CS 754 Assignment 3

By Waqar Mirza and Aditya Kudre

March 23, 2022

Problem 1

The original barbara image is given below for comparison:

Figure 1: Original Barbara image

The following metric will be used to compare the reconstruction to the original:
RMSE $:= ||X(:) - \hat{X}(:)||_2/||X(:)||_2$
where $X(:)$ is the true image and $\hat{X}(:)$ is the reconstruction.

1(a)

Note that in this case our matrix A is the inverse 2D DCT matrix.
Here we will display the noise corrupted image and reconstructed image along with their respective RMSE values:

RMSE of corrupted image = 0.0127

Figure 2: Corrupted image (noise variance = 3)

RMSE of reconstructed image = 0.0119

Figure 3: Reconstructed image

Note that there isn't a very significant improvement due to reconstruction in this case of low noise.

1(b)

Let us first show the reconstruction:

RMSE of reconstructed image = 0.2873

Figure 4: Reconstructed image

Now we observe that the reconstruction is of much lower intensity than the original image but it still resembles the original image. It is also observed that the intensity and RMSE varies each time the code is run, indicating that it depends on the matrix phi which is randomly chosen each time. Thus we get a downscaled version of the original image in our reconstruction. The reconstructed image was scaled up with a scaling factor which minimised the RMSE of the scaled image. This gave the following results for this particular reconstruction:
RMSE of scaled reconstruction = 0.0493
Scaling factor = 1.4

Figure 5: Scaled reconstructed image

Problem 2 :

We are given $y = X\beta^* + w$
and estimate $\hat{\beta}$ is found by minimising $J(\beta)$.

(a) The restricted eigenvalue condition refers to the following condition for the model matrix X (w.r.t a $C \subset \mathbb{R}^p$) :
There exists some $\gamma > 0$, such that

$$\frac{1}{N} \frac{v^T X^T X v}{\|v\|_2^2} \geq \gamma \quad \forall \text{ nonzero } v \in C.$$

In the above condition we are lower bounding the restricted eigenvalues (on the set C) of the model matrix, ~~in~~
That is, we restrict our v to belong to a subset $C \subset \mathbb{R}^p$.

(b) We define

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

$$= J(\beta^* + v)$$

Since $\hat{\beta}$ is the $\arg\min_{\beta} J(\beta)$, we know:

$$J(\hat{\beta}) \leq J(\beta^*)$$

and $J(\beta) = G(\beta - \beta^*)$.

$$\Rightarrow G(\hat{\beta} - \beta^*) \leq G(\beta^* - \beta^*) = G(0)$$

taking $\hat{v} := \hat{\beta} - \beta^*$

$$\Rightarrow G(\hat{v}) \leq G(0)$$

(c) As $G(\hat{v}) \leq G(0)$

$$\Rightarrow \frac{1}{2N} \|y - X(\beta^* + \hat{v})\|_2^2 + \lambda_N \|\beta^* + \hat{v}\|_1$$

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

note that $\|y - X(\beta^* + \hat{v})\|_2^2$

$$= \|(y - X\beta^*) - X\hat{v}\|_2^2$$

$$= \|y - X\beta^*\|_2^2 + \|X\hat{v}\|_2^2 - 2(y - X\beta^*)^T X\hat{v}$$

$y - X\beta^* = w$, so

$$\Rightarrow \frac{1}{2N}\left(\|w\|_2^2 + \|X\hat{v}\|_2^2 - 2w^T X\hat{v}\right)$$

$$+ \lambda_N \|\beta^* + \hat{v}\|_1 \leq \frac{1}{2N}\|w\|_2^2 + \lambda_N \|\beta^*\|_1$$

$$\Rightarrow \frac{1}{2N}\|X\hat{v}\|_2^2 \leq \frac{1}{N} w^T X\hat{v} + \lambda_N\left(\|\beta^*\|_1 - \|\beta^* + \hat{v}\|_1\right)$$

($\frac{1}{2N}\|w\|_2^2$ gets cancelled on both sides) (11.21)

