Midsem: CS 754, Advanced Image Processing, 23/2/2024

Instructions: There are 120 minutes for this exam. This exam is worth 10% of your final grade. Attempt all questions. Write brief answers - **ideally, no more than 4 sentences per sub-question.** Wherever necessary, please write equations with the meaning of all terms clearly stated. You can quote results/theorems done in class directly without proving/justifying them. Each question carries 10 points.

Useful information:

1. Given a matrix $\boldsymbol{\Phi} \in \mathbb{R}^{m \times n}, m < n$, the s-order restricted isometry constant δ_s of $\boldsymbol{\Phi}$ is the smallest number for which the following is true for any s-sparse vector $\boldsymbol{f}$: $(1-\delta_s)\|\boldsymbol{f}\|_2^2 \leq \|\boldsymbol{\Phi f}\|_2^2 \leq (1+\delta_s)\|\boldsymbol{f}\|_2^2$. If $\delta_s \in [0,1)$, then $\boldsymbol{\Phi}$ is said to obey the restricted isometry property (RIP) of order s.

2. Let $\boldsymbol{\theta}^\star$ be the result of the following minimization problem: (P1) $\min \|\boldsymbol{\theta}\|_1$ such that $\|\boldsymbol{y} - \boldsymbol{\Phi\Psi\theta}\|_2 \leq \varepsilon$, where $\boldsymbol{y}$ is an m-element measurement vector of the form $\boldsymbol{y} = \boldsymbol{\Phi x} + \boldsymbol{\eta}$, $\boldsymbol{\Phi}$ is a $m \times n$ measurement matrix $(m < n)$, $\boldsymbol{\Psi}$ is a $n \times n$ orthonormal basis in which n-element signal $\boldsymbol{x}$ has a sparse representation of the form $\boldsymbol{x} = \boldsymbol{\Psi\theta}$. Note that ε is an upper bound on the magnitude of the noise vector $\boldsymbol{\eta}$. Theorem 3 we studied in class states the following: If $\boldsymbol{\Phi}$ obeys the restricted isometry property with isometry constant $\delta_{2s} < \sqrt{2} - 1$, then we have $\|\boldsymbol{\theta} - \boldsymbol{\theta}^\star\|_2 \leq C_1 s^{-1/2}\|\boldsymbol{\theta} - \boldsymbol{\theta_s}\|_1 + C_2\varepsilon$ where C_1 and C_2 are functions of only δ_{2s} and where $\forall i \in \mathcal{S}, [\boldsymbol{\theta_s}]_i = \theta_i; \forall i \notin \mathcal{S}, [\boldsymbol{\theta_s}]_i = 0$. Here $\mathcal{S}$ is a set containing the s largest magnitude elements of $\boldsymbol{\theta}$.

Questions:

1. Consider the CASSI architecture for compressive acqusition of hyperspectral images. Also consider the technique of color filter arrays (CFAs) for acquisition of RGB images, which can be interpreted as a form of compressive imaging. Explain the basic difference between the forward models of image acquisition in CASSI and CFAs, with the help of appropriate equations. Your answer should not merely state that one model is for hyperspectral and the other is for RGB images. What are the difficulties if the forward model of CFAs were to be used for hyperspectral image acquisition? Justify. [6+4 = 10 points]

   Solution: The forward model for CASSI is given by: [2 points]

$$s(x,y) = \sum_{j=1}^{N_\lambda} f_j(x-l_j,y) C_{CASSI}(x-l_j,y), \tag{1}$$

   where s is the snapshot image, (x,y) is a spatial location with $1 \leq x \leq N_x, 1 \leq y \leq N_y$, j is an index for the wavelengths (totally N_λ in number), f_j is the jth wavelength slice of the underlying hyperspectral datacube $f \in \mathbb{R}^{N_x \times N_y \times N_\lambda}$, C_{cASSI} is the binary code due to the coded aperture and l_j is the shift due to the prism in the jth slice. Notice that the codes for different wavelength-slices are shifts of each other.

