CS754 Assignment 3 Report

Arpon Basu
Shashwat Garg

Spring 2022

Contents

Introduction

Welcome to our report on CS754 Assignment 3. We have tried to make this report comprehensive and self-contained. We hope reading this would give you a proper flowing description of our work, methods used and the results obtained.

Also note that we installed the `Image Processing Toolbox` in MATLAB for this assignment. Thus the grader is urged to install it if she wishes to run the code on her on her machine. Also note that some of our code may take a while to run because of the intensive nature of the computations involved.

Hope you enjoy reading the report. Here we go!

1 Problem 1

For this problem we implemented the ISTA algorithm as was mentioned in the lecture slides and used it for both parts. For further perusal, we encourage the grader to check the code files submitted, which have been written very intuitively and commented well.

Figure 1: Original 256×256 image for this question

(a) In this part, we read the image, vectorized it, and added some noise to it (since noise of variance 3 was required, we took a Gaussian Random Variable vector with variance 1 and multiplied it by $\sqrt{3}$ to make the variance 3). Since natural images are assumed to be sparse in orthonormal bases like the 2D-DCT, the matrix for our ISTA cost function was the DCT matrix itself.

Thus our cost function was

$$J(\theta) = \|y - U\theta\|_2^2 + \lambda\|\theta\|_1$$

where y is the vectorized noisy image.
In our code, we took $\lambda = 1$, and we also took $\alpha = 2$ (note that since U is orthonormal, the largest eigenvalue of U^*U is 1 itself. We added an additional one to avoid numerical stability issues).
Also, since the basis matrix for the 256×256 images was too large to store in memory, we performed our computations patch-wise, in overlapping patches of 8×8 and a stride of 1. In addition, to help average out overlapping patches, we padded images with a border of zeros so that the average count within the main image would be same for every pixel within the same image (otherwise border pixels occur in lesser patches than middle ones). We got a **RMSE error of 1.17%** after all the overlapping images were summed up and averaged accordingly. (Note that the RMSE error of the noisy image with the original one was 1.27 %, so our algorithm definitely gave us an improvement).
Attached below is the reconstructed image for the graders perusal:

Figure 2: Reconstructed image for the first part

(b) For the second part, we reused the same ISTA and other utility function codes from the first part. However, here the matrix was slightly different, ie:-

$$J(\theta) = \|y - \Phi U\theta\|_2^2 + \lambda\|\theta\|_1$$

, and thus $A = \Phi U$, where $\Phi \in \mathbb{R}^{32\times64}$ is a matrix of i.i.d unit Gaussian RVs.
The hyperparameter λ was kept 1000 here, for better results. However, this time α in the ISTA algorithm had to be chosen to be the largest eigenvalue of A^*A. We also added a 5

to that to avoid numerical stability issues.
Most of the other code for this part was the same as part a, including the idea of padding to make averaging easier.
On running this code, we obtained a **RMSE error of 6.8 %**.

Figure 3: Reconstructed image for the second part

2 Problem 2

(a) The restricted eigenvalue condition is defined as follows:
The matrix $\boldsymbol{X} \in \mathbb{R}^{n \times p}$ is said to satisfy the Restricted Eigenvalue Property over a constraint set $\mathcal{C}$ with coefficient γ if

$$\frac{1}{N \cdot \|\nu\|_2^2} \nu^T X^T X \nu \geq \gamma \; \forall \; \nu \in \mathcal{C} \setminus \{\mathbf{0}\}$$

$$\implies \frac{1}{N \cdot \|\nu\|_2^2} (X\nu)^T X \nu \geq \gamma \; \forall \; \nu \in \mathcal{C} \setminus \{\mathbf{0}\}$$

$$\implies \frac{1}{N} \|X\nu\|_2^2 \geq \gamma \|\nu\|_2^2 \; \forall \; \nu \in \mathcal{C}$$

