

Question 1, 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. Implement the ALM algorithm for robust PCA. For this, create a matrix $\mathbf{M} \in \mathbb{R}^{n_1 \times n_2}$ which is the sum of a low rank matrix \mathbf{L} of rank r and a sparse matrix \mathbf{S} with $s = f_s n_1 n_2$ non-zero elements where $f_s \in [0, 1]$. Create \mathbf{L} using truncated SVD of Gaussian random matrices. The non-zero entries of \mathbf{S} should be drawn from $\mathcal{N}(0, 9)$ and they should exist at randomly chosen indices. For this experiment, let $n_1 = 800, n_2 = 900$. Vary $r \in \{10, 30, 50, 75, 100, 125, 150, 200\}$ and $f_s \in \{0.01, 0.02, 0.03, 0.04, 0.06, 0.08, 0.1, 0.15\}$. Each value of (r, f_s) , execute the algorithm 15 times and record the success probability. Plot the success probability as an image whose X and Y axes are r and f_s respectively (lower probability in darker color and higher probability in brighter color). Plot a colorbar using the `colorbar` function in MATLAB. Note that we define a reconstruction to be successful if $\|\mathbf{L} - \hat{\mathbf{L}}\|_F / \|\mathbf{L}\|_F \leq 0.001$ and $\|\mathbf{S} - \hat{\mathbf{S}}\|_F / \|\mathbf{S}\|_F \leq 0.001$ where $\hat{\mathbf{L}}, \hat{\mathbf{S}}$ are estimates of \mathbf{L}, \mathbf{S} respectively. (Note, typically a single run of RPCA took just about 1.5-2 seconds on my desktop.) For any one successful and one unsuccessful (r, f_s) configuration, plot the ground truth and estimated low-rank and sparse matrices as separate grayscale images. Note that for RPCA, we are minimizing $\|\mathbf{L}\|_* + \lambda \|\mathbf{S}\|_1$ subject to $\mathbf{M} = \mathbf{L} + \mathbf{S}$ where $\lambda = 1/\sqrt{\max(n_1, n_2)}$. [30 points]

Soln:

Results

Note: Please run the `driver_code.m` file in MATLAB while keeping the `alm_rpca.m` function in the same directory. The function is called internally by the driver script.

Part 1: Success Fraction Across (r, f_s) Configurations

Figure 1: Success probability of ALM-RPCA for various rank r and sparsity levels f_s . Darker colors indicate lower success, while brighter colors indicate higher success.

Part 2: Visualizing Reconstruction Quality

For $r = 200$ and $f_s = 0.15$, the algorithm failed in one of the iterations. The original and reconstructed low-rank and sparse matrices are shown below:

In contrast, for $r = 100$ and $f_s = 0.15$, the algorithm succeeded. The corresponding visualizations are:

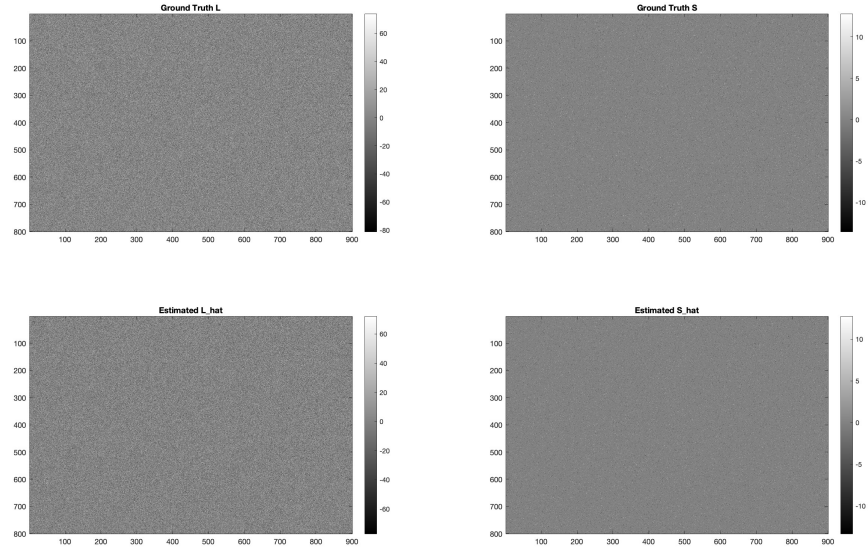


Figure 2: Reconstruction failure: $r = 200$, $f_s = 0.15$. Top: Ground truth L and S . Bottom: Estimated \hat{L} and \hat{S} .

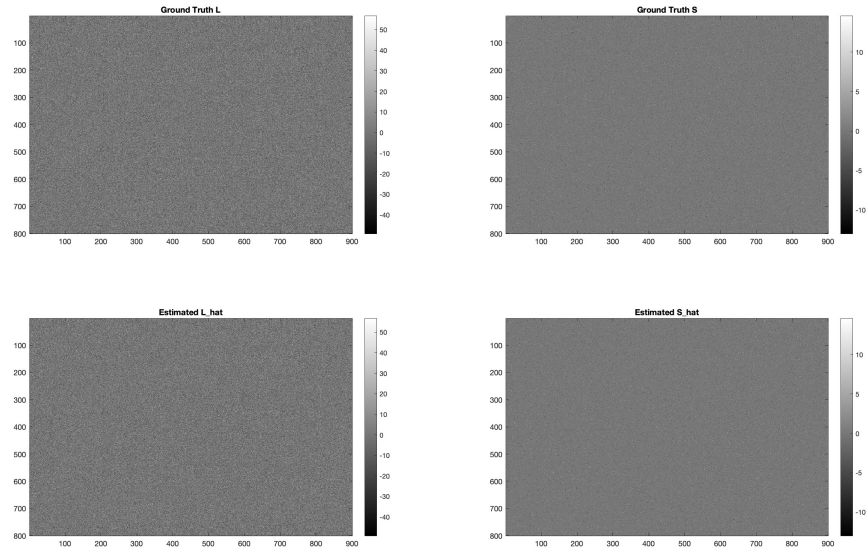


Figure 3: Successful reconstruction: $r = 100$, $f_s = 0.15$. Top: Ground truth L and S . Bottom: Estimated \hat{L} and \hat{S} .