   The forward model for CFAs is given by: [2 points]

$$s(x,y) = \sum_{j=1}^{N_\lambda} f_j(x,y) C_{CFA}(x,y,j), \tag{2}$$

   with the same notation as earlier. We see that there are no wavelength-dependent shifts in CFA [1 points]. Moreover, for any (x,y), there exists a unique j for which $C_{CFA}(x,y,j) = 1$, and for all other values of j, we must have $C_{CFA}(x,y,j) = 0$. [1 point]
   The forward model of CFAs (typical for RGB images) is not suitable for CASSI (typical for hyperspectral) because the former measures only one wavelength out of the available set at any pixel. Interpolating the values at other wavelengths is possible for RGB images, but much more difficult for hyperspectral images which have a large number of wavelengths. [4 points]

2. In compressed sensing, we are given measurement vector $\boldsymbol{y} \in \mathbb{R}^m$ of an unknown sparse signal $\boldsymbol{x} \in \mathbb{R}^n$ of the form $\boldsymbol{y} = \boldsymbol{\Phi x}$, where $\boldsymbol{\Phi} \in \mathbb{R}^{m \times n}$ is the known sensing matrix ($m < n$). We know that $\boldsymbol{x}$ can be inferred accurately from $\boldsymbol{y}, \boldsymbol{\Phi}$ if the matrix $\boldsymbol{\Phi}$ meets some conditions such as RIP. A curious (and therefore, very good :-)) student asks: How will you estimate $\boldsymbol{x}$ if you knew in advance that most of the elements of $\boldsymbol{x}$ are not zero in value, but equal to say $\alpha \neq 0$ in value, in the case where α is known? Your job is to answer the student's question. Now if α is unknown, how would you estimate $\boldsymbol{x}$? Would the theoretical guarantees for estimating $\boldsymbol{x}$ in the sparse case also apply in these two cases? Justify. [4+4+1+1=10 points]
Answer: Define $\boldsymbol{z} := \boldsymbol{x} - \alpha\boldsymbol{1}$, which will be sparse. Then $\boldsymbol{\Phi z} = \boldsymbol{y} - \alpha\boldsymbol{\Phi 1}$. Denoting $\boldsymbol{y'} := \boldsymbol{y} - \alpha\boldsymbol{\Phi 1}$, we see that $\boldsymbol{y'} = \boldsymbol{\Phi z}$. We can estimate $\boldsymbol{z}$ from $\boldsymbol{y'}, \boldsymbol{\Phi}$ using P1 and then add α to obtain $\boldsymbol{x}$. [4 points]
In the above case, the same theoretical guarantees as for normal CS will apply. [1 point]

If α is unknown, we consider $\boldsymbol{y} = \boldsymbol{\Phi z} + \alpha\boldsymbol{1} = (\boldsymbol{\Phi}|\boldsymbol{1})(\boldsymbol{z};\alpha)$ where $(\boldsymbol{z};\alpha)$ is a vector in $\mathbb{R}^{n+1}$ and $(\boldsymbol{\Phi}|\boldsymbol{1}) \in \mathbb{R}^{m\times(n+1)}$. We can estimate $(\boldsymbol{z};\alpha)$ using P1, from which we can estimate $\boldsymbol{x}$. Another approach: Even if $\boldsymbol{x}$ is not sparse, its difference vector $\boldsymbol{\delta x}$ such that $\delta x_i = x_{i+1} - x_i$ will be sparse. Hence we can obtain $\boldsymbol{x}$ up to an additive constant by minimizing $\|\boldsymbol{y} - \boldsymbol{\Phi x}\|_2^2 + \lambda\|\boldsymbol{\delta x}\|_1$ (which is similar to total variation). The additive constant can be resolved if one of the rows of $\boldsymbol{\Phi}$ was all ones. [4 points for either of these approaches]
In the latter approach, the theoretical bounds for normal CS apply. In the first approach, however, the bounds will apply if $(\boldsymbol{\Phi}|\boldsymbol{1})$ obeys the RIP. [1 point]

3. In the estimator (P1) defined above, explain how ε is to be set in practice, if you know that the noise distribution is: (a) Uniform$(-r, r)$ with known r; (b) $\mathcal{N}(0, \sigma^2)$ with known σ. Also, what is the motivation behind minimizing $\|\boldsymbol{x}\|_1$ instead of $\|\boldsymbol{x}\|_0$? [3+4+3 = 10 points]
Answer: (a) $\varepsilon = r\sqrt{m}$; (b) $\varepsilon = 3\sigma\sqrt{m}$.
Minimizing $\|\boldsymbol{x}\|_0$ is computationally expensive (NP hard) [1.5 points], whereas minimizing $\|\boldsymbol{x}\|_1$ is computationally more manageable while providing the same guarantees as that of the former problem [1.5 points - this point of maintaining quality of approximation is important, otherwise you lose the 1.5 points] (albeit at the expense of some more measurements: $O(s \log n)$ for ℓ_1 as compared to $2s$ for ℓ_0 for recovering s-sparse signals).

