CS754: Advanced Image Processing
Assignment 3

Karan Taneja, 15D070022
Sucheta Ravikanti, 160040100

Question 4

(a) The restricted eigenvalue condition is a condition on the eigenvalues of the matrix $X^T X$ which ensures that the loss function $\mathbf{f}$ satisfies restricted strong convexity at β^* and states the following

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

for all $\nu \in C$ non-zero.

(b) The function $G(\nu)$ is defined as follows

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

Noting that $\widehat{\nu}$ minimizes $G(\nu)$ by construction, we have $G(\widehat{\nu}) \leq G(0)$

(c) From the previous part, we get

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

Also, we know the following.

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

Substituting the above relation in the result obtained in the expression for $G(\nu)$

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

$$G(0) \;=\; \frac{1}{2N}||w||_2^2 \;+\; \lambda_N ||\beta^*||_1$$

From the above result and 4(b)

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

Expanding the LHS we get,

$$\frac{||X\nu||_2^2 \ - \ 2\langle w, X\widehat{\nu}\rangle}{2N} \ \leq \ \lambda_N \left(||\beta^*||_1 \ - \ ||\beta^* \ + \ \nu||_1\right)$$

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

(d) $S = S(\beta^*)$ represents the set of indices which have non-zero support on β^*. $\beta^*_{S^c} = 0$. Therefore, we have $||\beta^*||_1 = ||\beta^*_S||_1$. From the previous results and reverse triangle inequality in the second inequality,

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

$$||\beta^*||_1 \ - \ ||\beta^* \ + \ \nu||_1 \ \leq \ ||\widehat{v}_S||_1 \ - \ ||\widehat{v}_{S^c}||_1$$

Substituting the above result in 4(c)

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

Holder's Inequality: Let (S, $\sum$, μ) be a measure space and let $p, q \in [1, \infty]$ with $\frac{1}{p} + \frac{1}{q} = 1$. Then, for all measurable real-or complex-valued functions f and g on S, we have (quoted from Wikipedia)

$$||fg||_1 \ \leq \ ||f||_p \ ||g||_q$$

Using the **Holder's Inequality** by putting $p = \infty, q = 1$, putting $f = w^T X, g = \widehat{\nu}$ and considering the fact that $w^T X\widehat{\nu}$ is a scalar, $w^T X\widehat{\nu} = |w^T X\widehat{\nu}|$

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

We know that since the infinite norm is merely the maximum absolute element in a vector, the transpose also would have the same infinite norm.

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

(e) From **Example 11.1**, the following is true with a high probability

$$\frac{||X^T w||_\infty}{N} \ \leq \ \frac{\lambda_N}{2}$$

$$||\widehat{\nu}||_1 \ = \ ||\widehat{\nu}_S||_1 \ + \ ||\widehat{\nu}_{S^c}||_1$$

From the 4(d) and above results, we get

$$\frac{||X\nu||_2^2}{2N} \ \leq \ \frac{\lambda_N}{2} \left(||\widehat{\nu}_S||_1 \ + \ ||\widehat{\nu}_{S^c}||_1\right) \ + \ \lambda_N \left(||\widehat{v}_S||_1 \ - \ ||\widehat{v}_{S^c}||_1\right)$$

k is the sparsity level (number of non-zero elements and also the cardinality of the set S) of the vector β^*

$$||\widehat{\nu}_S||_1 \ = \ (\widehat{\nu}_{S_1} + \widehat{\nu}_{S_2} + ...\widehat{\nu}_{S_k}) \ = \ \left(\widehat{\nu}_{S_1}.1 + ... + \widehat{\nu}_{S_k}.1 + \widehat{\nu}_{{S^c}_1}.0 + ... + \widehat{\nu}_{{S^c}_{(n-k)}}.0\right)$$

