

# CS 754 - Advanced Image Processing

## Assignment 3 - Report

Shaan ul Haque - 180070053  
Mantri Krishna Sri Ipsit - 180070032

March 23, 2021

### Question 1

#### 1.a

We would like the loss function defined by:

$$f_N(\beta) = \frac{\|\mathbf{y} - \mathbf{X}\beta\|_2^2}{2N}$$

to be strictly convex. For this we would like that the double derivative matrix,  $\nabla^2 f(\beta) = \mathbf{X}^T \mathbf{X}/N$ , wrt to  $\beta$  have all its eigenvalues be positive or be away from zero. Since  $\mathbf{X} \in \mathbf{R}^{p \times N}$ , the rank of  $\mathbf{X}^T \mathbf{X}$  is equal to  $\min\{N, p\}$  which makes the matrix rank deficit.

To relax the notion of convexity, we allow the function to be strictly convex in only some subset  $\mathbf{C} \subset \mathbf{R}^p$ . Thus, a function is strictly convex at  $\beta^*$  wrt to a set  $\mathbf{C}$  if:

$$\frac{v^T \nabla^2 f(\beta) v}{\|v\|_2^2} \geq \gamma \quad \forall \text{ non-zero } v \in \mathbf{C}$$

which, for our case gives (also called as *restricted eigenvalues*):

$$\frac{v^T \mathbf{X}^T \mathbf{X} v}{N \|v\|_2^2} \geq \gamma \quad \forall \text{ non-zero } v \in \mathbf{C}$$

Thus, we would like the minimum value of the LHS in the above equation to be greater than or equal to  $\gamma$ . Since the minimum value of the LHS is the minimum eigenvalue ( $\lambda_{min}$ ) with the corresponding eigenvector constrained in the subset  $\mathbf{C}$ , we can also write the above equation as:

$$\lambda_{min} \geq \gamma \quad \forall \text{ non-zero } v \in \mathbf{C}$$

This is called as restricted eigenvalue condition.

#### 1.b

It is given that  $\hat{\nu}$  is the minimizer of the equation:

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

Thus, from the above assumption we can write  $G(\hat{\nu}) \leq G(0)$

### 1.c

Using the inequality obtained in previous part:

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

Now since  $\mathbf{y} = \mathbf{X}\beta^* + \mathbf{w}$ , we substitute  $\mathbf{w}$  in the above inequality to get:

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

opening the brackets, we get:

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

which after simplification gives:

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

The above inequality after shifting terms right and left becomes our required expression:

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

### 1.d

Since the expression  $\frac{\mathbf{w}^T \mathbf{X} \hat{\nu}}{N}$  can also be written as  $\frac{\hat{\nu}^T \mathbf{X}^T \mathbf{w}}{N}$  because this quantity is a scalar and transpose of scalar is same as the scalar, we get:

$$\hat{\nu}^T \mathbf{X}^T \mathbf{w} = \langle \hat{\nu}, \mathbf{X}^T \mathbf{w} \rangle$$

Let for the sake of clarity assume  $\mathbf{X}^T \mathbf{w} = \mathbf{y}$ . Thus, the dot product becomes:

$$\langle \hat{\nu}, \mathbf{y} \rangle = \sum_{i=1}^p \hat{\nu}_i y_i \leq \sum_{i=1}^p |\hat{\nu}_i y_i| \leq \max_i \{|y_i|\} \sum_{i=1}^p |\hat{\nu}_i|$$

Writing the above equation in norm form we get ( $\|a\|_\infty = \max_n |a_n|$ ):

$$\langle \hat{\nu}, \mathbf{y} \rangle = \sum_{i=1}^p \hat{\nu}_i y_i \leq \|\mathbf{y}\|_\infty \|\hat{\nu}\|_1$$

Hence, after dividing by N both sides we get:

$$\frac{\hat{\nu}^T \mathbf{X}^T \mathbf{w}}{N} \leq \frac{\|\mathbf{X}^T \mathbf{w}\|_\infty}{N} \|\hat{\nu}\|_1$$

Let  $x_S \in R^{|S|}$  denote the subvector indexed by elements of set S, with  $x_{S^c}$  defined in an analogous manner.