(d) As β^* has support on set S, $\beta^*_{S^c} = 0$

So $\|\beta^* + \hat{v}\|_1 = \sum_{i \in S} |(\beta_i^* + \hat{v}_i)| + \sum_{i \in S^c} |\beta_i^* + \hat{v}_i|$

$$= \sum_{i \in S} |\beta_i^* + \hat{v}_i| + \sum_{i \in S^c} |\hat{v}_i|$$

$$= \|\beta_S^* + \hat{v}_S\|_1 + \|\hat{v}_{S^c}\|_1$$

$$\geq \|\beta_S^*\|_1 - \|\hat{v}_S\|_1 + \|\hat{v}_{S^c}\|_1$$

Substituting in ~~eqn~~ 11.21, we get:

$$\Rightarrow \frac{\|X\hat{v}\|_2^2}{2N} \le \frac{W^T X \hat{v}}{N} + \lambda_N(\|\beta^*\|_1 - \|\beta^* + \hat{v}\|_1)$$

and $\|\beta^*\|_1 - \|\beta^* + \hat{v}\|_1$

$$\le \|\beta^*\|_1 - (\|\beta^*_S\|_1 - \|\hat{v}_S\|_1 + \|\hat{v}_{S^c}\|_1)$$

$$= \|\hat{v}_S\|_1 - \|\hat{v}_{S^c}\|_1 + (\|\beta^*\|_1 - \|\beta^*_S\|_1)$$

$$\Rightarrow \|\beta^*\|_1 - \|\beta^* + \hat{v}\|_1 \le \|\hat{v}_S\|_1 - \|\hat{v}_{S^c}\|_1$$

(since $\|\beta^*\|_1 = \|\beta^*_S\|_1$)

So
$$\frac{\|X\hat{v}\|_2^2}{2N} \le \frac{W^T X \hat{v}}{N} + \lambda_N(\|\hat{v}_S\|_1 - \|\hat{v}_{S^c}\|_1)$$

Now we will use Hölder's inequality

For vectors a_1 and a_2,

$$a_1^T a_2 \le |a_1^T a_2| = |\sum_i a_{1i} a_{2i}| \le \sum_i |a_{1i} a_{2i}| \cancel{= \|a_1 a_2\|_1}$$

$$\Rightarrow a_1^T a_2 \le \|a_1 \odot a_2\|_1 \le \|a_1\|_p \|a_2\|_q$$

with $\frac{1}{p} + \frac{1}{q} = 1$, $p, q \in [1, \infty]$ (Hölder's inequality)

so choose $p = 1$, $q = \infty$.

we get
$$\boxed{a_1^T a_2 \le \|a_1\|_1 \|a_2\|_\infty}$$

$$\Rightarrow \frac{W^T X \hat{v}}{N} \le \frac{\|X^T W\|_\infty \|\hat{v}\|_1}{N}$$

So, we get:

$$\Rightarrow \frac{\|X\hat{v}\|_2^2}{2N} \le \frac{\|X^T W\|_\infty \|\hat{v}\|_1}{N} + \lambda_N(\|\hat{v}_S\|_1 - \|\hat{v}_{S^c}\|_1) \tag{11.22}$$

(e) We have assumed that $\frac{1}{N}\|X^T w\|_\infty \le \frac{\lambda_N}{2}$.

So; from 11.22:

$$\frac{\|X\hat{v}\|_2^2}{2N} \le \frac{\|X^T w\|_\infty}{N}\|\hat{v}\|_1 + \lambda_N(\|\hat{v}_S\|_1 - \|\hat{v}_{S^c}\|_1)$$

$$\le \frac{\lambda_N}{2}\|\hat{v}\|_1 + \lambda_N(\|\hat{v}_S\|_1 - \|\hat{v}_{S^c}\|_1)$$

$$= \frac{\lambda_N}{2}(\|\hat{v}_S\|_1 + \|\hat{v}_{S^c}\|_1) + \lambda_N(\|\hat{v}_S\|_1 - \|\hat{v}_{S^c}\|_1)$$

$$= \frac{3}{2}\lambda_N\|\hat{v}_S\|_1 - \frac{\lambda_N}{2}\|\hat{v}_{S^c}\|_1$$

$$\le \frac{3}{2}\lambda_N\|\hat{v}_S\|_1$$

[since $\frac{\lambda_N}{2}\|\hat{v}_{S^c}\|_1 \ge 0$]

now note:

$$\|\hat{v}_S\|_1 = \sum_{i\in S}|\hat{v}_i|$$

$$\le |S|\sqrt{\frac{\sum_{i\in S}|\hat{v}_i|^2}{|S|}}$$ [RMS $\ge$ AM]

$$= \sqrt{|S|}\,\|\hat{v}_S\|_2 = \sqrt{k}\,\|\hat{v}_S\|_2$$

$$\Rightarrow \|\hat{v}_S\|_1 \le \sqrt{k}\,\|\hat{v}_S\|_2 = \sqrt{k}\,\|\hat{v}\|_2$$

