CS754 Assignment 3

Hitesh Kumar Punna, Sudhansh Peddabomma

190050093, 190050118

Question 1.

(a) The Restricted eigenvalue condition states that, there exists a γ such that the following inequality is satisfied

$$\frac{v^T X^T X v}{N * \|v\|_2^2} \geq \gamma$$

for all non zero $v \in C$ where $C \subset \mathbb{R}^p$ and perturbation vectors $v \in \mathbb{R}^p$. Here X= sensing matrix of dimension $N \times P$, N = Number of Observations and P = size of the Original Image. (Note: Here $\mathbb{R}^p$ is set of vectors in real space of dimension "px1")

(b)

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

and we know that we from minimizing the cost function

$$f(\beta) = minimize_{\beta \in \mathbb{R}^p} \frac{1}{2N}\|y - X\beta\|_2^2 + \lambda_N \|\beta\|_1$$

Let $\hat{\beta}$ be the LASSO estimate. Hence we say that for all $\beta^\star \in R^p$,

$$f(\hat{\beta}) \leq f(\beta^\star)$$

as we can see $f(\beta) = G(\beta - \beta^\star) \Rightarrow G(\hat{\beta} - \beta^\star) \leq G(0)$ for all $\beta^\star \in \mathbb{R}^p$ Therefore as $\hat{v} = \hat{\beta} - \beta^\star$

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

(c) As $G(\hat{v}) \leq G(0)$, we can say that

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

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

and as $y - X\beta^\star = w$

$$\frac{1}{2N}(||w - X\hat{v}||_2^2 - ||w||_2^2) \leq \lambda_N(||\beta^\star||_1 - ||\hat{v}+\beta^\star||_1)$$

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

as $w^T X\hat{v}$ is a 1x1 matrix , $\hat{v}^T X^T w = w^T X\hat{v}$

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

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

(d) Holder's inequality states that $||fg||_1 \leq ||f||_1||g||_\infty$, for all measurable functions f, g

From this we can say that $||w^T X\hat{v}||_1 \leq ||w^T X||_\infty||\hat{v}||_1 \Rightarrow ||w^T X\hat{v}||_1 \leq ||X^T w||_\infty||\hat{v}||_1$ (as $||A||_p = ||A^T||_p \forall p \in N$) as $w^T X\hat{v}$ is a 1x1 matrix, $w^T X\hat{v} = ||w^T X\hat{v}||_1$

From the previous equations we can say that,

$$\frac{1}{2N}||X\hat{v}||_2^2 \leq \frac{1}{N}||X^T w||_\infty||\hat{v}||_1 + \lambda_N(||\beta^\star||_1 - ||\hat{v} + \beta^\star||_1)$$

Now since $\beta^\star_{S^c} = 0$, we have $||\beta^\star||_1 = ||\beta^\star_S||_1$ (as $\beta^\star$ is "S" sparse), and

$$||\beta^\star + \hat{v}||_1 = ||\beta^\star_S + \hat{v}_S||_1 + ||\hat{v}_{S^c}||_1 \geq ||\beta^\star||_1 - ||\hat{v}_S||_1 + ||\hat{v}_{S^c}||_1$$

$$||\beta^\star||_1 - ||\hat{v} + \beta^\star||_1 \leq ||\hat{v}_S||_1 - ||\hat{v}_{S^c}||_1$$

$$\frac{1}{2N}||X\hat{v}||_2^2 \leq \frac{1}{N}||X^T w||_\infty||\hat{v}||_1 + \lambda_N(||\hat{v}_S||_1 - ||\hat{v}_{S^c}||_1)$$

(e) We use our Assumption from the Theorem 11.14b, which is $\frac{1}{N}||X^T w||_\infty \leq \frac{\lambda_N}{2}$,from this we can say

$$\frac{1}{2N}||X\hat{v}||_2^2 \leq \frac{\lambda_N}{2}||\hat{v}||_1 + \lambda_N(||\hat{v}_S||_1 - ||\hat{v}_{S^c}||_1)$$

