

# Assignment 4: CS 754, Advanced Image Processing

Due: 4th April 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. We will adopt a zero-tolerance policy against any violation.**

**Submission instructions:** You should ideally type out all the answers in Word (with the equation editor) or using Latex. 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: A4-IdNumberOfFirstStudent-IdNumberOfSecondStudent.zip. (If you are doing the assignment alone, the name of the zip file is A4-IdNumber.zip). Upload the file on moodle BEFORE 11:55 pm on the due date. The cutoff is 10 am on 5th April after which 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.

## Instructions for Coding Questions

1. Make a subfolder in the submission folder. Name the folder 'media'.
2. The directory structure should look like :

```
A4-<Roll_No_1>-<Roll_No_2>-<Roll_No_3>
|
|_____media
|_____<other_file_1>
|_____<other_file_2>
|_____
|_____
|_____<other_file_n>
```

3. Read ANY image/video in ANY code from this folder(media) itself.
4. ALL the images/videos required for ANY code should be present in the folder 'media' itself, if your final compressed submission folder size DOES NOT EXCEED THE MOODLE SIZE LIMIT.
5. The TAs will copy all the images/video to the folder 'media' at the time of evaluation, if your final compressed submission folder DOES EXCEED THE MOODLE SIZE LIMIT. In this case leave the 'media' folder blank.
6. Please ensure that all the codes run at the click of a single go (RUN button) in MATLAB.
7. Please ensure that all the asked result images/videos, plots and graphs pop up at the click of a single go (RUN button) in MATLAB, while running the corresponding code for any question.
8. The result images/videos, plots and graphs should match those present in the report.

