

# Assignment 1: CS 754, Advanced Image Processing

Due: 9th Feb before 11:55 pm

**Remember the honor code while submitting this (and every other) assignment. All members of the group should work on and *understand* all parts of the assignment. Exchange of answers between groups is not allowed. We will adopt a zero-tolerance policy against any violation, and we will expressly check for plagiarism.**

**Submission instructions:** You should ideally type out all the answers in Latex or else in MS Word with the equation editor. In either case, prepare a pdf file. Create a single zip or rar file containing the report, code and sample outputs and name it as follows: A1-IdNumberOfFirstStudent-IdNumberOfSecondStudent.zip. (If you are doing the assignment alone, the name of the zip file is A1-IdNumber.zip). Upload the file on moodle BEFORE 11:55 pm on 9th Feb. Beyond the cutoff time of 7 am on 10th Feb, no assignments will be accepted. Note that only one student per group should upload their work on moodle. Please preserve a copy of all your work until the end of the semester. *If you have difficulties, please do not hesitate to seek help from me.*

1. Let  $\theta^*$  be the result of the following minimization problem (BP):  $\min \|\theta\|_1$  such that  $\|y - \Phi\Psi\theta\|_2 \leq \varepsilon$ , where  $y$  is an  $m$ -element measurement vector,  $\Phi$  is a  $m \times n$  measurement matrix ( $m < n$ ),  $\Psi$  is a  $n \times n$  orthonormal basis in which  $n$ -element signal  $x$  has a sparse representation of the form  $x = \Psi\theta$ . Notice that  $y = \Phi x + \eta$  and  $\varepsilon$  is an upper bound on the magnitude of the noise vector  $\eta$ .

Theorem 3 we studied in class states the following: If  $\Phi$  obeys the restricted isometry property with isometry constant  $\delta_{2s} < \sqrt{2} - 1$ , then we have  $\|\theta - \theta^*\|_2 \leq C_1 s^{-1/2} \|\theta - \theta_s\|_1 + C_2 \varepsilon$  where  $C_1$  and  $C_2$  are functions of only  $\delta_{2s}$  and where  $\forall i \in \mathcal{S}, \theta_{si} = \theta_i; \forall i \notin \mathcal{S}, \theta_{si} = 0$ . Here  $\mathcal{S}$  is a set containing the  $s$  largest magnitude elements of  $\theta$ .

A curious student asks the following questions: ‘(1) It appears that the upper bound on  $\|\theta - \theta^*\|_2$  is reduced as  $s$  increases, which goes against the very premise of compressed sensing. How do we address this apparent discrepancy? (2) It also appears that the error bound is independent of  $m$ . How do you address this? (3) Now consider that I gave you another theorem (called Theorem 3A), which is the same as Theorem 3 except that it requires that  $\delta_{2s} < 0.3$ . Out of Theorem 3 and Theorem 3A, which is the more useful theorem? Why? (4) It appears that if I set  $\varepsilon = 0$  in BP, I can always reduce the upper bound on the error even if the noise vector  $\eta$  has non-zero magnitude. What am I missing?’

Your job is to answer all four of the student’s questions. [5+5+5+5=20 points]

2. In class, we studied a video compressive sensing architecture from the paper ‘Video from a single exposure coded snapshot’ published in ICCV 2011 (See [http://www.cs.columbia.edu/CAVE/projects/single\\_shot\\_video/](http://www.cs.columbia.edu/CAVE/projects/single_shot_video/)). Such a video camera acquires a ‘coded snapshot’  $E_u$  in a single exposure time interval  $u$ . This coded snapshot is the superposition of the form  $E_u = \sum_{t=1}^T C_t \cdot F_t$  where  $F_t$  is the image of the scene at instant  $t$  within the interval  $u$  and  $C_t$  is a randomly generated binary code at that time instant, which modulates  $F_t$ . Note that  $E_u$ ,  $F_t$  and  $C_t$  are all 2D arrays. Also, the binary code generation as well as the final summation all occur within the hardware of the camera. Your task here is as follows:
  - (a) Read the ‘cars’ video in the homework folder in MATLAB using the ‘mmread’ function which has been provided in the homework folder and convert it to grayscale. Extract the first  $T = 3$  frames of the video.
  - (b) Generate a  $H \times W \times T$  random code pattern whose elements lie in  $\{0, 1\}$ . Compute a coded snapshot using the formula mentioned and add zero mean Gaussian random noise of standard deviation 2 to it. Display the coded snapshot in your report.

