SI231 - Matrix Computations, 2021 Fall
Solution of Homework Set #5
Prof. Yue Qiu

Acknowledgements:

1) Deadline: **2021-12-18 23:59:59**

2) **Late Policy details** can be found on piazza.

3) Submit your pdf homework in **Homework 5** on **Gradscope**. Make sure that you have correctly selected pages for each problem. If not, you probably will get 0 point.

4) No handwritten homework is accepted. You need to write LaTeX.

5) Use the given template and give your solution in English. Solution in Chinese is not allowed.

6) Your homework should be uploaded in the PDF format, and the naming format of the file is not specified.

7) For the calculation problems, you are highly required to write down your solution procedures in detail. And **all values must be represented by integers, fractions or square root**, floating points are not accepted.

8) When handing in your homework in gradescope, package all your codes into **your_student_id+hw5_code.zip**(must be zip) and upload it in **Homework 5 programming part**. In the package, you also need to include a file named README.txt/md to clearly identify the function of each file. Make sure that your codes can run and are consistent with your homework.

I. SINGULAR VALUE DECOMPOSITION

Problem 1. (10 + 10 points)

1) We have n × n real matrix $\mathbf{B}$, prove that $\mathbf{B}\mathbf{B}^{\dagger}$ is the orthogonal projection onto $\mathcal{R}(\mathbf{B})$ using singular value decomposition.

2) We have n × n real matrix $\mathbf{B}$, prove $\|\mathbf{B}\|_F$ equal to $\|\Sigma\|_F$ with $\mathbf{B} = \mathbf{U}\Sigma\mathbf{V}^T$. (Hint: $\|\mathbf{B}\|_F = \sqrt{\Sigma_{i=1}^n \Sigma_{j=1}^n \mathbf{b}_{i,j}^2}$)

Solution:

Problem 2. (10 + 10 points)

1) Assume the singular value decomposition of matrix $\mathbf{A}$ is $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H$. Solve the singular value decomposition of matrix $\begin{bmatrix} \mathbf{A}_R & -\mathbf{A}_I \\ \mathbf{A}_I & \mathbf{A}_R \end{bmatrix}$ via $\mathbf{U}_R$, $\mathbf{U}_I$, $\mathbf{V}_R$, $\mathbf{V}_I$.

   (**Hint**: $\mathbf{A}_R$ = Re[$\mathbf{A}$], $\mathbf{U}_R$ = Re[$\mathbf{U}$] and $\mathbf{V}_R$ = Re[$\mathbf{V}$]. $\mathbf{A}_I$ = Im[$\mathbf{A}$], $\mathbf{U}_I$ = Im[$\mathbf{U}$] and $\mathbf{A}_I$ = Im[$\mathbf{A}$].)

2) Consider a 3×2 matrix $\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$, compute the singular value decomposition of matrix $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$.

Solution:

Problem 3. (2 + 3 points) Consider two linear systems $\mathbf{A}\mathbf{x} = \mathbf{0}$ and $\mathbf{B}\mathbf{x} = \mathbf{0}$. Give $\mathbf{A} = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}$ and

$\mathbf{B} = \begin{bmatrix} 1 & -1 & 1000000 & 1000000 \\ 0 & 0 & 1 & 1 \end{bmatrix}$

1) Use **Matlab** or **Python** to calculate the 2 norm condition number of $\mathbf{A}$ and $\mathbf{B}$.

2) If there is a small perturbation that makes $\mathbf{A}\mathbf{x} = \begin{bmatrix} 0 \\ 0.01 \end{bmatrix}$ and $\mathbf{B}\mathbf{x} = \begin{bmatrix} 0 \\ 0.01 \end{bmatrix}$. Please illustrate the influence of this small perturbation on the two linear systems respectively through analyzing the form of matrix $\mathbf{A}$ and $\mathbf{B}$.

Solution:

II. LOW RANK APPROXIMATION

Problem 1. (10 + 7 + 10 points)

In this problem, we will learn how to use SVD to compress, or reduce the dimension of the data. The real data is with low rank internal, but often corrupted, or contaminated by noise, which leads to the full rank of the data matrix $\mathbf{A} \in \mathbb{R}^{D \times N}$. To reduce the storage consumption of the real data (note that D is extremely large in practical), we seek to find a low rank approximation to $\mathbf{A}$, that is, we want to solve the problem

$$\min_{\mathbf{X} \in \mathbb{R}^{D \times N}} \quad \|\mathbf{X} - \mathbf{A}\|_F^2 \tag{1}$$

$$\text{s.t.} \quad \text{rank}(\mathbf{X}) \leq d \tag{2}$$

where $\| \cdot \|_F$ is the Frobenius norm, $d << D$ is unknown. By the Theorem of PCA via SVD, the above problem is equivalent to the following problem (suppose the singular values of $\mathbf{A}$ are given by $\sigma_1(\mathbf{A}) \geq \sigma_2(\mathbf{A}) \geq \cdots \geq \sigma_K(\mathbf{A}) \geq 0$ and $K = \min\{D, N\}$):

$$\min_{d=1,\ldots,K} \quad \sum_{i=d+1}^{K} \sigma_i^2(\mathbf{A}) \tag{3}$$

However, this is not a good criterion, since the optimal solution is given by $d^* = \text{rank}(\mathbf{A})$.

