**UNIVERSIDADE FEDERAL DO PARANÁ**

**Carlos Ronchi**

# Lista 3 - CM116

Lista de exercícios para a disciplina de tópicos em matemática aplicada I.

Professor Dr. Geovani Grapiglia.

**Curitiba**
**Junho de 2017**

*Solução.* (Exercício 1)

Para este exercício vamos usar os seguintes lemas.

**Lema 0.1.** *Seja $A \in \mathbb{R}^{m\times n}$, então* $\text{Ker}(A^TA) = \text{Ker}(A)$.

*Demonstração.* Seja $x \in \text{Ker}(A)$, então $A^TAx = A^T0 = 0$. Portanto, $x \in \text{Ker}(A^TA)$.

Agora, seja $x \in \text{Ker}(A^TA)$, então $A^TAx = 0$. Multiplicando por $x^T$ em ambos os lados, temos que

$$\|Ax\| = 0.$$

Portanto, $Ax = 0$ e assim $x \in \text{Ker}(A)$. Concluimos então que $\text{Ker}(A^TA) = \text{Ker}(A)$. □

**Lema 0.2.** *Seja $A \in \mathbb{R}^{m\times n}$, então $A^TA$ possui autovalores não-negativos.*

*Demonstração.* Seja $v$ um autovetor associado a $\lambda$ da matriz $A^TA$, então

$$A^TAv = \lambda v \to v^TA^TAv = \lambda v^Tv.$$

Como $A^TA$ é uma matriz semi definida positiva, temos que

$$\lambda v^Tv \geq 0.$$

Além disso, como $v$ é um autovetor, $v^Tv > 0$, portanto,

$$\lambda \geq 0.$$

□

Vamos agora provar o SVD. Seja $A \in \mathbb{R}^{m\times n}$ uma matriz qualquer e suponha que $m \geq n$. Seja a matriz $A^TA \in \mathbb{R}^{n\times n}$ e sejam $x_1, \ldots, x_n$ autovetores ortonormais e $\lambda_1, \ldots, \lambda_n$ os seus respectivos autovalores. Vamos ordenar os autovalores de forma que $\lambda_1 \geq \ldots \lambda_n$. E seja $r$ tal que

$$\lambda_1 \geq \cdots \geq \lambda_r > \lambda_{r+1} = \cdots = \lambda_n.$$

Defina agora os vetores

$$q_i = \frac{1}{\sqrt{\lambda_i}} Ax_i, \quad \text{para } i \leq r,$$

e chame de $\sigma_i = \sqrt{\lambda_i}$, para todo $i = 1, \ldots, n$.

Observe que $q_i^Tq_j = 0$ para todo $i \neq j$ e $q_i^Tq_i = 1$. Portanto, vamos completar $\{q_1, \ldots, q_r\}$ de forma a obter uma base $\{q_1, \ldots, q_r, q_{r+1}, \ldots, q_m\}$ de $\mathbb{R}^m$. Sejam agora as matrizes

$$U = \begin{pmatrix} | & & | \\ q_1 & \ldots & q_m \\ | & & | \end{pmatrix}, \qquad V = \begin{pmatrix} | & & | \\ x_1 & \ldots & x_n \\ | & & | \end{pmatrix}.$$

Note que $x_j \in \text{Ker}(A^TA)$ para $j > r$, pois o autovalor associado a $x_j$ é 0. Pelo Lema 0.1, $Ax_j = 0$. Portanto,

$$(U^TAV)_{ij} = q_i^TAx_j = \begin{cases} \frac{1}{\sigma_i} x_iA^TAx_j & \text{se } j \leq r \\ 0 & \text{se } j > r \end{cases} = \begin{cases} \sigma_i\delta_{ij} & \text{se } j \leq r \\ 0 & \text{se } j > r \end{cases},$$

pois

$$\frac{1}{\sigma_i} x_i^TA^TAx_j = \frac{\lambda_j}{\sigma_i} x_i^Tx_j = \sigma_i\delta_{ij}.$$