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

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

$$\frac{1}{2N}||X\hat{v}||_2^2 \leq \frac{\lambda_N}{2}3||\hat{v}_S||_1$$

(as$||\hat{v}||_1 \leq \sqrt{k}||\hat{v}||_2$, here k= sparsity)

$$\frac{1}{2N}||X\hat{v}||_2^2 \leq \frac{3}{2}\sqrt{k}\lambda_N||\hat{v}_S||_2$$

and $||\hat{v}_S||_2 \leq ||\hat{v}||_2$

$$\frac{1}{2N}||X\hat{v}||_2^2 \leq \frac{3}{2}\sqrt{k}\lambda_N||\hat{v}||_2$$

(f) From Equation 11.1 we can say that, $||X\hat{v}||_2^2 \geq \gamma N||\hat{v}||_2^2$ (as $||A||_2^2 = A^T A$, where A is a $m \times 1$ matrix and $X\hat{v}$ is a $P \times 1$ matrix)
So from the equation in previous part we can say that,

$$\frac{1}{2}\gamma||\hat{v}||_2^2 \leq \frac{1}{2N}||X\hat{v}||_2^2 \leq \frac{3}{2}\sqrt{k}\lambda_N||\hat{v}||_2$$

$$||\hat{v}||_2^2 \leq \frac{3}{\gamma}\sqrt{k}\lambda_N||\hat{v}||_2$$

$$||\hat{v}||_2 \leq \frac{3}{\gamma}\sqrt{k}\lambda_N$$

$$||\hat{v}||_2 \leq \frac{3}{\gamma}\sqrt{\frac{k}{N}}\sqrt{N}\lambda_N$$

Hence Theorem 11.14b is Proved.

(g) 1. The bound $\frac{1}{N}||X^T w||_\infty \leq \frac{\lambda_N}{2}$ is used in proving Theorem 11.23 ($\frac{1}{2N}||X\hat{v}||_2^2 \leq \frac{3}{2}\sqrt{k}\lambda_N||\hat{v}||_2$) from Theorem 11.22 ($\frac{1}{2N}||X\hat{v}||_2^2 \leq \frac{1}{N}||X^T w||_\infty||\hat{v}||_1 + \lambda_N(||\hat{v}_S||_1 - ||\hat{v}_{S^c}||_1)$) as we did in the part (e) of this question.

2. And the bound $\frac{1}{N}||X^T w||_\infty \leq \frac{\lambda_N}{2}$ also helps in establishing the cone constraint $||\hat{v}_{S^c}||_1 \leq 3||\hat{v}_S||_1$ in the following way,

   and Our assumption that $\frac{1}{N}||X^T w||_\infty \leq \frac{\lambda_N}{2}$ we have the following,

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

   from this we can say

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

   So,

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

   from this we have $\alpha = 3$

(h) While Restricting the Eigenvalues using

$$\frac{v^T X^T X v}{N * ||v||_2^2} \geq \ \gamma$$

for all non zero $v \in C$, we need to choose an efficient Constraint Set C such that it's easy for computation.

Suppose the parameter vector $\beta^\star$ is sparse, and is supported on the subset $S = S(\beta^\star)$.

Defining the lasso error $\hat{v} = \hat{\beta} - \beta^\star$, let $\hat{v}_S \in R^{|S|}$ denote the sub vector indexed by elements of S, with v_{S^c} defined in an analogous manner.

Now we define the Set "C" by using a cone constraint as,

$$C(S;\alpha) := \{v \in R_p|||v_{S^c}||_1 \leq \alpha||v_S||_1\}$$

Having different choices for "α" will make the Set "C" more easy for computing.

α can be chosen in two ways,

1. Regular form:-which has $\alpha = 1$
2. Constrained form using Lasso Norm:-which has $\alpha = 3$

Using this cone constraint we can restrict the lasso error $\hat{v} = \hat{\beta} - \beta^\star$ to smaller set of values, and for a given regularization parameter λ_N (as we use $\frac{1}{N}||X^T w||_\infty \leq \frac{\lambda_N}{2}$). This Cone Constraint helps to prove the bound for theorem 11.14b

$$||\hat{v}||_2 \leq \frac{3}{\gamma}\sqrt{\frac{k}{N}}\sqrt{N}\lambda_N$$

This allows us to implement the estimator using efficient algorithmic methods, such as interior-point methods, which provide polynomial-time bounds on computational time.