Thus our definition condenses to: **The matrix $\boldsymbol{X} \in \mathbb{R}^{n \times p}$ is said to satisfy the Restricted Eigenvalue Property over a constraint set $\mathcal{C}$ with coefficient γ if**

$$\boldsymbol{\frac{1}{N} \|X\nu\|_2^2 \geq \gamma \|\nu\|_2^2 \;\; \forall \nu \in \mathcal{C}}$$

(b) We have that

$$G(\nu) := \frac{1}{2N} \|y - X(\beta^* + \nu)\|_2^2 + \lambda_N \|\beta^* + \nu\|_1$$

$$J(\boldsymbol{x}) := \frac{1}{2N}\|y - X\boldsymbol{x}\|_2^2 + \lambda_N\|\boldsymbol{x}\|_1$$

Thus

$$G(0) := \frac{1}{2N}\|y - X\beta^*\|_2^2 + \lambda_N\|\beta^*\|_1 = J(\beta^*)$$

On the other hand, for $\widehat{\nu} := \widehat{\beta} - \beta^*$, we have

$$G(\widehat{\nu}) := \frac{1}{2N}\|y - X\widehat{\beta}\|_2^2 + \lambda_N\|\widehat{\beta}\|_1 = J(\widehat{\beta})$$

But $\widehat{\beta}$ was, **by construction** (as mentioned on Pg# 308, third last line), the optimizer of the LASSO function $J(\boldsymbol{x})$, and consequently $G(\widehat{\nu}) = J(\widehat{\beta}) \leq J(\beta^*) = G(0)$, as desired.
(c) We just have to piece together some equations and inequalities to derive equation 11.21. They go as follows:

$$G(\widehat{\nu}) \leq G(0)$$

$$y = X\beta^* + w$$

$$\widehat{\beta} = \beta^* + \widehat{\nu}$$

The first inequality $G(\widehat{\nu}) \leq G(0)$, combined with the fact that $\widehat{\beta} = \beta^* + \widehat{\nu}$ yields

$$G(\widehat{\nu}) \leq G(0)$$

$$\implies \frac{1}{2N}\|y - X\widehat{\beta}\|_2^2 + \lambda_N\|\widehat{\beta}\|_1 \leq \frac{1}{2N}\|y - X\beta^*\|_2^2 + \lambda_N\|\beta^*\|_1$$

$$\implies \frac{1}{2N}\|y - X(\beta^* + \widehat{\nu})\|_2^2 + \lambda_N\|\beta^* + \widehat{\nu}\|_1 \leq \frac{1}{2N}\|y - X\beta^*\|_2^2 + \lambda_N\|\beta^*\|_1$$

Now, using the fact that $y = X\beta^* + w$, we get

$$\frac{1}{2N}\|y - X(\beta^* + \widehat{\nu})\|_2^2 + \lambda_N\|\beta^* + \widehat{\nu}\|_1 \leq \frac{1}{2N}\|y - X\beta^*\|_2^2 + \lambda_N\|\beta^*\|_1$$

$$\implies \frac{1}{2N}\|y - X\beta^* - X\widehat{\nu}\|_2^2 + \lambda_N\|\beta^* + \widehat{\nu}\|_1 \leq \frac{1}{2N}\|w\|_2^2 + \lambda_N\|\beta^*\|_1$$

$$\implies \frac{1}{2N}\|w - X\widehat{\nu}\|_2^2 \leq \frac{1}{2N}\|w\|_2^2 + \lambda_N(\|\beta^*\|_1 - \|\beta^* + \widehat{\nu}\|_1)$$

