Assignment 2: CS 754, Advanced Image Processing

Due: 16th Feb before 11:55 pm

1. Refer to a copy of the paper 'The restricted isometry property and its implications for compressed sensing' in the homework folder. Your task is to open the paper and answer the question posed in each and every green-colored highlight. The task is the complete proof of Theorem 3 done in class. [32 points = 2 points for each of the 16 questions]
   Solution:

   (a) Q1: We know that if $\delta_{2s}2$ is the order-$2s$ RIC of a matrix $\boldsymbol{\Phi}$, then we have $(1-\delta_{2s})\|\boldsymbol{x}\|^2 \leq \|\boldsymbol{\Phi x}\|^2 \leq (1+\delta_{2s})\|\boldsymbol{x}\|^2$ for any $2s$-sparse vector $\boldsymbol{x}$. If $\delta_{2s} = 1$, then we could have $\boldsymbol{Ax} = \mathbf{0}$ for $\underline{\text{some}}$ $2s$-sparse $\boldsymbol{x}$, which would mean that there exist some set of $2s$ columns from $\boldsymbol{A}$ which are linearly]emphnot independent.

   (b) Q2: This holds due to the triangle inequality satisfied by the vector 2-norm and also because of the constraint imposed in the optimization problem.

   (c) Q3: This is true because $\|\mathbf{h}_{T_j}\|_2 = \sqrt{\sum_{i=1}^{s} h^2_{T_j,i}} \leq \sqrt{\sum_{i=1}^{s} \|\mathbf{h}_{T_j}\|^2_\infty} = s^{1/2}\|\mathbf{h}_{T_j}\|_\infty$. Also, notice that $s^{1/2}\|\mathbf{h}_{T_j}\|_\infty \leq \dfrac{s^{1/2}}{s} \sum_{i=1}^{s} \|\mathbf{h}_{T_j}\|_\infty \leq s^{-1/2}\|\mathbf{h}_{T_{j-1}}\|_1$. Also $\underline{\text{any}}$ element of $\mathbf{h}_{T_j}$ (including $\|\mathbf{h}_{T_j}\|_\infty$) is less than or equal to $\underline{\text{any}}$ element of $\mathbf{h}_{T_{j-1}}$.

   (d) Q4: $\sum_{j\geq 2} \|\mathbf{h}_{T_j}\|_2 \leq s^{-1/2} \sum_{j\geq 1} \|\mathbf{h}_{T_j}\|_1 = s^{-1/2}\|\mathbf{h}_{T_0^c}\|_1$. The last equality is because $T_0^c = T_1 \cup T_2 \cup .....$ The first inequality holds due to a simple summation starting from the previous relation $\|\boldsymbol{h}_{\boldsymbol{T_j}}\|_2 \leq s^{-1/2}\|\boldsymbol{h}_{\boldsymbol{T_{j-1}}}\|_1$.

   (e) Q5: We have $\|\mathbf{h}_{(T_0\cup T_1)^c}\|_2 = \|\sum_{j\geq 2} \mathbf{h}_{T_j}\|_2$ as $\forall j, \mathbf{h}_{T_j}$ have disjoint support. The next inequality follows by triangle inequality and the last one is because we earlier proved that $\sum_{j\geq 2} \|\mathbf{h}_{T_j}\|_2 \leq s^{-1/2}\|\boldsymbol{h}_{\boldsymbol{T_0^c}}\|_1$.

   (f) Q6: Reverse Triangle inequality on $|x_i + h_i|$ in two different directions.

   (g) Q7: Directly uses the previous equation from the paper and re-arranges the terms.

   (h) Q8: This is almost directly given in the paper via equations 11 and 12.

   (i) Q9: This is due to the Cauchy Schwartz inequality.

   (j) Q10: This comes from Lemma 2.1 in the paper, but extended to vectors that have magnitude greater than 1. See lecture slides for more details.

   (k) Q11: Consider a 2-element vector $\mathbf{w} = (\|\mathbf{h_{T_0}}\|_2\|\mathbf{h_{T_1}}\|_2)$. Then $\|\mathbf{w}\|_1 = \|\mathbf{h_{T_0}}\|_2 + \|\mathbf{h_{T_1}}\|_2$. We know that $\|\mathbf{w}\|_1 \leq \sqrt{2}\|\mathbf{w}\|_2 = \sqrt{2}\|\mathbf{h_{T_0\cup T_1}}\|_2$ since the support sets T_0 and T_1 are disjoint.