So:

$$\frac{\|X\hat{v}\|_2^2}{2N} \le \frac{3}{2}\lambda_N\|\hat{v}_S\|_1 \le \frac{3}{2}\sqrt{k}\,\lambda_N\|\hat{v}\|_2$$

(11.23)

Page No.:
Date:
YOUVA

(f) Lemma 11.1 tells us that if $\lambda_N \geq 2\left\|\frac{X^T w}{N}\right\|_\infty > 0$, then $\hat{\nu} := \hat{\beta} - \beta^*$ for ~~the~~ a lasso solution $\hat{\beta}$ satisfies $\hat{\nu} \in C(S;3)$, (here $C(S;\alpha) := \{z \mid \|z_{S^c}\|_1 \leq \alpha \|z_S\|_1, z \in \mathbb{R}^p\}$)

That is, $\hat{\nu}$ belongs to the cone set with following cone constraint:

$$\|\hat{\nu}_{S^c}\|_1 \leq 3\|\hat{\nu}_S\|_1 .$$

This allows us to apply the γ-restricted eigenvalue condition, since the theorem assumes that X satisfies the γ-RE condition over $C(S;3)$.

As $\hat{\nu} \in C(S;3)$,

$$\frac{1}{N} \frac{\hat{\nu}^T X^T X \hat{\nu}}{\|\hat{\nu}\|_2^2} \geq \gamma$$

$$\Rightarrow \frac{1}{N}\|X\hat{\nu}\|_2^2 \geq \gamma \|\hat{\nu}\|_2^2$$

Combining with 11.23:

$$\frac{\gamma}{2}\|\hat{\nu}\|_2^2 \leq \frac{\|X\hat{\nu}\|_2^2}{2N} \leq \frac{3}{2}\sqrt{k}\,\lambda_N \|\hat{\nu}\|_2$$

$$\Rightarrow \|\hat{\nu}\|_2 \leq \frac{3}{\gamma}\sqrt{k}\,\lambda_N = \frac{3}{\gamma}\sqrt{\frac{k}{N}}\sqrt{N}\,\lambda_N$$

since $\hat{\nu} = \hat{\beta} - \beta^*$, we get 11.14b

$$\|\hat{\beta} - \beta^*\|_2 \leq \frac{3}{\gamma}\sqrt{\frac{k}{N}}\sqrt{N}\,\lambda_N \qquad (11.14b)$$

Page No.:
Date:
YOUVA

(g) The bound $\lambda_N \geq 2\frac{\|X^T w\|_\infty}{N}$ appears in the following parts of the proof:

(i) In order to obtain equation 11.23 from 11.22, we use $\frac{1}{N}\|X^T w\|_\infty \leq \frac{\lambda_N}{2}$ to get

$$\frac{\|X\hat{v}\|_2^2}{2N} \leq \frac{\|X^T w\|_\infty \|\hat{v}\|_1}{N} + \lambda_N(\|\hat{v}_S\|_1 - \|\hat{v}_{S^c}\|_1)$$
$$\leq \frac{\lambda_N}{2}(\|\hat{v}_S\|_1 + \|\hat{v}_{S^c}\|_1) + \lambda_N(\|\hat{v}_S\|_1 - \|\hat{v}_{S^c}\|_1)$$

1

allowing us to obtain equation 11.23.

(ii) To prove Lemma 11.1, the condition $\lambda_N \geq 2\|\frac{X^T w}{N}\|_\infty$ is assumed.

This gives us the inequality 1 written above:

$$0 \leq \frac{\|X\hat{v}\|_2^2}{2N} \leq \frac{\lambda_N}{2}(\|\hat{v}_S\|_1 + \|\hat{v}_{S^c}\|_1) + \lambda_N(\|\hat{v}_S\|_1 - \|\hat{v}_{S^c}\|_1)$$

(This was just proved in part (i) above).

So

$$\frac{3\lambda_N}{2}(\|\hat{v}_S\|_1) - \frac{\lambda_N}{2}\|\hat{v}_{S^c}\|_1 \geq 0$$

$$\Rightarrow \quad 3\|\hat{v}_S\|_1 \geq \|\hat{v}_{S^c}\|_1 \Rightarrow \hat{v} \in C(S;3)$$

this proving the Lemma.