For the other term we have  $\beta_{S^c}^* = 0$ , as  $\beta^*$  is assumed to be —S—sparse vector and:

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

Using above two inequalities, we can obtain eq. 11.22 as:

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

### 1.e

Firstly, it is assumed that  $\frac{\|\mathbf{X}^T \mathbf{w}\|_\infty}{N} \leq \frac{\lambda_N}{2}$ . Using this:

$$\begin{aligned} \frac{\|\mathbf{X} \hat{\nu}\|_2^2}{2N} &\leq \frac{\lambda_N}{2} (\|\hat{\nu}_S\|_1 - \|\hat{\nu}_{S^c}\|_1) + \lambda_N (\|\hat{\nu}_S\|_1 + \|\hat{\nu}_{S^c}\|_1) \\ &= \frac{3}{2} \lambda_N \|\hat{\nu}_S\|_1 - \frac{1}{2} \lambda_N \|\hat{\nu}_{S^c}\|_1 \leq \frac{3}{2} \lambda_N \|\hat{\nu}_S\|_1 \leq \frac{3}{2} \sqrt{k} \lambda_N \|\hat{\nu}\|_2 \quad (\text{using Cauchy-Schwartz inequality}) \end{aligned}$$

### 1.f

Using the restricted Eigenvalue condition and assuming the lasso error  $\hat{\beta} - \beta^*$  is in the set  $\mathbf{C}$  (defined in part a), we have:

$$\frac{v^T \mathbf{X}^T \mathbf{X} v}{\|v\|_2^2} \geq \gamma \quad \forall \text{ non-zero } v \in \mathbf{C}$$

which gives:

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

From the bound derived in previous part we have:

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

Thus, we have:

$$\|\hat{\beta} - \beta^*\|_2 = \|\hat{\nu}\|_2 \leq \frac{3}{\gamma} \sqrt{\frac{k}{N}} \sqrt{N} \lambda_N$$

### 1.g

It is an assumption that we make in order to prove the cone constraint that  $\|\hat{\nu}_{S^c}\|_1 \leq \alpha \|\hat{\nu}_S\|_1$ . Specifically, we used the inequality to prove the bound in part e. Using the inequality in part e, we have:

$$\frac{\|\mathbf{X} \hat{\nu}\|_2^2}{2N} \leq \frac{3}{2} \lambda_N \|\hat{\nu}_S\|_1 - \frac{1}{2} \lambda_N \|\hat{\nu}_{S^c}\|_1$$

Since the LHS is always greater than or equal to 0, we get:

$$0 \leq \frac{3}{2} \lambda_N \|\hat{\nu}_S\|_1 - \frac{1}{2} \lambda_N \|\hat{\nu}_{S^c}\|_1 \implies \|\hat{\nu}_{S^c}\|_1 \leq 3 \|\hat{\nu}_S\|_1 \quad (\text{Cone constraint})$$

### 1.h

In the first part we saw that in order to minimize the loss function it must be strictly convex whereas in practice it is convex in some direction while non-concave in others. Thus, we relaxed the condition and allowed strict convexity only in a subset  $\mathbf{C} \subset \mathbf{R}^p$ . The lasso regression constraint that  $\|\beta\|_1 \leq R$  with or without regularization leads the subset to be cone, i.e.,  $\|\hat{\nu}_S\|_1 \leq \alpha \|\hat{\nu}_{S^c}\|_1$ .

## 1.i

In both the theorems we are required to calculate the minimum eigenvalue of our dictionary matrix but in this theorem our search for the minimum eigenvalue is deterministic, i.e., in a set with cone constraint and moreover here we do not need the explicit calculation of the RIC. One fascinating thing to observe here is that this theorem also emphasizes on the interaction of the dictionary with noise vector, greater the correlation greater is the error whereas theorem 3 doesn't take any such thing into account. Apart from this, the rate this theorem achieves, including the logarithmic factor, is known to be minimax optimal, meaning that it cannot be substantially improved upon by any estimator.