   (l) Q12: We need to be very careful here as so many steps are involved! From the RIP of $\boldsymbol{\Phi}$ for order $2s$ with RIC δ_{2s}, we know that $(1-\delta_{2s})\|\mathbf{h}_{T_0\cup T_1}\|_2^2 \leq \|\boldsymbol{\Phi}\mathbf{h}_{T_0\cup T_1}\|_2^2 =$
   $\langle \boldsymbol{\Phi}\mathbf{h_{T_0\cup T_1}}, \boldsymbol{\Phi}\mathbf{h}\rangle - \langle \boldsymbol{\Phi}\mathbf{h_{T_0\cup T_1}}, \sum_{j\geq 2} \boldsymbol{\Phi}\mathbf{h_{T_j}}\rangle$(as shown in Q14)
   $\leq |\langle \boldsymbol{\Phi}\mathbf{h_{T_0\cup T_1}}, \boldsymbol{\Phi}\mathbf{h}\rangle| + |\langle \boldsymbol{\Phi}\mathbf{h_{T_0\cup T_1}}, \sum_{j\geq 2} \boldsymbol{\Phi}\mathbf{h_{T_j}}\rangle|$
   $\leq 2\epsilon\sqrt{1+\delta_{2s}}\|\mathbf{h}_{T_0\cup T_1}\|_2 + |\langle \boldsymbol{\Phi}\mathbf{h_{T_0\cup T_1}}, \sum_{j\geq 2} \boldsymbol{\Phi}\mathbf{h_{T_j}}\rangle|$
   (using Cauchy-Schwartz inequality, equation 9 from the paper and right side of RIP)
   $\leq 2\epsilon\sqrt{1+\delta_{2s}}\|\mathbf{h}_{T_0\cup T_1}\|_2 + |\langle \boldsymbol{\Phi}\mathbf{h_{T_0}} + \boldsymbol{\Phi}\mathbf{h_{T_1}}, \sum_{j\geq 2} \boldsymbol{\Phi}\mathbf{h_{T_j}}\rangle|$
   (as T_0 and T_1 are disjoint sets and hence $\boldsymbol{\Phi}\mathbf{h}_{T_0\cup T_1} = \boldsymbol{\Phi}\mathbf{h}_{T_0} + \boldsymbol{\Phi}\mathbf{h}_{T_1}$)
   $\leq 2\epsilon\sqrt{1+\delta_{2s}}\|\mathbf{h}_{T_0\cup T_1}\|_2 + \delta_{2s}(\|\mathbf{h_{T_0}}\|_2 + \|\mathbf{h_{T_1}}\|_2)\|\sum_{j\geq 2} \|\mathbf{h_{T_j}}\|_2$ from Lemma 2.1 of the paper
   $\leq 2\epsilon\sqrt{1+\delta_{2s}}\|\mathbf{h}_{T_0\cup T_1}\|_2 + \delta_{2s}\sqrt{2}(\|\mathbf{h_{T_0\cup T_1}}\|_2)\|\sum_{j\geq 2} \|\mathbf{h_{T_j}}\|_2$.

(m) Q13: This follows straightforwardly from the previous step. Just divide the leftmost and rightmost sides by $\|\boldsymbol{h}_{T_0 \cup T_1}\|_2$, and from equation 10 of the paper.

(n) Q14: Follows from straightforward algebra using equation 12 of the paper.

(o) Q15: Follows from triangle inequality.

(p) Q16: This follows in a very straightforward way using Lemma 2.2 from the paper which shows that $\|\boldsymbol{h}_{\boldsymbol{T_0}}\|_1 \leq \rho \|\boldsymbol{h}_{\boldsymbol{T_0^c}}\|_1$. The paper has already derived a bound for $\|\boldsymbol{h}_{\boldsymbol{T_0^c}}\|_1$. This produces the bound for $\|\boldsymbol{h}\|_1$. (Food for thought: can this be easily extended for the noisy case as well? Why (not?))

