Advanced Image Processing HW3

Anilesh Bansal (22b0928) Parth Pujari (210100106)

1 Question 1

Theorem 11.1b) Suppose that the model matrix X satisfies the restricted eigen-value property with parameter $\gamma > 0$ over $C(S;3)$ then given a regularization parameter $\lambda_N \geq 2||X^T w||_\infty / N > 0$, any estimate $\hat{\beta}$ from the regularized lasso satisfies the bound

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

a) A $m \times n$ matrix X is said to satisfy the restricted eigenvalue property with respect to some subset $C \subseteq \mathbb{R}^n$ if

$$\frac{1}{m}\nu^T X^T X \nu \geq \gamma ||\nu||_2^2 \text{ for all } \nu \in C$$

The idea behind this is to ensure that if our $\hat{\nu} = \hat{\beta} - \beta^*$, then the error function is not flat around $\beta*$, especially in the directions of the support of $\beta*$. Thus we can choose C accordingly and ensure that the function has a minimum convexity in the directions that we want based on whether or not the matrix X satisfies the restricted eigenvalue property or not for that C.

b) Since our estimate $\hat{\beta}$ is the solution that minimizes the Lasso-Error.

Loss$(\hat{\beta}) \leq$ Loss $(\beta^*) \implies G(\hat{\nu}) \leq G(0)$

c)

$$\begin{aligned}
\frac{1}{2N}||y - X(\beta^* + \hat{\nu})||_2^2 + \lambda_N ||\beta^* + \hat{\nu}||_1 &\leq \frac{1}{2N}||y - X\beta^*||_2^2 + \lambda_N ||\beta^*||_1 \\
\frac{1}{2N}||w - X\hat{\nu}||_2^2 &\leq \frac{1}{2N}||w||_2^2 + \lambda_N(||\beta^*||_1 - ||\beta^* + \hat{\nu}||_1) \\
\frac{1}{2N}||w||_2^2 + \frac{1}{2N}||X\hat{\nu}||_2^2 - \frac{1}{2N}2w^T X\hat{\nu} &\leq \frac{1}{2N}||w||_2^2 + \lambda_N(||\beta^*||_1 - ||\beta^* + \hat{\nu}||_1) \\
\frac{1}{2N}||X\hat{\nu}||_2^2 &\leq \frac{1}{N}w^T X\hat{\nu} + \lambda_N(||\beta^*||_1 - ||\beta^* + \hat{\nu}||_1)
\end{aligned}$$

d)

$$\begin{aligned}
||\beta^* + \hat{\nu}||_1 = ||\beta_S^* + \hat{\nu}_S||_1 + ||\hat{\nu}_{S^c}||_1 &\geq ||\beta_S^*||_1 - ||hat\nu_S||_1 + ||\hat{\nu}_{S^c}||_1 \\
\implies ||\hat{\nu}_S||_1 - ||\hat{\nu}_{S^c}||_1 &\geq ||\beta^*||_1 - ||\beta^* + \hat{\nu}||_1
\end{aligned}$$

Further, Holder's Inequality states that for $1 \leq p, q \leq \infty$ if $1/p + 1/q = 1$ then

$$\sum_{k=1}^{n} a_k b_k \leq \left(\sum_{k=1}^{n} a_k^p\right)^{\frac{1}{p}} \left(\sum_{k=1}^{n} b_k^q\right)^{\frac{1}{q}}$$

Using the Inequality for $a = X^T w, b = hat\nu, p = \infty$ and $q = 1$ we get that

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

Substituting in the above equations we get that

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

e) Note that for any d-dimensional vector x, By Cauchy-Schwarz Inequality -

$$||x||_1 = \sum_1^d 1.|x_i| \leq \sqrt{\sum_1^d 1}\sqrt{\sum_1^d x_i^2} \implies ||x||_1 \leq \sqrt{d}||x||_2$$