The biggest drawback of this theorem is that it doesn't consider compressible signals. The sparsity constraint for the sensed signal is very tight and does not allow room for signals which are less sparse but some elements are very small. Theorem 3 handles such situation well.

## 1.j

The common thread between 'Dantzig selector' and the LASSO is accounting for the interaction of dictionary with the noise vector in the error bound. For 'Dantzig selector' the maximum correlation  $\|A^T e\|_\infty$  is assumed to be less than some constant  $\lambda$ . A similar kind of constraint was assumed for LASSO also, as assumed for proving the sub part 1.e.

## 1.k

Square-root lasso's biggest advantage is that without knowing or pre-estimating the noise variance  $\sigma^2$ , we can get near-oracle performance. The regression parameter in the lasso regression depends on the noise variance whose estimation in certain cases might be non-trivial. Moreover, in all  $l_2$  error constraints, noise is assumed to be Gaussian in nature. But the theorem in paper doesn't require the knowledge of the noise variance, on top of that it doesn't even assume noise to be Gaussian in nature.

## Question 2

The original images of the slices used are shown below. They have been zero-padded to size  $255 \times 255$ .

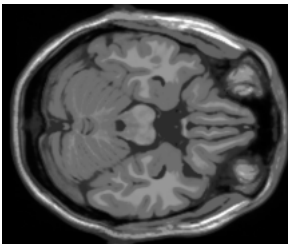


Figure 1: Slice 50

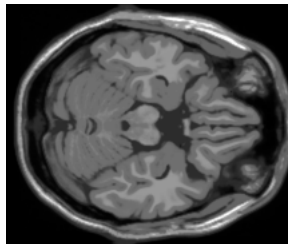


Figure 2: Slice 51

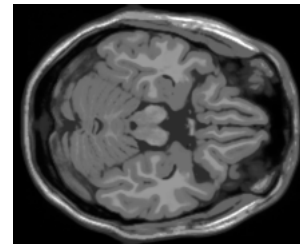
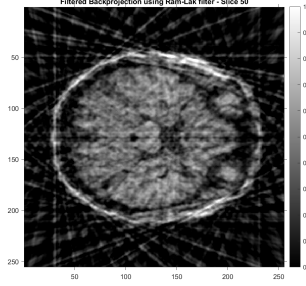


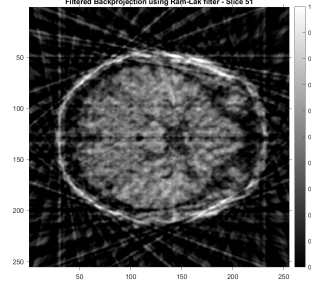
Figure 3: Slice 52

## 2.a

Here the tomographic reconstruction is performed using the **Filtered Backprojection** using the **Ram-Lak** filter. The reconstruction results for slices 50 and 51 are as follows:

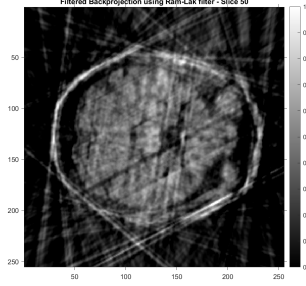


Ram-Lak FBP - Slice 50

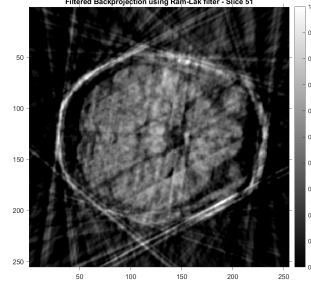


Ram-Lak FBP - Slice 51

Figure 4: Using 18 uniformly spaced angles for measurement



Ram-Lak FBP - Slice 50



Ram-Lak FBP - Slice 51

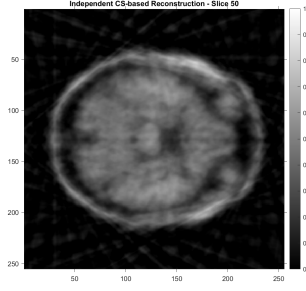
Figure 5: Using 36 randomly chosen angles for measurement