Applying **Cauchy Schwarz Inequality**, we get

$$\leq \ \left(\widehat{\nu}^2_{S_1} + ... + \widehat{\nu}^2_{S_k} + \widehat{\nu}^2_{{S^c}_1} + ... + \widehat{\nu}^2_{{S^c}_{(n-k)}}\right)^{\frac{1}{2}} \ \left(1^2 + ... + 1^2 + 0^2 + ... + 0^2\right)$$

$$= \ ||\widehat{\nu}||_2 \sqrt{k}$$

$$||\widehat{\nu}_S||_1 \ \leq \ \sqrt{k} \ ||\widehat{\nu}||_2$$

Using the above relation, we get

$$\frac{||X\nu||_2^2}{2N} \ \leq \ \frac{\lambda_N}{2}\left(\sqrt{k} \ ||\widehat{\nu}||_2 \ + \ ||\widehat{\nu}_{S^c}||_1\right) \ + \ \lambda_N\left(\sqrt{k} \ ||\widehat{\nu}||_2 \ - \ ||\widehat{v}_{S^c}||_1\right)$$

$$= \ \frac{3\lambda_N}{2}\left(\sqrt{k} \ ||\widehat{\nu}||_2\right) \ - \ \frac{\lambda_N}{2}\left(||\widehat{v}_{S^c}||_1\right)$$

$$\leq \ \frac{3\lambda_N}{2}\left(\sqrt{k} \ ||\widehat{\nu}||_2\right)$$

Therefore, putting everything together we get the following

$$\frac{||X\nu||_2^2}{2N} \ \leq \ \frac{3\lambda_N}{2}\left(\sqrt{k} \ ||\widehat{\nu}||_2\right)$$

(f) From the previous exercise, we get

$$\frac{||X\nu||_2^2}{2N} \ \leq \ \frac{3\lambda_N}{2}\left(\sqrt{k} \ ||\widehat{\nu}||_2\right)$$

Lemma 11.1 states that if $\lambda_N \geq 2||\frac{X^T w}{N}||_\infty$, the error vector $\widehat{\nu}$ associated with the lasso solution $\widehat{\beta}$ belongs to the cone set $C(S;3)$. Now we can apply the restricted eigenvalue condition to $\widehat{\nu}$ belonging to the cone set as defined above. We will therefore get the following for $\widehat{\nu}$ belonging to the cone set.

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

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

Combining this and the earlier exercise, we get

$$\frac{\gamma}{2} \ ||\widehat{\nu}||_2^2 \ \leq \ \frac{3\lambda_N}{2}\left(\sqrt{k} \ ||\widehat{\nu}||_2\right)$$

Rearranging and substituting, we get

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

(g) The inequality $\lambda_N \geq 2||\frac{X^T w}{N}||_\infty$ shows up in finding the cone set $C(S;\alpha)$ for $\widehat{\nu}$ in the regularized lasso setting we are considering. From 4(d) part, we get

$$0 \;\leq\; \frac{||X\nu||_2^2}{2N} \;\leq\; \frac{||w^T X||_\infty \; ||\widehat{\nu}||_1}{N} \;+\; \lambda_N \left(||\widehat{v}_S||_1 \;-\; ||\widehat{v}_{S^c}||_1\right)$$

$$0 \;\leq\; \frac{\lambda_N}{2}||\widehat{\nu}||_1 \;+\; \lambda_N \left(||\widehat{v}_S||_1 \;-\; ||\widehat{v}_{S^c}||_1\right)$$

$$0 \;\leq\; \frac{3\lambda_N}{2}||\widehat{v}_S||_1 - \frac{\lambda_N}{2}||\widehat{\nu}_{S^c}||_1$$

$$||\widehat{\nu}_{S^c}||_1 \;\leq\; 3||\widehat{\nu}_S||_1$$

This is the cone constraint $C(S;3)$

(h) The condition of strong convexity cannot be achieved on the whole space because of the rank deficiency of the matrix $X^T X$, hence we relax and impose the strong convexity on a certain subset $C \in R^p$ for possible vectors $\widehat{\nu}$. The cost function in $y = X\beta^* + w$ setting satisfies the restricted strong convexity with respect to the set C if there exists a $\gamma > 0$ such that

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

Putting a constraint on λ_N gives a condition or constraint (cone constraint) the lasso error will satisfy, and this is exactly what's relevant to our setting and problem.