4. Consider two s-sparse signals $\boldsymbol{f_1}$ and $\boldsymbol{f_2}$, both in $\mathbb{R}^n$. You have access to their compressive measurements $\boldsymbol{y_1} = \boldsymbol{\Phi f_1}, \boldsymbol{y_2} = \boldsymbol{\Phi f_2}$ (both in $\mathbb{R}^m, m < n$) respectively, but you do not have access to the signals themselves. *Without* reconstructing these signals, how will you determine approximate values of the following quantities *directly* from $\boldsymbol{y_1}, \boldsymbol{y_2}$? Assume that $\boldsymbol{\Phi}$ obeys RIP of order $2s$. [3+5+2 = 10 points]

   (a) $\|\boldsymbol{f_1} - \boldsymbol{f_2}\|_2^2$
   (b) $\boldsymbol{f_1}^t\boldsymbol{f_2}$, if you knew that $\|\boldsymbol{f_1}\|_2^2 = \|\boldsymbol{f_2}\|_2^2 = 1$
   (c) List any one practical application of being able to estimate $\|\boldsymbol{f_1} - \boldsymbol{f_2}\|_2^2$ from $\boldsymbol{y_1}, \boldsymbol{y_2}$ in machine learning or image retrieval.

   Answer:

   (a) Since the sensing matrix satisfies RIP of order $2s$ and because $\boldsymbol{f_1} \pm \boldsymbol{f_2}$ is a $2s$-sparse vector, we have $(1 - \delta_{2s})\|\boldsymbol{f_1} - \boldsymbol{f_2}\|_2^2 \leq \|\boldsymbol{\Phi f_1} - \boldsymbol{\Phi f_2}\|_2^2 \leq (1 + \delta_{2s})\|\boldsymbol{f_1} - \boldsymbol{f_2}\|_2^2$ 1.5 points. Hence, we know that $\|\boldsymbol{f_1} - \boldsymbol{f_2}\|_2^2$ lies between $\dfrac{\|\boldsymbol{\Phi f_1} - \boldsymbol{\Phi f_2}\|_2^2}{1 + \delta_{2s}}$ and $\dfrac{\|\boldsymbol{\Phi f_1} - \boldsymbol{\Phi f_2}\|_2^2}{1 - \delta_{2s}}$ which is a good estimate if δ_{2s} is small 1 point. Note that both upper and lower bounds need to be given. Otherwise, only 2 points are to be awarded..
   (b) $\boldsymbol{f_1}^t\boldsymbol{f_2}$, if you knew that $\|\boldsymbol{f_1}\|_2^2 = \|\boldsymbol{f_2}\|_2^2 = 1$: Since the sensing matrix satisfied RIP of order $2s$, we know that $(1 - \delta_{2s}) \leq \dfrac{\|\boldsymbol{\Phi f_1} - \boldsymbol{\Phi f_2}\|^2}{\|\boldsymbol{f_1} - \boldsymbol{f_2}\|^2} = \dfrac{\|\boldsymbol{\Phi f_1} - \boldsymbol{\Phi f_2}\|^2}{2 - 2\boldsymbol{f_1}^t\boldsymbol{f_2}} \leq (1 + \delta_{2s})$. Now, we note that $(\boldsymbol{\Phi f_1})^t\boldsymbol{\Phi f_2} = \dfrac{|\|\boldsymbol{\Phi f_1} + \boldsymbol{\Phi f_2}\|_2^2 - \|\boldsymbol{\Phi f_1} - \boldsymbol{\Phi f_2}\|_2^2|}{4} \leq \dfrac{(1 + \delta_{2s})(1 + \boldsymbol{f_1}^t\boldsymbol{f_2}) - (1 - \delta_{2s})(1 + \boldsymbol{f_1}^t\boldsymbol{f_2})}{2} = \boldsymbol{f_1}^t\boldsymbol{f_2} + \delta_{2s}$. Likewise, we also note that $(\boldsymbol{\Phi f_1})^t\boldsymbol{\Phi f_2} = \dfrac{|\|\boldsymbol{\Phi f_1} + \boldsymbol{\Phi f_2}\|_2^2 - \|\boldsymbol{\Phi f_1} - \boldsymbol{\Phi f_2}\|_2^2|}{4} \geq \dfrac{(1 - \delta_{2s})(1 + \boldsymbol{f_1}^t\boldsymbol{f_2}) - (1 + \delta_{2s})(1 + \boldsymbol{f_1}^t\boldsymbol{f_2})}{2} = \boldsymbol{f_1}^t\boldsymbol{f_2} - \delta_{2s}$.

Thus we have $\boldsymbol{f_1}^t\boldsymbol{f_2} \in [(\boldsymbol{\Phi f_1})^t\boldsymbol{\Phi f_2} - \delta_{2s}, (\boldsymbol{\Phi f_1})^t\boldsymbol{\Phi f_2} + \delta_{2s}]$. Both upper and lower bounds need to be given. Only stating that $\boldsymbol{f_1}^t\boldsymbol{f_2}$ is approximately equal to $\boldsymbol{\Phi f_1}^t\boldsymbol{\Phi f_2}$ will fetch only 1 point.