## 2.b

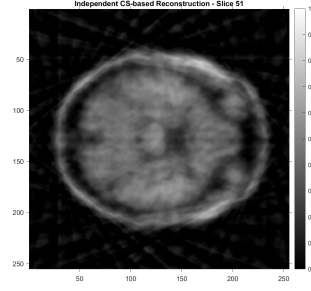
Here the tomographic reconstruction is performed by using the **Independent Compressive Sensing** based reconstruction for each slice by solving an optimization problem of the form

$$J(x) = \|y - Ax\|_2^2 + \lambda \|x\|_1$$

The reconstruction results for slices 50 and 51 are as follows:

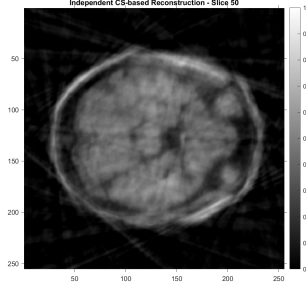


Independent CS - Slice 50

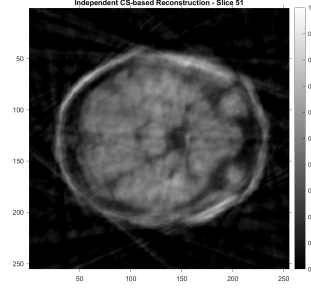


Independent CS - Slice 51

Figure 6: Using 18 uniformly spaced angles for measurement



Independent CS - Slice 50



Independent CS - Slice 51

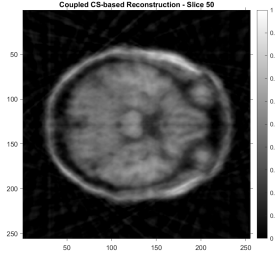
Figure 7: Using 36 randomly chosen angles for measurement

## 2.c

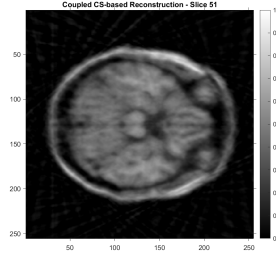
Here the tomographic reconstruction is performed by using the **Coupled Compressive Sensing** based reconstruction by solving an optimization problem of the form:

$$J\left(\begin{bmatrix} \beta_1 \\ \Delta\beta_1 \end{bmatrix}\right) = \left\| \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} - \begin{bmatrix} R_1 U & 0 \\ R_2 U & R_2 U \end{bmatrix} \begin{bmatrix} \beta_1 \\ \Delta\beta_1 \end{bmatrix} \right\|_2^2 + \lambda \left\| \begin{bmatrix} \beta_1 \\ \Delta\beta_1 \end{bmatrix} \right\|_1$$

The reconstruction results for slices 50 and 51 are as follows:

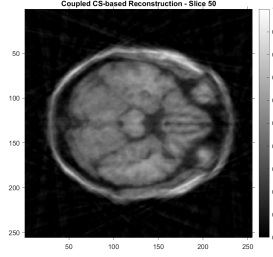


Coupled CS - Slice 50

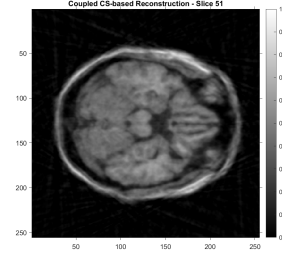


Coupled CS - Slice 51

Figure 8: Using 18 uniformly spaced angles for measurement



Coupled CS - Slice 50



Coupled CS - Slice 51

Figure 9: Using 36 randomly chosen angles for measurement

## 2.d - Using 3 slices

Here the tomographic reconstruction is performed by using the **Coupled Compressive Sensing** based reconstruction **using 3 slices** by solving an optimization problem of the form:

$$J\left(\begin{bmatrix} \beta_1 \\ \Delta\beta_1 \\ \Delta\beta_2 \end{bmatrix}\right) = \left\| \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} - \begin{bmatrix} R_1 U & 0 & 0 \\ R_2 U & R_2 U & 0 \\ R_3 U & 0 & R_3 U \end{bmatrix} \begin{bmatrix} \beta_1 \\ \Delta\beta_1 \\ \Delta\beta_2 \end{bmatrix} \right\|_2^2 + \lambda \left\| \begin{bmatrix} \beta_1 \\ \Delta\beta_1 \\ \Delta\beta_2 \end{bmatrix} \right\|_1$$