(i) From the Exercise 11.1 we can see that

$$\lambda_N = 2\sigma\sqrt{T\frac{\log p}{N}} \text{ for some } T > 2 \text{ , and } ||\hat{\beta} - \beta^\star||_2 \leq \frac{c\sigma}{\gamma}\sqrt{\frac{Tk\log p}{N}}$$

Whereas the bound given by our theorem 3 is $||\theta - \theta^\star||_2 \leq \frac{C_0}{\sqrt{S}}||\theta - \theta_S||_1 + C_1\epsilon$ (here s=k=sparsity)
Lasso Estimation is advantageous over Theorem 3 in the following ways.

1. Cases for which the Support set ($S(\beta^\star)$) is known, (i.e k is known) but each $\beta_i^\star$, $\forall i \in S(\beta^\star)$ is not known. In such cases Lasso Estimation gives error bound ($\hat{\beta} - \beta^\star$) in the order of $\frac{k}{N}$, and no method can achieve squared l2-error that decays more quickly than $\frac{k}{N}$, hence it is more efficient than Theorem3.

2. On increasing the number of observations N , there is no significant change in the error bound for Theorem 3 , as the RIP constant(δ_s) depends on sparsity but does not change much with change in Number of observations, So the error bound remains unaffected.

   On the other hand,the error bound would be significantly decrease for LassoEstimate.As error bound is

$$||\hat{\beta} - \beta^\star||_2 \leq \frac{c\sigma}{\gamma}\sqrt{\frac{Tk\log p}{N}}$$

   it decreases as N increases. So, it would be easier to attain the solution.

3. We need to check restricted eigen value condition over C(S;3) in this theorem while we need to check over all $x \in R^p$ for theorem 3.Therefore this theorem is better than Theorem 3.

4. In fact, the rate in Lasso Estimation ($||\hat{\beta} - \beta^\star||_2 \leq \frac{c\sigma}{\gamma}\sqrt{\frac{Tk\log p}{N}}$) including the logarithmic factor ($\log p$)is known to be "minimax optimal", meaning that it cannot be substantially improved upon by any estimator.

Advantages of Theorem3 over this theorem is

– Theorem 3 is better than this theorem in the aspect that it does not have any probability term .Therefore we cant set σ and T arbitrarily because that may decrease the probability value

(j)
1. The Similarity between the 'Dantzig selector' and the LASSO is is the Order of Error bound which is O($\sqrt{k\sigma^2\log n}$) the bound is also called as "near oracle estimator."

2. Both bounds are assuming Gaussian Noise.

3. Both are assuming unit normalized columns in Sensing matrix.

4. While they both provide very similar guarantees, there are certain circumstances where the Dantzig selector is preferable, like the Corollary 1.1 and Corollary 1.2, which is used in the book. We can see that on decreasing 'k' (sparsity) there is very small change in error bound in Corollary 1.1, where as there is a significant decrease in the upper bound for Corollary 1.2, By this we can say that Dantzig selector is preferable in certain cases over LASSO.

5. It can also be seen that results such as Corollary 1.2 guarantee that the Dantzig selector achieves an error $||\hat{x} - x||_2$ which is bounded by a constant times $k\sigma^2\log n$, with high probability. Note that since we typically require $m > k\log n$, this can be substantially lower than the expected noise power $E||e||_2 = m\sigma^2$, illustrating the fact that sparsity-based techniques are highly successful in reducing the noise level.

(k) Lasso Estimate is

$$\hat{\beta} \in argmin_{\beta \in R^p} Q(\hat{\beta}) + \frac{\lambda}{N}||\beta||_1$$

and the penalty level is $\lambda = 2\sigma c n^{1/2} X^{-1}(1-\alpha/2p)$
Square Lasso Estimate is

$$\hat{\beta} \in argmin_{\beta \in R^p} \{Q(\hat{\beta})\}^{1/2} + \frac{\lambda}{N}||\beta||_1$$

and the penalty level is $\lambda = c n^{1/2} X^{-1}(1-\alpha/2p)$

1. In both Square Root LASSO and LASSO we have a factor σ referring to the standard deviation of the Noise.Estimation of σ is non-trivial when p is large, particularly when $p \gg n$, and remains an outstanding practical and theoretical problem.The estimator we propose in this paper, the square-root lasso, eliminates the need to know or to pre estimate σ.

2. The penalty level in Square Root LASSO is independent of σ, in contrast to LASSO, and hence is pivotal with respect to this parameter.