(h) The fact is that for a suitable choice of λ_N, i.e., $\lambda_N \geq 2 \frac{\|X^T w\|_\infty}{N}$

it turns out that the error $\hat{\nu} = \hat{\beta} - \beta^*$ satisfies the cone constraint

$$\|\hat{\nu}_{S^c}\|_1 \leq 3 \|\hat{\nu}_S\|_1 \Rightarrow \hat{\nu} \in \mathcal{C}(S;3)$$

So we may impose the γ-RE condition for X only on $\mathcal{C}(S;3) \subset \mathbb{R}^p$, loosening the constraints on X (as X doesn't need to satisfy γ-RE on entire $\mathbb{R}^p$).

Thus, the theorem assumes that X satisfies γ-RE condition on $\mathcal{C}(S;3)$ and as $\hat{\nu} \in \mathcal{C}(S;3)$, this allows us to prove 11.14b from 11.23 to get the required result.

Effectively we obtain that satisfying the RE condition on $\mathcal{C}(S;3)$ is sufficient for the result to hold, giving us the sufficient constraint on X.

(i) Lets call this result theorem A:

$$\|\hat{\beta} - \beta^*\|_2 \leq \frac{c\sigma}{\gamma} \sqrt{\frac{\tau k \log p}{N}}$$

with probability $\geq 1 - 2e^{-\frac{1}{2}(\tau-2)\log p}$. where we obtained $\hat{\beta}$ by solving the lasso.

Now if we solved the BP problem and obtained $\hat{\beta}$, as the estimator for β^*, theorem 3 gives us the following bounds:

$$\|\hat{\beta}_1 - \beta^*\|_2 \leq C_1 \varepsilon$$

where

$$\hat{\beta}_1 = \arg\min \|\beta\|_1 \text{ such that } \|y - X\beta\|_2 \leq \varepsilon$$

for $\varepsilon \geq \|w\|_2$.

now as $w_i \sim N(0, \sigma^2)$ $\forall i$, we have using gaussian tail bound:

$$Pr\left[|w_i| \geq \sigma t\right] \leq 2 e^{-\frac{(t\sigma)^2}{2\sigma^2}} = 2 e^{-t^2/2}$$

$$\Rightarrow Pr\left[|w_i| \leq \sigma t\right] \geq 1 - 2 e^{-t^2/2}$$

So $Pr\left[\|w\|_2 \leq \varepsilon\right] \geq Pr\left[|w_i| \leq \frac{\varepsilon}{\sqrt{N}} \ \forall i\right]$

$$\Rightarrow Pr\left[\|w\|_2 \leq \varepsilon\right] \geq \prod_i Pr\left[|w_i| \leq \frac{\varepsilon}{\sqrt{N}}\right] \geq \left(1 - 2e^{-t^2/2}\right)^N$$

$$\geq 1 - 2N e^{-t^2/2}$$

$$\Rightarrow Pr\left[\|w\|_2 \leq \varepsilon\right] \geq 1 - 2N e^{-\frac{\varepsilon^2}{2N\sigma^2}} \quad \left(\text{for } t = \frac{\varepsilon}{\sigma\sqrt{N}}\right)$$

$$= 1 - 2 e^{-\left(\frac{\varepsilon^2}{2N\sigma^2} - \log N\right)}$$

so let $\varepsilon = \sigma\sqrt{\tau N \log N}$ for some $\tau > 2$

$$\Rightarrow Pr\left(\|w\|_2 \leq \varepsilon\right] \geq 1 - 2 e^{-\frac{1}{2}(\tau - 2)\log N}$$

So we have

$$\|\hat{\beta}_1 - \beta^*\|_2 \leq C_1 \sigma \sqrt{\tau N \log N}$$

with probability $\geq 1 - 2 e^{-\frac{1}{2}(\tau-2)\log N}$

(here C_1 is solely a function of δ_{2s})

So for this case of gaussian noise, theorem A gives much tighter bounds on the error for lasso solution as compared to theorem 3. for BP. Thus it guarantees a better estimate than theorem 3.

This is an advantage of theorem A over 3.

Moreover theorem ~~3~~ A also gives us ~~an~~ a dependance of the bound on p, which gives us an idea of how much p should be to get a good recovery, i.e., how much we can compress.