Portanto, $U^TAV$ é dada por

$$\begin{array}{c} \overbrace{\hphantom{\sigma_1 \quad 0 \quad \cdots \quad 0}}^{\text{n}} \\ \text{n}\left\{ \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_n \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \end{bmatrix} \right\} \text{m-n} \end{array}$$

Denotando por $\Sigma$ a matriz acima, temos que

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

Seja agora $m < n$. Considere a matriz simétrica $AA^T$. Do memso modo que a anrerior, temos que seus autovalores são todos não negativos. Sejam $v_1, \ldots, v_m$ os autovalores ortonormais associados a $\lambda_1, \ldots, \lambda_m$ autovalores. Ordene $\lambda_1 \geq \cdots \geq \lambda_m$. Denote por $\sigma_i = \sqrt{\lambda_i}$ para todo $i = 1, \ldots, m$. Seja $r$ tal que

$$\lambda_1 \geq \cdots \geq \lambda_r > \lambda_{r+1} = \cdots = \lambda_m.$$

Para $i = 1, \ldots, r$, denote por

$$q_i = \frac{1}{\sigma_i} A^T v_i.$$

Note que os $q_i$'s são ortonormais. Completando a base e utilizando o processo de ortogonalização de Gram-Schimdt, temos uma base $\{q_1, \ldots, q_r, q_{r+1}, \ldots, q_n\}$ ortornormal de $\mathbb{R}^n$. Sejam agora as matrizes

$$U = \begin{pmatrix} | & & | \\ v_1 & \ldots & v_m \\ | & & | \end{pmatrix}, \qquad V = \begin{pmatrix} | & & | \\ q_1 & \ldots & q_n \\ | & & | \end{pmatrix}.$$

Além disso, de modo análogo ao Lema 0.1, temos que $\mathrm{Ker}(AA^T) = \mathrm{Ker}(A^T)$, basta substituir $A$ por $A^T$. Portanto,

$$(U^TAV)_{ij} = v_i^T A q_j = \begin{cases} \frac{1}{\sigma_i} v_i AA^T v_j & \text{se } j \leq r \\ 0 & \text{se } j > r \end{cases} = \begin{cases} \sigma_j \delta_{ij} & \text{se } j \leq r \\ 0 & \text{se } j > r \end{cases},$$

pois $v_i \in \mathrm{Ker}(A^T)$ e

$$\frac{1}{\sigma_j} v_i^T AA^T v_j = \frac{\lambda_i}{\sigma_j} v_i^T v_j = \sigma_j \delta_{ij}.$$

Portanto, a matriz $U^TAV$ é dada por

$$\begin{array}{c c} & \overbrace{\hphantom{\sigma_1 \quad 0 \quad \cdots \quad 0}}^{\text{m}} \; \overbrace{\hphantom{0 \quad \cdots \quad 0}}^{\text{n-m}} \\ \text{m}\left\{ \vphantom{\begin{matrix} 0\\0\\0\\0\\0 \end{matrix}} \right. & \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & 0 & \cdots & 0 \\ 0 & 0 & \cdots & \sigma_m & 0 & \cdots & 0 \end{bmatrix} \end{array}.$$

□

*Solução.* (Exercício 2)
Suponha $m < n$, então

$$Ax = U\Sigma V^T = \begin{pmatrix} | & & | \\ u_1 & \ldots & u_m \\ | & & | \end{pmatrix} \begin{pmatrix} \sigma_1 & & & & 0 \\ & \sigma_2 & & & 0 \\ & & \ddots & & 0 \\ & & & \sigma_m & 0 \end{pmatrix} \begin{pmatrix} - & v_1^T & - \\ & \vdots & \\ - & v_n^T & - \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

$$= \begin{pmatrix} | & & | \\ u_1 & \ldots & u_m \\ | & & | \end{pmatrix} \begin{pmatrix} \sigma_1 v_1^T x \\ \vdots \\ \sigma_m v_m^T x \end{pmatrix}$$

