Question 5, Assignment 5: CS 754, Spring 2024-25

Amitesh Shekhar
IIT Bombay
22b0014@iitb.ac.in

Anupam Rawat
IIT Bombay
22b3982@iitb.ac.in

Toshan Achintya Golla
IIT Bombay
22b2234@iitb.ac.in

April 17, 2025

1. Explain how you will minimize the following cost functions efficiently. In each case, mention any one application in image processing where the problem arises. [4 x 5 = 20 points]
   (a) $J_1(\boldsymbol{A_r}) = \|\boldsymbol{A} - \boldsymbol{A_r}\|_F^2$, where $\boldsymbol{A}$ is a known $m \times n$ matrix of rank greater than r, and $\boldsymbol{A_r}$ is a rank-r matrix, where $r < m, r < n$.
   (b) $J_2(\boldsymbol{R}) = \|\boldsymbol{A} - \boldsymbol{RB}\|_F^2$, where $\boldsymbol{A} \in \mathbb{R}^{n\times m}, \boldsymbol{B} \in \mathbb{R}^{n\times m}, \boldsymbol{R} \in \mathbb{R}^{n\times n}, m > n$ and $\boldsymbol{R}$ is constrained to be orthonormal. Note that $\boldsymbol{A}$ and $\boldsymbol{B}$ are both known.
   (c) $J_3(\boldsymbol{A}) = \|\boldsymbol{C} - \boldsymbol{A}\|_F^2 + \lambda\|\boldsymbol{A}\|_1$, where matrix $\boldsymbol{C}$ is known.
   (d) $J_4(\boldsymbol{A}) = \|\boldsymbol{C} - \boldsymbol{A}\|_F^2 + \lambda\|\boldsymbol{A}\|_*$, where matrix $\boldsymbol{C}$ is known.

(a) First of all, without loss of generality, let's assume that $m \leq n$.

Now let's consider the SVD of A as $A = UDV^\top$

Now let's use the equation: $\|X\|_F^2 = \text{trace}(X^\top X)$

Thus, we can rewrite the objective function as:

$$
\begin{aligned}
J(A_r) &= \text{trace}((A - A_r)^\top (A - A_r)) \\
&= \text{trace}(VV^\top (A - A_r)^\top UU^\top (A - A_r)) \\
&= \text{trace}(V^\top (A - A_r)^\top UU^\top (A - A_r)V) \\
&= \text{trace}((D - Z)^\top (D - Z)) \\
&= \|Z - D\|_F^2
\end{aligned}
$$

, where $Z = U^\top A_r V$ is a rank-r matrix.

The minimum of this objective function is achieved when Z is diagonal, with the first r diagonal elements being the first r diagonal elements of D and the rest being zero. This is because the singular values are in decreasing order in D, so picking the largest r values would minimize the Frobenius norm.

Let M be an $m \times m$ matrix, with the first r elements being 1, and the rest being 0. Then, $Z = MD$ is the required rank-r matrix. Hence, $A_r = UMDV^\top$

In image processing, the problem of denoising is a common problem. In this problem, we are given a noisy image, and we need to find the original image. This can be formulated as an optimization problem, where we need to minimize the Frobenius norm of the difference between the noisy image and the original image. This is similar to the first problem, where we need to find the rank-r matrix that is closest to the original matrix.

(b) We will again use the following equation $\|X\|_F^2 = \text{trace}(X^\top X)$

We can rewrite the objective function as:

$$
\begin{aligned}
J(R) &= \text{trace}((A - RB)^\top (A - RB)) \\
&= \text{trace}(A^\top A + B^\top R^\top RB - B^\top R^\top A - A^\top RB) \\
&= \text{trace}(A^\top A) + \text{trace}(B^\top R^\top RB) - 2\,\text{trace}(A^\top RB) \\
&= \text{trace}(A^\top A) + \text{trace}(B^\top B) - 2\,\text{trace}(A^\top RB) \qquad [R \text{ is orthonormal}]
\end{aligned}
$$

The first 2 terms are constants, and minimizing a negative quantity is equivalent to maximizing the positive quantity.

Thus, we need to maximise

$$\begin{aligned}\text{trace}(A^\top RB) &= \text{trace}(BA^\top R)\\ &= \text{trace}(ZR)\end{aligned}$$

, where $Z = BA^\top$. Now consider the SVD of Z as $Z = UDV^\top$. The equation then simplifies to

$$\begin{aligned}\text{trace}(ZR) &= \text{trace}(UDV^\top R)\\ &= \text{trace}(DV^\top RU)\\ &= \text{trace}(DQ)\end{aligned}$$

, where Q is orthonormal. The maximum of this is achieved when $Q_{ii} = 1$ all along its diagonal. Since Q is orthonormal, $Q = I$.

Thus, $R = VU^\top$ is the required orthonormal matrix. However, there is a subtle problem. R won't be a rotation matrix if $\det(R) = -1$ instead of 1. And simply flipping the sign of R won't help, as the sign of the objective function will change.