Now, $\|w - X\widehat{\nu}\|_2^2 = (w - X\widehat{\nu})^T \cdot (w - X\widehat{\nu}) = \|w\|_2^2 + \|X\widehat{\nu}\|_2^2 - 2w^T X\widehat{\nu}$. Substituting this in the equation above yields

$$\frac{1}{2N}(\|w\|_2^2 + \|X\widehat{\nu}\|_2^2 - 2w^T X\widehat{\nu}) \leq \frac{1}{2N}\|w\|_2^2 + \lambda_N\{\|\beta^*\|_1 - \|\beta^* + \widehat{\nu}\|_1\}$$

$$\implies \frac{1}{2N}(\|X\widehat{\nu}\|_2^2 - 2w^T X\widehat{\nu}) \leq \lambda_N\{\|\beta^*\|_1 - \|\beta^* + \widehat{\nu}\|_1\}$$

$$\implies \frac{1}{2N}\|X\widehat{\nu}\|_2^2 - \frac{1}{N}w^T X\widehat{\nu} \leq \lambda_N\{\|\beta^*\|_1 - \|\beta^* + \widehat{\nu}\|_1\}$$

$$\implies \frac{\|X\widehat{\nu}\|_2^2}{2N} \leq \frac{w^T X\widehat{\nu}}{N} + \lambda_N\{\|\beta^*\|_1 - \|\beta^* + \widehat{\nu}\|_1\}$$

as desired.
(d) We use a version of Hölder's inequality for vectors as follows:

Let $p, q \in [1, \infty]$ be two real numbers such that $\frac{1}{p} + \frac{1}{q} = 1$. Let $\boldsymbol{x} = (x_1, x_2, ..., x_n)$ and $\boldsymbol{y} = (y_1, y_2, ..., y_n)$ be two vectors. Then

$$\sum_{\boldsymbol{k=1}}^{\boldsymbol{n}} |\boldsymbol{x_k y_k}| \leq (\sum_{\boldsymbol{k=1}}^{\boldsymbol{n}} |\boldsymbol{x_k}|^{\boldsymbol{p}})^{\frac{\boldsymbol{1}}{\boldsymbol{p}}} (\sum_{\boldsymbol{k=1}}^{\boldsymbol{n}} |\boldsymbol{y_k}|^{\boldsymbol{q}})^{\frac{\boldsymbol{1}}{\boldsymbol{q}}} = \boldsymbol{\|x\|_p \|y\|_q}$$

We augment this inequality for our purposes: Note that $\boldsymbol{x} \cdot \boldsymbol{y} = \sum_{k=1}^{n} x_k y_k \leq \sum_{k=1}^{n} |x_k y_k| \leq (\sum_{k=1}^{n} |x_k|^p)^{\frac{1}{p}} (\sum_{k=1}^{n} |y_k|^q)^{\frac{1}{q}}$, thus showing that $\boldsymbol{x} \cdot \boldsymbol{y} \leq (\sum_{k=1}^{n} |x_k|^p)^{\frac{1}{p}} (\sum_{k=1}^{n} |y_k|^q)^{\frac{1}{q}} = \|\boldsymbol{x}\|_p \|\boldsymbol{y}\|_q$.
In our context let $\boldsymbol{x} = X^T w$ and $\boldsymbol{y} = \nu$, and let $p = \infty$, $q = 1$. Then, $(X^T w) \cdot \nu = (X^T w)^T \nu = w^T X \nu$. Then we have that

$$(X^T w) \cdot \nu \leq \|X^T w\|_\infty \|\nu\|_1$$

$$\implies w^T X \nu \leq \|X^T w\|_\infty \|\nu\|_1$$

Thus,

$$\frac{w^T X \widehat{\nu}}{N} + \lambda_N \{\|\widehat{\nu_S}\|_1 - \|\widehat{\nu_{S^c}}\|_1\} \leq \frac{\|X^T w\|_\infty}{N} \|\nu\|_1 + \lambda_N \{\|\widehat{\nu_S}\|_1 - \|\widehat{\nu_{S^c}}\|_1\}$$

as desired.
(e) Let S be the set of indices where β^* is non-zero, and let $S^c := [n] \setminus S$. Now, we've already shown that

$$\frac{\|X\widehat{\nu}\|_2^2}{2N} \leq \frac{w^T X \widehat{\nu}}{N} + \lambda_N \{\|\beta^*\|_1 - \|\beta^* + \widehat{\nu}\|_1\}$$

