Assignment 3: CS 754, Advanced Image Processing

Due: 13th 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 date/time (beyond which no submission will be accepted) is 14th March 9 am. 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. 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 ρ is the offset and θ is the angle of projection.
   (b) Rotation: Let $g'(r, \psi) = g(r, \psi - \psi_0)$. Then prove that $R(g')(\rho, \theta) = R(g)(\rho, \theta - \psi_0)$. [4+4=8 points]

2. State the Fourier slice theorem for 3D images (and their associated 2D projections) with the meaning of all symbols carefully explained. You are allowed to look this up on the web, but you have to carefully write down the meaning of all terms. [6 points]

3. In this task, you will you use the well-known package L1_LS from `https://stanford.edu/~boyd/l1_ls/`. This package is popularly used for compressed sensing, but here you will use it for the purpose of tomographic reconstruction in a compressed sensing framework. 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 with default parameters, as implemented in the 'iradon' function in MATLAB, (b) independent CS-based reconstruction for each slice by solving an optimization problem of the form $J(\boldsymbol{x}) = \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|^2 + \lambda\|\boldsymbol{x}\|_1$, (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 $\boldsymbol{A}$ or a function handle representing that matrix, as well as the corresponding adjoint operator $\boldsymbol{A}^T$. Use the 2D-DCT and Haar Wavelet basis for the image representation in parts (b) and (c). 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 for part (c) with 2D-DCT as well as Haar wavelet bases. [FBP: 4 points, independent CS for 2 slices: 4 (DCT) + 4 (Haar), coupled CS with two slices: 6 (DCT) + 6 (Haar) points, coupled CS with three slices: 6 (DCT) + 6 (Haar) points = 36 points]

4. 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`, 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: $J(\boldsymbol{\beta}) = \frac{1}{2N}\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|^2 + \lambda_N \|\boldsymbol{\beta}\|_1$ where λ_N is a regularization parameter that satisfies a certain statistically motivated lower bound, $\boldsymbol{\beta} \in \mathbb{R}^p$ is the unknown sparse signal, $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{w}$ is a measurement vector with N values, $\boldsymbol{w}$ is a zero-mean i.i.d. Gaussian noise vector whose each element has standard deviation σ and $\boldsymbol{X} \in \mathbb{R}^{N \times p}$ is a sensing matrix whose every column is unit normalized. This particular estimator (i.e. minimizer of $J(\boldsymbol{x})$ for $\boldsymbol{x}$) is called the LASSO (least absolute shrinkage and selection operator) in the statistics literature, which has made fundamental contributions to compressed sensing. Your task is as follows:

(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 $\lambda_N \geq 2\frac{\|\boldsymbol{X}\boldsymbol{w}\|_\infty}{N}$ 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 λ_N assuming that the noise vector $\boldsymbol{w}$ is i.i.d. zero-mean Gaussian with standard deviation σ. 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 the bounds for LASSO that you have derived via Theorem 11.1(b)?

(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 x 8 + 9 + 5 + 6 = 36 points]

5. Here is our Google search question again. You know of the applications of tomography in medicine and virology/structural biology. Your job is to search for a journal/conference paper from any other field which requires the use of tomographic reconstruction - examples being agriculture, mechanical engineering, geology, oceanography, botany, etc. Mention the title, author list, year of publication and name of the journal/conference in your report. State the mathematical problem defined in the paper with the associated cost function. Take care to explain the meaning of all key terms clearly. State the method of optimization that the paper uses to optimize the cost function. [14 points]