Since $\frac{||X^T w||_\infty}{N} \leq \frac{\lambda_N}{2}$ is necessary as per the statement of the theorem

$$\begin{aligned}
\frac{1}{2N}||X\hat{\nu}||_2^2 &\leq \frac{||X^T w||_\infty}{N}||\hat{\nu}||_1 + \lambda_N(||\hat{\nu}_S||_1 - ||\hat{\nu}_{S^c}||_1) \\
&= \frac{\lambda_N}{2}(||\hat{\nu}_S||_1 + ||\hat{\nu}_{S^c}||_1) + \lambda_N(||\hat{\nu}_S||_1 - ||\hat{\nu}_{S^c}||_1) \\
&= \frac{3\lambda_N}{2}||\hat{\nu}_S||_1 - \frac{\lambda_N}{2}||\hat{\nu}_{S^c}||_1 \\
&\leq \frac{3\lambda_N}{2}||\hat{\nu}_S||_1 \\
&\leq \frac{3\lambda_N}{2}||\hat{\nu}||_1 \leq \frac{3\lambda_N}{2}\sqrt{k}||\hat{\nu}||_2
\end{aligned}$$

f) According to Lemma 11.1, given that $\lambda_N \geq 2||X^T w||_\infty/N > 0$, the error $\hat{\nu} = \hat{\beta} - \beta^*$ associated with any lasso solution $\hat{\beta}$ belongs to the cone set $C(S;3)$

Thus we can apply the restricted eigen-value condition to $\hat{\nu}$ to get $\frac{1}{N}||X\hat{\nu}||_2^2 \geq \gamma||\hat{\nu}||_2^2$

$$\begin{aligned}
&\implies \frac{1}{2}\gamma||\hat{\nu}||_2^2 \leq \frac{3\lambda_N}{2}\sqrt{k}||\hat{\nu}||_2 \\
&\implies ||\hat{\nu}||_2 \leq \frac{3}{\gamma}\sqrt{k}\lambda_N \\
&\implies ||\hat{\beta} - \beta^*||_2 \leq \frac{3}{\gamma}\sqrt{\frac{k}{N}}\sqrt{N}\lambda_N
\end{aligned}$$

g) It shows up in the (e) part of the derivation wherein we bounded $\frac{||X^T w||_\infty}{N}$ with $\frac{\lambda_N}{2}$

It also shows up in the proof for the Lemma that the lasso error $\hat{\nu}$ satisfies the cone constraint. This is because we got the equation

$$\begin{aligned}
\frac{1}{2N}||X\hat{\nu}||_2^2 &\leq \frac{3\lambda_N}{2}||\hat{\nu}_S||_1 - \frac{\lambda_N}{2}||\hat{\nu}_{S^c}||_1 \\
\implies 0 &\leq \frac{3\lambda_N}{2}||\hat{\nu}_S||_1 - \frac{\lambda_N}{2}||\hat{\nu}_{S^c}||_1 \\
\implies ||\hat{\nu}_{S^c}||_1 &\leq 3||\hat{\nu}_S||_1
\end{aligned}$$

which means that $\hat{\nu}$ satisfies the cone-constraint and lies in $C(S;3)$. This is necessary to complete the proof in the next steps.

h) Since the restricted eigenvalue property cannot be satisfied for all $\nu \in \mathbb{R}^p$ whenever $N < p$, we need a set of vectors $C \subset \mathbb{R}^p$ such that even if the matrix X satisfies REP for all vectors in set C, then that gives a good bound error.

Since the error cannot be strongly convex in all directions, we require it to be strongly convex in atleast the directions along support of β^*.