Now,

$$\|\beta^* + \widehat{\nu}\|_1 = \|\beta^*_S + \widehat{\nu_S}\|_1 + \|\beta^*_{S^c} + \widehat{\nu_{S^c}}\|_1 = \|\beta^*_S + \widehat{\nu_S}\|_1 + \|\widehat{\nu_{S^c}}\|_1$$

$$\geq \|\beta^*_S\|_1 - \|\widehat{\nu_S}\|_1 + \|\widehat{\nu_{S^c}}\|_1 = \|\beta^*\|_1 - \|\widehat{\nu_S}\|_1 + \|\widehat{\nu_{S^c}}\|_1$$

$$\implies \|\beta^*\|_1 - \|\beta^* + \widehat{\nu}\|_1 \geq \|\widehat{\nu_S}\|_1 - \|\widehat{\nu_{S^c}}\|_1$$

since $\beta^*_{S^c} = \mathbf{0}$, and applying the triangle inequality.
From the previous part, we also know that

$$\frac{w^T X \widehat{\nu}}{N} + \lambda_N \{\|\widehat{\nu_S}\|_1 - \|\widehat{\nu_{S^c}}\|_1\} \leq \frac{\|X^T w\|_\infty}{N} \|\nu\|_1 + \lambda_N \{\|\widehat{\nu_S}\|_1 - \|\widehat{\nu_{S^c}}\|_1\}$$

$$\implies \frac{w^T X \widehat{\nu}}{N} + \lambda_N \{\|\widehat{\nu_S}\|_1 - \|\widehat{\nu_{S^c}}\|_1\} \leq \frac{\lambda_N}{2} \|\nu\|_1 + \lambda_N \{\|\widehat{\nu_S}\|_1 - \|\widehat{\nu_{S^c}}\|_1\}$$

since $\lambda_N \geq 2\frac{\|X^T w\|_\infty}{N}$ by the premise of the theorem. Finally, note that by the additivity of the l_1-norm,

$$\|\widehat{\nu}\|_1 = \|\widehat{\nu_S}\|_1 + \|\widehat{\nu_{S^c}}\|_1$$

yielding

$$\frac{w^T X \widehat{\nu}}{N} + \lambda_N \{\|\widehat{\nu_S}\|_1 - \|\widehat{\nu_{S^c}}\|_1\} \leq \frac{\lambda_N}{2} \{\|\widehat{\nu_S}\|_1 + \|\widehat{\nu_{S^c}}\|_1\} + \lambda_N \{\|\widehat{\nu_S}\|_1 - \|\widehat{\nu_{S^c}}\|_1\}$$

$$= \frac{3\lambda_N}{2}\|\widehat{\nu_S}\|_1 - \frac{\lambda_N}{2}\|\widehat{\nu_{S^c}}\|_1 \leq \frac{3\lambda_N}{2}\|\widehat{\nu_S}\|_1$$

For the final step, let $|S| = k$. Then by the RMS-AM inequality on vectors,

$$\frac{\|\boldsymbol{x}\|_1}{n} \leq \frac{\|\boldsymbol{x}\|_2}{\sqrt{n}}$$

where $\boldsymbol{x} \in \mathbb{R}^n$.
Thus $\|\widehat{\nu_S}\|_1 \leq \sqrt{|S|}\|\widehat{\nu_S}\|_2 \leq \sqrt{|S|}\|\widehat{\nu}\|_2 = \sqrt{k}\|\widehat{\nu}\|_2$. Combining this with the inequality above yields

$$\frac{w^T X\widehat{\nu}}{N} + \lambda_N\{\|\widehat{\nu_S}\|_1 - \|\widehat{\nu_{S^c}}\|_1\} \leq \frac{3\lambda_N}{2}\|\widehat{\nu_S}\|_1 \leq \frac{3\lambda_N}{2}\sqrt{k}\|\widehat{\nu}\|_2$$

as desired.
(f) We have

$$\frac{w^T X\widehat{\nu}}{N} + \lambda_N\{\|\widehat{\nu_S}\|_1 - \|\widehat{\nu_{S^c}}\|_1\} \leq \frac{3\lambda_N}{2}\sqrt{k}\|\widehat{\nu}\|_2$$