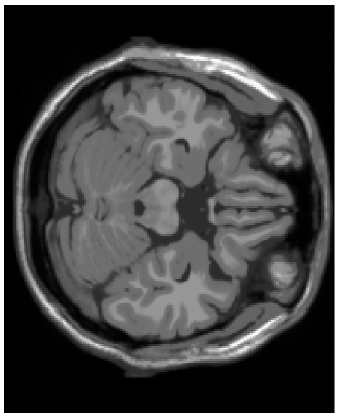
Question 2.

Instructions to run the code.

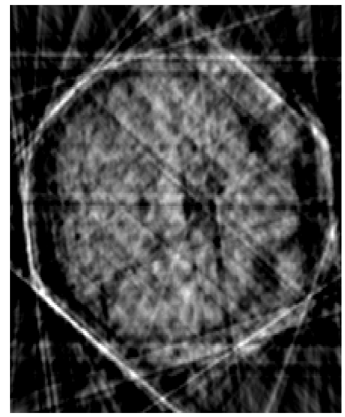
- Copy `Q2.m` to the folder containing the images and `l1_ls` package.
- The code uses slices 50 and 51 for the first 3 parts and 50, 51 and 52 for the last part.

1. Filtered Back Projection

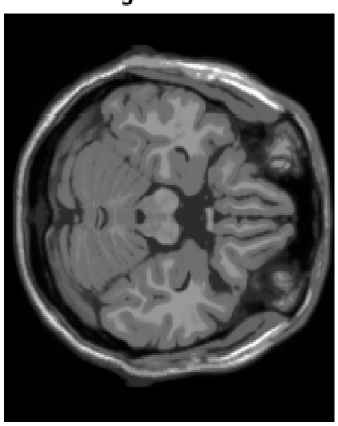
Original slice 1



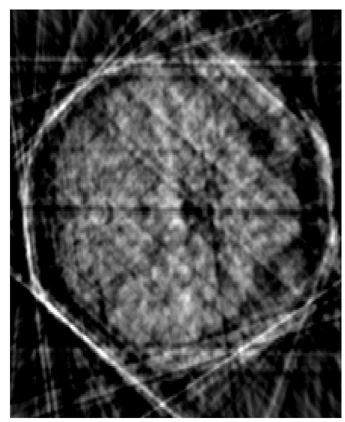
Reconstructed slice 1



Original slice 2

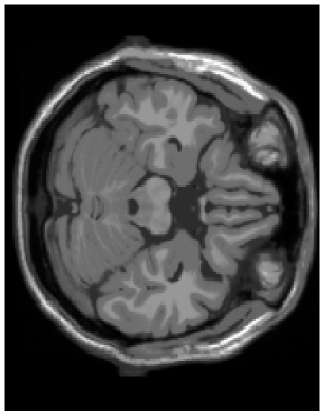


Reconstructed Slice 2

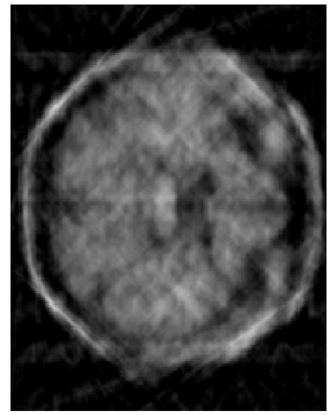


2. Independent CS reconstruction

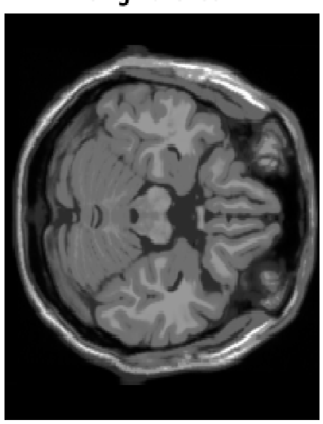
Original slice 1



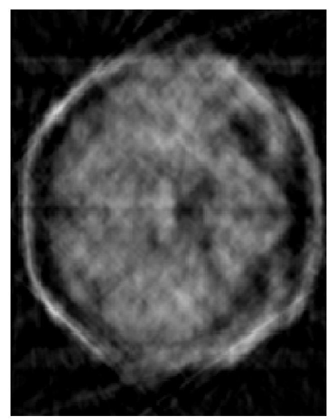
Reconstructed slice 1



Original slice 2



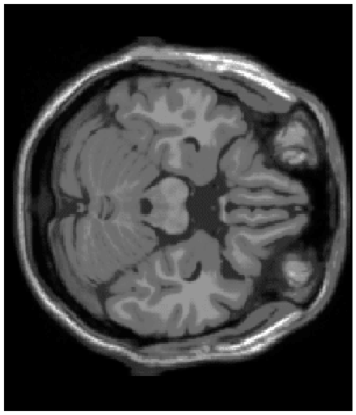
Reconstructed Slice 2



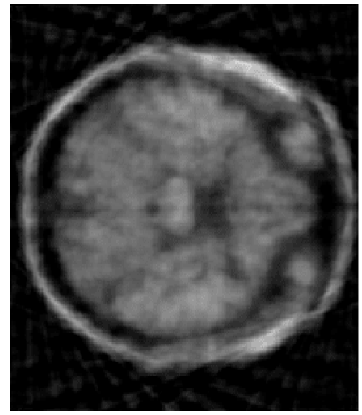
3. Coupled CS reconstruction

   - For 2 slices,

