Decomposição em Valores Singulares

Prof. Alexandre Salles da Cunha e Profa. Ana Paula Couto

UFMG
UNIVERSIDADE FEDERAL
DE MINAS GERAIS

UFMG - ICEx
DEPARTAMENTO DE CIÊNCIA DA
COMPUTAÇÃO

Assumimos que $A \in \mathbb{R}^{m \times n}$, com posto $r(A) = r$, não necessariamente completo.

Fatoração completa e reduzida

Da mesma maneira que apresentamos uma fatoração completa e uma reduzida para $A = QR$, apresentaremos uma fatoração reduzida e uma completa para $A = U\Sigma V^T$ ou $AV = U\Sigma$. Começaremos com a fatoração reduzida.

SVD Reduzida

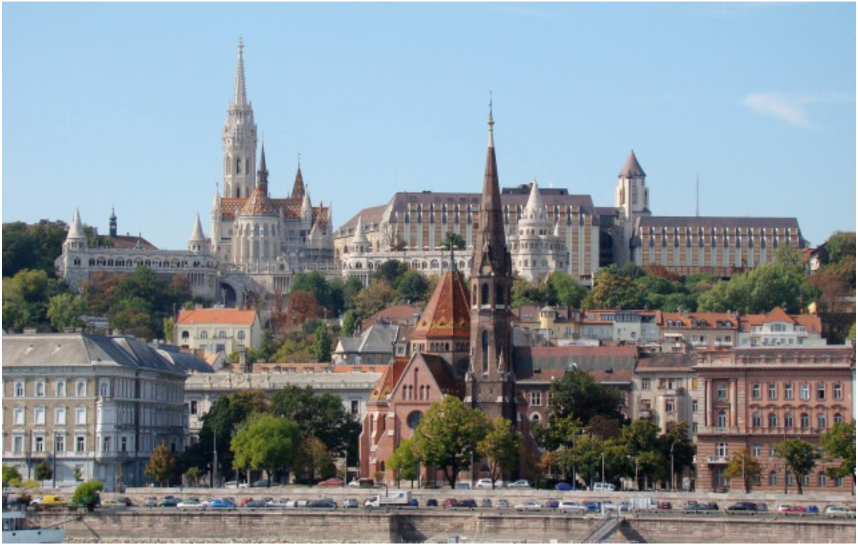
$$A = U\Sigma V^T \rightarrow AV = U\Sigma \rightarrow Av_i = \sigma_i u_i : i = 1, \ldots, r.$$

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

1. A é uma soma de r matrizes (bastante particulares !) de rank 1.
2. V é uma matriz $n \times r$, cujas colunas são os vetores singulares à direita de A. V possui colunas ortonormais, isto é, $V^T V = I_r$.
3. U é uma matriz $m \times r$, cujas colunas são os vetores singulares à esquerda de A. U também possui colunas ortonormais, isto é, $U^T U = I_r$.
4. Σ é uma matriz diagonal, $r \times r$, cujos elementos na diagonal, $\sigma_i : i = 1, \ldots, r$, são os valores singulares de A.
5. $\sigma_1 \geq \sigma_2 \cdots \geq \sigma_r > 0$: indicam a importância do termo $\sigma_i u_i v_i^T$ em A.

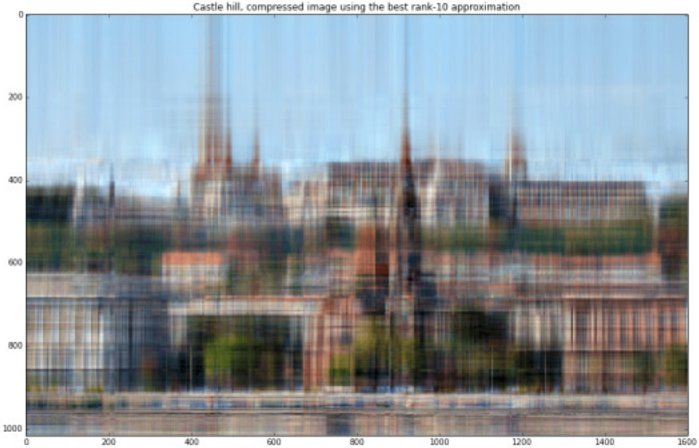
Uma aplicação comum: Low rank approximation (imagem completa)

https://www.balabit.com/blog/image-compression-using-singular-value-decomposition/



Aproximação da imagem com os $k = 50$ maiores valores singulares

Aproximação da imagem com os $k = 10$ maiores valores singulares



- Para algum $i = 1, \ldots, r$, A é aplicada em um vetor singular $v_i \in \mathbb{R}^n$ (com norma 2 unitária).
- A imagem desta transformação linear ($Av_i \in \mathbb{R}^m$) é $\sigma_i u_i$, um vetor em $\mathcal{C}(A)$ (espaço coluna de A).
- u_i também tem norma 2 unitária.
- Logo, σ_i representa a norma desta transformação Av_i.
- σ_1 é a norma espectral de A.