Since the way lasso is defined always makes it satisfy the cone constraint that

$$||\hat{\nu}_{S^c}||_1 \leq \alpha||\hat{\nu}_S||_1 \text{ for some } \alpha > 0$$

Choosing the set C as the vectors that satisfy this cone-constraint is a good choice.

i) Both of these theorem's are quite similar in the sense that both of them require some conditions on the eigenvalues of the matrix X. Theorem 3 requires the RIC or order 2s $\delta_{2s} < \sqrt{2} - 1$ which basically relates to eigen-values as well since $\delta_S = max(\lambda_{max} - 1, 1 - \lambda_{min})$. This theorem requires minimum eigen value to be $\geq \gamma$. However there are a couple of differences as well.

   The given theorem is better over Theorem 3 because the squared L2-error is proportional to $\frac{k\sigma^2}{N}$ which is the best-possible one can hope to achieve even if one knew the support vector beforehand. Thus other than the log(p) factor, error bound of the given theorem is best that one can hope to achieve. Also it doesn't require one to do an explicit calculation for the RIC of the matrix X, which we know is computationally hard.

   However, theorem 3 is better for those signals which aren't exactly s-sparse, but have values close to 0. The given theorem fails to work well in this case since it requires the values to be exactly 0 and close to 0 doesn't work. However in Theorem 3, since the error is given by

$$||\hat{\beta} - \beta^*||_2 \leq \frac{C_0}{\sqrt{S}}||\beta^* - \beta_S||_1 + C_1\epsilon$$

   , even compressible signals give a really low error bound in Theorem 3.

j) Both the Dantzig-Selector and LASSO have similar error bounds in the sense that they are bounded by a constant times $k\sigma^2 logn$ with high probability. This value doesn't depend on m and since we often require $m > klogn$ it leads to better error bounds than just the expected noise.

   Further this error bound is near-optimal as well. Even if somehow we had an oracle and knew exactly the locations of the k nonzero components, the error would be of the order of $k\sigma^2$.

   In fact, this rate including the logarithmic factor—is known to be minimax optimal, meaning that it cannot be substantially improved upon by any estimator.

   This is because of the similarity in the optimization problem of Dantzig Selector and LASSO, wherein both of them account for the interaction of noise with matrix X. Dantzig Selector requires maximum correlation between X and noise e to be less than some constant, while LASSO assumes noise as iid from some prior distribution and maximises posterior using prior and likelihood, which leads to the regularization term.

2 Question 2

Title of Paper: "Three-dimensional seismic velocity structure of Long Valley Caldera from local earthquake tomography."
Venue: Journal of Geophysical Research: Solid Earth
Year of Publication: 1996
The paper addresses the reconstruction of three-dimensional (3D) seismic velocity structure of the Long Valley Caldera region using seismic tomography techniques. Seismic tomography involves the imaging of the Earth's interior by analyzing the travel times of seismic waves recorded at different seismograph stations. In this context, the mathematical problem involves determining variations in seismic velocity within the Earth's crust beneath the Long Valley Caldera region based on seismic arrival times recorded at various seismic stations.
The technique developed to solve the problem is an extension of the method of Aki and Lee (1976), for determination of three-dimensional velocity structure and hypocenter locations from local earthquakes as refined by Thurber (1981) who added the parameter separation algorithm of Pavlis and Booker (1980) and the ray tracing method of Thurber and Ellsworth (1980). The extension used involves the construction of an approximate inverse to a very large system of linear equations. This approximation is similar to the algebraic reconstruction in tomography.
The paper employs iterative inversion techniques to solve the tomographic problem and reconstruct the 3D seismic velocity structure of the Long Valley Caldera region. These techniques involve iteratively adjusting model

parameters (such as seismic velocities) to minimize the difference between observed and predicted seismic arrival times.