2. Consider compressive measurements $\boldsymbol{y} = \boldsymbol{\Phi x} + \boldsymbol{\eta}$ of a purely sparse signal $\boldsymbol{x}$, where $\|\boldsymbol{\eta}\|_2 \leq \epsilon$. When we studied Theorem 3 in class, I had made a statement that the solution provided by the basis pursuit problem for a purely sparse signal comes very close (i.e. has an error that is only a constant factor worse than) an oracular solution. An oracular solution is defined as the solution that we could obtain if we knew in advance the indices (set S) the non-zero elements of the signal $\boldsymbol{x}$. This homework problem is to understand my statement better. For this, do as follows. In the following, we will assume that the inverse of $\boldsymbol{\Phi_S^T \Phi_S}$ exists, where $\boldsymbol{\Phi_S}$ is a submatrix of $\boldsymbol{\Phi}$ with columns belonging to indices in S.

   (a) Express the oracular solution $\tilde{\boldsymbol{x}}$ using a pseudo-inverse of the sub-matrix $\boldsymbol{\Phi_S}$. [5 points]
   Solution: $\tilde{\boldsymbol{x}} = \boldsymbol{\Phi_S}^{\dagger}\boldsymbol{y}$.

   (b) Now, show that $\|\tilde{\boldsymbol{x}} - \boldsymbol{x}\|_2 = \|\boldsymbol{\Phi}_{\boldsymbol{S}}^{\dagger}\boldsymbol{\eta}\|_2 \leq \|\boldsymbol{\Phi}_{\boldsymbol{S}}^{\dagger}\|_2 \|\boldsymbol{\eta}\|_2$. Here $\boldsymbol{\Phi}_{\boldsymbol{S}}^{\dagger} \triangleq (\boldsymbol{\Phi_S^T \Phi_S})^{-1}\boldsymbol{\Phi_S^T}$ is standard notation for the pseudo-inverse of $\boldsymbol{\Phi_S}$. The largest singular value of matrix $\boldsymbol{X}$ is denoted as $\|\boldsymbol{X}\|_2$. [3 points]
   Solution: We have $\|\boldsymbol{x} - \tilde{\boldsymbol{x}}\|_2 = \|(\boldsymbol{\Phi_S}^T\boldsymbol{\Phi_S})^{-1}\boldsymbol{\Phi_S}^T(\boldsymbol{\Phi x} + \boldsymbol{\eta}) - \boldsymbol{x}\|_2 = \|\boldsymbol{\Phi_S}^{\dagger}\boldsymbol{\eta}\|_2 \leq \|\boldsymbol{\Phi_S}^{\dagger}\|_2\|\boldsymbol{\eta}\|_2$. The last inequality follows by the definition of the matrix operator norm (i.e. the largest singular value of the matrix).

   (c) Argue that the largest singular value of $\boldsymbol{\Phi}_{\boldsymbol{S}}^{\dagger}$ lies between $\dfrac{1}{\sqrt{1+\delta_{2k}}}$ and $\dfrac{1}{\sqrt{1-\delta_{2k}}}$ where $k = |S|$ and δ_{2k} is the RIC of $\boldsymbol{\Phi}$ of order $2k$. [4 points]
   Solution: If $\boldsymbol{\Phi}$ obeys RIP of order $2k$ with RIC δ_{2k}, then the singular values of $\boldsymbol{\Phi_S}$ must satisfy $\sqrt{1-\delta_{2k}} \leq \lambda_{min} \leq \lambda_{max} \leq \sqrt{1+\delta_{2k}}$. (This is by the very definition of singular value.) Hence the largest singular value of $\boldsymbol{\Phi_S}^{\dagger}$ (which is equal to $\dfrac{1}{\lambda_{min}}$) must lie in the range from $\dfrac{1}{\sqrt{1+\delta_{2k}}}$ to $\dfrac{1}{\sqrt{1-\delta_{2k}}}$.

   (d) This yields $\dfrac{\epsilon}{\sqrt{1+\delta_{2k}}} \leq \|\boldsymbol{x} - \tilde{\boldsymbol{x}}\|_2 \leq \dfrac{\epsilon}{\sqrt{1-\delta_{2k}}}$. Argue that the solution given by Theorem 3 is only a constant factor worse than this solution. [3 points]