Seja agora $x = v_i$, então

$$Av_i = \begin{pmatrix} | & & | \\ u_1 & \dots & u_m \\ | & & | \end{pmatrix} \begin{pmatrix} \sigma_1 v_1^T v_i \\ \vdots \\ \sigma_m v_m^T v_i \end{pmatrix}$$

$$= \begin{pmatrix} | & & | \\ u_1 & \dots & u_m \\ | & & | \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ \vdots \\ \sigma_i \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

$$= \begin{pmatrix} \sigma_i u_{i1} \\ \sigma_i u_{i2} \\ \vdots \\ \sigma_i uii \\ \vdots \\ \sigma_i u_{im} \end{pmatrix} = \sigma_i u_i$$

Já para $A^T u_i$, temos que

$$A^T x = V\Sigma^T U^T = \begin{pmatrix} | & & | \\ v_1 & \dots & v_n \\ | & & | \end{pmatrix} \begin{pmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_m \\ 0 & \dots & 0 & 0 \end{pmatrix} \begin{pmatrix} - & u_1^T & - \\ & \vdots & \\ - & u_m^T & - \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix}$$

$$= \begin{pmatrix} | & & | \\ v_1 & \dots & v_n \\ | & & | \end{pmatrix} \begin{pmatrix} \sigma_1 u_1^T x \\ \vdots \\ \sigma_m u_m^T x \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

Portanto, para $x = u_i$

$$A^T u_i = \sigma_i v_i.$$

Para calcular os valores de $\sigma_i$, $u_i$ e $v_i$,

$$A^T A v_i = \sigma_i A^T u_i = \sigma_i^2 v_i.$$

Logo, $v_i$ é o autovetor de $A^T A$ associado ao autovalor $\sigma_i^2$ da matriz $A^T A$. E

$$AA^T u_i = \sigma_i A v_i = \sigma_i^2 u_i.$$

Logo, $u_i$ é o autovetor de $AA^T$ associado ao autovalor $\sigma_i^2$ da matriz $AA^T$. Isto pode ser visto também através da demonstração do SVD pelo exercício 1. $\square$

*Solução.* (Exercício 3)

(a) De fato, como $A = U\Sigma V^T$, e como $U$, $V^T$ são matrizes ortogonais, temos que posto$(U\Sigma V^T) =$

posto($\Sigma$). E como posto($\Sigma$) $= r$, já que

$$\Sigma = \begin{pmatrix} \sigma_1 & & & & & & \\ & \sigma_2 & & & & & \\ & & \ddots & & & & \\ & & & \sigma_r & & & \\ & & & & 0 & & \\ & & & & & \ddots & \\ & & & & & & 0 \\ & & & \vdots & & & \\ 0 & & & & & & 0 \end{pmatrix}$$

$$\Rightarrow \text{posto}(A) = \text{posto}(\Sigma) = r.$$

(b) Seja $x \in \mathbb{R}^n$, então

$$x = b_1 v_1 + \dots b_n v_n,$$

onde $v_1, \dots, v_n$ são os vetores singulares à direita da decomposição SVD da matriz A. Então, pelo exercício 2

$$\begin{aligned} &Ax = A(b_1 v_1 + \dots b_n v_n) = b_1 \sigma_1 u_1 + \dots b_r \sigma_r u_n \\ &\Rightarrow \text{Im}(A) \in \text{span}\{u_1, \dots, u_r\} \end{aligned}$$

Seja $x \in \text{span}\{u_1, \dots, u_r\}$. Pelo exercício 2, $u_i = \frac{1}{\sigma_i} A v_i$, portanto, $x \in \text{Im}(A)$.

$$\text{Im}(A) = \text{span}\{u_1, \dots, u_r\}.$$

(c) Seja $x \in \text{Ker}(A)$. Como $x \in \mathbb{R}^n$, $x = b_1 v_1 + \dots b_n v_n$.

$$A(b_1 v_1 + \dots + b_n v_n) = 0 \Rightarrow b_1 \sigma_1 u_1 + \dots + b_r \sigma_r u_r = 0$$

