vetor do espaço gerado por A (S) que está mais próximo de b , o qual é a projeção de b nesse espaço, de forma que o vetor erro (diferença entre b e a projeção) seja ortogonal a S . Por essa ortogonalidade:

$$A^T(Ax - b) = 0 \Rightarrow A^T Ax = A^T b \Rightarrow x = (A^T A)^{-1} A^T b$$

Método dos mínimos quadrados

A análise feita anteriormente é válida, mas a computação de $A^T A$ é **origem de grande perda de precisão dos dados**, logo deve ser evitada.

Abordagem do problema com SVD:

Sendo $A = U\Sigma V^T$ o SVD de A , tem-se que:

$$Ax - b = U\Sigma V^T x - b = U(\Sigma V^T x) - U(U^T b)$$
$$\boxed{Ax - b = U(\Sigma y - c)} \text{ onde } y = V^T x \text{ e } c = U^T b.$$

Como U é uma matriz ortogonal, ela preserva a norma de vetores. Logo:
 $\|U(\Sigma y - c)\| = \|\Sigma y - c\|.$

Definição

Para minimizar o erro $\|Ax - b\|$, basta minimizar $\|y - c\|$ de tal forma que:

$$\boxed{\|Ax - b\| = \|\Sigma y - c\|}$$
$$y = V^T x \Rightarrow \boxed{x = Vy}$$

Método dos mínimos quadrados

Por que transformar a tarefa de minimizar $\|Ax - b\|$ em minimizar $\|\Sigma y - c\|$?

Pois a natureza diagonal de Σ torna a tarefa simples:

Sendo A $m \times n$ com rank k :

$$\Sigma y = \begin{bmatrix} \sigma_1 y_1 \\ \vdots \\ \sigma_k y_k \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad \text{então} \quad \Sigma y - c = \begin{bmatrix} \sigma_1 y_1 - c_1 \\ \vdots \\ \sigma_k y_k - c_k \\ -c_{k+1} \\ -c_{k+2} \\ \vdots \\ -c_m \end{bmatrix}$$

Método dos mínimos quadrados

Por inspeção, quando $y_i = c_i/\sigma_i$ para $1 \leq i \leq k$, as primeiras k -entradas do vetor erro são zeradas, e logo o erro assume menor valor:

$$\left[\sum_{i=k+1}^m c_i^2 \right]^{1/2}$$

Observação: quando $k = m$, todas as entradas do vetor erro são zeradas e, então, o erro é 0. Isso ocorre pois se A tem rank m , então seu espaço coluna é todo o R^m e portanto o problema é resolvido com uma solução perfeita.

Definição

No problema do método dos mínimos quadrados, para minimizar o erro $\|Ax - b\|$, sendo $A = U\Sigma V^T$ o SVD de A , a escolha de x é:

$$x = V\Sigma^+ U^T b$$

Σ^+ : Transposta de Σ com as entradas diferentes de 0 da diagonal invertidas.

Definição

Dado um vetor $b \in R^m$, há um único vetor $x \in R^n$ com norma mínima que aproxima Ax de b ao máximo. A matriz que transforma b em x é chamada de inversa generalizada de Moore-Penrose, sendo denotada por A^+ .

$$x = A^+ b$$

No contexto anterior, Σ^+ é o inverso generalizado de Σ , como definido na solução pelo método dos mínimos quadrados para Σ e c . Logo:

Se $A = U\Sigma V^T$, então a inversa generalizada de A é: $A^+ = V\Sigma^+ U^T$

Discussão sobre precisão

Se $A^T A$ for 'quase singular', ou seja, se houver uma quase dependência entre suas colunas, um de seus valores singulares será muito próximo de 0. Nesse caso, a solução pelas equações normais não é confiável, pois se o menor autovalor é muito próximo de 0, os outros podem cobrir um range muito grande de magnitudes. É justamente essa diferença que causa a perda de precisão nos cálculos numéricos.

