EPFL | MGT-418 : Convex Optimization | Project 5

Answers – Fall 2019

Recommender Systems

Answer to Question 1.1

By the singular value decomposition theorem, we can write any $X \in \mathbb{R}^{m\times n}$ as $X = U\Sigma V^\top$, where $U \in \mathbb{R}^{m\times m}$ and $V \in \mathbb{R}^{n\times n}$ are orthogonal matrices (*i.e.*, $U^\top U = UU^\top = \mathbb{I}_{m\times m}$, $V^\top V = VV^\top = \mathbb{I}_{n\times n}$), and $\Sigma \in \mathbb{R}^{m\times n}$ is a non-square diagonal matrix whose diagonal elements are the singular values of X.

Building on the singular value decomposition of X, consider the candidate $\hat{Y} = U\mathbb{I}_{m\times n}V^\top$, where $\mathbb{I}_{m\times n}$ denotes the m-by-n identity matrix, that is, the m-by-n diagonal matrix whose diagonal elements are all ones. By construction, all singular values of $\hat{Y}$ are equal to one, and so $\|\hat{Y}\| = \sigma_{\max}(\hat{Y}) = 1$, which shows that indeed $\|\hat{Y}\| \le 1$. Recalling that $\mathrm{tr}(AB) = \mathrm{tr}(BA)$ for any pair of matrices A and B whenever both products are well-defined, and that the trace inner product of $A, B \in \mathbb{R}^{m\times n}$ evaluates to $\mathrm{tr}(A^\top B) = \sum_{i=1}^m \sum_{j=1}^n A_{ij}B_{ij}$, a direct calculation further asserts that

$$\mathrm{tr}(X^\top \hat{Y}) = \mathrm{tr}(V\Sigma^\top U^\top U\mathbb{I}_{m\times n}V^\top) = \mathrm{tr}(\Sigma^\top \underbrace{U^\top U}_{\mathbb{I}_{m\times m}} \mathbb{I}_{m\times n} \underbrace{V^\top V}_{\mathbb{I}_{n\times n}}) = \mathrm{tr}(\Sigma^\top \mathbb{I}_{m\times n}) = \textstyle\sum_{i=1}^r \sigma_i(X) = \|X\|_*.$$

Answer to Question 1.2

Consider again the singular value decomposition $X = U\Sigma V^\top$. Recalling again that $\mathrm{tr}(AB) = \mathrm{tr}(BA)$ for any pair of matrices A and B whenever both products are well-defined, and that the trace inner product of $A, B \in \mathbb{R}^{m\times n}$ evaluates to $\mathrm{tr}(A^\top B) = \sum_{i=1}^m \sum_{j=1}^n A_{ij}B_{ij}$, we can conclude that

$$\mathrm{tr}(X^\top Y) = \mathrm{tr}(V\Sigma^\top U^\top Y) = \mathrm{tr}(\Sigma^\top U^\top YV) = \textstyle\sum_{i=1}^r \sigma_i(X)(U^\top YV)_{ii} = \sum_{i=1}^r \sigma_i(X)(u_i^\top Yv_i),$$

where u_i and v_i denote the i-th column of U and V, respectively. Observe that, since U and V are orthogonal matrices, their rows and columns have unit Euclidean length, and thus in particular it holds that $\|u_i\|_2 = \|v_i\|_2 = 1$ for all $i = 1, \dots, r$. From the hint, we thus know that for all $i = 1, \dots, r$

$$u_i^\top Yv_i \le \sup_{\|v\|_2=1} \sup_{\|u\|_2=1} u^\top Yv = \|Y\|.$$

Since singular values are non-negative and because $\|Y\| \le 1$, the last observation allows us to upper bound the expression developed before and assert that

$$\mathrm{tr}(X^\top Y) = \textstyle\sum_{i=1}^r \sigma_i(X)(u_i^\top Yv_i) \le \sum_{i=1}^r \sigma_i(X)\|Y\| \le \sum_{i=1}^r \sigma_i(X) = \|X\|_*.$$

Answer to Question 1.3

Using the recalled facts and keeping in mind that norms are non-negative, we indeed obtain that

$$\|Y\| \le 1 \;\Leftrightarrow\; \|Y\|^2 \le 1 \;\Leftrightarrow\; \lambda_{\max}(Y^\top Y) \le 1 \;\Leftrightarrow\; Y^\top Y \preceq \mathbb{I} \;\Leftrightarrow\; 0 \preceq \mathbb{I} - Y^\top \mathbb{I} Y \;\Leftrightarrow\; 0 \preceq \begin{bmatrix} \mathbb{I} & Y^\top \\ Y & \mathbb{I} \end{bmatrix}.$$

Answer to Question 2.1

The matrix of all zeros is a solution of problem (3) that satisfies Slater's condition. Substituting $\bar{Y} = 0$ into the semidefinite constraint of problem (3) yields the identity matrix. The latter is positive definite (as all of its eigenvalues are equal to one) and thus strictly feasible in the semidefinite constraint.