Como os vetores $u_1, \dots, u_m$ formam uma base para $\mathbb{R}^m$ e $\sigma_1, \dots, \sigma_r \neq 0$, temos que

$$b_1 = \dots = b_r.$$

Portanto, $x \in \text{span}\{v_{r+1}, \dots, v_n\}$. Do mesmo modo, temos que $A v_i = \sigma_i u_i$. Se $i > r$, então $\sigma_i = 0$.

$$A v_i = 0 \Rightarrow v_i \in \text{Ker}(A) \Rightarrow \text{Ker}(A) = \text{span}\{v_{r+1}, \dots, v_n\}.$$

(d) De fato, note que pelos exercícios anteriores, se $m < n$, então

$$\begin{aligned} A = U \Sigma V^T &= \begin{pmatrix} | & & | \\ u_1 & \dots & u_m \\ | & & | \end{pmatrix} \begin{pmatrix} \sigma_1 & & & & 0 \\ & \sigma_2 & & & 0 \\ & & \ddots & & 0 \\ & & & \sigma_m & 0 \end{pmatrix} \begin{pmatrix} - & v_1^T & - \\ & \vdots & \\ - & v_n^T & - \end{pmatrix} \\ &= \begin{pmatrix} | & & | \\ u_1 & \dots & u_m \\ | & & | \end{pmatrix} \begin{pmatrix} \sigma_1 v_1^T x \\ \vdots \\ \sigma_m v_m^T x \end{pmatrix} \\ &= \sum_{i=1}^{m} \sigma_i u_i v_i^T = \sum_{i=1}^{r} \sigma_i u_i v_i^T \end{aligned}$$

□

*Solução.* (Exercício 4)

(a) Esta é a norma de Froebenius de uma matriz A, dada por

$$\|A\|_F^2 = tr(AA^T),$$

onde $tr$ é o operador traço. Além disso, se $U$ for uma matriz ortogonal, então

$$\|UA\|_F^2 = tr(UAA^TU^T) = tr(AA^TU^TU) = tr(AA^T) = \|A\|_F^2 .$$

Ainda mais, isto vale para multiplicação à direita, ou seja

$$\|AU\|_F^2 = tr(AUU^TA^T) = tr(AA^T) = \|A\|_F^2 .$$

Portanto,

$$\|A\|_F^2 = \left\|U\Sigma V^T\right\|_F^2 = \|\Sigma\|_F^2 = \sigma_1 + \cdots + \sigma_p,$$

$p = \min\{m, n\}$.

(b) Note que se $x$ é um vetor cuja norma é igual a um e $V$ é uma matriz ortogonal, então $\|Vx\| = \|x\| = 1$. Portanto,

$$\begin{aligned}
\|A\|_2 &= \sup_{\|x\|=1} \|Ax\| = \sup_{\|x\|=1} \left\|U\Sigma V^T x\right\| \\
&= \sup_{\|x\|=1} \left\|\Sigma V^T x\right\| = \sup_{\|V^T x\|=} \left\|\Sigma V^T x\right\| = \sup_{\|y\|=1} \|\Sigma y\| \\
&= \|\Sigma\|_2
\end{aligned}$$

Suponha que $m < n$, então

$$\Sigma = \begin{pmatrix} \sigma_1 & & & & 0 \\ & \sigma_2 & & & 0 \\ & & \ddots & & 0 \\ & & & \sigma_m & 0 \end{pmatrix}$$

Logo,

$$\|\Sigma\|_2 = \sup_{\|x\|=1} \|\Sigma x\| = \sup_{\|x\|=1} \left|\sqrt{\sigma_1^2 x_1^2 + \cdots + \sigma_m^2 x_m^2}\right| \leq \sup_{\|x\|=1} \sigma_1 \|x\| = \sigma_1,$$

já que $\sigma_1 > 0$. Agora seja $x = (1, 0, \ldots, 0) \in \mathbb{R}^n$. Então, $\|x\|_2 = 1$. Portanto, $\|\Sigma x\|_2 = \sigma_1$. Ou seja, $\|\Sigma\|_2 \geq \sigma_1$.

$$\|A\|_2 = \|\Sigma\|_F = \sigma_1.$$

de modo análogo temos que para $m \geq n$.