(c) List any one practical application of being able to estimate $\|\boldsymbol{f_1} - \boldsymbol{f_2}\|_2^2$ from $\boldsymbol{y_1}, \boldsymbol{y_2}$ in machine learning or image retrieval: This offers a method dimensionality reduction which preserves distances between pairs of vectors approximately. Hence it allows for more efficient image or signal retrieval as it allows for searching in a lower dimensional space. [2 points]

5. Consider that you are given the Radon projections of an image $f(x, y)$ (defined on domain Ω), in directions $\theta_1, \theta_2, ..., \theta_K; K > 1$. *Without* reconstructing the image, state how you will infer the following properties of the image *directly* from the projections? [5+5=10 points]

   (a) $\sum_{(x,y)\in\Omega} f(x, y)$

   (b) A slice of the 2D Fourier transform of f in direction θ_1 in the frequency plane and passing through the origin of the frequency plane

   Answer:

   (a) $\sum_{(x,y)\in\Omega} f(x, y) = \sum_\rho g(\rho, \theta_1) = ... = \sum_\rho g(\rho, \theta_N)$.

   (b) The 1D Fourier of any projection vector gives the desired slice through the 2D Fourier transform of the image, by the Fourier slice theorem.

6. Can a Gaussian random matrix be used as a pooling matrix for RTPCR group testing? Why (not)? Can a Gaussian random matrix be used as the sensing matrix for the snapshot-based video compressed sensing camera? Why (not)? [5+5=10 points]
   Answer: A random Gaussian matrix cannot be used as a pooling matrix because a pooling matrix needs to be binary (a sample either participates in a pool or it doesn't) [2 points] and sparse [2 points] for simplicity of pooling and cannot contain any negative values [1 point]. A Gaussian random cannot be used in the snapshot-based video compressed sensing camera because the sensing matrix in this camera needs to be a column-wise concatenation of diagonal matrices [2 points] which have ones or zeros on its diagonal [2 points] to modulate the incoming sub-frames using an array of DMD mirrors of the LCoS device [1 point]. A Gaussian random matrix does not satisfy this criterion.

7. Consider the OMP algorithm to estimate sparse signal $\boldsymbol{x}$ from its compressive measurements $\boldsymbol{y} = \boldsymbol{\Phi x}$ given a sensing matrix $\boldsymbol{\Phi}$. Why does OMP never re-select a column of $\boldsymbol{\Phi}$ that was selected in some previous iteration? [6 points] If you knew that $\boldsymbol{x}$ had exactly one non-zero element, how many measurements are *necessary* to uniquely compute $\boldsymbol{x}$ from $\boldsymbol{y}, \boldsymbol{\Phi}$, using the best possible method that exists? Assume that $\boldsymbol{\Phi}$ obeys RIP of appropriate order. [4 points]
   Answer: If the current support set estimated by OMP is T, then OMP re-estimates the coefficients by minimization of $\|\boldsymbol{y} - \boldsymbol{\Phi_T \alpha}\|_2^2$. The resulting residual vector $\boldsymbol{r} = \boldsymbol{y} - \boldsymbol{\Phi_T \alpha}$ is necessarily orthogonal to every column of $\boldsymbol{\Phi_T}$. In each iteration, elements only get added to T and never removed. Hence, the columns selected in T will remain perpendicular to $\boldsymbol{r}$. The columns that get selected for inclusion in T are those for which the absolute normalized dot product is the highest, and thus columns from $\boldsymbol{\Phi_T}$ will never get selected. 6 points for the correct answer. Detailed calculations needs not be shown, but two aspects need to mentioned: (1) the residual is orthogonal to all selected columns vectors with indices in T, and (2) the next column vector is chosen based on highest absolute dot product with the residual. If only one of these aspects is mentioned, then only 3 points are to be awarded.

   The best algorithm to compute $\boldsymbol{x}$ in case it had just one non-zero element is to compute $\beta_i = y_1/\Phi_{1,i}$ for every $i \in [n]$. Determine for which i, we have $y_2/\Phi_{2,i} = \beta_i$. There cannot exist more than one such index because $\boldsymbol{\Phi}$ obeys order-2 RIP and hence it obeys the nullspace property for order 2 if it contains 2 measurements. If you used LASSO or OMP, you will need $O(\log n)$ measurements. 4 points for this correct answer. If the student mentions using OMP or LASSO, only two points will be awarded.