Here

- $y_1, y_2, y_3$  are the radon transforms of 3 slices  $x_1, x_2, x_3$  under consideration.
- $\beta_1$  is the 2D-DCT of first slice  $x_1$ , vectorized.
- $\beta_1 + \Delta\beta_1$  is the 2D-DCT of second slice  $x_2$ , vectorized.
- $\beta_1 + \Delta\beta_2$  is the 2D-DCT of second slice  $x_3$ , vectorized.
- $R_1$  is the radon transform matrix for the random angles used to compute tomographic projections of  $x_1$ .
- $R_2$  is the radon transform matrix for the random angles used to compute tomographic projections of  $x_2$ .
- $R_3$  is the radon transform matrix for the random angles used to compute tomographic projections of  $x_3$ .
- $U$  is the 2D-DCT basis matrix

The reconstruction results for slices 50, 51 and 52 are as follows:

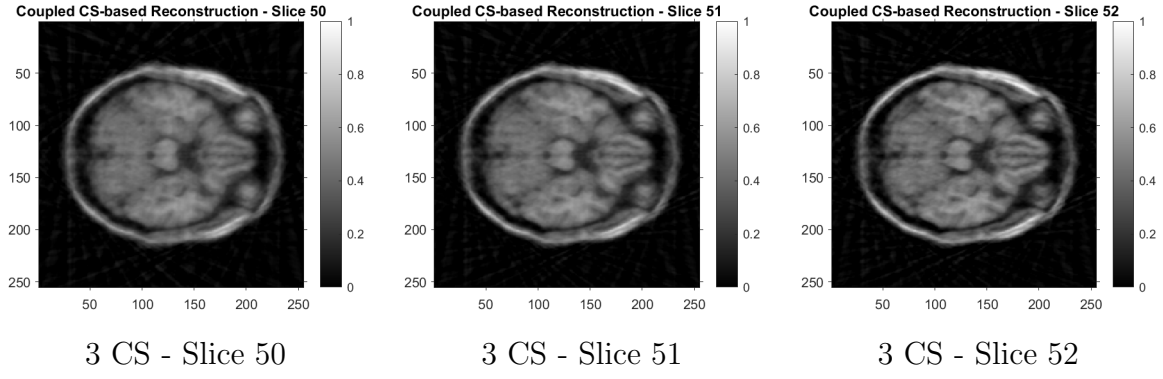


Figure 10: Using 18 uniformly spaced angles for measurement

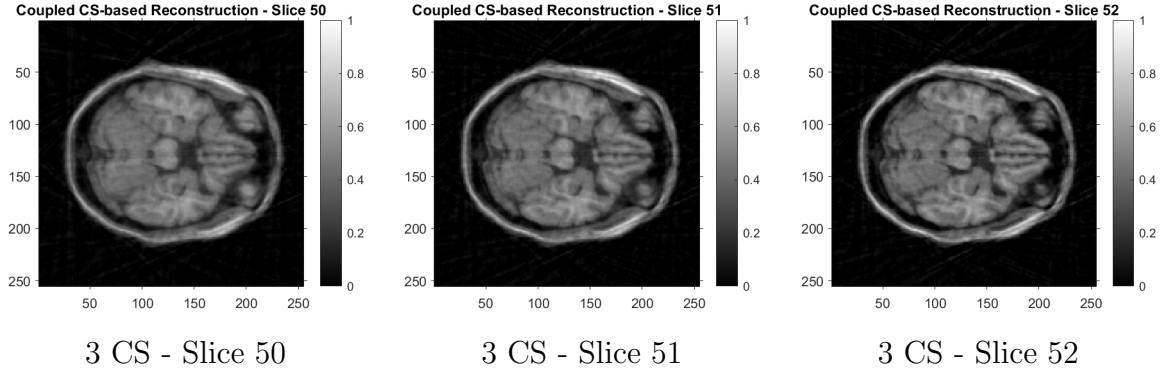


Figure 11: Using 36 uniformly chosen angles for measurement

The performance in all the above cases improves as the number of directions is increased.