   Solution: The (upper) error bound given by Theorem 3 for purely sparse signals is $\dfrac{4\sqrt{1+\delta_{2s}}}{1-\delta_{2s}(\sqrt{2}+1)}$. This upper bound is worse than the oracle solutions. To see this, notice that the denominator of the oracle solution is larger (since it involves a square root of a value between 0 and 1). A comment about the lower bound: Just as we have $\|\boldsymbol{x} - \tilde{\boldsymbol{x}}\|_2 \leq \|\boldsymbol{\Phi_S}^{\dagger}\|_2\varepsilon$, we also have $\|\boldsymbol{x} - \tilde{\boldsymbol{x}}\|_2 = \|\boldsymbol{\Phi}_{\boldsymbol{S}}^{\dagger}\boldsymbol{\eta}\|_2 \geq \|\boldsymbol{\Phi}_{\boldsymbol{S}}^{\dagger}\|_{min}\epsilon$, where $\|\boldsymbol{\Phi}_{\boldsymbol{S}}^{\dagger}\|_{min}$ represents the least singular value of $\boldsymbol{\Phi}_{\boldsymbol{S}}^{\dagger}$. The lower bound argument is not required and no points to be deducted for missing out on it.

3. If $s < t$ where s and t are positive integers, prove that $\delta_s \leq \delta_t$ where δ_s, δ_t stand for the restricted isometry constant (of any sensing matrix) of order s and t respectively. [8 points]
Solution: We have $\delta_s < \delta_t$ for $s < t$ because $\delta_s = \max_{\Gamma_1 \subset [n], |\Gamma_1| \leq s} \|\boldsymbol{A}_{\boldsymbol{\Gamma_1}}{}^T\boldsymbol{A}_{\boldsymbol{\Gamma_1}} - \boldsymbol{I_s}\|_2$, where $[n] = \{1, 2, ..., n\}$ and $\|\boldsymbol{B}\|_2$ for a matrix means its operator norm (i.e. its largest singular value). In English, this means that δ_s is the largest singular value of any sub-matrix of $\boldsymbol{A}$ with at the most s columns. Likewise, δ_t is the largest singular value of any sub-matrix of $\boldsymbol{A}$ with at the most t columns. Since the maximum is being taken over a larger set of sub-matrices, it will potentially be greater than the earlier one (or equal to the earlier one, but never lower than the earlier one). Hence $\delta_s \leq \delta_t$.
Marking scheme: Correct expression of the RIC in terms of eigenvalues of $\boldsymbol{A}_{\boldsymbol{\Gamma_1}}{}^T\boldsymbol{A}_{\boldsymbol{\Gamma_1}}$ will fetch 4 points. 4 points for the rest of the argument.

4. Here is our obligatory Google search question :-). Your task is to search the web for papers that used some technique of sensing matrix design (eg: coherence minimization, RIC minimization, or any other) to improve

the performance of a practical compressive imaging system. (Hint: Look at the archives of journals such as IEEE Transactions on Computational Imaging, IEEE Transactions on Image Processing, Applied Optics or the webpages of authors such as David Brady or Gonzalo Arce). To answer this question, do the following:

(a) Mention the title, venue, author list publication year of the paper. Put a link to it.

(b) Briefly describe the imaging system in the paper; you may refer to figures from the paper itself or refer to lecture slides.

(c) Write the mathematical expression for the matrix quality measure being optimized in the paper, along with various contraints on the matrix (eg: non-negative elements, block diagonal, etc.).

(d) Mention the optimization technique.

(e) Briefly describe the improvements due to this design as compared to a random design. You may refer to tables or graphs from the paper itself.

[3 +3 + 3 + 3 + 3 =15 points]
Solution: I am considering the paper 'Colored Coded Aperture Design in Compressive Spectral Imaging via Minimum Coherence' published in IEEE Trans. Comp. Imaging in June 2017. It is authored by Parada-Mayorga and Arce. Link: `https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=7895113`.
The imaging system used in the paper is basically CASSI, with one change. The original system used a coded aperture that is binary, i.e. it blocks light of all wavelengths or allows all to pass. The aperture used in the system selectively allows light of certain bandwidth ranges to pass or else blocks that range. These are called 'colored coded apertures' (read the very first paragraph of the paper, after the abstract).
The matrix quality measure being optimized over is the mutual coherence of the matrix $A \triangleq H\psi$ where H stands for the CASSI sensing matrix and ψ is the image representation matrix (sparisfying). The matrix H has a structrue given in figure 2. The appropriate non-zero entries of this matrix are obtained from the coded aperture pattern given by $T(x, y, \lambda)$. The constraints imposed are that the entries of H are non-negative with $\sum_{\lambda} T(x, y, \lambda) = 1$. An additional constraint is placed on the number of color filters at any pixel to avoid too much cost. There is another criterion which is an upper bound on coherence, denoted as ϕ. See equation 13 of the paper.
The optimization technique used is a combinatorial technique to design various sub-matrices.
The designed matrix gives superior results to random binary as seen in Figures 6, 7, 8.
Another related paper, a simpler one, for CASSI is `https://www.eecis.udel.edu/~arce/files/Publications/ArceMagazine06678264-2.pdf`. This one optimizes Boolean CASSI codes on RIC of a fixed order and obtains superior results to random binary codes. See figure 6 of the paper.