(i) **Theorem 3:** Suppose the matrix $A = \Phi\Psi$ of size m by n (where sensing matrix Φ has size m by n, and basis matrix Y has size n by n) has RIP property of order 2S where $\delta_{2S} < 0.41$. Let the solution of the following be denoted as q^*, (for signal $f = \Psi\theta$, measurement vector $y = \Phi\Psi\theta$):

$$\min ||\theta||_1; \;\; ||y - \Phi\Psi\theta||_2^2 \;\leq\; \epsilon$$

Then we have:

$$||\theta^* - \theta||_2 \;\leq\; \frac{C_0}{\sqrt{S}}||\theta - \theta_S||_1 \;+\; C_1\epsilon$$

Example 11.1: Classic linear Gaussian model

$$||\widehat{\theta} - \theta^*||_2 \;\leq\; \frac{c\sigma}{\gamma}\sqrt{\frac{\tau k \log n}{m}}$$

with probability at least $1 - 2e^{-\frac{1}{2}(\tau-2)\log p}$ and $w \in R^m$ is Gaussian with iid $N(0, \sigma^2)$ entries.

Advantages of 11.1 over Theorem 3:

- 11.1 does not involve finding the RIP of the measurement matrix. Getting RIP is a costly operation and theorem 11.1 has an edge in this sense.
- Fixing ϵ is not a trivial task, this will be avoided by using 11.1
- If the support set of β is known, we need to technically find only those k free parameters. Inspite of knowing the support set, no model can achieve the bound that decays by $\frac{k}{m}$ in 2-squared error. The rate, along with the logarithmic factor is also known to be **minimax optimal**, it means that no estimator can substantially improve upon this. So, this bound gives good guarantees while being agnostic to the support set and the RIP of the measurement matrix.

Advantages of Theorem 3 over the bounds of 11.1(b):

- The regularization parameter λ_N has to be set everytime for a different setting unlike Theorem 3, where the theorem is more general.
- The sparsity level has to be known in the case of theorem 11.1(b), whereas the theorem 3 makes no such assumptions on the sparsity of the signal. Any sparse signal would follow this bound.
- This theorem handles the compressible nature of natural signals. A lot natural images are not exacty sparse in their orthonormal basis, but there is a decay in the coefficient values observed. Such signals aren't handled any differently by the theorem 11.1(b)

(j) **Dantiz Estimator:**

$$||A^T e||_\infty \;\leq\; \lambda$$

$$||\widehat{x} - x||_2 \;\leq\; \frac{C_0 \sigma_k(x)_1}{\sqrt{k}} + C_3\sqrt{k}\lambda$$

Where C_0, C_3 are the increasing constants of δ_{2S}

LASSO:

$$||X^T w||_\infty \;\leq\; \frac{N\lambda_N}{2}$$

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

where N is the number of measurements, γ is the resitricted eigenvalue constant, p is the dimension of signal in both the theorems.
The λ in Dantiz estimator bound acts as λ_N in the LASSO case. One common thing is, both the lambdas give an upper bound on the largest value of the $||X^T w||$, the interaction of the measurement matrix with the noise vector, subject to some constants. The bounds are similar with respect to the error term ($= \sqrt{k}\lambda$).

(k) The paper **"Square-root Lasso: pivotal recovery of sparse signals**

via conic programming" proposes a modification to the lasso, called the square-root lasso which has several advantages over the standard lasso in certain problem environments. The setting we consider here is that of a high dimensional sparse linear regression model, where the number of regressors, **p** are larger than **n** with only **s** regressors being significant.

LASSO Estimator: Using the right penalty level and when the errors are **gaussian with mean 0, variance 1**, we get

$$||\widehat{\beta} - \beta^*||_2 \;\leq\; \sigma \left(s \frac{\log(\frac{2p}{\alpha})}{n} \right)^{\frac{1}{2}}$$

The attractive features of this are:

1. Lasso Estimator is a convex function so computation wise, it has an edge.
2. The number of regressors scale logarithmically.
3. Having knowing the identities T of significant regressors, rates as good as $\sigma \left(\frac{s}{n}\right)^{\frac{1}{2}}$ canbe achieved.

Square-Root LASSO Estimator: The cost function used is the following:

$$||y - X\beta||_2 + \frac{\lambda}{n}||\beta||_1$$

Using the right penalty level and by using moderate deviation theory (that helps us in dispensing any assumptions made on noise), we get

$$||\widehat{\beta} - \beta^*||_2 \;\leq\; \sigma \left(s \frac{\log(\frac{2p}{\alpha})}{n} \right)^{\frac{1}{2}}$$

The advantages this method has over the previous method are:

1. Even this achieves the same level of performance as the normal LASSO estimator, despite the fact that we are agnostic to the value of σ.
2. From the moderate deviation theory, we can say that even without imposing any constraint on noise, penalty level will be valid asymptotically.
3. Taking the square root will not affect the global convexity of the cost function, making it computationally easy. Besides, leveraging knowledge from LP and the fact that this problem takes a second order cone, efficient methods such as interior point methods that claim to produce polynomial time bounds can be used.