(c) Seja $m \geq n$. Então,

$$\Sigma = \begin{pmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_n \\ 0 & 0 & \ldots & 0 \end{pmatrix}$$

Assim,

$$\begin{aligned}
\min_{x\neq 0} \frac{\|Ax\|}{\|x\|} &= \min_{x\neq 0} \|\Sigma x\| \|x\| \\
&= \min_{x\neq 0} \frac{\left|\sqrt{\sigma_1^2 x_1^2 + \cdots + \sigma_n^2 x_n^2}\right|}{\|x\|} \\
&\geq \sigma_n \min_{x\neq 0} \frac{\|x\|}{\|x\|} = \sigma_n.
\end{aligned}$$

Por outro lado, seja $x = (0, 0, \dots, 0, 1) \in \mathbb{R}^m$, então

$$\min_{y \neq 0} \frac{\|Ay\|_2}{\|y\|_2} \leq \frac{\|\Sigma x\|_2}{\|x\|_2} = \sigma_n.$$

$$\Rightarrow \min_{x \neq 0} \frac{\|Ax\|_2}{\|x\|_2} = \sigma_n.$$

□

*Solução.* (Exercício 5)

Note que $\|A - A_k\|_2 = \sigma_{k+1}$, pois

$$\|A - A_k\|_2 = \left\| \sum_{j=k+1}^{n} \sigma_j u_j v_J^T \right\| = \sigma_{k+1}$$

pelo exercício 4-b. Portanto, resta mostrar que

$$\min_{\text{posto}(B)=k} \|A - B\|_2 = \|A - A_k\|.$$

Como $posto(A_k) = k$, já que $k < r$, temos que

$$\min_{\text{posto}(B)=k} \|A - B\|_2 \leq \|A - A_k\|.$$

Resta mostrar que

$$\min_{\text{posto}(B)=k} \|A - B\|_2 \geq \|A - A_k\|.$$

Seja $U\Sigma V^T$ a decomposição SVD da matriz A. Seja $B$ uma matriz de posto igual a $k$. Como $k < r$, temos pelo teorema do núcleo e da imagem que $\dim(\text{Ker}(B)) = n - k$. Além disso, restrinja a matriz V para $V_k$, sendo $V_k$ a matriz das primeiras $k+1$ colunas de V, portanto, $dim(V_k) = k+1$. Logo,

$$dim(\text{Ker}(B)) + dim(V_k) = n - k + k + 1 = n + 1.$$

Como a soma das dimensões é maior que $n$ e ambos espaços estão contidos em $\mathbb{R}^n$, temos que

$$\text{Ker}(B) \cap V_k \neq \emptyset.$$

Tome $x \in \text{Ker}(B) \cap V_k$, tal que $\|x\|_2 = 1$. Então

$$\|A - B\|_2 \geq \|Ax - Bx\|_2 = \|A_{k+1}x\|_2.$$

Como $x \in V_k$, $x = c_1 v_1 + \cdots + c_{k+1} v_{k+1}$, então

$$\begin{aligned} A_{k+1}x &= \sum_{i=1}^{k+1} \sigma_i u_i v_i^T x = \sum_{i=1}^{k+1} \sigma_i u_i v_i^T (c_1 v_1 + \cdots + c_{k+1} v_{k+1}) \\ &= \sum_{i=1}^{k+1} \sigma_i c_i u_i. \end{aligned}$$

Utilizando o resultado acima, temos que

$$\|A_{k+1}x\|_2^2 = \left\| \sum_{i=1}^{k+1} \sigma_i c_i u_i \right\|_2 \geq \sigma_{k+1} \left\| \sum_{i=1}^{k+1} c_i u_i \right\|_2.$$

E note que

$$1 = \|x\|_2^2 = \left\langle \sum_i c_i v_i, \sum_j c_j v_j \right\rangle = \sum_{i=1}^{k+1} c_i^2.$$

Mas como os vetores $u_1, \ldots, u_{k+1}$ são ortonormais, temos que

$$\left\| \sum_{i=1}^{k+1} c_i u_i \right\|_2^2 = \left\langle \sum_i c_i u_i, \sum_j c_j u_j \right\rangle = \sum_{i=1}^{k+1} c_i^2.$$