Dentro dos limites da precisão computacional, o valor de $y = (A^T A)^{-1}x$ está correto, mas não serve como uma aproximação boa o bastante para $A^T A y = x$, pois os dígitos perdidos na computação de y tinham significado essencial para a restauração de x .

Definição

O número de condicionamento de uma matriz é dado pelo quociente λ_1/λ_n , com os lambdas ordenados de forma decrescente (λ_n é o menor autovalor).

Discussão sobre precisão

Definição

Um número de condicionamento grande significa maior chance de instabilidade numérica na computação envolvendo essa matriz. Para matrizes quadradas, esse número também pode ser interpretado como a recíproca da distância para a matriz singular mais próxima.

Equações normais vs SVD

Na solução utilizando-se do SVD, os valores singulares são explicitamente invertidos em Σ^+ , então há perda de precisão. Mas, se para A o número de condicionamento é σ_1/σ_n , para $A^T A$ o número de condicionamento é de $(\sigma_1/\sigma_n)^2$. Logo:

Para alcançar uma precisão próxima daquela fornecida pelo SVD de A , a computação com $A^T A$ necessita de aproximadamente o dobro de casas decimais.

Na compressão de dados, o objetivo é representar uma matriz A $m \times n$ de dados numéricos utilizando bem menos números do que as mn entradas originais.

Dessa forma, a matriz não é vista como uma transformação linear e nem como um objeto algébrico, mas sim como uma tabela de valores dos quais se deseja extrair as características mais importantes.

Redundância: O posto (rank) de uma matriz é igual ao número de colunas (ou linhas) independentes. Logo, uma matriz de posto pequeno tem uma grande quantidade de redundância. Ou seja, suas entradas podem ser representadas de forma mais simples.

Definição

Uma matriz de posto 1 tem apenas uma coluna independente, ou seja, todas as outras colunas são múltiplos dessa. **Toda matriz de posto 1 pode ser escrita na forma de um produto diádico (outer product).** Sendo B uma matriz dessa, u um vetor que é base do espaço coluna de B , e v composto por coeficientes tal que a i -ésima coluna de B é dada por $v_i u$:

$$B = [v_1 u \dots v_n u] = uv^T$$

É evidente que, antes representada por mn elementos, B foi comprimida ao ser representada suficientemente por $m + n$ valores.

Portanto, é natural que se busque a melhor aproximação com posto 1 de qualquer matriz, para realizar a compressão.

Definição - Norma de Frobenius

A norma de Frobenius de uma matriz ($|X|$) é definida como a soma dos quadrados de todas as suas entradas.

Essa norma é como a norma euclidiana abstraindo a matriz para um vetor em R^{mn} . O produto interno entre matrizes é definido da mesma forma.

Definição - Melhor aproximação de posto 1

A matriz B é a melhor aproximação de posto 1 de A se $B - A$ tem norma de Frobenius mínima.

Compressão de dados

O produto interno entre duas matrizes de posto 1 (xy^T e uv^T) é:

$$xy^T \cdot uv^T = [xy_1 \dots xy_n] \cdot [uv_1 \dots uv_n] = \sum_i xy_i \cdot uv_i = (x \cdot u)(y \cdot v)$$

Lembrando da expansão do produto diádico:

$$XY = \sum_i x_i y_i^T$$

Juntando essas duas expressões, temos que:

$$\begin{aligned} XY \cdot XY &= \left(\sum_i x_i y_i^T \right) \cdot \left(\sum_j x_j y_j^T \right) = \sum_{ij} (x_i \cdot x_j) (y_i \cdot y_j) \\ |XY|^2 &= \sum_i |x_i|^2 |y_i|^2 + \sum_{i \neq j} (x_i \cdot x_j) (y_i \cdot y_j) \end{aligned}$$

Da expressão anterior, é possível concluir que:

- Se X é ortogonal, o segundo somatório é igual a 0;
- Se X é ortonormal: $|XY|^2 = \sum_i |y_i|^2 = |Y|^2$.

Concluí-se então que, multiplicando pela direita ou pela esquerda por uma matriz ortonormal, a norma permanece constante.