On the other hand, Theorem 3 has some ~~advant~~ advantages over Theorem A. Theorem 3 is more robust as Theorem A requires β^* to be k-sparse and fails if β^* has few values outside set S ~~which~~ which are non-zero but small. However theorem 3 tells us:

$$\|\hat{\beta}_1 - \beta^*\|_2 \leq \frac{C_0}{\sqrt{k}} \|\beta^* - \beta^*_S\|_1 + C_1 \varepsilon$$

so theorem 3 works well for compressible signals where theorem A might fail as they may not be strictly sparse.

Note: An advantage of theorem A over theorem 3 is that the bound in A ~~dec~~ varies as $\sqrt{k}$ so it decreases quite fast as k is made small (more sparse). However in theorem 3 the dependance on k isn't guaranteed to be as drastic as δ_{2k} doesn't vary as fast (not necessarily).

(j) The bounds placed by Dantzig selector:

Measurements $y = Ax + e$, If A satisfies RIP of order $2k$ with $\delta_{2k} < \sqrt{2} - 1$,
Let $\lambda \geq \|A^T e\|_\infty$. If $B(y) = \{x : \|A^T(Ax - y)\|_\infty \leq \lambda\}$

let $\hat{x} = \arg\min_x \|x\|_1$, $x \in B(y)$
we have:

$$\|\hat{x} - x\|_2 \leq C_0 \frac{\|x - x_S\|_1}{\sqrt{k}} + C_3 \sqrt{k}\,\lambda$$

[C_0, C_3 are increasing functions of δ_{2k}]
(here S is set of indices of k largest elements)
Now to compare this with the Lasso bounds lets take x to be k sparse, we get:

$$\|\hat{x} - x\|_2 \leq C_3 \sqrt{k}\,\lambda$$

The bounds for lasso with true vector β^* and estimate $\hat{\beta}$, with $\lambda_N \geq 2\|X^T w\|_\infty / N$

$$\|\hat{\beta} - \beta^*\|_2 \leq \frac{3}{\gamma} \sqrt{k}\,\lambda_N .$$

we can see that the two bounds are very similar, i.e. they are proportional to $\sqrt{k}\,\lambda$, and the choice for λ is also very similar (lower bounded by $c\|X^T \eta\|_\infty$ where η is some error term and X is the model matrix for some constant c). The $\propto \sqrt{k}$ is a ~~also~~ similar feature of the two bounds. (k - sparsity)
So both these bounds are $O(\sqrt{k})$.

CS-754 Homework-3

Waqar Mirza and Aditya Kudre

March 2022

1 Problem 3

Here are the zero-padded images of the two slices (slice-50 and slice-51) we are using in this problem :

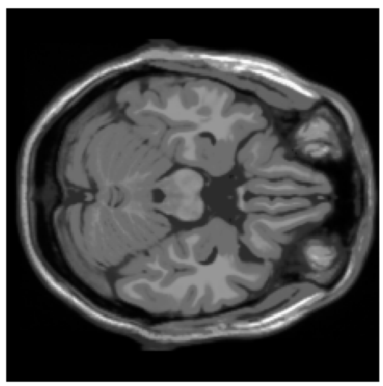

Figure 1: Padded Image for Slice 50

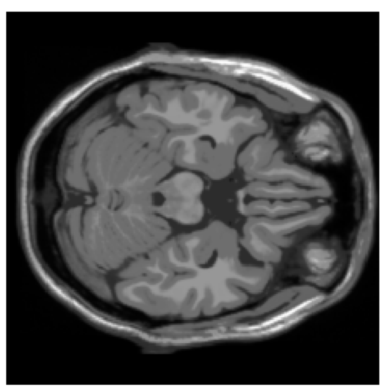

Figure 2: Padded Image for Slice 51

We will be taking the radon transforms of these two slices and reconstructing them using 3 different techniques : (a) Back-projection with the Ram-Lak Filter, (b) CS based reconstruction which involves solving this type of problem :

$$min_x ||Ax - y||_2^2 + \lambda ||x||_1 \tag{1}$$

and (c) a coupled CS based reconstruction that takes two consecutive slices and exploits the similarity between them.