Answer to Question 2.2

There are several ways to determine the Lagrangian dual problem of problem (3). Here, we directly dualize the given maximization problem.

Consider the matrices $\Lambda_1 \in \mathbb{S}^n$, $\Lambda_2 \in \mathbb{S}^m$ and $\Lambda_3 \in \mathbb{R}^{m\times n}$ in order to associate a multiplier matrix

$$\begin{bmatrix} \Lambda_1 & \Lambda_3^\top \\ \Lambda_3 & \Lambda_2 \end{bmatrix} \succeq 0$$

with the semidefinite constraint of problem (3). With this, the Lagrangian takes the form

$$\begin{aligned} L(Y, \Lambda_1, \Lambda_2, \Lambda_3) &= \operatorname{tr}(X^\top Y) + \operatorname{tr}\left(\begin{bmatrix} \Lambda_1 & \Lambda_3^\top \\ \Lambda_3 & \Lambda_2 \end{bmatrix}^\top \begin{bmatrix} \mathbb{I} & Y^\top \\ Y & \mathbb{I} \end{bmatrix} \right) \\ &= \operatorname{tr}(X^\top Y) + \operatorname{tr}\left(\begin{bmatrix} \Lambda_1 + \Lambda_3^\top Y & \Lambda_1 Y^\top + \Lambda_3^\top \\ \Lambda_3 + \Lambda_2 Y & \Lambda_3 Y^\top + \Lambda_2 \end{bmatrix} \right) \\ &= \operatorname{tr}(X^\top Y) + \operatorname{tr}(\Lambda_1 + \Lambda_3^\top Y) + \operatorname{tr}(\Lambda_3 Y^\top + \Lambda_2) \\ &= \operatorname{tr}(X^\top Y) + \operatorname{tr}(\Lambda_3^\top Y) + \operatorname{tr}(\Lambda_3 Y^\top) + \operatorname{tr}(\Lambda_1) + \operatorname{tr}(\Lambda_2) \\ &= \operatorname{tr}(X^\top Y) + 2\operatorname{tr}(\Lambda_3^\top Y) + \operatorname{tr}(\Lambda_1) + \operatorname{tr}(\Lambda_2) \\ &= \operatorname{tr}((X + 2\Lambda_3)^\top Y) + \operatorname{tr}(\Lambda_1) + \operatorname{tr}(\Lambda_2). \end{aligned}$$

Observe that $L(Y, \Lambda_1, \Lambda_2, \Lambda_3)$ is a linear function of Y. Thus, the dual objective function evaluates to

$$g(\Lambda_1, \Lambda_2, \Lambda_3) \;=\; \sup_{Y \in \mathbb{R}^{m\times n}} L(Y, \Lambda_1, \Lambda_2, \Lambda_3) \;=\; \begin{cases} \operatorname{tr}(\Lambda_1) + \operatorname{tr}(\Lambda_2) & \text{if } X + 2\Lambda_3 = 0 \\ +\infty & \text{otherwise.} \end{cases}$$

Using the implicit constraint $X + 2\Lambda_3 = 0 \Leftrightarrow \Lambda_3 = -\frac{1}{2}X$, the Lagrangian dual problem finally reads

$$\begin{aligned} \underset{\Lambda_1 \in \mathbb{S}^n, \Lambda_2 \in \mathbb{S}^m}{\text{minimize}} \quad & \operatorname{tr}(\Lambda_1) + \operatorname{tr}(\Lambda_2) \\ \text{subject to} \quad & \begin{bmatrix} \Lambda_1 & -\frac{1}{2}X^\top \\ -\frac{1}{2}X & \Lambda_2 \end{bmatrix} \succeq 0. \end{aligned}$$

Answer to Questions 3.1 and 3.2

An implementation of problem (5) enriched by the given range constraints can be found in the Matlab script `p5a3.m`. On a 3.5 GHz i7 computer with 16GB RAM, a single run of the script for the datasets `20c50m.mat` and `50c200m.mat` takes 5 seconds and 40 minutes, respectively. The results can be found below. The performance is more pronounced for the second dataset. Note that observing only 2.5% of the ratings, the low-rank SDP approach is able to recover about 30% of the ratings correctly and over 70% of the ratings within a precision of plus-minus one.

Dataset	f_{obs} [%]	$f_{\pm 0}$ [%]	$f_{\pm 1}$ [%]	$f_{\pm 2}$ [%]
20c50m	10.0	35.0	75.5	93.6
50c200m	2.5	30.4	73.6	93.3