We also know that

$$\frac{\|X\widehat{\nu}\|_2^2}{N} \geq \gamma\|\widehat{\nu}\|_2^2$$

since X satisfies the restricted eigenvalue property.
Thus

$$\frac{\gamma}{2}\|\widehat{\nu}\|_2^2 \leq \frac{\|X\widehat{\nu}\|_2^2}{2N} \leq \frac{w^T X\widehat{\nu}}{N} + \lambda_N\{\|\widehat{\nu_S}\|_1 - \|\widehat{\nu_{S^c}}\|_1\} \leq \frac{3\lambda_N}{2}\sqrt{k}\|\widehat{\nu}\|_2$$

$$\implies \frac{\gamma}{2}\|\widehat{\nu}\|_2^2 \leq \frac{3\lambda_N}{2}\sqrt{k}\|\widehat{\nu}\|_2$$

$$\implies \|\widehat{\nu}\|_2 \leq \frac{3\lambda_N}{\gamma}\sqrt{k} = \frac{3}{\gamma}\sqrt{\frac{k}{N}}\sqrt{N}\lambda_N$$

Finally, recalling that $\widehat{\nu} := \widehat{\beta} - \beta^*$, we get that

$$\boldsymbol{\|\widehat{\beta} - \beta^*\|_2 \leq \frac{3}{\gamma}\sqrt{\frac{k}{N}}\sqrt{N}\lambda_N}$$

as desired.
(g) When we show $\frac{\|X\widehat{\nu}\|_2^2}{2N} \leq \frac{3}{2}\sqrt{k}\lambda_N\|\widehat{\nu}\|_2^2$, en route to ultimately establishing the upper bound on $\|\widehat{\nu}\|_2$, we need to use the fact $\frac{1}{N}\|X^T w\|_\infty \leq \frac{\lambda_N}{2}$, to set up an inequality on the noise-dependent part of the term. It's here that we need to utilise the fact that $\lambda_N \geq 2\frac{\|X^T w\|_\infty}{N}$.

(h) Note that when we defined the **Restricted Eigenvalue Property** in the first part, we only defined it over a certain constraint set $\mathcal{C}$. That was because while properties of strong convexity are very desirable while proving mathematical bounds, the rectangular matrix X itself is such that the cost function ($\propto X^T X$) is only convex in certain directions while concave in others, and thus arises the the set $\mathcal{C} \subseteq \mathbb{R}^p$ which denotes the regions of convexity of our cost function.
Finally, when we actually come to applying the theorem, the regression constraint (that all feasible points $\|\beta\|_1 \leq R$) forces $\mathcal{C}$ to be a cone, ie:- $\|\widehat{\nu_S}\|_1 \leq \alpha\|\widehat{\nu_{S^c}}\|_1$, for some $\alpha > 0$, and where $\widehat{\nu} = \beta - \beta^*$.