## Questions

1. This question addresses a very practical implementation concern. Consider a signal  $\mathbf{x}$  which is sparse in the 1D-DCT basis  $\Psi \in \mathbb{R}^{n \times n}$  and contains  $n$  elements. Let us suppose that the signal is compressively sensed in the form  $\mathbf{y} = \Phi \mathbf{x} + \boldsymbol{\eta} = \Phi \Psi \boldsymbol{\theta} + \boldsymbol{\eta}$  where  $\mathbf{y}$ , the measurement vector, has  $m$  elements and  $\Phi$  is the  $m \times n$  sensing matrix. Also  $\boldsymbol{\theta}$  is a sparse vector of  $n$  coefficients. Here  $\boldsymbol{\eta}$  is a vector of noise values that are distributed by  $\mathcal{N}(0, \sigma^2)$ . One way to recover  $\boldsymbol{\theta}$  (and thereby also  $\mathbf{x}$ ) from  $\mathbf{y}, \Phi$  is to solve the LASSO problem, based on minimizing  $J(\boldsymbol{\theta}) \triangleq \|\mathbf{y} - \Phi \Psi \boldsymbol{\theta}\|^2 + \lambda \|\boldsymbol{\theta}\|_1$ . A crucial issue is to how to choose  $\lambda$ . One purely data-driven technique is called cross-validation. In this technique, out of the  $m$  measurements, a random subset of (say) 90 percent of the measurements is called the reconstruction set  $\mathcal{R}$ , and the remaining measurements constitute the validation set  $\mathcal{V}$ . Thus  $\mathcal{V}$  and  $\mathcal{R}$  are always disjoint sets. The signal  $\mathbf{x}$  is reconstructed using measurements only from  $\mathcal{R}$  (and thus only the corresponding rows of  $\Phi$ ) using one out of many different values of  $\lambda$  chosen from a set  $\Lambda$ . Let the estimate using the  $g^{th}$  value from  $\Lambda$  be denoted  $\hat{\mathbf{x}}_g$ . The corresponding validation error is computed using  $VE(g) \triangleq \sum_{i \in \mathcal{V}} (y_i - \Phi^i \hat{\mathbf{x}}_g)^2 / |\mathcal{V}|$ . The value of  $\lambda$  for which the validation error is the least is chosen to be the optimal value of  $\lambda$ . Your job is to implement this technique for the case when  $n = 500, m = 300, \|\boldsymbol{\theta}\|_0 \in \{5, 10, 15, 20\}, \sigma = 0.025 \times \sum_{i=1}^m |\Phi^i \mathbf{x}| / m$ . Choose  $\Lambda = \{0.0001, 0.0005, 0.001, 0.005, 0.01, 0.05, 0.1, 0.5, 1, 2, 5, 10, 15, 20, 30, 50, 100\}$ . Draw the non-zero elements of  $\boldsymbol{\theta}$  at randomly chosen location, and let their values be drawn randomly from Uniform(0, 1000). The sensing matrix  $\Phi$  should be drawn from  $\pm 1/\sqrt{m}$  Bernoulli with probability of  $+1/\sqrt{m}$  being 0.5. Now do as follows. Use the popular CVX package (MATLAB version) for implementing the LASSO (or you may use your own previous ISTA code).
  - (a) Plot a graph of  $VE$  versus the logarithm of the values in  $\Lambda$  for each value of  $\|\boldsymbol{\theta}\|_0$ . Also plot a graph of the RMSE versus the logarithm of the values in  $\Lambda$ , where RMSE is given by  $\|\hat{\mathbf{x}}_g - \mathbf{x}\|_2 / \|\mathbf{x}\|_2$ . Comment on the plots. Do the optimal values of  $\lambda$  from the two plots agree? (Also see the last question in this list).
  - (b) What would happen if  $\mathcal{V}$  and  $\mathcal{R}$  were not disjoint but coincident sets?
  - (c) The validation error is actually a proxy for actual mean squared error. Note that you can never determine the mean squared error since the ground truth  $\mathbf{x}$  is unknown in an actual application. Which theorem/lemma from the paper <https://ieeexplore.ieee.org/document/6854225> (On the theoretical analysis of cross-validation in compressed sensing) refers to this proxying ability? Explain how.
  - (d) In your previous assignment, there was a theorem from the book by Tibshirani and others which gave you a certain value of  $\lambda$ . What is the advantage of this cross-validation method compared to the choice of  $\lambda$  using that theorem? Explain.
  - (e) A curious student proposes the following method to choose  $\lambda$ : Pick the value from  $\Lambda$  for which  $\|\mathbf{y} - \Phi \hat{\mathbf{x}}_g\|_2^2$  is the closest possible to  $m\sigma^2$ . This technique is motivated by the very definition of variance, and is often called Morozov's discrepancy principle. Implement this method as well, and plot a graph of RMSE and  $|\|\mathbf{y} - \Phi \hat{\mathbf{x}}_g\|_2^2 - m\sigma^2|$  versus  $\log \lambda$ . What are the advantages and disadvantages of this method as compared to cross-validation?
  - (f) Read the paper 'On cross-validated Lasso in high dimensions' published in the Annals of Statistics. The paper can be found at <https://projecteuclid.org/journals/annals-of-statistics/volume-49/issue-3/On-cross-validated-Lasso-in-high-dimensions/10.1214/20-AOS2000.full>. What is the meaning of the symbol  $K$  in the paper? How do the bounds in theorem 4.1 of this paper compare to the bounds of the LASSO that you studied in the previous assignment? [8+3+3+3+(5+5)+8=35 points]
2. Consider that there are  $n$  coupons, each of a different colour. Suppose we sample coupons uniformly at random and with replacement. Then answer the following questions from first principles. Do not merely quote results pertaining to any distribution other than Bernoulli. [3+3+3+3+(4+4=8)+2.5+2.5=25 points]
  - (a) What is the probability  $q_j$  of getting a new coupon on the  $j$ th trial, assuming that new, unique coupons were obtained in all the previous trials? What is  $q_1$ ?