- Nas matrizes reais simétricas, podemos ortogonalizar os autovetores. Estas matrizes admitem uma fatoração espectral

$$S = Q^T \Lambda Q.$$

- A fatoração SVD de uma matriz quadrada generaliza a fatoração espectral de matrizes reais simétricas, fornecendo uma "diagonalização" para estas matrizes retangulares.

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

- Diferente das matrizes simétricas, que têm $\mathcal{C}(A) = \mathcal{C}(A^T)$ (espaço coluna e linha são iguais), nas matrizes retangulares (ou quadradas não simétricas), estes espaços são diferentes.

  Portanto, precisamos de bases diferentes para representá-los:

  - Os vetores $u_i : i = 1, \ldots, r$ fornecem uma base ortonormal para $\mathcal{C}(A)$.
  - Os vetores $v_i : i = 1, \ldots, r$ fornecem uma base ortonormal para o espaço linha de A, isto é, $\mathcal{C}(A^T)$.

Autovetores de $A^T A$.

- $A^T A$ é uma matriz simétrica semipositiva definida, de posto $r = r(A)$ e, portanto, possui r autovalores positivos (contando as multiplicidades).
- A multiplicidade do autovalor $\lambda_i = 0$ para $A^T A$ é $n - r$, que é a dimensão do nulo de A.
- Veja a fatoração espectral de $A^T A$:

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

- Os vetores singulares $v_i : i = 1, \ldots, r$ são os autovetores de $A^T A$, associados aos autovalores $\lambda_i > 0$
- $\sigma_i = \sqrt{\lambda_i} : i = 1, \ldots, r$.

Autovetores de AA^T.

- AA^T é uma matriz simétrica semipositiva definida, de posto $r = r(A)$ e, portanto, possui r autovalores positivos (contando as multiplicidades).
- A multiplicidade do autovalor $\lambda_i = 0$ para AA^T é $m - r$, que é a dimensão do nulo de A^T.
- Veja a fatoração espectral de AA^T:

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

- Os vetores singulares $u_i : i = 1, \ldots, r$ são os autovetores de AA^T, associados aos autovalores $\lambda_i > 0$.
- Os r autovalores positivos de $A^T A$ e de AA^T são os mesmos.
- $\sigma_i = \sqrt{\lambda_i} : i = 1, \ldots, r$.

- O aspecto central que garante a fatoração SVD de A é o fato de que as matrizes simétricas admitem uma fatoração espectral em que seus autovetores são ortogonalizados. Este aspecto garante que os autovetores de A^TA e AA^T forneçam os vetores singulares necessários na fatoração SVD de A.

- Veja que os u_i, u_j que resultam de Av_i, Av_j são automaticamente ortogonais, dado que v_i, v_j são autovetores (ortogonais) de A^TA:

$$\begin{aligned}
Av_i &= \sigma_i u_i \\
u_i^T u_j &= \left(\frac{Av_i}{\sigma_i}\right)^T \left(\frac{Av_j}{\sigma_j}\right) \\
&= \frac{v_i^T(A^TAv_j)}{\sigma_i\sigma_j} \\
&= \frac{\sigma_j^2}{\sigma_i\sigma_j} v_i^T v_j \\
&= 0
\end{aligned}$$

Além dos r vetores singulares

- $u_i : i = 1, \dots, r$, que fornecem uma base ortonormal para $\mathcal{C}(A)$ e de
- $v_i : i = 1, \dots, r$, que fornecem uma base ortonormal para $\mathcal{C}(A^T)$,

a fatoração completa oferece bases ortonormais para os dois espaços nulos:

- $v_{r+i} : i = 1, \dots, n - r$, para $\mathcal{N}(A)$
- $u_{r+i} : i = 1, \dots, m - r$, para $\mathcal{N}(A^T)$

Desta forma, na fatoração completa, as matrizes V, U são quadradas de ordem n e m, respectivamente, e a matriz Σ é $m \times n$, pois ganha $m - r$ linhas de zeros e $n - r$ colunas de zeros, garantindo a conformabilidade de $A = U\Sigma V^T$.