(i) The advantages of the given theorem over Theorem 3 are:-

- The only property of matrix $\boldsymbol{X}$ that is needed to apply this theorem is it's coefficient γ of restricted eigenvalue property, while for applying Theorem 3 we need to know the RIP-coefficient δ of $\boldsymbol{X}$. Note that while γ can be easily calculated (it can be shown that γ is linked to the smallest singular value of $\boldsymbol{X}$), calculating δ for a general matrix is NP-hard.
- This theorem focuses on the matrix-noise interaction as opposed to looking at noise independently, as in Theorem 3.
- The bound in this theorem is provably optimal ("minimax-optimal"), in the sense that no other estimator can improve upon it further. Such guarantees are unavailable for Theorem 3.
- This theorem gives us a deterministic bound which is true for all cases, and it also gives the freedom to choose λ_N in such a way that one can guarantee better bounds on a probabilistic basis. In theorem 3, once we fix our sensing matrix, we can't improve upon our bound further.
- The fact that we can adjust λ_N according to the level of accuracy required is a flexibility unavailable to us in Theorem 3, because the probabilistic factor in Theorem 3 is subsumed within the design of a matrix with low δ. Thus, for example, if better bounds are desired in Theorem 3, one has to change the sensing matrix itself. However, in this theorem, it suffices to simply change the regularization parameter λ_N.

Advantages of Theorem 3 over the given theorem are:-

- Theorem 3 works well even for compressible signals due to it's oracular upper bound. This theorem demands strict sparsity, which may be unrealistic in many situations.

(j) One of the common threads between the Dantzig selector and LASSO is the their similar treatment of noise models, ie:-

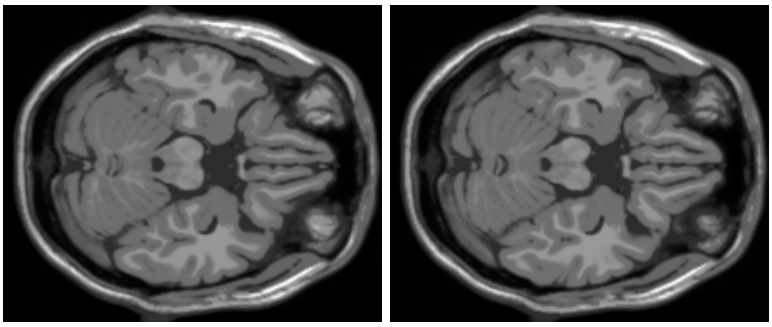
- Both don't assume any certain distribution (as in Gaussian, or Uniform, etc.)for the noise w (or e).
- Both instead deal with all types of noise in that they propose upper bounds on $\|X^T w\|_\infty$ (for LASSO) or $\|A^T e\|_\infty$ (for the Dantzig selector).
- In both types of theorems, in case the norm of the matrix-noise interaction can't be upper bounded (ie:- if the support of our noise vector is unbounded), both of them provide (exponential) probabilistic guarantees on the theorems to still hold.
- Both of them deal with the matrix-noise vector, than with the noise itself. This gives the bounds more adaptability in similar scenarios but with different sensing matrices.

3 Problem 3

Here, we implement tomographic reconstruction using simple techniques like the RamLak filter, as well the method of compressive reconstruction and coupled compressive reconstruction.

The code is well commented to elaborate every step in the process. Kindly refer to that for a detailed description of our work.

We use uniformly distributed angles from 0 to 180 for this problem and in part(c), we take two sets of uniformly distributed angles. The images used are as follows-

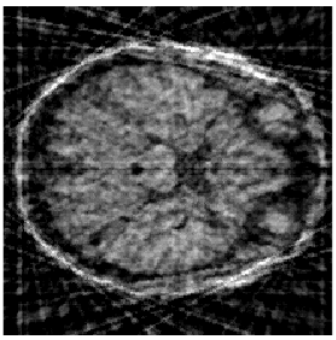


3.1 Ram-Lak Filter

We begin with reading in the image files, converting them into a square shaped image and calculating the radon transform using the radon() function in MATLAB.

This part just involves calling the iradon function with RamLak as the argument, which performs the necessary reconstruction and returns an image with MSE as 15.4%

The image is as follows-

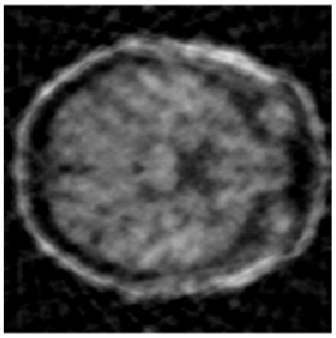


3.2 Using Compressive Sensing

This involves using compressive sensing to reconstruct the image using the Radon transforms. The fundamental is that the Radon Tranformation can be represented as a matrix multiplication and the image-vector can be represented as a sparse vector in the 2d-dct basis. This gives us a problem struture just like the compressive sensing problem and now we can use our familiar tools like Lasso, L1-norm minimization to get to the correct results.