Portanto, juntando os resultados acima, temos que

$$\|A - B\|_2 \geq \sigma_{k+1}$$

Assim, obtemos o resultado desejado.

$$\min_{\text{posto}(B)=k} \|A - B\|_2 = \|A - A_k\|_2 = \sigma_{k+1}.$$

□

*Solução.* (Exercício 6)

Considere a SVD de $A \in \mathbb{R}^{m \times n}$ dada no Exercício 1. Portanto, pelo exercício 4, queremos mostrar que

$$\sigma_1 = \max_{\substack{y \in \mathbb{R}^m \\ x \in \mathbb{R}^n}} \frac{y^T A x}{\|x\|_2 \|y\|_2}.$$

Suponha que a matriz $A$ tenha posto maior ou igual a 1, então $\sigma_1 \neq 0$.

Agora, considere $x$ como o autovetor unitário associado a $\lambda_1$ da matriz $A^T A$. Seja agora $y = Ax/\sigma_1$, onde $\sigma_1 = \sqrt{\lambda_1}$. Então

$$\|y\|_2^2 = y^T y = \frac{1}{\sigma_1^2} x^T A^T A x = 1.$$

Logo,

$$y^T A x = \frac{x^T A^T A x}{\sigma_1} = \sigma_1.$$

Acabamos de mostrar que

$$\max_{\substack{y \in \mathbb{R}^m \\ x \in \mathbb{R}^n}} \frac{y^T A x}{\|x\|_2 \|y\|_2} \geq \sigma_1.$$

Vamos mostrar que vale a outra desigualdade agora. Seja $r = \text{posto}(A)$, então

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

Como os vetores $v_1, \ldots, v_n$ formam uma base de $\mathbb{R}^n$ e $u_1, \ldots, u_m$ de $\mathbb{R}^m$, para todo $x \in \mathbb{R}^n$ e $y \in \mathbb{R}^m$,

$$x = d_1 v_1 + \ldots d_n v_v, \quad y = c_1 u_1 + \cdots + c_m u_m.$$

Logo,

$$\begin{aligned}
\frac{y^T A x}{\|y\|_2 \|x\|_2} &= \frac{\left(\sum_{i=1}^m c_i u_i\right)^T \left(\sum_{i=1}^r \sigma_i u_i v_i^T\right) \left(\sum_{i=1}^n d_i v_i\right)}{\left(\sum_{i=1}^m c_i^2\right)^{\frac{1}{2}} \left(\sum_{i=1}^n d_i^2\right)^{\frac{1}{2}}} \\
&= \frac{\left(\sum_{i=1}^m c_i u_i\right)^T \left(\sum_{i=1}^r \sigma_i u_i d_i\right)}{\left(\sum_{i=1}^m c_i^2\right)^{\frac{1}{2}} \left(\sum_{i=1}^n d_i^2\right)^{\frac{1}{2}}} \\
&= \frac{\sum_{i=1}^r \sigma_i c_i d_i}{\left(\sum_{i=1}^m c_i^2\right)^{\frac{1}{2}} \left(\sum_{i=1}^n d_i^2\right)^{\frac{1}{2}}} \\
&\leq \sigma_1 \frac{\sum_{i=1}^r c_i d_i}{\left(\sum_{i=1}^m c_i^2\right)^{\frac{1}{2}} \left(\sum_{i=1}^n d_i^2\right)^{\frac{1}{2}}}
\end{aligned} \tag{1}$$

Como $r \leq m, n$, considere os vetores

$$c' = (c_1, \ldots, c_r), \quad d' = (d_1, \ldots, d_r).$$

Além disso, considere $c = (c_1, \ldots, c_m)$ e $d = (d_1, \ldots, d_n)$. Então

$$\sum i = 1^r c_i^2 \leq \sum_{i=1}^{m} c_i^2 \Rightarrow \frac{1}{(\sum i = 1^r c_i^2)^{\frac{1}{2}}} \geq \frac{1}{(\sum i = 1^m c_i^2)^{\frac{1}{2}}}.$$