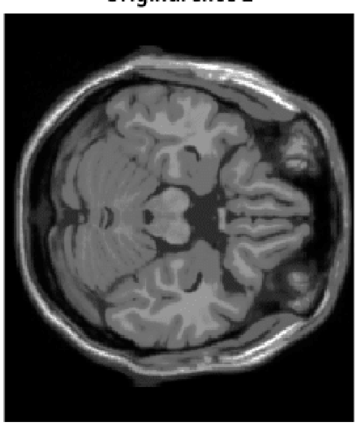
Original slice 1



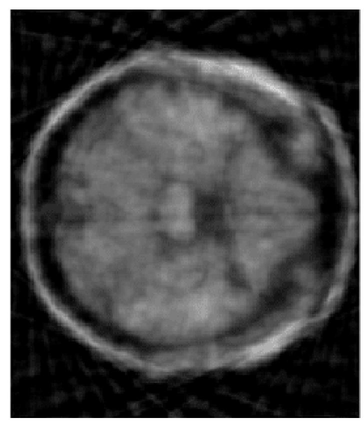
Reconstructed slice 1



Original slice 2



Reconstructed Slice 2



   - For 3 slices, The model considered was as follows,

$$E(\beta) = \|Y - A\beta\| + \lambda \|\beta\|$$

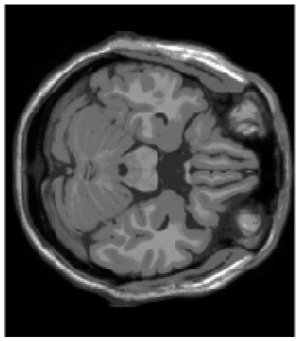
     where,

     – $\beta = \begin{bmatrix} X_1 & X_2 - X_1 & X_3 - X_2 \end{bmatrix}^T$ has dimensions $3 * 217 \times 1$

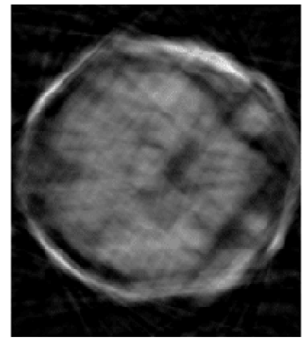
     – $A = \begin{bmatrix} R_1U & 0 & 0 \\ R_2U & R_2U & 0 \\ R_3U & R_3U & R_3U \end{bmatrix}$ where R_i(309*18 $\times$217*217)$representstheradontransformationmatrixalongangles_i$ and U represents the 217×217 2D DCT transformation matrix.

- $Y = \begin{bmatrix} Y_1 & Y_2 & Y_3 \end{bmatrix}$ where $Y_i(307 * 18 \times 1$ for 18 angles)represents the radon projection of the image i along angles$_i$.
- λ is a regularisation parameter.