A representação do SVD pelo produto diádico (outer product expansion) pode ser vista como uma decomposição de uma matriz em várias matrizes de posto 1.

$$A = \sum_i \sigma_i u_i v_i^T \Rightarrow |A|^2 = \sum_i |\sigma_i u_i v_i^T|^2 = \sum_i \sigma_i^2$$

Sendo A_1 uma matriz de posto 1:

$$|A - A_1| = |U\Sigma V^T - A_1| = |\Sigma - U^T A_1 V|$$

Podemos escrever:

$$\Sigma - U^T A_1 V = \alpha xy^T$$

com x e y unitários e α positivo.

$$|\Sigma - \alpha xy^T|^2 = |\Sigma|^2 + \alpha^2 - 2\alpha \Sigma \cdot xy^T$$

Focando no último termo da expressão:

$$\begin{aligned}\Sigma \cdot xy^T &= \sum_{i=1}^k \sigma_i x_i y_i \\ &\leq \sum_{i=1}^k \sigma_i |x_i| |y_i| \\ &\leq \sigma_1 \sum_{i=1}^k |x_i| |y_i| = \sigma_1 x^* \cdot y^*\end{aligned}$$

Pela desigualdade de Cauchy - Schwarz: $x^* \cdot y^* \leq |x^*| \cdot |y^*| \leq |x| \cdot |y| = 1$
Voltando para a expressão anterior:

$$\Sigma \cdot xy^T \leq \sigma_1 x^* \cdot y^* \leq \sigma_1$$

Voltando a argumentação principal:

$$|\Sigma - \alpha xy^T|^2 \geq |\Sigma|^2 + \alpha^2 - 2\alpha\sigma_1 = |\Sigma|^2 + (\alpha - \sigma_1)^2 - \sigma_1^2$$

A expressão anterior é minimizada quando $\alpha = \sigma_1$ e x e y são iguais a $e_1 = (1, 0, \dots, 0)$.

Logo, para que seja obtido o menor erro:

$$A_1 = \alpha U_{xy}^T V^T = \sigma_1 (Ue_1)(Ve_1)^T v = \sigma_1 u_1 v_1^T$$

Definição

A melhor aproximação de posto 1 de A é o primeiro termo da outer product expansion do seu SVD: $A_1 = \sigma_1 u_1 v_1^T$.

Cujo erro é igual a: $|\sum_{i=2} \sigma_i u_i v_i^T| = \sigma_2^2 + \dots + \sigma_k^2$

Definição

Um algoritmo para encontrar sucessivamente as melhores aproximações de A é o seguinte: Primeiramente, encontra-se a melhor aproximação de posto 1 A_1 de A , a qual tem erro E_1 . Em seguida, encontra-se a melhor aproximação de posto 1 de E_1 tal que: $E_2 = E_1 - A_2 = A - (A_1 + A_2)$, ou seja, $A_1 + A_2$ é a melhor aproximação de posto 2 de A , e assim por diante para os outros postos, encontrando sempre a melhor aproximação de posto 1 para as matrizes de erro.

Definição

A matriz A pode ser decomposta da seguinte maneira:

$$A = \sum_i A_i$$

Onde $A_i = \sigma_i u_i v_i^T$. De maneira análoga:

$$|A|^2 = |S_r|^2 + |E_r|^2$$

Onde $S_r = \sum_{i=1}^r \sigma_i u_i v_i^T$ é a melhor aproximação de posto r de A , e $E_r = \sum_{i=r+1}^k \sigma_i u_i v_i^T$ é o erro dessa aproximação.

Definição

A porcentagem de erro da melhor aproximação de A para certo posto r é dada por:

$$\frac{|E_r|}{|A|}(\%) = \sqrt{\frac{\sum_{i=r+1}^k \sigma_i^2}{\sum_{i=1}^k \sigma_i^2}} (\times 100)$$

Logo, o posto r é escolhido como o menor possível que esteja dentro da margem de erro necessária.

Fim