### Question 3

For the sake of notational simplicity, denote  $R(\rho, \theta)$  as the radon transform of  $g(x, y)$  at a translation  $\rho$  and at an angle  $\theta$ .

$$R(\rho, \theta) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) \delta(x \cos \theta + y \sin \theta - \rho) dx dy$$

*Please turn over*



### 3.a

$$R(g(x - x_0, y - y_0))(\rho, \theta) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x - x_0, y - y_0) \delta(x \cos \theta + y \sin \theta - \rho) dx dy$$

Replace  $x \rightarrow u + x_0, y \rightarrow v + y_0$

$$\begin{aligned} &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(u, v) \delta(u \cos \theta + v \sin \theta - (\rho - x_0 \cos \theta - y_0 \sin \theta)) dx dy \\ &= R(g(u, v))(\rho - x_0 \cos \theta - y_0 \sin \theta, \theta) \\ &= R(g(x, y))(\rho - x_0 \cos \theta - y_0 \sin \theta, \theta) \end{aligned}$$

### 3.b

$$g'(r, \psi) = g(r, \psi - \psi_0)$$

Consider the radon transform in polar coordinates:

$$R(g)(\rho, \theta) = \int_0^{2\pi} \int_0^{\infty} g(r, \psi) \delta(r \cos(\psi - \theta) - \rho) r dr d\psi$$

*Please turn over*

Now,

$$\begin{aligned}
 R(g')(\rho, \theta) &= \int_0^\infty \int_0^{2\pi} g'(r, \psi) \delta\left(r \cos(\psi - \theta) - \rho\right) r dr d\psi \\
 &= \int_0^\infty \int_0^{2\pi} g(r, \psi - \psi_0) \delta\left(r \cos(\psi - \theta) - \rho\right) r dr d\psi
 \end{aligned}$$

Replace  $\psi \rightarrow \psi_0 + \alpha$

$$\begin{aligned}
 &= \int_0^\infty \int_{-\psi_0}^{2\pi-\psi_0} g(r, \alpha) \delta\left(r \cos(\alpha + \psi_0 - \theta) - \rho\right) r dr d\alpha \\
 &= \int_0^\infty \int_0^{2\pi} g(r, \alpha) \delta\left(r \cos(\alpha - (\theta - \psi_0)) - \rho\right) r dr d\alpha
 \end{aligned}$$

Replace  $\alpha \rightarrow \psi$

$$\begin{aligned}
 &= \int_0^\infty \int_0^{2\pi} g(r, \psi) \delta\left(r \cos(\psi - (\theta - \psi_0)) - \rho\right) r dr d\psi \\
 &= R(g)(\rho, \theta - \psi_0)
 \end{aligned}$$

### 3.c

The convolution of two signals  $f(x, y)$  and  $k(x, y)$  is given by:

$$(f * k)(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(u, v) k(x - u, y - v) du dv$$

*Please turn over*

The radon transform of the convolution will be

$$\begin{aligned}
R_\theta(f * k) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (f * k)(x, y) \delta(x \cos \theta + y \sin \theta - \rho) dx dy \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(u, v) k(x - u, y - v) \delta(x \cos \theta + y \sin \theta - \rho) du dv dx dy
\end{aligned}$$