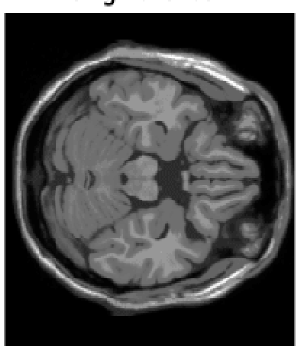
Original slice 1



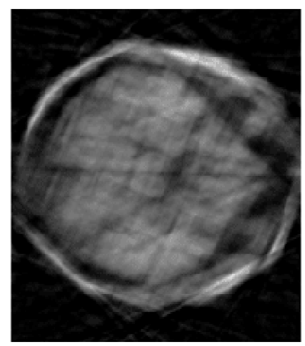
Reconstructed slice 1



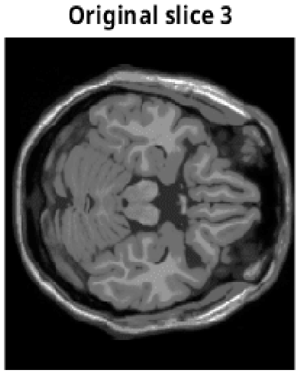
Original slice 2



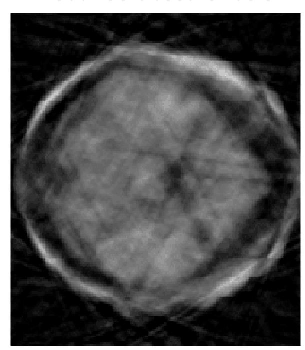
Reconstructed Slice 2



Original slice 3



Reconstructed Slice 3



Question 3.

(a) Translation. $R(g(x-x_o, y-y_o))(\rho,\theta) = R(g(x,y))(\rho - x_o\cos\theta - y_o\sin\theta, \theta)$

We know that,

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

Let $g'(x,y) = g(x-x_o, y-y_o)$. Therefore,

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

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

Substituting $x' = x - x_o$ and $y' = y - y_o$, we have

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

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

$$R(g(x-x_o,y-y_o))(\rho,\theta) = R(g(x,y))(\rho - x_o\cos\theta - y_o\sin\theta, \theta)$$

□

(b) Rotation. $g'(r,\psi) = g(r,\psi-\psi_o)$. Then, $R(g')(\rho,\theta) = R(g)(\rho, \psi_o - \theta)$

We have,

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

Converting to polar coordinates. That is, $dxdy = rdrd\psi$, $x = r\cos\psi$, and $y = r\sin\psi$

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

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

Substituting $\phi = \psi - \psi_o$, we get

$$R(g')(\rho,\theta) = \int_{-\psi_o}^{2\pi-\psi_o}\int_0^{\infty} g(r,\phi)\cdot\delta(r\cos(\phi-(\theta-\psi_o))-\rho)\cdot r\,dr\,d\phi$$

$$= \int_0^{2\pi-\psi_o}\int_0^{\infty} g(r,\phi)\cdot\delta(r\cos(\phi-(\theta-\psi_o))-\rho)\cdot r\,dr\,d\psi + \int_{-\psi_o}^{0}\int_0^{\infty} g(r,\phi)\cdot\delta(r\cos(\phi-(\theta-\psi_o))-\rho)\cdot r\,dr\,d\phi$$

Substitute, $\psi = 2\pi + \phi$ in the 2nd term. Now, $g(r,\psi) = g(r,\psi-2\pi)$ and $\cos(\phi) = \cos(-2\pi+\phi)$. Therefore,

$$= \int_0^{2\pi-\psi_o}\int_0^{\infty} g(r,\phi)\cdot\delta(r\cos(\phi-(\theta-\psi_o))-\rho)\cdot r\,dr\,d\psi + \int_{2\pi-\psi_o}^{2\pi}\int_0^{\infty} g(r,\psi-2\pi)\cdot\delta(r\cos(\psi-2\pi-(\theta-\psi_o))-\rho)\cdot r\,dr\,d\psi$$

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

□

(c) Convolution.
Image $f(x,y)$ and kernel $k(x,y)$. Show $R_\theta(f*k) = R_\theta(f) * R_\theta(k)$. $*$ is the convolution operation.