Question 1

(a)

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

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

Performing change of variable from $(x - x_0)$ to p and $y - y_0$ to q, we get

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(p,q)\delta((p+x_0)\cos\theta + (q+y_0)\sin\theta - \rho)\ dp\ dq$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(p,q)\delta(p\cos\theta + q\sin\theta - (\rho - x_0\cos\theta - y_0\sin\theta))\ dp\ dq$$

$$R(g(x,y))(\rho - x_0\cos\theta - y_0\sin\theta, \theta)$$

(b)

$$g'(r,\Psi) \;=\; g(r, \Psi - \Psi_0)$$

This means the graph has been rotated by Ψ_0 radians/degrees in the counter-clockwise direction assuming it is the direction of increasing angle. Noting that the equation of line from $x\cos\theta + y\sin\theta = \lambda$ in the xy coordinates will be $r\cos\Psi\cos\theta + r\sin\Psi\sin\theta = \rho$, $r\cos(\Psi - \theta) = \rho$ in the polar coordinates and transforming an integral from $dxdy$ to $rd\Psi dr$, we get

$$\int_0^{\infty}\int_0^{2\pi} g(r,\Psi - \Psi_0)\delta(r\cos(\theta - \Psi) - \rho))\ r\ d\Psi\ dr$$

Putting $(\Psi - \Psi_0)$ as Φ, we get

$$\int_0^{\infty}\int_0^{2\pi} g(r,\Phi)\delta(r\cos(\theta - (\Phi + \Psi_0)) - \rho))\ r\ d\Phi\ dr$$

$$\int_0^{\infty}\int_0^{2\pi} g(r,\Phi)\delta(r\cos((\theta - \Psi_0) - \Phi) - \rho))\ r\ d\Phi\ dr$$

$$= \; R(g)(\rho, \theta - \Psi_0)$$

Question 2

Fourier Slice Theorem for 3D Images: Let us take the projection angle to be θ. Say the plane determined by θ (θ is normal to the plane) is the detector. Now say, the projection observed on this is $p(u,v,\theta)$, a 2D image as expected. Now, taking its fourier transform with respect to u, v we get, $P(\omega_u, \omega_v, \theta)$. Let us call $F(\omega_u, \omega_v, \omega_z)$ as the Fourier transform of the 3D object f. Take a slice passing through the origin and parallel to the detector $F_{central,\theta}(\omega_x, \omega_y, \omega_z)$ in the frequency plane.

$$P(\omega_u, \omega_v, \theta) \;=\; F_{central,\theta}(\omega_x, \omega_y, \omega_z)$$

It is further illustrated in the figure given below, the notations are the same.