Replace  $x - u \rightarrow a, y - v \rightarrow b$

$$\begin{aligned}
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(u, v) k(a, b) \delta((u + a) \cos \theta + (v + b) \sin \theta - \rho) du dv da db \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(u, v) k(a, b) du dv da db \int_{-\infty}^{\infty} \delta(u \cos \theta + v \sin \theta - \sigma) \\
&\quad \delta(a \cos \theta + b \sin \theta - (\rho - \sigma)) d\sigma \\
&= \int_{-\infty}^{\infty} \left( \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(u, v) \delta(u \cos \theta + v \sin \theta - \sigma) du dv \right) \\
&\quad \left( \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} k(a, b) \delta(a \cos \theta + b \sin \theta - (\rho - \sigma)) da db \right) d\sigma \\
&= \int_{-\infty}^{\infty} R(f)(\sigma, \theta) R(k)(\rho - \sigma, \theta) d\sigma \\
&= R_\theta(f) * R_\theta(k)
\end{aligned}$$

## Question 4

Mutual coherence is defined as:

$$\mu(\mathbf{A}) = \max_{i,j,i \neq j} \frac{|\mathbf{A}_i^T \mathbf{A}_j|}{\|\mathbf{A}_i\|_2 \|\mathbf{A}_j\|_2} = \max_{i,j,i \neq j} |\mathbf{A}_i^T \mathbf{A}_j|$$

where, the second equality assumes unit normalized columns. Now coming at RIC:

$$\delta_S = \max\{\lambda_{max} - 1, 1 - \lambda_{min}\}$$

where,  $\lambda_{max}$  and  $\lambda_{min}$  are the maximum and minimum Eigenvalue of  $\mathbf{X} = \mathbf{A}_\Gamma^T \mathbf{A}_\Gamma$ . Here subscript  $\Gamma$  denotes only those indices retained in column of the matrix where the sparse vector  $\boldsymbol{\theta}$  has non-zero elements. Gershgorin theorem states that every Eigenvalue of a matrix B lies between:

$$B_{ii} - r_i \leq \lambda \leq B_{ii} + r_i$$

where  $B_{ii}$  is the diagonal element at  $i$ th row while  $r_i$  is the sum of off-diagonal elements. Using this we have:

$$\mathbf{X}_{ii} - r_i \leq \lambda_{max} \leq \mathbf{X}_{ii} + r_i$$

Now, since every (i,j) element of  $\mathbf{X}$  is the dot product between the ith and jth column of  $\mathbf{A}_\Gamma$  and  $\mu(\mathbf{A})$  is maximum among them barring the special case when i=j where the value of  $\mathbf{X}_{ii} = 1$ , we get:

$$\lambda_{max} \leq 1 + \mu(\mathbf{A})(S - 1) \implies \lambda_{max} - 1 \leq 1 + \mu(\mathbf{A})(S - 1)$$

where, the RHS is the consequence of only  $|\Gamma| \leq S$  columns in  $\mathbf{A}_\Gamma$ . Now using the other side of inequality for  $\lambda_{min}$ , we get:

$$1 - \mu(\mathbf{A})(S - 1) \leq \lambda_{min} \implies 1 - \lambda_{min} \leq \mu(\mathbf{A})(S - 1)$$

Using the above two inequalities, we get:

$$\max\{\lambda_{max} - 1, 1 - \lambda_{min}\} \leq \mu(\mathbf{A})(S - 1) \implies \delta_S \leq \mu(\mathbf{A})(S - 1)$$

## Question 5

- **Title:** Tomographic reconstruction of the ionosphere using ground-based GPS data in the Australian region.
- **Venue:** Physics Department, La Trobe University, Bundoora, Vic 3086, Australia
- **Year of publication of the paper:** 2015

