Queen Mary
University of London

MTH793P **Advanced Machine Learning, Semester B, 2023/24**
Coursework 9

In this coursework we will prove a few statements that we used in the lecture.

Robust PCA

1. Let $X \in \mathbb{R}^{m\times n}$, and consider its SVD: $X = U \cdot \Sigma \cdot V^T$. Let $A, B \in \mathbb{R}^{m\times n}$ be two different matrices. Prove that $B = U^T \cdot A \cdot V$ **if and only if** $A = U \cdot B \cdot V^T$.
2. Let $X \in \mathbb{R}^{m\times n}$, and suppose that $\sigma_1, \ldots, \sigma_r$ are the singular values of X ($r = \min(m, n)$). Let $U \in \mathbb{R}^{m\times m}$ and $V \in \mathbb{R}^{n\times n}$ be two orthogonal matrices. Define $\tilde{X} = U \cdot X \cdot V^T$. Prove that X and $\tilde{X}$ have the same singular values.
3. Let $X \in \mathrm{R}^{m\times n}$, and recall the definition of the singular thresholding operator

$$D_\tau(X) = US_\tau(\Sigma)V^T,$$

where $U\Sigma V^T$ is the SVD decomposition of X. Prove that:

   (a) $\|D_\tau(X)\|_* \le \|X\|_*$, where $\|\cdot\|_*$ is the nuclear norm.

   (b) $\mathrm{rank}(D_\tau(X)) \le \mathrm{rank}(X)$.

Under what conditions will we have $\|D_\tau(X)\|_* = \|X\|_*$, $\mathrm{rank}(D_\tau(X)) = \mathrm{rank}(X)$?

Matrix Completion

Let $M \in \mathrm{R}^{m\times n}$. Recall that Ω represents the indexes of known values in M, and $P_\Omega(\cdot)$ is the projection on these locations.

4. Let for any $X, Y \in \mathbb{R}^{m\times n}$ show that $\langle X, P_\Omega(Y)\rangle = \langle P_\Omega(X), Y\rangle$.

Solution

1. By definition, we know that $U \in \mathbb{R}^{m\times m}$ satisfies $U^T U = I_{m\times m}$, and $V \in \mathbb{R}^{n\times n}$ satisfies $V^T V = I_{n\times n}$. This implies that $U^{-1} = U^T$ and $V^{-1} = V^T$.

   Suppose that $B = U^T \cdot A \cdot V$, then

   $$U \cdot B \cdot V^T = U \cdot U^T \cdot A \cdot V \cdot V^T = I_{m\times m} \cdot A \cdot I_{n\times n} = A.$$

   Similarly, if $A = U \cdot B \cdot V^T$, then

   $$U^T \cdot A \cdot V = A = U^T \cdot U \cdot B \cdot V^T \cdot V = B.$$

2. We take the SVD $X = U_X \cdot \Sigma_X V_X^T$, then the diagonal of Σ_X consists of $\sigma_1, \ldots, \sigma_r$. Now

   $$\tilde{X} = U \cdot X \cdot V^T = (U \cdot U_X) \cdot \Sigma_X \cdot (V \cdot V_X)^T.$$

   Note that $\tilde{U} = U \cdot U_X \in \mathbb{R}^{m\times m}$ statisfies

   $$\tilde{U}^T \tilde{U} = (U \cdot U_X)^T \cdot (U \cdot U_X) = U_X^T \cdot (U^T \cdot U) \cdot U_X = U_X^T \cdot U_X = I_{m\times m}.$$

   Similarly, $\tilde{V}^T \cdot \tilde{V} = I_{n\times n}$. Therefore,

   $$\tilde{X} = (U \cdot U_X) \cdot \Sigma_X \cdot (V \cdot V_X)^T$$

   is the SVD decomposition of $\tilde{X}$, from which we conclude that the singular values are $\sigma_1, \ldots, \sigma_r$.

3. (a) Suppose that the singular values of X are $\sigma_1, \ldots, \sigma_r$. In that case the singular values of $D_\tau(X)$ are $S_\tau(\sigma_1), \ldots S_\tau(\sigma_r)$. By the definition of S_τ, and since the singular values are non-negative, we know that $S_\tau(\sigma_i) \leq \sigma_i$. Therefore

   $$\|D_\tau(X)\|_* = \sum_{i=1}^{r} S_\tau(\sigma_i) \leq \sum_{i=1}^{r} \sigma_i = \|X\|_*.$$

   We have equality if and only if $S_\tau(\sigma_i) = \sigma_i$ for all i. This can happen only if $\sigma_i = 0$. In other words, this is true only if $X = 0$.

   (b) Suppose that $\sigma_1 \geq \ldots \geq \sigma_r > 0$, $r = \min(m, n)$ are the singular values of X. If $\text{rank}(X) = d$ then $\sigma_i = 0$ for all $i > d$. The new singular values are given by $\hat{\sigma}_i = S_\tau(\sigma_i)$. Therefore, we have that $\hat{\sigma}_i = 0$ for all $i > d$ implying that $\text{rank}(D_\tau(X)) \leq d$. To get equality, we need $\hat{\sigma}_d > 0$, which happens if and only if $\sigma_d > \tau$.

4. Recall that for any $X, Y \in \mathbb{R}^{m\times n}$ we have

   $$\langle X, Y \rangle = \text{Tr}(X^T Y) = \sum_{i,j} X_{ij} Y_{ij}.$$

   Now

   $$\langle X, P_\Omega(Y) \rangle = \sum_{(i,j)\in\Omega} X_{ij} Y_{ij} = \langle P_\Omega(X), Y \rangle$$