The problem of determining the optimal dimension d is in fact a "model selection" problem, which is due to the fact that choice of d balances the complexity of the model and the storage of the data. There are many model selection criterion, we will learn two of them.

1) The first way is to bound the residual of approximation, that is, suppose $\hat{\mathbf{X}}$ is the best $\text{rank} - d$ approximation of $\mathbf{A}$, given a threshold $\tau > 0$, we want to minimize the residual with respect to Frobinius-norm, that is

$$\min_{d=1,\ldots,K} \quad \|\mathbf{A} - \hat{\mathbf{X}}\|_F^2 \leq \tau \tag{4}$$

the problem can be explicitly written as

$$d^* = \min_{d=1,\ldots,K} \left\{ d \,\middle|\, \sum_{i=d+1}^{K} \sigma_i^2(\mathbf{A}) \leq \tau \right\} \tag{5}$$

2) Note that the first criterion (5) depends on specific problem (singular values of data matrix are not invariant with respect to linear transformations), it is hard to chose a reasonable threshold. To generalize our criterion, we consider the normalized version of (5), this new criterion, *a.k.a* **variance explained ratio** in machine learning, is given by

$$d^* = \min_{d=1,\ldots,K} \left\{ d \,\middle|\, \frac{\sum_{i=d+1}^{K} \sigma_i^2(\mathbf{A})}{\sum_{i=1}^{K} \sigma_i^2(\mathbf{A})} \leq \tau \right\} \tag{6}$$

Now, suppose the data matrix $\mathbf{A} \in \mathbb{R}^{D \times N}$ is data1/data1.mat. You are required to

1) (5 + 5 points)Let $rank(\mathbf{A}) = r$ and $\mathbf{A}_k = \sum_{i=1}^{k} \sigma_i \mathbf{u}_i \mathbf{v}_i^T$ be the sum truncated after k terms ($k < r$). Show that $\|\mathbf{A} - \mathbf{A}_k\|_2^2 = \sigma_{r+1}^2$ and $\|\mathbf{A} - \mathbf{A}_k\|_F^2 = \sum_{i=k+1}^{r} \sigma_i^2$

2) (5 + 2 points) Plot the squared singular values of $\mathbf{A}$ along with the threshold $\tau = 150$, what is the solution to (5)?

3) (5 + 5 points) Plot the figure of function

$$f(d) = \frac{\sum_{i=d+1}^{K} \sigma_i^2(\mathbf{A})}{\sum_{i=1}^{K} \sigma_i^2(\mathbf{A})}, \quad d = 1, \ldots, K \tag{7}$$

and fill the following table from the figure with respect to (6):

Table I

THE COMPRESSION RATE WITH RESPECT DIFFERENT THRESHOLD

τ	0.1	0.05	0.02	0.005
Compression rate				

where the compression rate is defined as

$$\text{compression rate} = \frac{\#\{\text{entries in } \hat{\mathbf{X}}\}}{\#\{\text{entries in } \mathbf{A}\}} = \frac{d^*(D + N + 1)}{DN} \tag{8}$$

Remarks

1) Please **insert your figures in your PDF**.

2) You can just mark the solution of problem 2) on your figure (**plot a vertical line with red color or mark the solution with marker**).

3) Please draw a continuous curve for problem 3), see *stairs* function in Matlab, *plt.step* function in Python for more details.

4) For simplicity, we omit some proof details, you may refer to *Generalized Principal Component Analysis, Section 2.1* for more details.

Solution:

III. REGULARIZED LEAST SQUARES

Problem 1. (11 + 11 + 6 points) Denote a rank-r matrix as $\mathbf{A} \in \mathbb{R}^{m \times n} (m \geq n > r)$, and the SVD of it as $\mathbf{U} \begin{bmatrix} \mathbf{\Sigma} \\ \mathbf{0} \end{bmatrix} \mathbf{V}^T$ with singluar values $\sigma_1 \geq \sigma_2 \geq ... \geq \sigma_r > 0$, where $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n] \in \mathbb{R}^{m \times n}$, $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n] \in \mathbb{R}^{n \times n}$ and $\mathbf{\Sigma} = \text{diag}(\sigma_1, \sigma_2, ...\sigma_r, 0, ..., 0) \in \mathbb{R}^{n \times n}$. Consider the Least Squares Problem and Regularized Least Squares Problem as follows

$$\min_{\mathbf{x}} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2, \tag{9}$$

$$\min_{\mathbf{x}} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 + \lambda \|\mathbf{x}\|_2^2, \tag{10}$$

where $\lambda > 0$.

1) Show the Least Squares solution of (9) is $\sum_{i=1}^{r} \frac{\mathbf{u}_i^T \mathbf{b}}{\sigma_i} \mathbf{v}_i$.
2) Show the Regularized Least Squares solution of (10) is $\sum_{i=1}^{r} \frac{\sigma_i(\mathbf{u}_i^T \mathbf{b})}{\sigma_i^2 + \lambda} \mathbf{v}_i$.
3) Observe the solution of (9) and (10), and analysis how does "Regularization" work? In particular, what might happen when $\sigma_i \ll \lambda$ and $\sigma_i \gg \lambda$?

Solution: