Optimization Models: Homework #4

Due on October 14, 2025 at 11:59pm

Zachary Brandt
zbrandt@berkeley.edu

Problem 1

Answer the following questions about SVD (note: all calculations should be done by hand):

i) Consider the matrix

$$A = \begin{bmatrix} 2 & -1 & 2 \\ -1 & 2 & 2 \\ 2 & 2 & -1 \end{bmatrix} \tag{1}$$

Show that the columns of A are orthogonal to each other. By using this fact, find a singular value decomposition of A.

ii) Find a singular value decomposition of the matrix

$$C = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 2 & 0 \\ 3 & 0 & 0 \end{bmatrix} \times \begin{bmatrix} 2 & -1 & 2 \\ -1 & 2 & 2 \\ 2 & 2 & -1 \end{bmatrix} \tag{2}$$

iii) Consider the optimization problem

$$\min_{B \in \mathbb{R}^{3\times 3}} \|C - B\|_F \qquad \text{s.t.} \qquad \text{rank}(B) \leq 2 \tag{3}$$

Find the optimal solution B^* and compute the error $\frac{||C-B^*||_F^2}{\|C\|_F^2}$.

Solution

i) If the dot product of each pair of columns in A is zero, then the columns are orthogonal.

$$a_1 = \begin{bmatrix} 2 \\ -1 \\ 2 \end{bmatrix}, \quad a_2 = \begin{bmatrix} -1 \\ 2 \\ 2 \end{bmatrix}, \quad a_3 = \begin{bmatrix} 2 \\ 2 \\ -1 \end{bmatrix}$$

$$\begin{aligned} a_1 \cdot a_2 &= 2 \cdot (-1) + (-1) \cdot 2 + 2 \cdot 2 = -2 - 2 + 4 = 0 \\ a_1 \cdot a_3 &= 2 \cdot 2 + (-1) \cdot 2 + 2 \cdot (-1) = 4 - 2 - 2 = 0 \\ a_2 \cdot a_3 &= (-1) \cdot 2 + 2 \cdot 2 + 2 \cdot (-1) = -2 + 4 - 2 = 0 \end{aligned}$$

Since the dot products are all equal to zero, the columns of A are orthogonal. Using this fact, the singular values of A are the norms of its columns, since $A^\top A$ is a diagonal matrix with the squared norms of the columns on the diagonal.

$$\begin{aligned} \sigma_1 &= \|a_1\| = \sqrt{2^2 + (-1)^2 + 2^2} = \sqrt{9} = 3 \\ \sigma_2 &= \|a_2\| = \sqrt{(-1)^2 + 2^2 + 2^2} = \sqrt{9} = 3 \\ \sigma_3 &= \|a_3\| = \sqrt{2^2 + 2^2 + (-1)^2} = \sqrt{9} = 3 \end{aligned}$$

The singular value decomposition of A is given by $A = \mathbf{U}\tilde{\mathbf{\Sigma}}\mathbf{V}^\top$, where $\tilde{\mathbf{\Sigma}}$ is a diagonal matrix with the singular values on the diagonal, and $\mathbf{U}$ and $\mathbf{V}$ are orthogonal matrices. The columns of $\mathbf{V}$ are the normalized eigenvectors of $A^\top A$. But, since $A^\top A$ is a scalar multiple of the identity matrix, any vector in $\mathbb{R}^3$ is an eigenvector, $A^\top A v_i = 9v_i$. Therefore, the standard basis vectors can be used as the columns of $\mathbf{V}$:

$$\mathbf{V} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Problem 1 continued on next page...

The matrix $\tilde{\mathbf{\Sigma}}$ is a diagonal matrix with the singular values on the diagonal:

$$\tilde{\mathbf{\Sigma}} = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{bmatrix}$$

The matrix $\mathbf{U}$ is obtained by normalizing the columns of A since $u_i = \frac{Av_i}{\sqrt{\lambda_i}} = \frac{a_i}{\sqrt{\|a_i\|}}$:

$$\mathbf{U} = \begin{bmatrix} \frac{2}{3} & -\frac{1}{3} & \frac{2}{3} \\ -\frac{1}{3} & \frac{2}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{2}{3} & -\frac{1}{3} \end{bmatrix}$$

