Recall: Given two matrices $A=(A_{ij}),\ B=(B_{ij}) \in \mathbb{R}^{m\times n}$, their <u>Frobenius inner product</u> is

$$\langle A, B\rangle_F \overset{\text{def}}{=} \operatorname{tr}(A^T B) = \operatorname{tr}(B^T A) = \sum_{i=1}^{m}\sum_{j=1}^{n} A_{ij} B_{ij}$$

(because
$$A^T B = \begin{pmatrix} A_{11} & A_{21} & \cdots & A_{m1} \\ A_{12} & A_{22} & \cdots & A_{m2} \\ \vdots & \vdots & & \vdots \\ A_{1n} & A_{2n} & \cdots & A_{mn} \end{pmatrix} \begin{pmatrix} B_{11} & B_{12} & \cdots & B_{1n} \\ B_{21} & B_{22} & \cdots & B_{2n} \\ \vdots & & & \vdots \\ B_{m1} & B_{m2} & \cdots & B_{mn} \end{pmatrix}$$

$$= \begin{pmatrix} A_{11}B_{11} + A_{21}B_{21} + \cdots + A_{m1}B_{m1} & & & * \\ & A_{12}B_{12} + A_{22}B_{22} + \cdots + A_{m2}B_{m2} & & \\ & & \ddots & \\ * & & & A_{1n}B_{1n} + A_{2n}B_{2n} + \cdots + A_{mn}B_{mn} \end{pmatrix}$$)

The induced <u>Frobenius norm</u> of A is

$$\|A\|_F^2 \overset{\text{def}}{=} \operatorname{tr}(A^T A) = \sum_{i=1}^{m}\sum_{j=1}^{n} |A_{ij}|^2$$

Remark: If we identify $A=(A_{ij}) \in \mathbb{R}^{m\times n}$,

with the vector

$$\vec{A} \overset{\text{def}}{=} \begin{pmatrix} A_{11} \\ A_{21} \\ \vdots \\ A_{m1} \\ A_{12} \\ A_{22} \\ \vdots \\ A_{m2} \\ \vdots \end{pmatrix} \begin{matrix} \left.\vphantom{\begin{matrix}A\\A\\A\\A\end{matrix}}\right\} \text{first column of } A \\ \left.\vphantom{\begin{matrix}A\\A\\A\\A\end{matrix}}\right\} \text{second column of } A \\ \} \text{ third column of } A \end{matrix} \in \mathbb{R}^{mn}$$

then $(A, B)_F = \vec{A} \cdot \vec{B}$ and $\|A\|_F^2 = \|\vec{A}\|^2$.

Prop: Let $A \in \mathbb{R}^{m \times n}$.

(1) If $U \in \mathbb{R}^{m \times m}$ is orthogonal, then

$$\|UA\|_F = \|A\|_F$$

(2) If $U \in \mathbb{R}^{n \times n}$ is orthogonal, then

$$\|AU\|_F = \|A\|_F.$$

In other words, multiplication by an orthogonal matrix preserves the Frobenius norm.

Proof: (1) $\|UA\|_F^2 = tr((UA)^T(UA))$

$= \mathrm{tr}(A^T \underbrace{U^T U}_{=I} A) = \mathrm{tr}(A^T A) = \|A\|_F^2$.

(2) similar. ■

Prop: Let $A \in \mathbb{R}^{m \times n}$ be a matrix with singular values $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$ ($r = \mathrm{Rank}(A)$), then

$$\|A\|_F^2 = \sigma_1^2 + \cdots + \sigma_r^2.$$

Proof: Let $A = U \Sigma V^T$ be the SVD with orthogonal U and V (hence V^T)

then
$$\|A\|_F^2 = \|U \Sigma V^T\|_F^2 \overset{\text{Prop (1)}}{=} \|\Sigma V^T\|_F^2$$
$$\overset{\text{Prop (2)}}{=} \|\Sigma\|_F^2 \overset{\text{def of } \|\cdot\|_F}{=} \sigma_1^2 + \cdots + \sigma_r^2$$ ■

Prop: (Von-Neumann Inequality)

For any $A, B \in \mathbb{R}^{m \times n}$ with singular values

$\sigma_1(A) \geq \sigma_2(A) \geq \cdots > 0$ and $\sigma_1(B) \geq \sigma_2(B) \geq \cdots > 0$.

Then

$$(A, B)_F \leq \sum_i \sigma_i(A)\sigma_i(B)$$

and "=" holds if and only if A. B have identical orthogonal matrices in their SVDs

i.e. $A = U\Sigma_A V^T$, $B = U\Sigma_B V^T$.

3.B.2 Theory:

We look for

$$\arg\min_A \|X - A\|_F \quad \text{s.t. } \operatorname{Rank}(A) = d.$$

Let $X = U_1\Sigma_1 V_1^T$ and $A = U_2\Sigma_2 V_2^T$ be the SVDs with orthogonal U_1, U_2, V_1, V_2.

$$\|X - A\|_F^2 = \|U_1\Sigma_1 V_1^T - U_2\Sigma_2 V_2^T\|_F^2$$

$$\overset{\substack{\text{multi by } U_1^T \text{ from left} \\ \text{multi by } V_1 \text{ from right}}}{=} \|\Sigma_1 - \underbrace{U_1^T U_2}_{\stackrel{\text{def}}{=} U} \Sigma_2 \underbrace{V_2^T V_1}_{\stackrel{\text{def}}{=} V^T}\|_F^2$$

$$U = U_1^T U_2 \qquad \| \Sigma_1 - U \Sigma_2 V^T \|_F^2$$
$$V = V_2^T V_1$$

Frobenius inner product $\left(\Sigma_1 - U\Sigma_2 V^T, \Sigma_1 - U\Sigma_2 V^T\right)_F$

$$= (\Sigma_1, \Sigma_1)_F - (\Sigma_1, U\Sigma_2 V^T)_F$$
$$- (U\Sigma_2 V^T, \Sigma_1)_F + (U\Sigma_2 V^T, U\Sigma_2 V^T)_F$$

$$= \|\Sigma_1\|_F^2 - 2(\Sigma_1, U\Sigma_2 V^T)_F + \|U\Sigma_2 V^T\|_F^2 \qquad (*)$$

To minimize $\|X-A\|_F^2$, it suffices to maximize $(\Sigma_1, U\Sigma_2 V^T)_F$.

By Von-Neumann's inequality:

$$(\Sigma_1, U\Sigma_2 V^T)_F \le \sum_i \sigma_i(\Sigma_1)\, \sigma_i(U\Sigma_2 V^T)$$
$$= \sum_i \sigma_i(X)\, \sigma_i(A)$$

and "=" holds if and only if $U = I_{m\times m}$ and $V = I_{n\times n}$.

Thus

$$\|X - A\|_F^2 \overset{(*)}{=} \|\Sigma_1\|_F^2 - 2(\Sigma_1, U\Sigma_2 V^T)_F + \|U\Sigma_2 V^T\|_F^2$$

$$\geq \|\Sigma_1\|_F^2 - 2 \sum_i \sigma_i(X)\sigma_i(A) + \|\Sigma_2\|_F^2$$

$$= \|\Sigma_1\|_F^2 - 2\sum_i \sigma_i(X)\sigma_i(A) + \sum_i \sigma_i^2(A)$$