We know,

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

Now,

$$R_\theta(f*k) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (f*k)(x,y)\cdot\delta(x\cos\theta + y\sin\theta - \rho)\,dx\,dy$$

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

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

$$R_\theta(f*k) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} R_\theta(k)(\rho - u\cos\theta - v\sin\theta)\cdot f(u,v)\,du\,dv$$

We shall use a neat trick here. Let us introduce a new variable η, $f(u\cos\theta - v\sin\theta) = \int_{-\infty}^{-\infty} f(\eta)\cdot\delta(u\cos\theta - v\sin\theta - \eta)\,d\eta$ by the *sifting* property of Dirac-delta function. Now,

$$R_\theta(f*k) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} R_\theta(k)(\rho-\eta)\cdot f(u,v)\cdot\delta(u\cos\theta - v\sin\theta - \eta)\,du\,dv\,d\eta$$

$$R_\theta(f*k) = \int_{-\infty}^{\infty} R_\theta(k)(\rho-\eta)\left(\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdot f(u,v)\cdot\delta(u\cos\theta - v\sin\theta - \eta)\,du\,dv\right)d\eta = \int_{-\infty}^{\infty} R_\theta(k)(\rho-\eta)\cdot R_\theta(f)(\eta)\,d\eta$$

Therefore,

$$R_\theta(f*k) = R_\theta(f) * R_\theta(k)$$

$\square$

Question 4.

Let Φ follow s-order restricted isometry property with s-order restricted isometry constant δ_s and mutual coherence μ. Prove $\delta_s \leq (s-1)\mu$.

Let A be an arbitrary $m \times n$ matrix whose columns are unit-normalised for the rest of the solution.

Lemma. *Gershgorin's disc theorem*
Every eigenvalue of a square matrix B lies in the union of Gershgorin discs $D(B_{ii}, r_i)$ where, $r_i = \sum_{j \neq i} B_{ij}$.

Mutual coherence of a matrix A is defined as

$$\mu(A) = \max_{i,j\ i\neq j} \frac{|A_i^t \cdot A_j|}{\|A_i\|_2 \|A_j\|_2} = \max_{i,j\ i\neq j} |A_i^t \cdot A_j|$$

If A follows RIP with s-order restricted isometry constant δ_s and S is the set of s-sparse vectors, then

$$\max_{x \in S} (1-\delta_s) \leq \frac{\|Ax\|_2^2}{\|x\|_2^2} \leq (1+\delta_s)$$

From this we also have,

$$\delta_s = \max(1-\lambda_{min}, \lambda_{max}-1)$$

where,

$$\lambda_{max} = \max_{\theta_\Gamma \in \mathbb{R}^s, |\Gamma| \leq s} \frac{\|A_\Gamma \theta_\Gamma\|^2}{\|\theta_\Gamma\|^2} \text{ and } \lambda_{min} = \min_{\theta_\Gamma \in \mathbb{R}^s, |\Gamma| \leq s} \frac{\|A_\Gamma \theta_\Gamma\|^2}{\|\theta_\Gamma\|^2}$$

We also know that λ_{max} and λ_{min} represent the maximum and minimum eigenvalues of $(A_\Gamma)^T(A_\Gamma)$. In general, the maximum eigenvalue of a symmetric matrix B is given by $max_x \frac{\|Bx\|_2^2}{\|x\|_2^2}$ (mentioned in Lecture notes on the page number 101, in CS-Theory.pdf)

We shall use the above equations to prove the result. Firstly, A^TA is a symmetric matrix whose maximum eigenvalue is given by $max_x \frac{\|A^TAx\|_2^2}{\|x\|_2^2}$. Let x be any s-sparse vector whose support is Γ. Let A_Γ be the matrix made from subset of columns of A whose indices belong to Γ. From Gershgorin's disc theorem,

$$\left((A_\Gamma^T A_\Gamma)_{ii} - \sum_{i \neq j} (A_\Gamma^T A_\Gamma)_{ij}\right) \leq \frac{\|A_\Gamma^T A_\Gamma x_\Gamma\|_2^2}{\|x\|_2^2} \leq \left((A_\Gamma^T A_\Gamma)_{ii} + \sum_{i \neq j} (A_\Gamma^T A_\Gamma)_{ij}\right)$$

