

# Assignment 3: CS 754, Advanced Image Processing

Due: 23rd March 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: A3-IdNumberOfFirstStudent-IdNumberOfSecondStudent.zip. (If you are doing the assignment alone, the name of the zip file is A3-IdNumber.zip). Upload the file on moodle BEFORE 11:55 pm on the due date. The cutoff (beyond which no submission will be accepted) is mentioned on moodle. 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. Download the book ‘Statistical Learning with Sparsity: The Lasso and Generalizations’ from [https://web.stanford.edu/~hastie/StatLearnSparsity\\_files/SLS\\_corrected\\_1.4.16.pdf](https://web.stanford.edu/~hastie/StatLearnSparsity_files/SLS_corrected_1.4.16.pdf), which is the website of one of the authors. (The book can be officially downloaded from this online source). Your task is to trace through the steps of the proof of Theorem 11.1(b). This theorem essentially derives error bounds on the minimum of the following objective function:  $J(\beta) = \frac{1}{2N} \|\mathbf{y} - \mathbf{X}\beta\|^2 + \lambda_N \|\beta\|_1$  where  $\lambda_N$  is a regularization parameter,  $\beta \in \mathbb{R}^p$  is the unknown sparse signal,  $\mathbf{y} = \mathbf{X}\beta + \mathbf{w}$  is a measurement vector with  $N$  values,  $\mathbf{w}$  is a zero-mean i.i.d. Gaussian noise vector whose each element has standard deviation  $\sigma$  and  $\mathbf{X} \in \mathbb{R}^{N \times p}$  is a sensing matrix whose every column is unit normalized. This particular estimator (i.e. minimizer of  $J(\mathbf{x})$  for  $\mathbf{x}$ ) is called the LASSO in the statistics literature. The theorem derives a statistical bound on  $\lambda$  also. Your task is split up in the following manner:
  - (a) Define the restricted eigenvalue condition (the answer’s there in the book and you are allowed to read it, but you also need to understand it).
  - (b) Starting from equation 11.20 on page 309 - explain why  $G(\hat{v}) \leq G(0)$ .
  - (c) Do the algebra to obtain equation 11.21.
  - (d) Do the algebra in more detail to obtain equation 11.22 (state the exact method of application of Holder’s inequality - check the wiki article on it, if you want to find out what this inequality states).
  - (e) Derive equation 11.23.
  - (f) Assuming Lemma 11.1 is true and now that you have derived equation 11.23, complete the proof for the final error bound for equation 11.14b.
  - (g) In which part of the proof does the bound  $\lambda_N \geq 2 \frac{\|\mathbf{X}^T \mathbf{w}\|_\infty}{N}$  show up? Explain.
  - (h) Why is the cone constraint required? You may read the rest of the chapter to find the answer.
  - (i) Read example 11.1 which tells you how to put a tail bound on  $\lambda_N$  assuming that the noise vector  $\mathbf{w}$  is zero-mean Gaussian with standard deviation  $\sigma$ . Given this, state the advantages of this theorem over Theorem 3 that we did in class. You may read parts of the rest of the chapter to answer this question. What are the advantages of Theorem 3 over this particular theorem?

- (j) Now read Theorem 1.10 till corollary 1.2 and comments on it concerning an estimator called the ‘Dantzig selector’, in the tutorial ‘Introduction to Compressed Sensing’ by Davenport, Duarte, Eldar and Kutyniok. You can find it here: <http://www.ecs.umass.edu/~mduarte/images/IntroCS.pdf>. What is the common thread between the bounds on the ‘Dantzig selector’ and the LASSO?
- (k) Read the abstract and introduction (section 1) of the paper ‘Square-root lasso: pivotal recovery of sparse signals via conic programming’ by Belloni et al, published in the journal Biometrika. See <https://www.jstor.org/stable/23076172>. This paper proposes an estimator called the square-root LASSO. What is the advantage of the square-root LASSO over the LASSO? [ $2 \times 8 + 6 + 4 + 4 = 30$  points]
2. In this task, you will use the well-known package L1\_LS from [https://stanford.edu/~boyd/l1\\_ls/](https://stanford.edu/~boyd/l1_ls/). This package is often used for compressed sensing solution, but here you will use it for the purpose of tomographic reconstruction. The homework folder contains images of two slices taken from an MR volume of the brain. Create measurements by parallel beam tomographic projections at any 18 randomly angles chosen from a uniform distribution on  $[0, \pi)$ . Use the MATLAB function ‘radon’ for this purpose. Now perform tomographic reconstruction using the following method: (a) filtered back-projection using the Ram-Lak filter, as implemented in the ‘iradon’ function in MATLAB, (b) independent CS-based reconstruction for each slice by solving an optimization problem of the form  $J(\mathbf{x}) = \|\mathbf{y} - \mathbf{Ax}\|^2 + \lambda\|\mathbf{x}\|_1$ , (c) a coupled CS-based reconstruction that takes into account the similarity of the two slices using the model given in the lectures notes on tomography. For parts (b) and (c), use the aforementioned package from Stanford. For part (c), make sure you use a different random set of 18 angles for each of the two slices. The tricky part is careful creation of the forward model matrix  $\mathbf{A}$  or a function handle representing that matrix, as well as the corresponding adjoint operator  $\mathbf{A}^T$ . Use the 2D-DCT basis for the image representation. Modify the objective function from the lecture notes for the case of three similar slices. Carefully define all terms in the equation and re-implement it. [ $3+7+8+7 = 25$  points]
3. Prove the following properties of the Radon transform:
- (a) Shifting:  $R(g(x - x_0, y - y_0))(\rho, \theta) = R(g(x, y))(\rho - x_0 \cos \theta - y_0 \sin \theta, \theta)$ . Here  $\rho$  is the offset and  $\theta$  is the angle of projection.
- (b) Rotation: Let  $g'(r, \psi) = g(r, \psi - \psi_0)$ . Then prove that  $R(g')(\rho, \theta) = R(g)(\rho, \theta - \psi_0)$ .  ~~$R(g')(\rho, \theta) = R(g)(\rho, \psi_0 - \theta)$~~
- (c) Convolution: Given image  $f(x, y)$  and kernel  $k(x, y)$ , show that  $R_\theta(f * k) = R_\theta(f) * R_\theta(k)$ , where  $*$  is the convolution operation. In other words, the Radon transform of the convolution of two signals is equal to the convolution of the Radon transform of the individual signals (in the same angles). [ $8+5+7=20$  points]
4. For a sensing matrix with columns having unit magnitude, with  $s$ -order restricted isometry constant  $\delta_s$  and mutual coherence  $\mu$ , prove that  $\delta_s \leq (s - 1)\mu$ . (Hint: Use Gershgorin’s disc theorem.) [10 points]
5. Here is our Google search question again. You know of the applications of tomography in medicine (CT scanning) and virology/structural biology. Your job is to search for a journal paper from any other field which requires the use of tomographic reconstruction (examples: seismology, agriculture, gemology). State the title, venue and year of publication of the paper. State the mathematical problem defined in the paper. Take care to explain the meaning of all key terms clearly. State the method of optimization that the paper uses to solve the problem. [15 points]