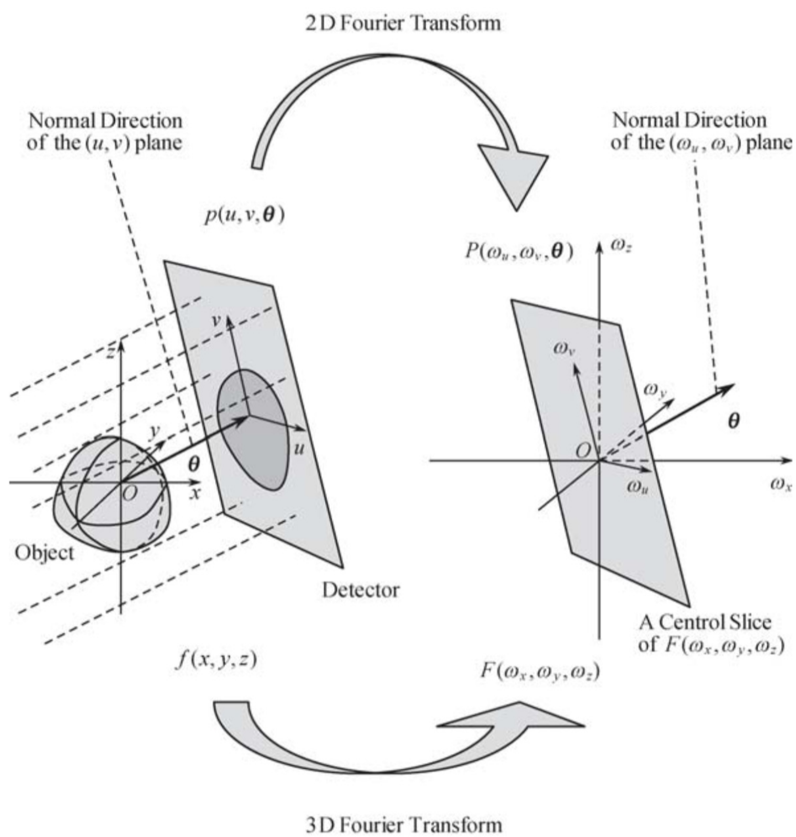

Figure 1: **3D Fourier Slice Transform**
Image Ref: https://link.springer.com/chapter/10.1007/978-3-642-05368-9_5

Question 5

Title: Application of computer tomography in electronic technology
Author List: M. Danczak, K.J. Wolter, R. Rieske, H. Roth
Year of Publication: 8-11 May 2003
Name of the Conference: 26th International Spring Seminar on Electronics Technology: Integrated Management of Electronic Materials Production, 2003.

What the paper aimed to do: Talks about the new non-invasive or non-destructive technology that is, the cone-beam Computer Tomography. This method uses a volumetric beam to scan the subject as compared to the fan-shaped beam used in medical applications. Main applications of cone-beam CT scans is in the visualization of inner and outer structures in the electronic components, packaging and interconnection technologies eg: Flip Chip (FC) and micro ball grid arrays. One issue they addressed is that that of the beam hardening that is to be performed because the radiation is evidently polychromatic and for the best imaging results, monochromatic radiation is required.

Reconstruction Methods and Cost Functions:
The paper suggests the use of **Filtered Back Projection** method.

FBP: This is based on the **Fourier Slice Theorem**. The main idea in this is assembling all the fourier transforms of the projections at various angles and inverting to get an estimate of the subject. Precisely, it is filtering the sinogram data obtained and then back-projecting.
Say the function we are considering is $f(x, y, z)$ and the measurement is y. R_h acts as a Radon operator with the filter h embedded in it in the form of a matrix. The cost function associated will be

$$||y - R_h f||_2^2$$

Take a point (x, y, z) and find the appropriate (a, b) (function of (x, y, z) and θ), an equivalent of $\rho = x \cos\theta_k + y \sin\theta_k$ in 2D reconstruction case, this is $p_\theta(a, b)$, whose fourier transform is $P_\theta(\omega_a, \omega_b)$. Putting these terms in mind, the following is the expression for reconstruction using FBP

$$f(x, y, z) \;=\; \int_0^{\pi} \widehat{p}_\theta(a, b) d\theta$$

where the hat version is a filtered version, could be a ramp filter or Ram-Lak filter or Ram-Lak with hamming window. Let us say the filter is $H(\omega_a, \omega_b)$

$$\widehat{p}_\theta(a, b) \;=\; \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} P_\theta(\omega_a, \omega_b)\; H(\omega_a, \omega_b)\; e^{j2\pi(\omega_a a + \omega_b b)} d\omega_a d\omega_b$$

$$f(x, y, z) \;=\; \int_0^{\pi} \left(\int_{\infty}^{\infty}\int_{-\infty}^{\infty} P_\theta(\omega_a, \omega_b)\; H(\omega_a, \omega_b)\; e^{j2\pi(\omega_a a + \omega_b b)} d\omega_a d\omega_b \right) d\theta$$

The 2-D case can be written as follows

$$f(x, y) \;=\; \int_0^{\pi} \left(\int_{-\infty}^{\infty} H(\mu)\; G(\mu, \theta)\; e^{j2\pi\mu\rho}\; d\mu \right) d\theta$$

where $\mu = u^2 + v^2$ and $\rho = x \cos\theta + y \sin\theta$

References

1. https://link.springer.com/chapter/10.1007/978-3-642-05368-9_5
2. https://drive.google.com/file/d/1lhKb3ustquCuICHnESHhznLHc0M2j83N/view
3. https://ieeexplore.ieee.org/document/1260495/authorsauthors
4. http://www.ecs.umass.edu/~mduarte/images/IntroCS.pdf
5. https://web.stanford.edu/~hastie/StatLearnSparsity_files/SLS_corrected_1.4.16.pdf
6. Slides