De modo análogo para $d$ e $d'$

$$\frac{1}{(\sum i = 1^r d_i^2)^{\frac{1}{2}}} \geq \frac{1}{(\sum i = 1^n d_i^2)^{\frac{1}{2}}}.$$

E pela desigualdade de Cauchy-Schwarz, $|\langle c, d \rangle| \leq \|c\|_2 \|d\|_2$. Portanto,

$$\frac{|\langle c, d \rangle|}{\|c\|_2 \|d\|_2} \leq 1.$$

Substituindos os valores e desigualdades acima na Equação (1), temos que

$$\sigma_1 \frac{\sum_{i=1}^{r} c_i d_i}{(\sum_{i=1}^{m} c_i^2)^{\frac{1}{2}} (\sum_{i=1}^{n} d_i^2)^{\frac{1}{2}}} \leq \sigma_1 \frac{\sum_{i=1}^{r} c_i d_i}{(\sum_{i=1}^{r} c_i^2)^{\frac{1}{2}} (\sum_{i=1}^{r} d_i^2)^{\frac{1}{2}}} \leq \sigma_1.$$

Como isto vale para qualqer $x$ e $y$,

$$\sigma_1 = \sigma_{\max}(A) = \max_{\substack{y \in \mathbb{R}^m \\ x \in \mathbb{R}^n}} \frac{y^T A x}{\|x\|_2 \|y\|_2}.$$

□

*Solução.* (Exercício 7)

Vamos usar SVD para provar. Seja $A \in \mathbb{R}^{m \times n}$ uma matriz qualquer, e considere a SVD de $A$ dada no exercício 1. Suponha que $posto(A) = r$, sendo que $r \leq \min m, n$. Suponha que $n \leq m$, a demonstração a seguir pode ser usada analogamente para o caso $n > m$.

Temos

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

Seja agora a sequência de matrizes $B_k$, onde

$$B_k = \sum_{i=1}^{n} \left( \sigma_i + \frac{1}{k} \right) u_i v_i^T.$$

Observe que as matrizes $B_k$ tem posto completo, já que $B_k = U \Sigma_k V^T$, onde $U, V$ são matrizes ortogonais e $\Sigma_k$ possui posto cheio, já que as colunas de $\Sigma_k$ são linearmente independentes.

$$\Sigma_k = \begin{pmatrix} \sigma_1 + \frac{1}{k} & 0 & \cdots & 0 \\ 0 & \sigma_2 + \frac{1}{k} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_n + \frac{1}{k} \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \end{pmatrix}.$$

Além disso, note que $A - B_k = U(\Sigma - \Sigma_k)V^T$, e portanto,

$$\|A - B_k\|_2 = \left\| U(\Sigma - \Sigma_k)V^T \right\|_2 = \|\Sigma - \Sigma_k\|,$$

já que $U$ e $V$ são matrizes ortogonais. Além disso, note que

$$\Sigma - \Sigma_k = \begin{pmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_n \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \end{pmatrix} - \begin{pmatrix} \sigma_1 + \frac{1}{k} & 0 & \cdots & 0 \\ 0 & \sigma_2 + \frac{1}{k} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_n + \frac{1}{k} \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \end{pmatrix} = \begin{pmatrix} \frac{1}{k} & 0 & \cdots & 0 \\ 0 & \frac{1}{k} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{k} \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \end{pmatrix}.$$

Sabemos que o valor singular $\sigma_1$ de uma matrix $A$ é a raíz quadrada do maior autovalor da matriz $A^T A$. Neste caso, o valor singular $\sigma_1$ da matriz $\Sigma - \Sigma_k$ é $\frac{1}{k}$ já que o maior autovalor da matriz $(\Sigma - \Sigma_k)^T(\Sigma - \Sigma_k)$ é $\frac{1}{k^2}$. Pelo exercício 4-b, temos que

$$\|\Sigma - \Sigma_k\|_2 = \sigma_1.$$

Portanto,

$$\|A - B_k\|_2 = \|\Sigma - \Sigma_k\|_2 = \sigma_1 = \frac{1}{k}.$$

Ou seja, para todo $\epsilon > 0$, existe $k_0$ tal que para todo $k > k_0$,

$$\|A - B_k\|_2 < \epsilon.$$

Assim, mostramos que qualquer matriz em $\mathbb{R}^{m\times n}$ é o limite de uma sequência de matrizes de posto completo. □