3 Question 3

$$\begin{aligned}
R_\theta g(\rho) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)\delta(xcos\theta + ysin\theta - \rho)dxdy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x-x_0, y-y_0)\delta(xcos\theta + ysin\theta - \rho)dxdy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x', y')\delta((x'+x_0)cos\theta + (y'+y_0)sin\theta - \rho)dxdy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x', y')\delta(x'cos\theta + y'sin\theta - \rho + x_0 sin\theta + y_0 cos\theta)dxdy \\
&= R_\theta f(\rho - (x_0, y_0).(cos\theta, sin\theta))
\end{aligned}$$

Where we substituted $x - x_0 = x'$ and $y - y_0 = y'$ respectively

4 Question 4

a)

$$\begin{aligned}
\int g(\rho,\theta)z(\rho)d\rho &= \int\int\int (f(x,y)\delta(xcos\theta + ysin\theta - \rho)dxdy)z(\rho)d\rho \\
&= \int\int\int f(x,y)\delta(xcos\theta + ysin\theta - \rho)dxdy)z(\rho)d\rho \\
&= \int\int f(x,y)\left[\int \delta(xcos\theta + ysin\theta - \rho)z(\rho)d\rho\right]dxdy \\
&= \int\int f(x,y)\left[z(xcos\theta + ysin\theta)\right]dxdy
\end{aligned}$$

b) For getting the Fourier slice theorem, we put $z(\rho) = e^{-j2\pi\mu\rho}$

Thus the Fourier Transform of the Radon Transform w.r.t ρ becomes -

$$\begin{aligned}
\int g(\rho,\theta)e^{-j2\pi\mu\rho}d\rho &= \int\int f(x,y)\left[e^{-j2\pi\mu(xcos\theta + ysin\theta)}\right]dxdy \\
&= \int\int f(x,y)\left[e^{-j2\pi(x\mu cos\theta + y\mu sin\theta)}\right]dxdy \\
&= F(\mu cos\theta, \mu sin\theta)
\end{aligned}$$

i.e. The Fourier Transform of $g(\rho,\theta)$ is the same as a slice of the Fourier Transform $F(u,v)$ of $f(x,y)$ along an axis passing through origin and parallel to θ

5 Question 5

The values of **A** and **b** are as follows:

A is a $n \times nT$ matrix where n is the number of pixels in each frame of the video and T is the number of frames.

It is a product of two matrices Φ and Ψ where Φ is an $n \times nT$ size matrix of the form : [diag(S1)|diag(S2)|...diag(S_T)] where Si is a diagonal matrix of size $n \times n$, with each diagonal element corresponding to a value in the binary pattern C_i.
Ψ is the 3D DCT matrix of size $nT \times nT$

b is the vectorized form of the coded snapshot. It is of size $n \times 1$.

While working with patches, we use the binary pattern of the corresponding 8×8 patches to form the matrix Φ which is now a diagonal matrix of size $64 \times T$. Similarly, **b** is the vectorized form of the 8×8 patch.

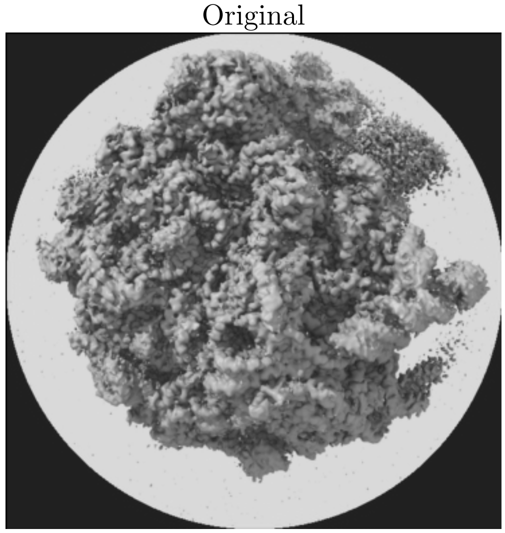
For the cars video:
Errors
T = 3 : 1.21
T = 5 : 1.6
T = 7 : 2.01