Therefore, the singular value decomposition of A is:

$$A = \begin{bmatrix} \frac{2}{3} & -\frac{1}{3} & \frac{2}{3} \\ -\frac{1}{3} & \frac{2}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{2}{3} & -\frac{1}{3} \end{bmatrix} \begin{bmatrix} 3 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}^\top$$

ii) The matrix C is given by the product of two matrices:

$$C = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 2 & 0 \\ 3 & 0 & 0 \end{bmatrix} \begin{bmatrix} 2 & -1 & 2 \\ -1 & 2 & 2 \\ 2 & 2 & -1 \end{bmatrix} = \begin{bmatrix} 2 & 2 & -1 \\ -2 & 4 & 4 \\ 6 & -3 & 6 \end{bmatrix}$$

The matrix $C^\top C$ is then:

$$C^\top C = \begin{bmatrix} 2 & -2 & 6 \\ 2 & 4 & -3 \\ -1 & 4 & 6 \end{bmatrix} \begin{bmatrix} 2 & 2 & -1 \\ -2 & 4 & 4 \\ 6 & -3 & 6 \end{bmatrix} = \begin{bmatrix} 44 & -22 & 26 \\ -22 & 29 & -4 \\ 26 & -4 & 53 \end{bmatrix}$$

The eigenvalues of $C^\top C$ are found by solving the characteristic polynomial $\det(C^\top C - \lambda I) = 0$. The eigenvalues are $\lambda_1 = 81$, $\lambda_2 = 36$, and $\lambda_3 = 9$. The singular values are the square roots of the eigenvalues:

$$\sigma_1 = \sqrt{81} = 9, \quad \sigma_2 = \sqrt{36} = 6, \quad \sigma_3 = \sqrt{9} = 3$$

The matrix $\tilde{\mathbf{\Sigma}}$ is then:

$$\tilde{\mathbf{\Sigma}} = \begin{bmatrix} 9 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 3 \end{bmatrix}$$

The columns of $\mathbf{V}$ are the normalized eigenvectors of $C^\top C$, and the columns of $\mathbf{U}$ are given by $u_i = \frac{Cv_i}{\sigma_i}$. The eigenvectors corresponding to λ_1, λ_2, and λ_3 are then the following respectively:

$$e_1 = \begin{bmatrix} 1 \\ -\frac{1}{2} \\ 1 \end{bmatrix}, \quad e_2 = \begin{bmatrix} -\frac{1}{2} \\ 1 \\ 1 \end{bmatrix}, \quad e_3 = \begin{bmatrix} -2 \\ -2 \\ 1 \end{bmatrix}$$

The matrix $\mathbf{V}$ is then formed by normalizing these eigenvectors:

$$\mathbf{V} = \begin{bmatrix} \frac{2}{3} & -\frac{1}{3} & -\frac{2}{3} \\ -\frac{1}{3} & \frac{2}{3} & -\frac{2}{3} \\ \frac{2}{3} & \frac{2}{3} & \frac{1}{3} \end{bmatrix}$$

Problem 1 continued on next page...

The matrix $\mathbf{U}$ is then formed by computing $u_i = \frac{Cv_i}{\sigma_i}$ for each i:

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

Therefore, the singular value decomposition of C is:

$$C = \begin{bmatrix} 0 & 0 & -1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 9 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 3 \end{bmatrix} \begin{bmatrix} \frac{2}{3} & -\frac{1}{3} & -\frac{2}{3} \\ -\frac{1}{3} & \frac{2}{3} & -\frac{2}{3} \\ \frac{2}{3} & \frac{2}{3} & \frac{1}{3} \end{bmatrix}^\top$$

iii) The optimal solution B^* to the optimization problem is given by the truncated singular value decomposition of C, keeping only the top two singular values and the first two columns of $\mathbf{U}$ and $\mathbf{V}$:

$$B^* = \mathbf{U_2}\tilde{\mathbf{\Sigma}}_\mathbf{2}\mathbf{V_2}^\top$$

where $\mathbf{U_2}$, $\tilde{\mathbf{\Sigma}}_\mathbf{2}$, and $\mathbf{V_2}$ are the matrices formed by taking the first 2 columns of $\mathbf{U}$, the first 2 singular values in $\tilde{\mathbf{\Sigma}}$, and the first 2 columns of $\mathbf{V}$, respectively.

$$\mathbf{U_2} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \tilde{\mathbf{\Sigma}}_\mathbf{2} = \begin{bmatrix} 9 & 0 \\ 0 & 6 \end{bmatrix}, \quad \mathbf{V_2} = \begin{bmatrix} \frac{2}{3} & -\frac{1}{3} \\ -\frac{1}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{2}{3} \end{bmatrix}$$