$\Rightarrow$ Continuamos fatorando $A = \sigma_1 u_1 v_1^T + \cdots + \sigma_r u_r v_r^T$.

- $A = \begin{bmatrix} 3 & 0 \\ 4 & 5 \end{bmatrix}$ com $r(A) = 2$
- $A^T A = \begin{bmatrix} 25 & 20 \\ 20 & 25 \end{bmatrix}$, $AA^T = \begin{bmatrix} 9 & 12 \\ 12 & 41 \end{bmatrix}$.
- Autovalores de $A^T A : \lambda_1 = \sigma_1^2 = 45, \lambda_2 = \sigma_2^2 = 5$
- Autovetores de $A^T A$:
  - Associado a $\lambda_1 = 45$: $v_1 = \frac{\sqrt{2}}{2} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$
  - Associado a $\lambda_2 = 5$: $v_2 = \frac{\sqrt{2}}{2} \begin{bmatrix} -1 \\ 1 \end{bmatrix}$
- Vetores u_1, u_2 calculados a partir de v_1, v_2:
  - $u_1 = \frac{1}{\sqrt{45}} A v_1 = \frac{\sqrt{10}}{10} \begin{bmatrix} 1 & 3 \end{bmatrix}^T$
  - $u_2 = \frac{1}{\sqrt{5}} A v_2 = -\frac{\sqrt{10}}{10} \begin{bmatrix} -3 & 1 \end{bmatrix}^T$
- A não é simétrica: os u's e os v's são diferentes.

- Verifique que $u_1 = \frac{\sqrt{10}}{10}\begin{bmatrix} 1 & 3 \end{bmatrix}^T$ e $u_2 = -\frac{\sqrt{10}}{10}\begin{bmatrix} -3 & 1 \end{bmatrix}^T$ são autovelores de $AA^T = \begin{bmatrix} 9 & 12 \\ 12 & 41 \end{bmatrix}$, para $\lambda_1 = 45$ e $\lambda_2 = 5$.
- Termos na fatoração:

$$U = \frac{\sqrt{10}}{10}\begin{bmatrix} 1 & -3 \\ 3 & 1 \end{bmatrix},$$

$$\Sigma = \begin{bmatrix} \sqrt{45} & 0 \\ 0 & \sqrt{5} \end{bmatrix} \text{ e}$$

$$V = \frac{\sqrt{2}}{2}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}.$$

Abordagem inocente:

1. Construa $A^T A$, a matriz de covariância de A.
2. Faça a decomposição espectral $A^T A = V \Lambda V^T$.
3. Seja Σ a matriz $m \times n$ diagonal (não negativa) com a raiz quadrada das entradas de Λ.
4. Resolva o sistema $U\Sigma = AV$, obtendo as colunas de U, via fatoração QR, por exemplo.

A abordagem acima reduziu o problema de fator $A = U\Sigma V^T$ a um problema de decomposição espectral da matriz de covariância, que é muito mais sensível a erros numéricos e perturbações.

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

- autovalores $\lambda = 0, 0, 0, 0$ (todos zero).
- único autovetor $[1\,0\,0\,0]^T$.
- valores singulares $\sigma = 3, 2, 1$.
- vetores singulares são colunas da matriz I.

$$A^TA = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 9 \end{bmatrix}, \; AA^T = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 9 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

- autovalores são $\sigma^2 = 9, 4, 1$.
- autovetores u são calculados a partir de AA^T e de v a partir de A^TA.

$$U = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \Sigma = \begin{bmatrix} 3 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, V = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

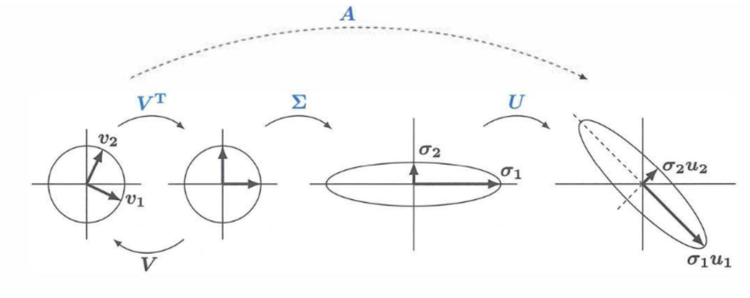
$$A = U\Sigma V^T = 3u_1 v_1^T + 2u_2 v_2^T + 1u_3 v_3^T.$$