In such a case, we can consider R as

$$R = VTU^\top$$

, where $T = \text{diag}(1, 1, \ldots, 1, -1)$ has shape $(n, 1)$. This way, we are only subtracting the smallest singular vector of R from the objective function and not changing the sign of the determinant.

So, we can define $W = \text{diag}(1, 1, \ldots, 1, \det(VU^\top))$.
Finally, we have

$$R = VWU^\top$$

In image registration, we are given two images, and we need to find the transformation that aligns the two images. This can be formulated as an optimization problem, where we need to minimize the Frobenius norm of the difference between the two images. This is similar to the second problem, where we need to find the rotation matrix that aligns the two images.

(c) **Minimizing $J_3(\boldsymbol{A}) = \|\boldsymbol{C} - \boldsymbol{A}\|_F^2 + \lambda\|\boldsymbol{A}\|_1$**

Objective: Minimize the cost function $J_3(\boldsymbol{A}) = \|\boldsymbol{C} - \boldsymbol{A}\|_F^2 + \lambda\|\boldsymbol{A}\|_1$, where $\boldsymbol{C}$ is a known matrix, and $\|\boldsymbol{A}\|_1$ is the element-wise L1 norm (sum of absolute values of all entries in $\boldsymbol{A}$).

Approach: This problem is separable in the entries of $\boldsymbol{A}$, meaning each entry A_{ij} can be optimized independently. The cost function can be rewritten as:

$$J_3(\boldsymbol{A}) = \sum_{i,j}(C_{ij} - A_{ij})^2 + \lambda\sum_{i,j}|A_{ij}|.$$

Thus, for each A_{ij}, we solve:

$$\min_{A_{ij}}(C_{ij} - A_{ij})^2 + \lambda|A_{ij}|.$$

This is the standard Lasso problem in one dimension, whose solution is given by the soft-thresholding operator:

$$A_{ij} = \text{sign}(C_{ij}) \cdot \max(|C_{ij}| - \lambda/2, 0).$$

Here, $\text{sign}(C_{ij})$ is the sign of C_{ij}, and $\max(|C_{ij}| - \lambda/2, 0)$ shrinks C_{ij} towards zero by $\lambda/2$, setting it to zero if $|C_{ij}| \leq \lambda/2$.

Efficient Minimization: Apply the soft-thresholding operator element-wise to $\boldsymbol{C}$:

$$\boldsymbol{A} = \text{SoftThreshold}_{\lambda/2}(\boldsymbol{C}),$$

where $\text{SoftThreshold}_\tau(x) = \text{sign}(x) \cdot \max(|x| - \tau, 0)$.

Application in Image Processing: This problem arises in **image denoising with sparsity constraints**, where $\boldsymbol{C}$ is a noisy image, and $\boldsymbol{A}$ is the denoised image. The L1 penalty promotes sparsity in the image (e.g., in wavelet or gradient domains), leading to piecewise smooth or sparse representations.

(d) **Minimizing** $J_4(\boldsymbol{A}) = \|\boldsymbol{C} - \boldsymbol{A}\|_F^2 + \lambda\|\boldsymbol{A}\|_*$

Objective: Minimize the cost function $J_4(\boldsymbol{A}) = \|\boldsymbol{C} - \boldsymbol{A}\|_F^2 + \lambda\|\boldsymbol{A}\|_*$, where $\boldsymbol{C}$ is a known matrix, and $\|\boldsymbol{A}\|_*$ is the nuclear norm (sum of singular values of $\boldsymbol{A}$).

Approach: The nuclear norm is a convex surrogate for rank minimization, and this problem is known as **matrix denoising** or **nuclear norm regularization**. The solution is obtained via singular value thresholding (SVT):

Compute the singular value decomposition (SVD) of $\boldsymbol{C}$:

$$\boldsymbol{C} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top, \quad \boldsymbol{\Sigma} = \text{diag}(\sigma_1, \sigma_2, \ldots, \sigma_r).$$

Apply the soft-thresholding operator to the singular values:

$$\boldsymbol{\Sigma}_\lambda = \text{diag}(\max(\sigma_i - \lambda, 0)).$$

Reconstruct the matrix:

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}_\lambda\boldsymbol{V}^\top.$$

Efficient Minimization: The key steps are:

- Compute the SVD of $\boldsymbol{C}$ (efficient for large matrices using randomized or truncated SVD methods).
- Threshold the singular values by λ.
- Reconstruct $\boldsymbol{A}$.

Application in Image Processing: This problem arises in **low-rank matrix completion** or **robust PCA**, where $\boldsymbol{C}$ is a corrupted or incomplete observation (e.g., missing pixels in an image), and $\boldsymbol{A}$ is the low-rank recovered matrix. The nuclear norm encourages $\boldsymbol{A}$ to be low-rank, which is useful for background modeling or structure recovery in images.