The optimal solution B^* is then:

$$B^* = \begin{bmatrix} 0 & 0 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 9 & 0 \\ 0 & 6 \end{bmatrix} \begin{bmatrix} \frac{2}{3} & -\frac{1}{3} \\ -\frac{1}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{2}{3} \end{bmatrix}^\top$$

$$= \begin{bmatrix} 0 & 0 & 0 \\ -2 & 4 & 4 \\ 6 & -3 & 6 \end{bmatrix}$$

The squared Frobenius norm of the difference $C - B^*$ is the sum of remaining squared singular value $3^2 = 9$. The squared Frobenius norm of C is the sum of squared singular values:

$$\|C\|_F^2 = \sigma_1^2 + \sigma_2^2 + \sigma_3^2 = 9^2 + 6^2 + 3^2 = 81 + 36 + 9 = 126$$

Therefore, the error is:

$$\frac{\|C - B^*\|_F^2}{\|C\|_F^2} = \frac{9}{126} = \frac{1}{14}$$

Problem 2

An exam with m questions is given to n students. The instructor collects all the grades in an $n \times m$ matrix G where G_{ij} shows the grade of student $i \in \{1, 2, ..., n\}$ for question $j \in \{1, 2...., m\}$. By analyzing the matrix G, the goal is to design a difficulty score for each question that shows the difficulty level of that question. As a naive approach, one may consider the average grade $\frac{\sum_{i=1}^{n} G_{ij}}{n}$ as the difficulty score of question j. To understand the issue with this difficulty score, assume that $n = m = 2$ and G is equal to

$$G = \begin{bmatrix} 50 & 100 \\ 50 & 0 \end{bmatrix} \tag{4}$$

where the minimum and maximum grades for each question are 0 and 100. In this example, both questions have the same average grade of 50. Both students have done poorly on question 1. For question 2, one student got the highest grade possible while the other student got 0 (which may imply that the student was not prepared for that question rather than the question being hard). Question 1 seems to be much harder than question 2 due to the distribution of the grades while the average grades cannot provide any useful information. To address this issue for arbitrary values of n and m, we propose an optimization model for the design of difficulty scores.

i) Consider the optimization problem

$$\min_{B \in \mathbb{R}^{n \times m}} \|G - B\|_F \qquad \text{s.t.} \qquad \text{rank}(B) \leq 1 \tag{5}$$

We decompose the optimal solution B^* as xy^T, where $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$. Assume that $x, y \geq 0$ (note: if no student receives a zero score on any question, then it can be proven that x and y are automatically nonnegative vectors). Assume that the error $\frac{\|G-B^*\|_F^2}{\|G\|_F^2}$ is small. Explain how a difficulty score can be designed for each question in terms of x and y.

ii) Consider the case with $n = 3$ and $m = 5$, together with the grade matrix

$$G = \begin{bmatrix} 100 & 90 & 100 & 80 & 70 \\ 80 & 70 & 60 & 70 & 80 \\ 60 & 50 & 40 & 50 & 60 \end{bmatrix} \tag{6}$$

Using Part (i), design a difficulty score for each question, and rank the questions from the hardest to the easiest based on their scores (note: you can use a computer code for SVD calculations).

Solution

i) The optimal solution $B^* = xy^\top$ is attained for $x = \sqrt{\sigma_1}u_1$ and $y = \sqrt{\sigma_1}v_1$, where σ_1 is the largest singular value of G, and u_1 and v_1 are the corresponding left and right singular vectors, respectively. Since the error is small, each predicted grade for student i on question j is given by $B_{ij}^* = x_i y_j$. Since $x \in \mathbb{R}^n$, each entry x_i can be interpreted as a measure of the ability of student i, for a constant y_j. Similarly, since $y \in \mathbb{R}^m$, each entry y_j can be interpreted as a measure of the difficulty of question j, for a constant x_i. Therefore, a difficulty score for each question could be of the form $z_j = 1/y_j$, where a higher value of z_j indicates a more difficult question.