1.1 Part (a)

Here are the images for each slice reconstructed by Back-projection with Ram-Lak Filter :

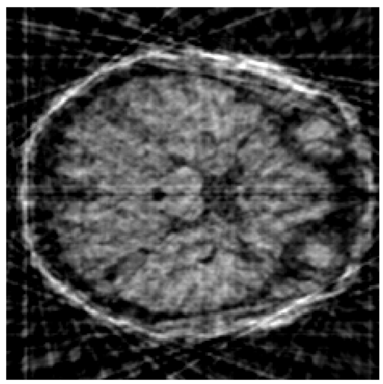

Figure 3: Slice 50 Image reconstructed by Back-projection with Ram-Lak Filter

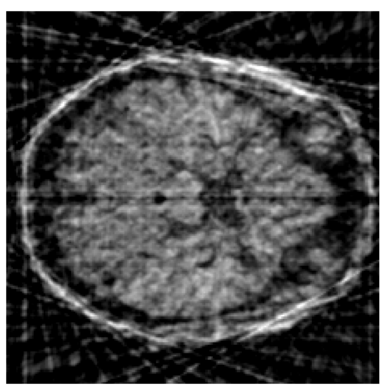

Figure 4: Slice 51 Image reconstructed by Back-projection with Ram-Lak Filter

1.2 Part (b)

Here are the images for each slice reconstructed by CS based reconstruction which involves solving the problem in equation (1) :

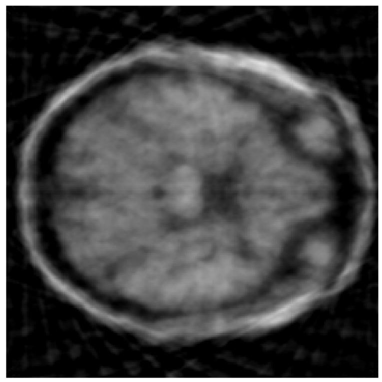

Figure 5: Slice 50 Image reconstructed by CS based reconstruction

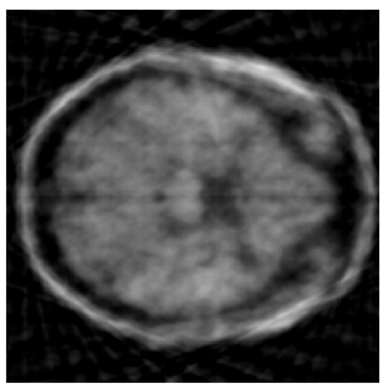

Figure 6: Slice 51 Image reconstructed by CS based reconstruction

1.3 Part (c)

Here, basically we are solving this problem :

$$min_{\beta} J(\beta) \tag{2}$$

where,

$$J(\beta) = ||y_1 - R_1 U \beta_1||_2^2 + ||y_2 - R_2 U \beta_2||_2^2 + \lambda||\beta_1||_1 + \lambda||\beta_2 - \beta_1||_1 \tag{3}$$

$$= ||y_1 - R_1U\beta_1||_2^2 + ||y_2 - R_2U(\beta_1 + \Delta\beta)||_2^2 + \lambda||\beta_1||_1 + \lambda||\Delta\beta||_1 \tag{4}$$

$$= ||y - A\beta||_2^2 + \lambda||\beta||_1 \tag{5}$$

where, $y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$, $A = \begin{pmatrix} R_1U & 0 \\ R_2U & R_2U \end{pmatrix}$ and $\beta = \begin{pmatrix} \beta_1 \\ \Delta\beta \end{pmatrix}$. Note that here, $\Delta\beta = \beta_2 - \beta_1$

Now, here are the images of the slices reconstructed using this coupled CS based method :

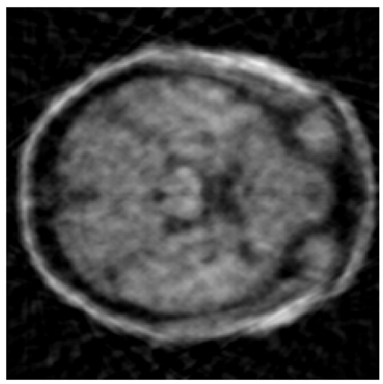

Figure 7: Slice 50 Image reconstructed by Coupled CS based reconstruction

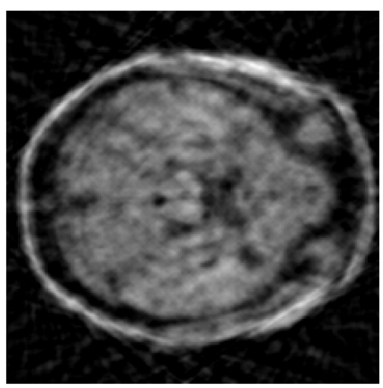

Figure 8: Slice 51 Image reconstructed by Coupled CS based reconstruction

For 3 slices, we have proposed the following CS based reconstruction problem :

