End-sem: CS 754, Advanced Image Processing, 1st 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. Define the problem of compressive low rank matrix recovery. Clearly state the meaning of all mathematical terms. Give a mathematical definition of the matrix restricted isometry property. [6 + 4 = 10 points]
Solution: Consider measurements of the form $y = \mathcal{A}(M)$ where $M \in \mathbb{R}^{n_1 \times n_2}$ is a low rank matrix and $\mathcal{A} : \mathbb{R}^{n_1 \times n_2} \to R^m$ is a linear measurement operator with $m < n_1 n_2$. The low rank matrix recovery problem is to recovery M from y and $\mathcal{A}$, possibly under noise in y.
The order r matrix-RIC of operator $\mathcal{A}$ is the smallest constant δ_r such that $(1-\delta_r)\|M\|_F^2 \leq \|\mathcal{A}(M)\|_2^2 \leq (1+\delta_r)\|M\|_F^2$ for all matrices of rank less than or equal to r. If $\delta_r < 1$, then $\mathcal{A}$ is said to obey matrix RIP.

2. In blind compressed sensing, recall that we consider N compressive measurements of the form $\boldsymbol{y_i} = \boldsymbol{\Phi_i \Psi \theta_i} + \boldsymbol{\eta_i}, 1 \leq i \leq N$. For each i, $\boldsymbol{y_i}$ is the compressive measurement for the signal $\boldsymbol{x_i} \triangleq \boldsymbol{\Psi \theta_i}$. 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]
Solution: Answer lies in slide 144 of the lecture slides on dictionary learning.

3. In parallel bean computed tomography, the projection measurements are represented as a single vector $\boldsymbol{y} \sim \text{Poisson}(I_o \exp(-\boldsymbol{Rf}))$, where $\boldsymbol{y} \in \mathbb{R}^m$ with $m =$ number of projection angles $\times$ number of bins per angle; I_o is the power of the incident X-Ray beam; $\boldsymbol{R}$ represents the Radon operator (effectively a $m \times n$ matrix) that computes the projections at the pre-specified known projection angles; and $\boldsymbol{f}$ represents the unknown signal (actually tissue density values) in $\mathbb{R}^n$. If $m < n$, write down a suitable objective function whose minimum would be a good estimate of $\boldsymbol{f}$ given $\boldsymbol{y}$ and $\boldsymbol{R}$ and which accounts for the Poisson noise in $\boldsymbol{y}$. State the motivation for each term in the objective function. Recall that if $z \sim \text{Poisson}(\lambda)$, then $P(z = k) = \lambda^k e^{-\lambda}/k!$ where k is a non-negative integer. Now suppose that apart from Poisson noise, there was also additive impulse noise (say, salt and pepper noise) in $\boldsymbol{y}$. How would you solve this problem (eg: appropriate preprocessing or suitable change of objective function)? [6+ 4 = 10 points]
Solution: The cost function to be minimized is $J(\boldsymbol{\theta}) = \lambda\|\boldsymbol{\theta}\|_1 + \sum_{i=1}^m [I_o \exp(-\boldsymbol{R^i \Psi \theta}) + y_i \boldsymbol{R^i \Psi \theta}]$ where $\boldsymbol{R^i}$ is the i^{th} row of $\boldsymbol{R}$. Here i is an index indicating the pair (projection bin, projection angle). Also we are representing the image $\boldsymbol{f}$ as $\boldsymbol{f} = \boldsymbol{\Psi\theta}$ where $\boldsymbol{\Psi}$ is a sparsifying basis. Since $m < n$, such a sparsity promoting prior is essential. However, you could also have a total variation prior on $\boldsymbol{f}$.
If there is signal-independent additive impulse noise besides Poisson noise, there are many possible solutions. One solution is to model the impulse noise by a heavy-tailed distribution such as the laplacian. Then the noise in $\boldsymbol{y}$ is modelled as the convolution of the Poisson and Laplacian distributions. This is because the pdf of the sum of two random variables is the convolution of their individual pdfs. Another solution is to detect bins with impulse noise and remove them from the data fidelity term for the reconstruction. One can do median filtering on $\boldsymbol{y}$ but realize that this will alter the noise statistics. Many students wrote this as the answer, but it is not fully accurate.