## Mathematical Problem

The reconstruction plane is discretized into two-dimensional pixels as shown in Fig. Usually, the grid of these two dimensional boxes is subdivided equidistantly with height and angular spacing. For the present study the grid boxes are 10 km in height and  $1^\circ$  ( 100 km) in width as shown in Fig. By assuming that the electron density is constant in each pixel, the TEC(Total Electron Content) along the ray path can be represented as a finite sum of shorter integrals along segments of the ray path length. Mathematically this can be expressed as:

$$STEC_i = \sum_{j=1}^M n_j d_{ij} + e_j$$

where,  $STEC_i$ (Slanting Total Electron Content) is the measurement for the ith projection,  $n_j$  is the electron density in the jth pixel,  $d_{ij}$  is the measurement constant,  $e_j$  is the noise while M is the total number of pixels along each projection. In matrix form, it can be written as:

$$\mathbf{Y}_{T \times 1} = \mathbf{D}_{T \times N} \mathbf{N}_{N \times 1} + \mathbf{E}_{T \times 1}$$

where  $\mathbf{Y}$  is a column of T measurements,  $\mathbf{N}$  is a column of the M unknown  $n_j$ s,  $\mathbf{E}$  is a column of T values representing the error due to data noise and discretization,  $\mathbf{D}$  is a  $T \times M$  matrix with  $d_{ij}$  being the length of link i that lies in pixel j, and thus  $d_{ij}$  is 1 if the ith ray traverses through the jth pixel and 0 otherwise. Due to high precision in measurements, the model is simplified into:

$$\mathbf{Y}_{T \times 1} = \mathbf{D}_{T \times N} \mathbf{N}_{N \times 1}$$

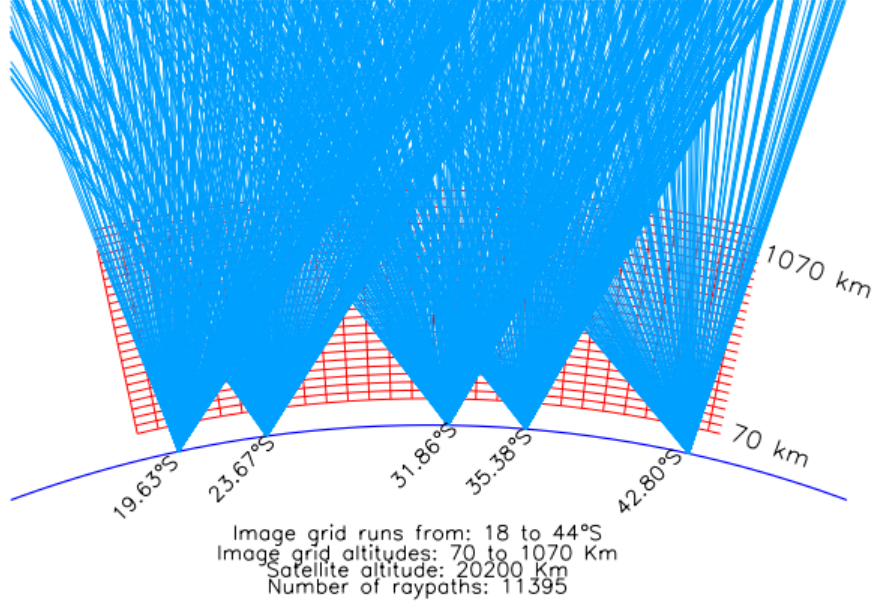


Figure 12: Experimental setup

## Method of optimization

One of the most commonly used inversion techniques is called algebraic reconstruction technique (ART). The ART algorithm, which can converge quickly in an iterative fashion compared to other reconstruction algorithms, is the preferable algorithm to use for ionospheric reconstruction in a region of interest with a limited widely spaced number of receivers, like the GPS receiver network in the Australian region. Due to these preferences, the ART algorithm has been used in this paper.

$$\mathbf{N}^{k+1} = \mathbf{N}^k + \lambda_k \frac{STEC_i - \sum_{j=1}^M d_{ij} n_j^k}{\sum_{j=1}^M d_{ij} d_{ij}} \mathbf{D}_i$$

where  $\mathbf{D}_i$  is the  $i$ th row of  $\mathbf{D}$ ,  $k$  is iteration number, and  $\lambda_k$  is the relaxation parameter. The relaxation parameter ensures that the correction remains stable, and it is a real number usually confined to the interval  $0 < \lambda_k < 2$ . The value of  $\lambda_k$  is chosen to be the same for all iterations. In this case  $\lambda_k = 0.005$  had been used for all iterations.