Also, $(A_\Gamma^T A_\Gamma)_{ij} = \sum_{k=1}^m (A_\Gamma)_{ki}(A_\Gamma)_{kj}$. Therefore,

$$(A_\Gamma^T A_\Gamma)_{ii} = \sum_{k=1}^m (A_\Gamma)_{ki}^2 = 1$$

. This is because, the columns of $A \implies A_\Gamma$ are unit-normalised. Also, $\mu = \max_{i,j\ i\neq j} |A_i^t \cdot A_j|$, and the number of columns in A_Γ is s. Therefore,

$$\sum_{i \neq j} (A_\Gamma^T A_\Gamma)_{ij} = \sum_{i \neq j} \left(\sum_{k=1}^m (A_\Gamma)_{ki}(A_\Gamma)_{kj}\right) = \sum_{i \neq j} |(A_\Gamma)_i \cdot (A_\Gamma)_j| \leq \mu \cdot (s-1)$$

Hence, we can modify the above equation to

$$(1 - \mu \cdot (s-1)) \leq \frac{\|A_\Gamma^T A_\Gamma x_\Gamma\|_2^2}{\|x\|_2^2} \leq (1 + \mu \cdot (s-1))$$

As x is an arbitrary s-sparse vector, we have

$$\max_y \frac{\left\|A_\Gamma^T A_\Gamma y\right\|_2^2}{\|y\|_2^2} = \lambda_{max} \leq (1 + \mu \cdot (s-1)) \implies \lambda_{max} - 1 \leq \mu \cdot (s-1)$$

Similarly,

$$\min_y \frac{\left\|A_\Gamma^T A_\Gamma y\right\|_2^2}{\|y\|_2^2} = \lambda_{min} \geq (1 - \mu \cdot (s-1)) \implies 1 - \lambda_{min} \leq \mu \cdot (s-1)$$

Therefore, we have

$$\delta_s = \max\left(1 - \lambda_{min}, \lambda_{max} - 1\right) \leq \mu(s-1) \implies \delta_s \leq \mu(s-1)$$

□

Question 5.

The following Paper comes under Astrophysics application for Tomography.

Title : TARDIS. I. A Constrained Reconstruction Approach to Modeling the z 2.5 Cosmic Web Probed by Ly Forest Tomography
Published at : The American Astronomical Society.
Published on: 2019 December 11
Reference Link: Here
Author: Benjamin Horowitz, Khee Gan Lee, Martin White, Alex Krolewski, and Metin Ata

Problem Statement:

Aim: In this paper, we apply initial density reconstruction to mock observations of IGM tomography using the Tomographic Absorption Reconstruction and Density Inference Scheme (TARDIS)
The current standard procedure for IGM tomography analysis is to create a Wiener-filtered absorption map from the observed Ly absorption features. In this work, they implement a different approach, finding the maximum a posteriori initial density field that gives rise to the observed density field, often known as a "constrained realization.This information can also help inform the astrophysical processes occurring in the region; for example, combining the flux information, matter velocity information, and a galaxy catalog will provide insights into galaxy formation environmental dependence.

Mathematical problem:

The equation for which we are trying to find solution is

$$d = R(s) + n$$

- The data in a vector is given by ' **d** ' of total dimension (N × L) x 1.
- We measure N skewers of flux, assuming perfect identification of the continuum spectra each of length L M is a certain known resolution value.
- Here $\boldsymbol{R} : M^3 \rightarrow N \times L$ is the (nonlinear) response operator. (or "R" is sensing matrix with dimension (N x L) x M^3)
- ' **s** ' is the estimated signal term with dimensions M^3 x 1.
- ' **n** ' is the noise term in the model dimensions (N x L) x 1.

Optimization:

For Optimizing we use the maximum likelihood estimator from a prior distribution give by:

$$L(s|d) = (2\pi)^{-(N+M)/2} det(SN_1)^{-1/2} exp[\frac{-1}{2} s^T S^{-1} s + (d - R(s))^T N_1^{-1} (d - R(s))]$$

- The Gaussian information is contained in covariance matrices, $\boldsymbol{S} = \langle \boldsymbol{s}\boldsymbol{s}^\dagger \rangle$ and $\boldsymbol{N_1} = \langle \boldsymbol{n}\boldsymbol{n}^\dagger \rangle$, for the estimated signal and noise components.
- Dimension for $S = M^3$ x M^3, and dimension for N_1= (N x L) x (N x L)
- L(s—d) is the maximum likelihood estimator from prior data.

This above equation reduces to Minimization of exponent term in the previous equation, which is given by χ^2 wrt to parameter ' s ' .

$$\chi^2 = \frac{-1}{2} s^T S^{-1} s + (d - R(s))^T N^{-1} (d - R(s))$$