For flame:
Error = 1.54

The generated videos are in the submission folder

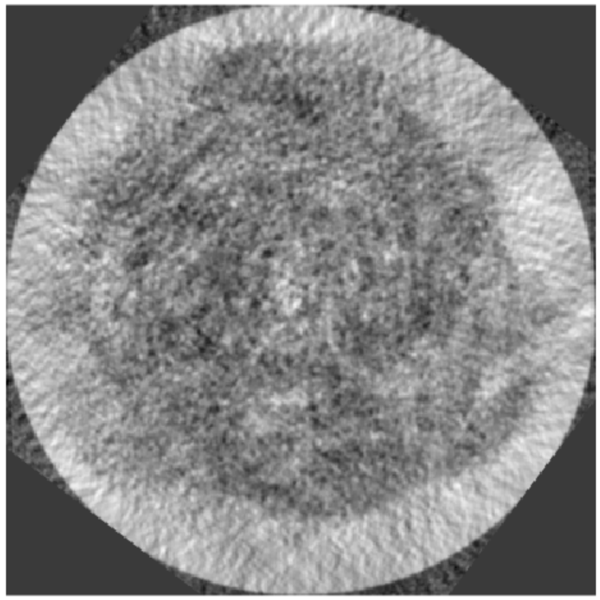
6 Question 6

Original

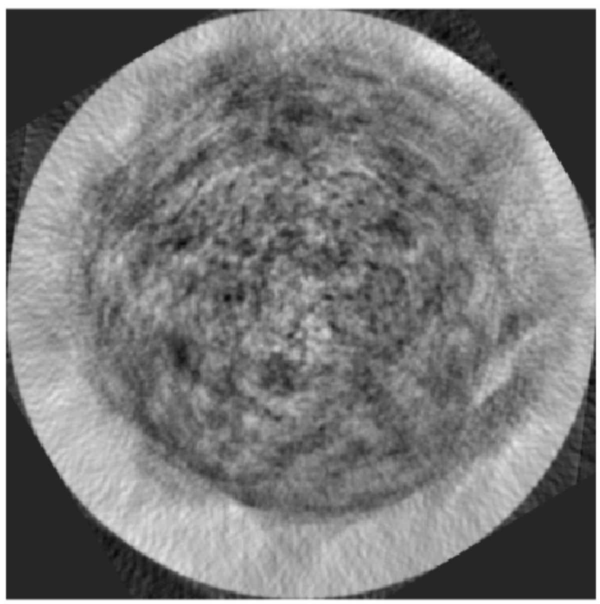


Reconstructions

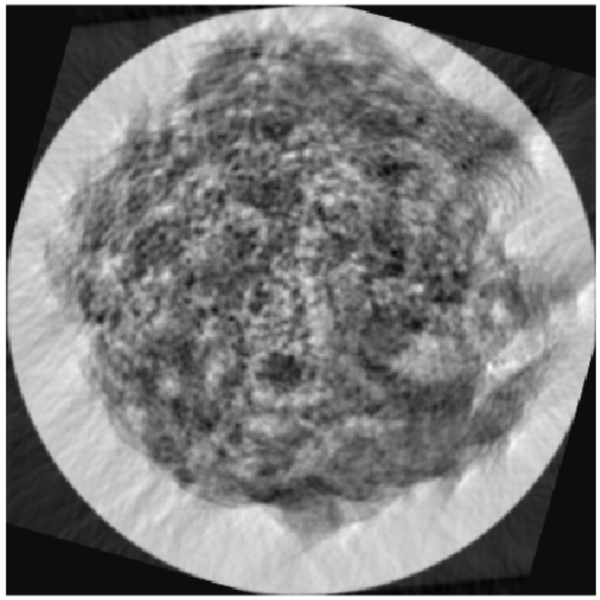
Reconstructed with N = 50 and RMSE = 0.31778