Suponha que o valor do elemento $A(4,1)$ seja igual à $1/60.000$:

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ \frac{1}{60.000} & 0 & 0 & 0 \end{bmatrix}$$

- pequena mudança gerou um variação grande nos autovalores de A: $\lambda = \frac{1}{10}, \frac{i}{10}, \frac{-1}{10}, \frac{-i}{10}$.
- autovalores são instáveis quando AA^T é muito diferente de A^TA. Se $A^TA = AA^T$ os autovetores de A são ortogonais e os autovalores de A são estáveis.
- os valores singulares são estáveis: não mudam mais do que a mudança em A.
- novos valores singulares: 3, 2, 1, $1/60.000$.
- matrizes U e V são as mesmas.

- SVD fatora a matriz em 3 partes: (ortogonal) x (diagonal) x (ortogonal).
- Geometricamente temos: (rotação) x (expansão) x (rotação).
- Matrizes U e V são rotações e reflexões.
- Matriz Σ transforma o círculo em elipse (mudança de escala).



Introduction to Linear Algebra - Gilbert Strang

A transformação linear definida por A corresponde à: sua aplicação em $v_i \in \mathcal{C}(A^T)$, obtendo um vetor $(\sigma_i u_i) \in \mathcal{C}(A)$.

E a transformação linear inversa ?

Vamos usar uma matriz auxiliar, denominada pseudoinversa de A, A^+, para representar esta transformação inversa, de $\mathcal{C}(A)$ para $\mathcal{C}(A^T)$:

$$A^+ u_i = \frac{v_i}{\sigma_i}, i = 1, \ldots, r(A)$$

Esta matriz A^+ que representa esta transformação sempre existe.

Desta definição, resulta:

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

onde $A^+ \in \mathbb{R}^{n\times m}$, $V \in \mathbb{R}^{n\times n}$, $\Sigma^+ \in \mathbb{R}^{n\times m}$, $U^T \in \mathbb{R}^{m\times m}$.

- Se A é quadrada e A^{-1} existe, ou seja, $m = n = r(A)$, $A^+ = A^{-1}$.
- Matrizes com $r(A) < m$ ou $r(A) < n$ possuem o mesmo posto de A e são definidas satisfazendo:

$$A^+ u_i = \frac{1}{\sigma_i} v_i, \text{ para } i \leq r \text{ e } A^+ u_i = 0, \text{ para } r+1 \leq i \leq m$$

Os vetores $u_1, \cdots, u_r \in \mathcal{C}(A)$ são mapeados nos vetores $v_1, \cdots, v_r \in \mathcal{C}(A^T)$. Os demais vetores $u_{r+1}, \cdots, u_m \in \mathcal{N}(A^T)$ e são mapeados no vetor nulo.

- A pseudoinversa da matriz Σ é obtida através da transposição de Σ e pela substituição de σ por $\frac{1}{\sigma}$.
- O produto $\Sigma^+\Sigma$ é uma aproximação da matriz I: invertemos os valores singulares na diagonal, sem modificar os zeros nas linhas e colunas.

$$\Sigma^+\Sigma = \begin{bmatrix} 1/2 & 0 & 0 \\ 0 & 1/3 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}$$

Encontre a pseudoinversa de $A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$

- Matriz singular.
- r(A) $= 1$.
- Valores singulares $\sigma_1 = 2$.

$$A^+ = V\Sigma^+U^T = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} 1/2 & 0 \\ 0 & 0 \end{bmatrix}\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} = \frac{1}{4}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

- Quando A tem posto completo, o sistema $A^T A\hat{x} = A^T b$ possui uma única solução. A partir da fatoração SVD temos:

$$
\begin{aligned}
A^T A\hat{x} &= A^T b \\
(U\Sigma V^T)^T (U\Sigma V^T)\hat{x} &= (U\Sigma V^T)^T b \\
V\Sigma^T U^T U\Sigma V^T \hat{x} &= (V\Sigma U^T) b \\
V\Sigma^2 V^T \hat{x} &= V\Sigma U^T b \\
V^T V\Sigma^2 V^T \hat{x} &= V^T V\Sigma U^T b \\
\Sigma^2 V^T \hat{x} &= \Sigma U^T b \\
(\Sigma^2)^{-1}\Sigma^2 V^T \hat{x} &= (\Sigma^2)^{-1}\Sigma U^T b \\
V^T \hat{x} &= \Sigma^{-1} U^T b \\
VV^T \hat{x} &= V\Sigma^{-1} U^T b \\
\Rightarrow \quad \hat{x} &= (V\Sigma^{-1} U^T) b
\end{aligned}
$$