*Solução.* (Exercício 8)

Seja $\Omega = \{x_1, \ldots, x_m\} \subset \mathbb{R}^n$ um conjunto de dados e $k \in (1, m)$ um número natural. O Algoritmo de Lloyd ($k$-médias) para clusterização consiste em achar centros, ou pontos, de modo que o conjunto de treinamento possa ser separado em $k$ partições, todas disjuntas. Ou seja, o objetivo é encontrar uma partição ótima

$$S = \{S_1, \ldots, S_m\}$$

de $\Omega$, isto é, encontrar $S_i$, chamados de *clusters*, tais que

$$S_i \cap S_j = \emptyset, \quad i \neq j$$
$$\bigcup_{i=1}^{k} S_i = \Omega.$$

Em termos de minimização, queremos resolver o seguinte problema de minização.

$$\min_{\substack{S \text{ partição} \\ \text{de } \Omega}} \sum_{i=1}^{k} \sum_{x_j \in S_i} \|x_j - \mu_i\|_2^2, \quad \text{onde } \mu_i = \frac{1}{|S_i|} \sum_{x_j \in S_i} x_j.$$

Então, o Algoritmo de Lloyd é dado a seguir.

□

*Solução.* (Exercício 9)

Seja um banco de dados $\{I_1, \ldots, I_m\}$, onde cada $I_i$ é uma matriz de imagens em escala cinza. Temos $m$ imagens, de tamanho $n \times n$, dividas em $p$ classes, onde cada classe representa a face de uma pessoa. Para cada matriz $I_j$, denote por $\Phi_j$ o vetor de concatenação das matrizes. Seja $A$ a matriz abaixo.

$$A = \begin{pmatrix} | & | & & | \\ \Phi_1 & \Phi_2 & \cdots & \Phi_m \\ | & | & & | \end{pmatrix} \in \mathbb{R}^{n^2 \times m}.$$

Para classificar uma nova imagem $I_{new} = \Phi_{new}$, calculamos a projeção ortogonal no espaço gerado pelas colunas da matriz $A$, assim podemos dizer de qual vetor coluna da matriz $A$, a imagem $\Phi_{new}$ está mais perto, e esta será a sua classe. Portanto, o procedimento é dado no Algoritmo abaixo.

□

**Algoritmo 1** Algoritmo de Lloyd

1: Escolha $k$ pontos $m_1^{(1)}, \ldots, m_k^{(1)}$ em $\Omega$, os quais serão usados como "médias iniciais". Defina $t = 1$.

2: Para cada $j = 1, \ldots, m$, encontre

$$i \in \underset{l=1,\ldots,k}{\operatorname{argmin}} \left\| x_j - m_l^{(t)} \right\|$$

e associe $x_j$ ao cluster $S_i$.

3: Se $t > 1$ e $S^t = S^{t-1}$, pare.

4: (Atualização das médias) Para cada $i = 1, \ldots, k$, defina

$$m_i^{t+1} = \frac{1}{|S_i|} \sum_{x_j \in S_i^{(t)}} x_j$$

5: Defina $t = t + 1$ e volte para o passo 2.

**Algoritmo 2** Auto-faces para reconhecimendo facial

1: Calcule a projeção ortogonal $\hat{\Phi}_{new}$ de $\Phi_{new}$ sobre o subespaço gerado por $\Phi_1, \ldots, \Phi_m$.

$$\hat{\Phi}_{new} = A(A^T A)^{-1} \Phi_{new}$$

2: Calcule

$$j = \underset{i=1,\ldots,m}{\operatorname{argmin}} \left\| \hat{\Phi}_{new} - \Phi_i \right\|$$

3: Dizemos então que a face em $I_{new}$ é a face da pessoa em $I_j$.