5. Consider the problem P1: $\min_{\boldsymbol{x}} \|\boldsymbol{x}\|_1$ s. t. $\|\boldsymbol{y} - \boldsymbol{\Phi x}\|_2 \leq \epsilon$. Also consider the LASSO problem which seeks to minimize the cost function $J(\boldsymbol{x}) = \|\boldsymbol{y} - \boldsymbol{\Phi x}\|_2^2 + \lambda \|\boldsymbol{x}\|_1$. If $\boldsymbol{x}$ is a minimizer of $J(.)$ for some value of $\lambda > 0$, then show that there exists some value of ϵ for which $\boldsymbol{x}$ is also the minimizer of the problem P1. [6 points] (Hint: Consider $\epsilon' = \|\boldsymbol{y} - \boldsymbol{\Phi x}\|_2$. Now use the fact that $\boldsymbol{x}$ is a minimizer of $J(.)$ to show that it is also a minimizer of P1 subject to an appropriate constraint involving ϵ'.)
Solution: Consider $\epsilon' = \|\boldsymbol{y} - \boldsymbol{\Phi x}\|_2$. Consider some vector $\boldsymbol{z}$ with n elements such that $\|\boldsymbol{Az} - \boldsymbol{y}\|_2 \leq \epsilon'$. As $\boldsymbol{x}$ is the minimizer of $J(.)$, we have $\lambda \|\boldsymbol{x}\|_1 + \|\boldsymbol{y} - \boldsymbol{Ax}\|^2 \leq \lambda \|\boldsymbol{z}\|_1 + \|\boldsymbol{y} - \boldsymbol{Az}\|^2 \leq \lambda \|\boldsymbol{z}\|_1 + \|\boldsymbol{y} - \boldsymbol{Ax}\|^2$ (3 points for this step). This yields $\|\boldsymbol{x}\|_1 \leq \|\boldsymbol{z}\|_1$ (1.5 points for this step). Hence $\boldsymbol{x}$ is a minimizer of P1 subject to the constraint involving ϵ' (1.5 points for this step).

6. Suppose there are n subjects being tested by Dorfman pooling and only $k \ll n$ out of these are infected. In the first round, assume that the n subjects are divided into groups of size g each. For simplicity, assume n/g is an integer. Derive a formula for the average number of tests required to be performed in Dorfman pooling. What is the worst case? What is the optimal group size in the worst case? [15 points]
Solution: The n subjects are divided into n/g groups, each of size g. So the first round requires n/g tests always, as there will be n/g different groups. There are k infected people out of n. The expected number is calculated as follows: We know the prevalence rate to be $p \triangleq k/n$. So the probability of any one individual being infected is p. Hence the probability of any one individual being non-infected is $1-p$. The probability of obtaining a group of g individuals who are all non-infected is $(1-p)^g$, and hence the probability of obtaining

a group of g individuals containing at least one infected individual is $1-(1-p)^g$. As the total number of groups is n/g, the expected number of groups with at least one infected member is $n/g \times [1-(1-p)^g]$. Hence the expected number of total tests is $n/g + g \times n/g \times [1-(1-p)^g] = n/g + n[1-(1-p)^g]$.
In the worst case, each of the k infected people will lie in a different group (i.e. from the n/g groups). So each of these k groups will have to be tested in the second round, and each member of these k groups will be tested individually. This will give rise to $k \times g$ more tests. So the total number is $n/g + kg$ for the worst case number of tests. If the worst case number has to be optimal, we set the derivative of this number (w.r.t. g) to zero, giving rise to $-n/g^2 + k = 0$, that is $g = \sqrt{n/k}$.
Remark: It is easy to get a closed form expression for the optimal g in the worst case, but not so in the average case.
Marking scheme: 5 points for expected number, 5 points for worst case, 5 point for optimal value of g in the worst case.