$$min_\beta J(\beta) \tag{6}$$

where,

$$J(\beta) = ||y_1 - R_1 U\beta_1||_2^2 + ||y_2 - R_2 U\beta_2||_2^2 + ||y_3 - R_3 U\beta_3||_2^2 + \tag{7}$$

$$\lambda||\beta_1||_1 + \lambda||\beta_2 - \beta_1||_1 + \lambda||\beta_3 - \beta_1||_1 \tag{8}$$

$$= ||y_1 - R_1 U\beta_1||_2^2 + ||y_2 - R_2 U[\beta_1 + (\Delta\beta)_2]||_2^2 + ||y_3 - R_3 U[\beta_1 + (\Delta\beta)_3]||_2^2 + \tag{9}$$

$$\lambda||\beta_1||_1 + \lambda||(\Delta\beta)_2||_1 + \lambda||(\Delta\beta)_3||_1 \tag{10}$$

$$= ||y - A\beta||_2^2 + \lambda||\beta||_1 \tag{11}$$

where, $y = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}$, $A = \begin{pmatrix} R_1 U & 0 & 0 \\ R_2 U & R_2 U & 0 \\ R_3 U & 0 & R_3 U \end{pmatrix}$ and $\beta = \begin{pmatrix} \beta_1 \\ (\Delta\beta)_2 \\ (\Delta\beta)_3 \end{pmatrix}$. Here, $(\Delta\beta)_2 = \beta_2 - \beta_1$ and $(\Delta\beta)_3 = \beta_3 - \beta_1$.

You can check the code implementation for this problem in the `codes/p3` folder.

2 Problem 4

2.1 Details of the Paper

- **Title :** Tomographic determination of velocity and depth in laterally varying media
- **Authors :** T. N. Bishop, K. P. Bube, R. T. Cutler, R. T. Langan, P. L. Love, J. R. Resnick, R. T. Shuey, D. A. Spindler, and H. W. Wyld
- **Journal :** Geophysics, Volume 50, Issue 6
- **Link :** https://library.seg.org/doi/10.1190/1.1441970

2.2 Mathematical Problem

Let x be the horizontal distance along the Earth's surface and z be the depth. We will consider a specific region in this $x-z$ plane which we will call as Region-Of-Interest (ROI). In the ROI, we have n_r number of reflectors each having a depths $Z_\rho(x)$, $\rho = 1, 2, 3, ..., n_r$ as a function of x. If drawn on a graph, these reflectors will essentially be smooth curves. The slowness function $w(x, z)$ is the reciprocal of the seismic velocity. We will be characterizing these functions by a finite set of parameters.

Now, we divide the ROI into rectangular boxes that make a matrix with n_x columns and n_z rows. The value of $w(x, z)$ at the center of the box in the

k^{th} row and l^{th} column is denoted as $w_{k,l}$. The slowness function in rest of the box apart from the center is obtained by using an interpolation function that depends on $w_{k-1,l}$, $w_{k+1,l}$, $w_{k,l-1}$, $w_{k,l+1}$ and w_{kl}. Thus, th slowness function is parametrized by $w_{k,l}$.

The functions $Z_\rho(x)$ are parametrized by depths at which the reflectors intersect the vertical boundaries of the columns. Let the depth for ρ^{th} reflector intersecting the m^{th} boundary be $Z_{\rho,m}$.

Finally, we have $M = n_x \cdot n_z + n_r \cdot (n_x + 1)$ parameters that can be written as an M-dimensional vector p.

This is the vector that we need to find given the measurements of travel time of rays that are sent into the ROI and measurements are taken at receivers. Let the travel time for a ray that emerges from source μ, hits the reflector ρ and returns to receiver ν be $T_{\mu,\rho,\nu}$. These travel times are represented as an N-dimensional vector. Let the travel times predicted from the model parametrized by p be $t(p)$ and actual observed travel times be t_d. Let $r(p) = t_d - t(p)$. Our objective for Tomographic Reconstruction is to minimize $||r(p)||$ (the usual L_2 norm of $r(p)$). Thus our mathematical problem is

$$min_p \phi(p) \tag{12}$$

where

$$\phi(p) = ||r(p)||^2 = [t_d - t(p)]^T [t_d - t(p)] \tag{13}$$

2.3 Optimization Algorithm

The Optimization Algorithm that will be used to solve the above problem is known as the Gauss-Newton method which is used to solve a general non-linear least squares problem (Reference : https://en.wikipedia.org/wiki/Gauss-Newton-algorithm). Here's a brief review of this Algorithm.