Overall, we get a reconstruction with an MSE of 6.1%

The reconstructed image is as follows-



3.3 Error Function for 2 consecutive slices

Here, we see an extension of the familiar compressive sensing problem. We see how instead of thinking like 2 separate CS problems, in case of consecutive slices, we can use the reducdancy in the frames to get an even better and more complete picture with less magnitude of error.

This improvement is mainly due to the fact that the set of angles of the two consecutive measurements will be different but both will give some information about each other. Thus, we get better results.

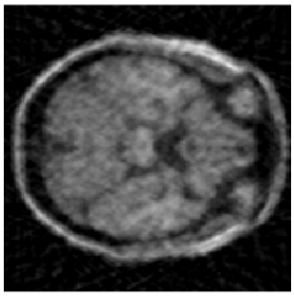
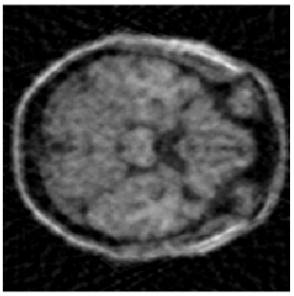
The formula used is as follows-

$$E(\beta_1, \beta_2) = ||y_1 - R_1 U \beta_1||^2 + ||y_2 - R_2 U \beta_2||^2 + \lambda||\beta_1||_1 + \lambda||\beta_2 - \beta_1||_1$$

$$E(\beta_1, \beta_2) = ||y_1 - R_1 U \beta_1||^2 + ||y_2 - R_2 U (\beta_1 + \Delta\beta_{21})||^2 + \lambda||\beta_1||_1 + \lambda||\Delta\beta_{21}||_1$$

Overall, we get a better reconstruction with an MSE of 3.2% and 2.8% for both images respectively. This is an upgrade over the previous MSE of 6.1%

The reconstructed image is as follows-

3.4 Error Function for 3 consecutive slices

We would use a similar approach as previous case. We center our sparsity measure at second slice and minimise the difference between the first-second and the second-third slices. This allows us to exploit the redundancy between consecutive frames.

The formula is as follows-

$E(\beta_1, \beta_2, \beta_3) = ||y_1 - R_1 U \beta_1||^2 + ||y_2 - R_2 U \beta_2||^2 + ||y_3 - R_3 U \beta_3||^2 + \lambda||\beta_2||_1 + \lambda||\beta_1 - \beta_2||_1 + \lambda||\beta_3 - \beta_2||_1$

$= ||y_1 - R_1 U(\beta_2 + \Delta\beta_{12})||^2 + ||y_2 - R_2 U \beta_2||^2 + ||y_3 - R_3 U(\beta_2 + \Delta\beta_{32})||^2 + \lambda||\beta_2||_1 + \lambda||\Delta\beta_{12}||_1 + \lambda||\Delta\beta_{32}||_1$

4 Problem 4

Paper Details	
Title of the Paper	Coastal Acoustic Tomography System and Its Field Application
Link of the paper	**Click Here**
Author List	Haruhiko Yamoaka, Arata Kaneko, Jae-Hun Park, Hong Zheng, Noriaki Gohda, Tadashi Takano, Xiao-Hua Zhu and Yoshio Takasugi
Publication Date	August 2002
Publication Venue	IEEE Journal of Oceanic Engineering, Volume 27, Issue 2

4.1 Introduction and Aim

This paper aims to map the structure of the "strongly nonlinear tidal currents in the coastal sea" by using multiple synchronised coastal acoustic tomography system (CATS). Using GPS clock signals and separate codes to distinguish between signals of individual systems, reconstruction of tidal process behaviour is done through an inverse analysis of the acoustic signals obtained by the sensors.