- (c) Given the coded snapshot and assuming full knowledge of  $C_t$  for all  $t$  from 1 to  $T$ , your task is to estimate the original video sequence  $F_t$ . For this you should rewrite the aforementioned equation in the form  $\mathbf{Ax} = \mathbf{b}$  where  $\mathbf{x}$  is an unknown vector (vectorized form of the video sequence). Mention clearly what  $\mathbf{A}$  and  $\mathbf{b}$  are, in your report.
  - (d) You should perform the reconstruction using Orthogonal Matching Pursuit (OMP). For computational efficiency, we will do this reconstruction patchwise. Write an equation of the form  $\mathbf{Ax} = \mathbf{b}$  where  $\mathbf{x}$  represents the  $i^{th}$  patch from the video and having size (say)  $8 \times 8 \times T$  and mention in your report what  $\mathbf{A}$  and  $\mathbf{b}$  stand for. For perform the reconstruction, assume that each  $8 \times 8$  slice in the patch is sparse or compressible in the 2D-DCT basis. Carefully work out the error term in the OMP algorithm, and explain this in your report!
  - (e) Repeat the reconstruction for all overlapping patches and average across the overlapping pixels to yield the final reconstruction. Display the reconstruction and mention the relative mean squared error between reconstructed and original data, in your report as well as in the code.
  - (f) Repeat this exercise for  $T = 5, T = 7$  and mention the relative mean squared error between reconstructed and original data again.
  - (g) **Note: To save time, extract a portion of about  $120 \times 240$  around the lowermost car in the cars video and work entirely with it. In fact, you can show all your results just on this part. Some sample results are included in the homework folder.**
  - (h) Repeat the experiment with any consecutive 5 frames of the ‘flame’ video from the homework folder. [35 points = 18 points for successful OMP implementation + 7 points for carefully presenting error term bound + 10 points for displaying of all results]
3. We will prove why the value of the coherence between  $m \times n$  measurement matrix  $\Phi$  (with all rows normalized to unit magnitude) and  $n \times n$  orthonormal representation matrix  $\Psi$  must lie within the range  $(1, \sqrt{n})$ . Recall that the coherence is given by the formula  $\mu(\Phi, \Psi) = \sqrt{n} \max_{i,j \in \{0,1,\dots,n-1\}} |\Phi^{i^t} \Psi_j|$ . Proving the upper bound should be very easy for you. To prove the lower bound, proceed as follows. Consider a unit vector  $\mathbf{g} \in \mathbb{R}^n$ . We know that it can be expressed as  $\mathbf{g} = \sum_{k=1}^n \alpha_k \Psi_k$  as  $\Psi$  is an orthonormal basis. Now prove that  $\mu(\mathbf{g}, \Psi) = \sqrt{n} \max_{i \in \{0,1,\dots,n-1\}} \frac{|\alpha_i|}{\sum_{j=1}^n \alpha_j^2}$ . Exploiting the fact that  $\mathbf{g}$  is a unit vector, prove that the minimal value of coherence is attained when  $\mathbf{g} = \frac{1}{\sqrt{n}} \sum_{k=1}^n \Psi_k$  and that hence the minimal value of coherence is 1. [10 points]
  4. Compressive sensing reconstructions involve estimating a sparse signal  $\mathbf{x} \in \mathbb{R}^n, n \gg 2$  from a vector  $\mathbf{y} \in \mathbb{R}^m (m \ll n)$  of compressed measurements of the form  $\mathbf{y} = \Phi \mathbf{x}$  where  $\Phi \in \mathbb{R}^{m \times n}$  is the measurement matrix (assume there is no noise). Now answer the following questions, from first principles. **Do not merely quote theorems or algorithms.**
    - (a) If  $\mathbf{x}$  has only 1 non-zero element and the other elements are zero, can you uniquely estimate  $\mathbf{x}$  if  $m = 1$ ? If yes, how? If not, why not? Now further suppose, you knew beforehand the index (but not the value) of the non-zero element of  $\mathbf{x}$ ? Does this help you any further? If yes, how? If not, why not?
    - (b) If  $\mathbf{x}$  has only 1 non-zero element and the other elements are zero, can you uniquely estimate  $\mathbf{x}$  if  $m = 2$ ? If yes, how? If not, why not?
    - (c) If  $\mathbf{x}$  has only 2 non-zero elements and the other elements are zero, can you uniquely estimate  $\mathbf{x}$  if  $m = 3$ ? If yes, describe an algorithm that is guaranteed to estimate it accurately. If not, explain why not, and explain whether there are any special instances of  $\Phi$  for which unique estimation is possible?
    - (d) Repeat part (c) with  $m = 4$ . [1+2+3+4=10 points]
  5. Read the paper ‘Coded aperture compressive temporal imaging’ which is a type of video compressive sensing camera developed at Duke University. The paper can be accessed here: <https://www.osapublishing.org/oe/fulltext.cfm?uri=oe-21-9-10526&id=253002>. A local copy is also placed in the homework folder. Answer the following questions:

- (a) Explain the similarities and differences between this architecture and the video compressed sensing architecture using coded snapshots (Hitomi et al, ICCV 2011) that we studied in class.
  - (b) What cost function is the reconstruction algorithm in the paper based on? Explain the meaning and dimensions of every term in the cost function. (Note: you are not expected to describe the algorithms that minimize this cost function, but only mention the cost function itself) [8+7=15 points]
6. Here is our mandatory Google search question. Note that this is the only question for which you can perform a google search to get the answer. A very interesting application of compressed sensing is in the area of optical or electron microscopy. Your task is to search for a research paper which applies compressed sensing in this application. Answer the following questions briefly:
- (a) Mention the title of the paper, where and when it was published and include a link to it.
  - (b) Very briefly describe the hardware architecture used in the paper. You may refer to figures from the paper itself.
  - (c) What reconstruction technique or cost function does the paper adopt for the sake of compressive reconstruction in this application? [3+4+4=10 points]