End-sem: CS 754, Advanced Image Processing, 4th May

Instructions: There are 180 minutes for this exam. This exam is worth 10% of your final grade. Attempt all **eight** questions. Write **brief** answers - lengthy answers are not expected. Each question carries 10 points.

1. Briefly explain any two applications of robust principal components analysis (RPCA). For each application, make sure to explain why the underlying matrix can be expressed as the sum of a low rank matrix $\boldsymbol{L}$ and sparse matrix $\boldsymbol{S}$. [5 + 5 = 10 points]

2. Clearly define the problem of compressive low rank matrix recovery, with clear definition of all mathematical terms. Give a mathematical definition of the matrix restricted isometry property. [6 + 4 = 10 points]

3. In blind compressed sensing, recall that we consider N compressive measurements of the form $\boldsymbol{y_i} = \boldsymbol{\Phi_i}\boldsymbol{\Psi}\boldsymbol{\theta_i} + \boldsymbol{\eta_i}, 1 \leq i \leq N$. For each i, $\boldsymbol{y_i}$ is the compressive measurement for the signal $\boldsymbol{x_i} \triangleq \boldsymbol{\Psi}\boldsymbol{\theta_i}$. We want to infer $\boldsymbol{\theta_i}$ as well as $\boldsymbol{\Psi}$ from the compressive measurements. The objective function that is optimized in this application is $J(\boldsymbol{\Psi}, \{\boldsymbol{\theta_i}\}_{i=1}^N) = \sum_{i=1}^N \|\boldsymbol{y_i} - \boldsymbol{\Phi_i}\sum_{k=1}^K \boldsymbol{\Psi_k}\theta_{ik}\|^2$ subject to the constraints $\forall i, \|\boldsymbol{\theta_i}\|_0 \leq T_0; \forall k \boldsymbol{\Psi_k}^t\boldsymbol{\Psi_k} = 1$. Why does the update of the dictionary columns $\{\boldsymbol{\Psi_k}\}_{k=1}^K$ require that the sensing matrices $\{\boldsymbol{\Phi_i}\}_{i=1}^N$ for the different signals $\{\boldsymbol{x_i}\}_{i=1}^N$ be different from each other? You may write an equation to support your answer. [10 points]

4. We have seen the following theorem for compressed sensing in class: Consider compressive measurements of the form $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{\theta} + \boldsymbol{\eta}$ where $\boldsymbol{y} \in \mathbb{R}^m, \boldsymbol{\theta} \in \mathbb{R}^n, \boldsymbol{A} \in \mathbb{R}^{m\times n}, m \ll n$. Suppose $\boldsymbol{A}$ obeys the Restricted isometry property with restricted isometry constant δ_{2s} (of order $2s$) such that $\delta_{2s} < \sqrt{2}-1$. Let $\boldsymbol{\theta}^\star$ be the solution to the following optimization problem (P1): $\min\|\boldsymbol{\theta}\|_1$ such that $\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{\theta}\|_2 \leq \varepsilon$ where $\|\boldsymbol{\eta}\|_2 \leq \varepsilon$. Then, we have the following error bound: $\|\boldsymbol{\theta}-\boldsymbol{\theta}^\star\|_2 \leq \dfrac{C_1}{\sqrt{s}}\|\boldsymbol{\theta}-\boldsymbol{\theta_s}\|_1 + C_2\varepsilon$ where C_1, C_2 are monotonically increasing functions of δ_{2s} (in the domain $[0,1]$). Also the vector $\boldsymbol{\theta_s}$ is defined such that $\forall i \in \mathcal{S}, \theta_s(i) = \theta_i, \forall i \notin \mathcal{S}, \theta_s(i) = 0$ where the set $\mathcal{S}$ consists of indices of the s largest absolute-value elements of $\boldsymbol{\theta}$.
If the elements of the noise vector $\boldsymbol{\eta}$ are i.i.d. random variables from the uniform distribution $\mathcal{U}(-r,+r)$ for known $r > 0$, the value of ε would be equal to $r\sqrt{m}$. This seems to imply that the upper bound on the recovery error increases with $\sqrt{m}$, which is counter-intuitive as simulations show that the recovery error *decreases* with m. Can you reconcile this apparent contradiction? Also the bounds seem to imply that as s increases, the first term of the error decreases, whereas we would expect signals that are sparser (i.e. have fewer number of high-valued components) to allow for better recovery. Can you reconcile this apparent contradiction? [5 + 5 = 10 points]

5. Suppose you wanted to compute the coherence $\mu(\boldsymbol{\Phi}, \boldsymbol{\Psi})$ between a Radon sensing matrix $\boldsymbol{\Phi}$ of size $m \times n, m < n$ and a $n \times n$ 2D-DCT representation matrix $\boldsymbol{\Psi}$. In applications in tomography, we deal with large images and hence n and m will be large in value. Hence it is impossible to store $\boldsymbol{\Phi}$ or $\boldsymbol{\Psi}$ in memory. How will you compute $\mu(\boldsymbol{\Phi}, \boldsymbol{\Psi})$ in such a case? Recall that $\mu(\boldsymbol{\Phi}, \boldsymbol{\Psi}) \triangleq \max_{i,j} \dfrac{|\boldsymbol{\Phi^i}\boldsymbol{\Psi_j}|}{\|\boldsymbol{\Phi^i}\|_2\|\boldsymbol{\Psi_j}\|_2}$ where $\boldsymbol{\Psi_j}$ is the j^{th} column-vector of $\boldsymbol{\Psi}$ $(1 \leq j \leq n)$, and $\boldsymbol{\Phi^i}$ is the i^{th} row-vector of $\boldsymbol{\Phi}$ $(1 \leq i \leq m)$. Assume you have access to a MATLAB function handle which efficiently computes the Radon transform of an image at specified angles. [10 points]

6. Explain the relative advantages and disadvantages of overcomplete dictionary representations as compared to orthonormal basis representations. [5 + 5 = 10 points]

7. Apart from sparsity of DCT or wavelet coefficients, briefly state any two statistical properties of natural images. We know that the negative log likelihood of a Laplacian random variable gives rise to an ℓ_1 term. Do the theoretical guarantees provided by the theorems for compressed sensing (refer to question 4) require the values in the unknown vector $\boldsymbol{\theta}$ to be Laplacian distributed? Explain. [5 + 5 = 10 points]

8. Consider that you learned a dictionary $\boldsymbol{D}$ to sparsely represent a certain class $\mathcal{S}$ of images - say handwritten alphabet or digit images. How will you convert $\boldsymbol{D}$ to another dictionary which will sparsely represent the following classes of images? Note that you are not allowed to learn the dictionary all over again, as it is time-consuming.

   (a) Class $\mathcal{S}_1$ which consists of images obtained by applying motion blur to the images in $\mathcal{S}$. Assume that the motion blur is represented as convolution with an oriented Gaussian kernel of a fixed known standard deviation σ and a fixed known blur direction $\boldsymbol{d}$.

   (b) Class $\mathcal{S}_2$ which consists of images obtained by applying an affine intensity transform to the images in $\mathcal{S}$. The affine transform has the form $I^i_{new}(x,y) = \alpha_i I^i_{old}(x,y) + \beta_i$ for unknown α_i and β_i (but constant throughout a given image, i.e. independent of x, y).

   (c) Class $\mathcal{S}_3$ which consists of images obtained by applying an intensity transformation $I^i_{new}(x,y) = (I^i_{old}(x,y))^2$ to the images in $\mathcal{S}$. [3 + 3 + 4 = 10 points]