- Para os casos onde $r(A) < n$: o sistema de equações normais terá infinita soluções. Utilizamos a pseudoinversa A^+ para obtermos uma solução x^+ :

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

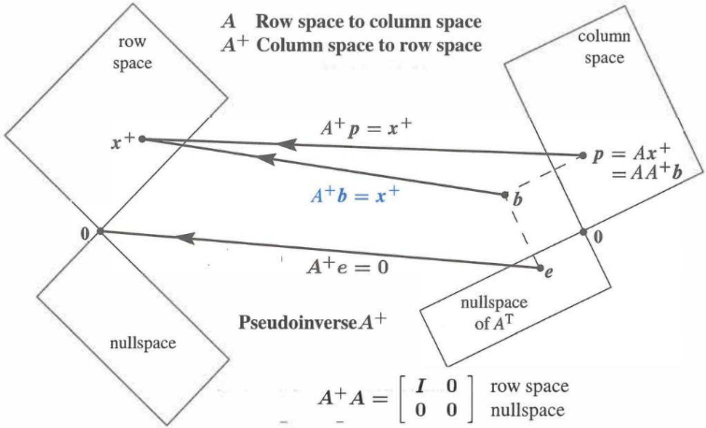


Figure 7.6: Ax^+ in the column space goes back to $A^+Ax^+ = x^+$ in the row space.

Introduction to Linear Algebra - Gilbert Strang

Problema

Qual a melhor aproximação para uma matriz A com posto $r(A) = r$ utilizando uma matriz de posto k?

Exemplo: Em aplicações de compressão de imagens, a aproximação de posto baixo resulta em uma versão compacta da matriz original. Se a matriz original A é descrita por mn elementos, podemos, por exemplo, decompor A em matrizes com $k(m+n)$ elementos. Em aplicações com imagens, valores de $k = \{100, 150\}$ são frequentemente usados para gerar imagens que são muito parecidas com a imagem original.

Se queremos a melhor aproximação da matriz A a partir de uma matriz de posto k, como podemos obtê-la?

A fatoração SVD decompõe a matriz A em componentes (matrizes de rank-1) ordenados por importância (valores singulares).

Low-Rank Approximation via SVD

Dada uma matriz A, selecionamos as componentes mais importantes (k maiores valores singulares) e definimos uma matriz aproximada A_k:

$$A_k = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \cdots + \sigma_k u_k v_k^T$$

com $r(A_k) = k$.

A melhor aproximação de uma matriz A a partir de uma outra matriz de posto k é a matriz A_k

Teorema de Eckart-Young

Se B é uma matriz de posto k então $||A - B|| \geq ||A - A_k||$.

- A aproximação fornecida por uma outra matriz B é, no máximo, tão boa quanto a aproximação dada por A_k.
- O teorema se aplica às três principais normas da matriz A: espectral, Frobenius, nuclear.

Teorema de Eckart-Young e Norma Espectral

Se $r(B) \leq k$ então $||A - B|| = \max \frac{||(A-B)x||}{||x||} \geq \sigma_{k+1}$.

A aproximação A_k para a matriz original A resulta em $||A - A_k|| = \sigma_{k+1}$. Qualquer matriz $B \neq A_k$ terá $||A - B|| \geq \sigma_{k+1}$.

Exemplo: Aproximar a matrix A por A_2.

$$A = \begin{bmatrix} 4 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Valores singulares de A: 4, 3, 2, 1.
Vetores singulares são colunas da matriz I.

$$A_2 = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T = \begin{bmatrix} 4 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$||A - A_2|| = 2$$

Como escolher o valor de k?

- Nos casos onde os valores singulares possuem uma divisão clara entre valores grandes (poucos) e valores pequenos (maioria), podemos escolher os top-k maiores valores.
- No entanto, na maioria dos casos, esta divisão não é clara. Nestes casos, diferentes estratégias podem ser utilizadas. Por exemplo, escolher um k tal que a soma dos top-k valores singulares é pelo menos c vezes maior do que a soma dos demais valores singulares.