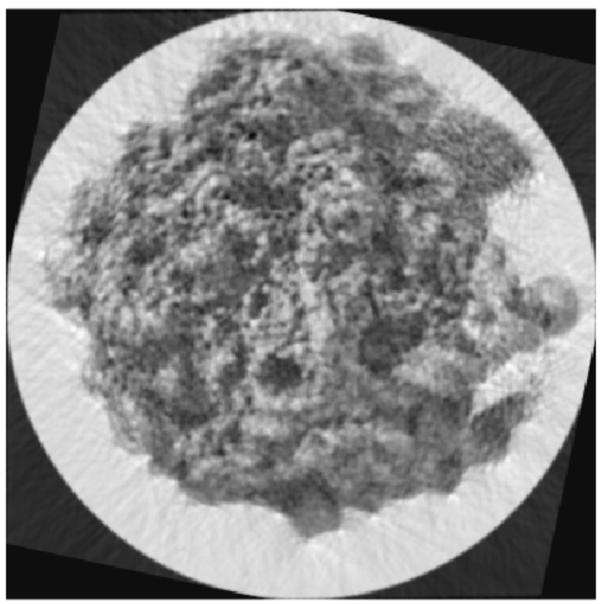
Reconstructed with N = 100 and RMSE = 0.2816



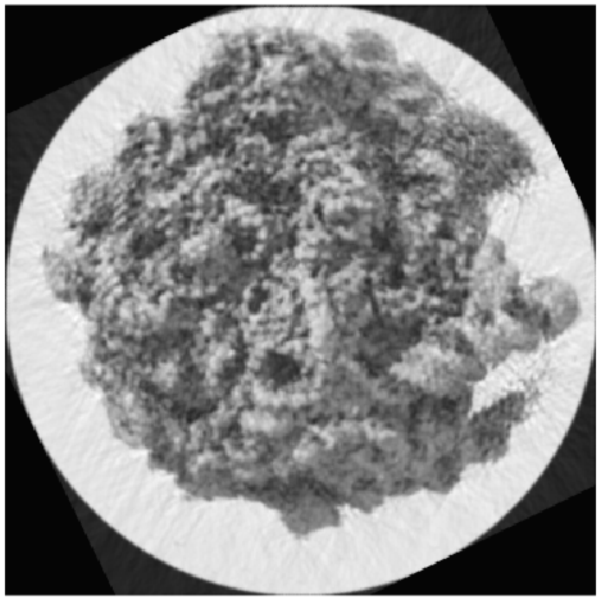
Reconstructed with N = 500 and RMSE = 0.20901



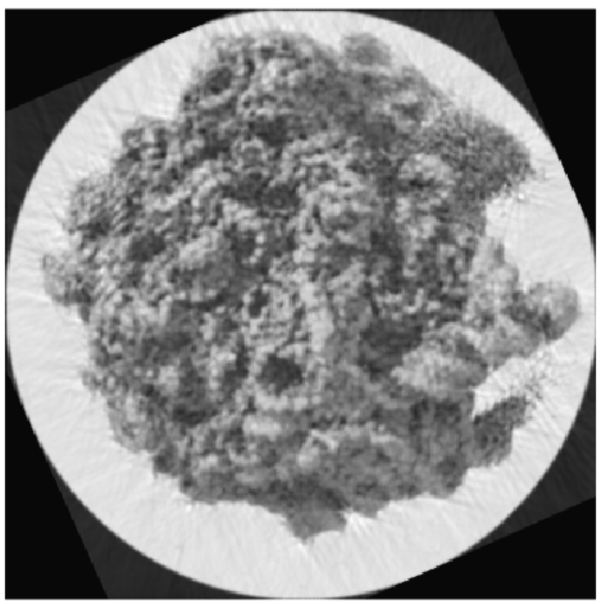
Reconstructed with N = 1000 and RMSE = 0.14549



Reconstructed with N = 2000 and RMSE = 0.12911



Reconstructed with N = 5000 and RMSE = 0.12814



Reconstructed with N = 10000 and RMSE = 0.079587