- (b) If you were to toss a coin independently many times, what is the probability that the first head appears on the  $k$ th trial? Assume that the probability of getting a heads on any trial is  $q$ .
  - (c) Let  $Y$  be the random variable which denotes the trial number on which the first head appears. Derive a formula for  $E(Y)$  in terms of  $q$ .
  - (d) Derive a formula for the variance of  $Y$ .
  - (e) Let  $Z_n$  be a random variable denoting the number of trials by which each of the  $n$  different coupons were selected at least once. Applying result in previous parts, what is the expected value of  $Z_n$ ? Derive an upper bound on the variance of  $Z_n$ . (You will need to use the following results:  $\sum_{i=1}^n 1/i \approx \log n + \gamma + O(1/n)$  where  $\gamma \approx 0.5772$  is a constant, and  $\sum_{i=1}^n 1/i^2 < \sum_{i=1}^{\infty} 1/i^2 < \pi^2/6$ ).
  - (f) Given the previous results, use Markov's inequality to upper bound  $P(Z_n \geq t)$  for some value  $t$ .
  - (g) Given the previous results, use Chebyshev's inequality to upper bound  $P(Z_n \geq t)$  for some value  $t$ .
3. Consider a matrix  $\mathbf{M}$  of size  $n_1 \times n_2$  having low but unknown rank  $r < \max(n_1, n_2)$ . Suppose you observe noiseless compressive measurements in the form of a matrix  $\mathbf{Y}$  of size  $m \times n_2$  where every column of  $\mathbf{Y}$  is obtained by taking dot-products of the corresponding column of  $\mathbf{M}$  with  $m$  different vectors whose elements are drawn from a zero-mean Gaussian distribution. Note that  $m < n_1, m < n_2$ . How will you determine the rank of  $\mathbf{M}$  given the measurements in  $\mathbf{Y}$ . [15 points]
4. Let  $\mathbf{x}$  be a real-valued vector of  $n$  elements. We have the relationship  $\mathbf{y} = \mathbf{A}\mathbf{x}$  where  $\mathbf{y}$  is a measurement vector with  $m$  elements, and  $\mathbf{A}$  is a  $m \times n$  sensing matrix, where  $A_{ij} = 0$  with probability  $1 - \gamma$  and  $A_{ij} = \mathcal{N}(0, 1/m\gamma)$  with probability  $\gamma$ . All entries of  $\mathbf{A}$  are drawn independently. Here  $\gamma \in (0, 1)$ . Note that we have no knowledge of  $k$  beforehand. This question seeks to explore a technique to estimate  $k$  directly from  $\mathbf{y}$  and  $\mathbf{A}$ . To this end, answer the following questions: [4+4+4+4+5+4=25 points]
- (a) Let  $d_i$  be the number of entries for which  $A_{ij}$  and  $x_j$  are both unequal to 0, where  $1 \leq j \leq n$ . What is the distribution of  $d_i$ , if the number of non-zero elements of  $\mathbf{x}$  is  $k$ , since the entries of  $\mathbf{A}$  are drawn independently?
  - (b) Prove that  $P(y_i = 0) = P(d_i = 0)$ .
  - (c) Let  $H$  be a random variable for the number of non-zero elements in  $\mathbf{y}$ . Then what is the distribution of  $H$ , if the number of non-zero elements of  $\mathbf{x}$  is  $k$ ?
  - (d) Express  $k$  in terms of  $P(d_i = 0)$  and hence write the maximum likelihood estimate of  $k$  given  $\mathbf{y}$ .
  - (e) Let  $\hat{P}$  be the estimate of  $P(d_i = 0)$ . Then  $\hat{P}$  is an approximately Gaussian random variable. Explain why. Using this, state how you will provide a confidence interval for the true  $k$  using its estimate  $\hat{k}$  derived so far. That is, you need to provide an interval of the form  $L(\hat{k}) \leq k \leq U(\hat{k})$  with some probability  $q$ .
  - (f) Now consider that you had knowledge of some prior distribution  $\pi(k)$  on  $k$ . How does your estimate of  $k$  now change?