4.2 Mathematical Formulation

5 Problem 5

We know that Radon Transform is given by-

$$R_\theta(f) = g(\rho,\theta) = \int_{-\infty}^{+\infty} f(\rho\cos\theta - z\sin\theta, \rho\sin\theta + z\cos\theta)dz$$

We can write the same as-

$$R_\theta(f) = g(\rho,\theta) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)\delta(x\cos\theta + y\sin\theta - \rho)dxdy$$

Let the scaled image be denoted by $h(x,y) = f(ax,ay)$. This is the same image as original, but scaled by a factor of a, in both x and y directions.

We can write the same Radon Transform as-

$$R_\theta(h) = g'(\rho,\theta) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} h(x,y)\delta(x\cos\theta + y\sin\theta - \rho)dxdy$$

$$R_\theta(h) = g'(\rho,\theta) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(ax,ay)\delta(x\cos\theta + y\sin\theta - \rho)dxdy$$

$$R_\theta(h) = g'(\rho,\theta) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x',y')\delta\left(\frac{x'\cos\theta + y'\sin\theta - a\rho}{a}\right)\frac{dx'}{a}\frac{dy'}{a}$$

Since $\delta(ax) = \delta(x)/a$, we get-

$$R_\theta(h) = g'(\rho,\theta) = \frac{1}{a}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x',y')\delta(x'\cos\theta + y'\sin\theta - a\rho)dx'dy'$$

$$R_\theta(h) = g'(\rho,\theta) = \frac{1}{a}g(a\rho,\theta)$$

Here, $g()$ is the original radon transform function and $g'()$ is the new obtained radon function. Thus, we can see that the Radon transform of the scaled image is also scaled by a factor of a in the size of projection, but the intensity of each projection has reduced by a as well.

6 Problem 6

We know that the Radon Transform is given by-

$$R_\theta(f) = g(\rho,\theta) = \int_{-\infty}^{+\infty} f(\rho\cos\theta - z\sin\theta, \rho\sin\theta + z\cos\theta)dz$$

We can write the same as-

$$R_\theta(f) = g(\rho,\theta) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)\delta(x\cos\theta + y\sin\theta - \rho)dxdy$$

Now, let $f(x, y) = \delta(x - x_0, y - y_0)$ for some given constants x_0, y_0. Also, for the sake of simplification, call $\delta(x\cos\theta + y\sin\theta - \rho)$ as $h(x, y, \rho, \theta)$.
Then the Radon transform of our function $f(x, y)$ becomes

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)\delta(x\cos\theta + y\sin\theta - \rho)dxdy$$

$$= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \delta(x - x_0, y - y_0)h(x, y, \rho, \theta)dxdy$$

Now, by a well known property of delta functions, the integration of a function multiplied by the delta function over any space (including the delta function's singularity) yields the evaluation of the function at the singularity point. In the context of our problem, we can state the above property as

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \delta(x - x_0, y - y_0)h(x, y)dxdy = h(x_0, y_0)$$

Applying this property verbatim on our integral above yields

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \delta(x - x_0, y - y_0)h(x, y, \rho, \theta)dxdy$$

$$= h(x_0, y_0, \rho, \theta) = \delta(x_0\cos\theta + y_0\sin\theta - \rho)$$

since ρ, θ are constant parameters within the integration.
Thus the Radon transform of the unit impulse function is another impulse function, ie:-

$$\boldsymbol{R_\theta(\delta(x - x_0, y - y_0))) = g(\rho, \theta) = \delta(x_0\cos\theta + y_0\sin\theta - \rho)}$$

$$\boldsymbol{R_\theta(\delta(x, y))) = g(\rho, \theta) = \delta(-\rho) = \delta(\rho)}$$