4. Explain the relative advantages and disadvantages of overcomplete dictionary representations as compared to orthonormal basis representations. [5 + 5 = 10 points]
Solution: Advantages: possibility of sparser representations for signals (claps + whistles example), tunability of number of columns. Disadvantage: possibility of overfitting due to very number of columns, lack of uniqueness of sparse codes, expensive sparse coding step. Some students seem to think that overcomplete dictionaries are necessarily learned and orthonormal bases are necessarily universal. This is wrong. Examples: overcomplete DCT/DFT; PCA is a learned orthonormal basis.

5. 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 a known derivative filter to the images in $\mathcal{S}$.

(b) Class $\mathcal{S}_2$ which consists of images obtained by rotating a subset of the images in class $\mathcal{S}$ by a known fixed angle α, and the other subset by another known fixed angle β.

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

Solution: Part a: create a new dictionary D_2 whose every column is obtained as follows: Reshape each column of D to form an image of the appropriate dimensions and convolve it with the derivative filter. Then reshape the resulting image to yield the column of D_2.

Part b: $D_2 = (D_{2\alpha} D_{2\beta})$, where the columns of $D_{2\alpha}$ are obtained by rotating the image-reshaped columns of D by α, followed by vectorization. You must take care of zero-padding since the image domain size may now increase. Another possible answer for this part is to pre-multiply the columns of D by a permutation matrix designed based on α or β. If you choose this route, you have to be careful to include appropriate interpolation in the design of the permutation matrix.

Part c: If $x = D\theta = \sum_k d_k \theta_k$, then $x.^2 = \sum_{k,l} d_k \cdot d_l \theta_k \theta_l$. With this in mind, the new dictionary is $(D2|D|\mathbf{1})$ where D_2 is a dictionary containing K^2 columns (if D has K columns) containing pointwise multiplication of the columns of D.

6. We have studied statistical compressed sensing in class for reconstruction of signal $\boldsymbol{x}$ from compressive measurements of the form $\boldsymbol{y} = \boldsymbol{\Phi x} + \boldsymbol{\eta}$. Here every element of $\boldsymbol{\eta}$ is randomly drawn from $\mathcal{N}(0, \sigma^2)$ where σ is known, and $\boldsymbol{y} \in \mathbb{R}^m, \boldsymbol{x} \in \mathbb{R}^n, \boldsymbol{\Phi} \in \mathbb{R}^{m \times n}, \boldsymbol{\eta} \in \mathbb{R}^m, m \ll n$. In addition, we assume $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ where $\boldsymbol{\mu} \in \mathbb{R}^n$ is the known mean vector, and $\boldsymbol{\Sigma} \in \mathbb{R}^{n \times n}$ is the known positive semi-definite covariance matrix. The MAP estimate of $\boldsymbol{x}$ is given as $(\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Phi}^T \boldsymbol{\Phi}/(2\sigma^2))^{-1} \boldsymbol{\Phi}^T \boldsymbol{y}/(2\sigma^2)$. What are the relative advantages and disadvantages of this technique for signal reconstruction compared to regular ℓ_1 norm based approaches for compressed sensing? [10 points]
Solution: The advantage is a simple closed-form expression which enables fast computation. The disadvantage is that such a prior is restrictive, and the class of images which are sparse in some basis such as DCT is larger than the class of images randomly drawn from $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

7. State briefly any one application of compressive RPCA. Write down the main objective function involved in the optimization problem and state the meaning of each term w.r.t. the particular application. [10 points]
Solution: Slides 89-93 of the lecture slides on matrix recovery.

8. How will you solve for the minimum of the following objective functions: (1) $J(\boldsymbol{A_r}) = \|\boldsymbol{A} - \boldsymbol{A_r}\|^2_F$, 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 $r < m, r < n$. (2) $J(\boldsymbol{R}) = \|\boldsymbol{A} - \boldsymbol{RB}\|^2_F$, 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.
In both cases, explain briefly any one situation in image processing where the solution to such an optimization problem is required. [2.5 + 2.5 + 2.5 + 2.5 = 10 points]
Solution: Part 1: A_r is obtained by computing the SVD of A, i.e. $A = USV^T$, and then $A_r = \sum_{k=1}^{r} s_{kk} u_k v_k^t$ where the r largest singular values and their corresponding singular vectors are considered in the summation. Application: Image denoising using singular value decomposition; PCA; K-SVD (for rank 1 approximation).
Part 2: $R = VU^T$ where $BA^T = USV^T$ (SVD of the LHS). This is orthogonal procrustes. Applications: Finding orthogonal transformation between two point-sets A, B. If A represents signals and B represents their coefficients in some unknown orthogonal basis R, then you use orthogonal procrustes to find R. This is used in the union of orthonormal bases method of dictionary learning.