For a local minimum of $\phi(p)$ it is essential that

$$\nabla \phi(p) = 0 \tag{14}$$

i.e. the gradient should be zero. Let $A(p)$ be the Jacobian of $t(p)$. Now the above condition implies that

$$A^T(p)[t(p) - t_d] = 0 \tag{15}$$

Our Algorithm finds the solution to this equation. This is done by starting at an approximation of p say $p^{(1)}$ and then iteratively finding $p^{(k+1)}$ from $p^{(k)}$. Let $\Delta p = p^{(k+1)} - p^{(k)}$, $A(p^{(k)}) = A^{(k)}$ and $r^{(k)} = r(p^{(k)})$. Now, expanding $t(p^{(k+1)})$ about $p^{(k)}$ in a Taylor Series up to a first order, we get

$$t(p^{(k+1)}) = t(p^{(k)}) + A^{(k)} \Delta p \tag{16}$$

and hence

$$A^{(k)T}[t(p^{(k+1)}) - t_d] = A^{(k)T}A^{(k)}\Delta p - A^{(k)T}r^{(k)} \tag{17}$$

The Algorithm will choose Δp such that

$$[A^{(k)T}A^{(k)} + \alpha(W^{(k)})^2]\Delta p = A^{(k)T}r^{(k)} \tag{18}$$

Here, α and the matrix $W^{(k)}$ are hyper-parameters that determine the rate of convergence of the Algorithm. If the Algorithm converges to a solution p^* then p^* is the solution to our problem.

3 Problem 5

The Radon Transform $R_\theta(f)$ of $f(x, y)$ is,

$$R_\theta(f) = g(\rho, \theta) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)\delta(xcos\theta + ysin\theta - \rho)dxdy \tag{19}$$

Hence, the Radon Transform of $f'(x, y) = f(ax, ay)$ will be,

$$R_\theta(f') = g'(\rho, \theta) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f'(x, y)\delta(xcos\theta + ysin\theta - \rho)dxdy \tag{20}$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(ax, ay)\delta(xcos\theta + ysin\theta - \rho)dxdy \tag{21}$$

Substituting $x' = ax$ and $y' = ay$ we get,

$$R_\theta(f') = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x', y')\delta(\frac{x'}{a}cos\theta + \frac{y'}{a}sin\theta - \rho)(\frac{1}{a^2})dx'dy' \tag{22}$$

$$= (\frac{1}{a^2})\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x', y')\delta[\frac{1}{a}(x'cos\theta + y'sin\theta - a\rho)]dx'dy' \tag{23}$$

$$= (\frac{1}{|a|^2})\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x', y')|a|\delta(x'cos\theta + y'sin\theta - a\rho)dx'dy' \tag{24}$$

$$= (\frac{1}{|a|})\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x', y')\delta(x'cos\theta + y'sin\theta - a\rho)dx'dy' \tag{25}$$

Comparing equations (19) and (25) we see that,

$$R_\theta(f') = g'(\rho, \theta) = \frac{g(a\rho, \theta)}{|a|} \tag{26}$$

4 Problem 6

4.1 Unit Impulse

The Radon Transform $R_\theta(\delta)$ of the unit impulse $\delta(x, y)$ is,

$$R_\theta(\delta) = g(\rho, \theta) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \delta(x, y)\delta(xcos\theta + ysin\theta - \rho)dxdy \tag{27}$$

$$= \delta(0 \cdot cos\theta + 0 \cdot sin\theta - \rho) = \delta(-\rho) = \delta(\rho) \tag{28}$$

Thus,

$$R_\theta(\delta) = g(\rho, \theta) = \delta(\rho) \tag{29}$$

4.2 Shifted Unit Impulse

Now, the Radon Transform for $f(x, y) = \delta(x - x_0, y - y_0)$ will be,

$$R_\theta(f) = g(\rho, \theta) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \delta(x - x_0, y - y_0)\delta(xcos\theta + ysin\theta - \rho)dxdy \tag{30}$$

$$= \delta(x_0cos\theta + y_0sin\theta - \rho) \tag{31}$$

Thus,

$$R_\theta(f) = g(\rho, \theta) = \delta(x_0cos\theta + y_0sin\theta - \rho) \tag{32}$$