ii) The compact, rank-1 singular value decomposition of G is given by

$$G \approx B^* = \sigma_1 u_1 v_1^\top = 280.216 \begin{bmatrix} -0.704 \\ -0.575 \\ -0.416 \end{bmatrix} \begin{bmatrix} -0.505 & -0.444 & -0.434 & -0.419 & -0.429 \end{bmatrix}$$

where $G^\top G$ is given by

$$G^\top G = \begin{bmatrix} 20000 & 17600 & 17200 & 16600 & 17000 \\ 17600 & 15500 & 15200 & 14600 & 14900 \\ 17200 & 15000 & 15200 & 14200 & 14200 \\ 16600 & 14200 & 14200 & 13800 & 14200 \\ 17000 & 14900 & 14200 & 14200 & 14900 \end{bmatrix}$$

and has an eigenvalue of $\lambda_1 = 78521.051$ with corresponding normalized eigenvector

$$v_1 = \begin{bmatrix} -0.505 \\ -0.444 \\ -0.434 \\ -0.419 \\ -0.429 \end{bmatrix}$$

The largest singular value of G is given by $\sigma_1 = \sqrt{\lambda_1} = 280.216$. The corresponding left singular vector is given by $u_1 = \frac{1}{\sigma_1} G v_1 \approx \begin{bmatrix} -0.420 \\ -0.336 \\ -0.252 \end{bmatrix}$

Therefore, the difficulty score for each question is given by

$$z = \begin{bmatrix} 1/(\sqrt{\sigma_1} v_1)_1 \\ 1/(\sqrt{\sigma_1} v_1)_2 \\ 1/(\sqrt{\sigma_1} v_1)_3 \\ 1/(\sqrt{\sigma_1} v_1)_4 \\ 1/(\sqrt{\sigma_1} v_1)_5 \end{bmatrix} = \begin{bmatrix} 1/(-4.749) \\ 1/(-4.176) \\ 1/(-4.084) \\ 1/(-3.942) \\ 1/(-4.037) \end{bmatrix} \approx \begin{bmatrix} -0.210 \\ -0.240 \\ -0.245 \\ -0.254 \\ -0.248 \end{bmatrix}$$

The questions ranked from hardest to easiest are 1, 2, 5, 3, 4.

Problem 3

(Coding) Consider a picture of the Berkeley logo named Berkeley1.png that you can download from bCourses → Files → HW.

i) Convert the image to a grayscale image, and store its data into a matrix A.

ii) Plot the singular values of A. To do so, draw the point (k, σ_k) in $\mathbb{R}^2$ for $k = 1, ..., \min(m, n)$, where m and n denote the dimensions of A and σ_k is the k^{th} largest singular value of A.

iii) Consider the optimization problem

$$\min_{B \in \mathbb{R}^{m \times n}} \|A - B\|_F, \qquad \text{s.t.} \qquad \text{rank}(B) \leq k \tag{7}$$

and consider the error $e_k = \frac{\|A-B^*\|_F^2}{\|A\|_F^2}$, where B^* is an optimal solution. Solve the above optimization problem in three scenarios of $k = 30$, $k = 80$ and $k = 100$. For each case, report the error e_k and the percentage $\frac{k}{\min(m,n)} \times 100$ (which shows what percentage of the singular values of A is used), and also draw the grayscale image corresponding to B^* (note: you do not need to report the matrix B^* in your homework submission and only the image together with the code is enough).

iv) Redo Parts (i), (ii), and (iii) for the Berkeley campus picture Berkeley2.png that you can download from bCourses → Files → HW.

Problem 4

Consider the least-squares problem

$$\min_x \|Ax - y\| \tag{8}$$

where $A \in \mathbb{R}^{2\times 3}$ and $y \in \mathbb{R}^2$. Assume that

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \end{bmatrix}, \qquad y = \begin{bmatrix} 4 \\ 7 \end{bmatrix} \tag{9}$$

i) Show that the set of all solutions is equal to

$$S = \left\{ x^* \in \mathbb{R}^3 \;\middle|\; x_1^* + x_2^* + x_3^* = \frac{18}{5} \right\} \tag{10}$$

ii) Assume that y is perturbed to $y + \Delta y$. Find the set of all solutions of the perturbed least-squares problem $\min_x \|Ax - (y + \Delta y)\|$.

iii) Assume that y is perturbed to $y + \Delta y$, where Δy can take any value in the set $\{\Delta y \mid \|\Delta y\| \leq 1\}$. Find the set of all possibilities for the minimum-norm solution of the perturbed least-squares problem $\min_x \|Ax - (y + \Delta y)\|$ (hint: write the ellipsoidal formula for Δx and compute the semi-axes and lengths of the ellipse; you can use a computer code for SVD calculations).

Solution