Assignment 3: CS 754, Advanced Image Processing

Shreyas Narahari-203050037
Nitish Gangwar-203050069

March 23, 2021

Answer 1:

(a).

We consider $\nu = \hat{\beta} - \beta^*$ as the lasso error lying in the restricted subset of C. The restricted eigen value condition is defined with respect to C and having a constant $\gamma > 0$ as,

$$\frac{\frac{1}{N}\nabla^2 f(\beta)\nu}{\|\nu\|_2^2}$$

A high dimension convex loss function may be flat in some dimensions thus we want to find out the subset in the space such that strong convexity condition is satisfied in it.

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

The above condition denotes the lower bound on restricted eigenvalues of the model matrix.

(b).

Considering the definition of $G(\nu)$ and the lasso error $\hat{\nu} := \hat{\beta} - \beta^*$

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

Substituting $\hat{\nu}$ in $G(\nu)$ we get,

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

Substituting $\hat{\beta} = \hat{\nu} + \beta^*$ in the above equation,

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

Now substituting 0 in $G(\nu)$ we get,

$$G(0) := \frac{1}{2N}\|y - X(\beta^*)\|_2^2 + \lambda_N\|\beta^*\|_1 \tag{2}$$

As here we consider β^* as true regression vector and $\hat{\beta}$ as the Lasso solution we get the inequality $\|y - X\hat{\beta}\|_2^2 \leq \|y - X\beta^*\|_2^2$ and the inequality $\|\hat{\beta}\|_1 \leq \|\beta^*\|_1$ is already known. From these inequalities and the equation 1 and 2 above we get,

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

Thus,

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

(c).

We consider the inequality from the above solution ie., $G(\nu) \leq G(0)$,

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

Rearranging we get,

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

Consider the relation $y = X\beta^* + w$ and $\nu = \hat{\beta} - \beta^*$, Substituting these relation in the L.H.S we get,

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

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

Rearranging we get,

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

(d).

We consider S to be subset of indices where we have values for the sparse vector. From definition we have $\|\beta^*\|_1 = \|\beta^*_S\|_1$ and $\beta^*_{S^c} = 0$ we get,

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

From triangle inequality we get,

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

Substituting the previous inequality in the inequality derived in the previous part [(c).] we get,

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

Holder's inequality states that for inner product $\|f^T g\|_1 \leq \|f\|_p \|g\|_q$ where f and g are vectors of n dimension is true when $p, q \in [1, \infty)$ and $\frac{1}{p} + \frac{1}{q} = 1$ hold. So, for this derivation we consider $p = \infty$ and $q = 1$, this gives us,

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

From the Holder's inequality definition we get,

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

(e).

Since, by assumption we have $\frac{1}{N}\|X^T w\|_\infty \leq \frac{\lambda_N}{2}$ and from definition we have $\|\hat{\nu}\|_1 = \|\hat{\nu}_S\|_1 + \|\hat{\nu}_{S^c}\|_1$,

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

Ignoring the negative term $\{-\|\hat{\nu}_{S^c}\|_1\}$ will not change the inequality and using the fact that $\|\hat{\nu}_S\|_1 \leq \sqrt{k}\|\hat{\nu}\|_2$ we get,

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

(f).

Taking Lemma 11.1 to be true allows us to apply the Restricted Eigenvalues property with the lower bound γ ,

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

Rearranging we get,

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

Combining this bound with the inequality proven in the previous equation and rearranging we get,

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

Dividing and multiplying the RHS by $\sqrt{N}$ and substituting $\hat{\nu} = \hat{\beta} - \beta^*$ and rearranging we get,

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

(g).

The bound $\lambda_N \geq 2\frac{||Xw||_\infty}{N}$ is being used in the equation 11.22 to finally reach to the equation 11.23 i.e.

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

now the following bound $\lambda_N \geq 2\frac{||Xw||_\infty}{N}$ is used to reach to the below given equation:

$$\frac{||X\hat{\nu}||_2^2}{2N} \leq \frac{\lambda_N}{2}\{||\hat{\nu_S}||_1 + ||\hat{\nu_{S^c}}||_1\} + \lambda_N\{||\hat{\nu_S}||_1 - ||\hat{\nu_{S^c}}||_1\} \leq \frac{3}{2}\sqrt{k}\lambda_N||\hat{\nu}||_2$$

since the term $\frac{\lambda_N}{2}$ is greater than the term $\frac{||Xw||_\infty}{N}$ so it will remain greater, that is why the following term $\frac{||Xw||_\infty}{N}$ is replaced by $\frac{\lambda_N}{2}$ in the above written equation. here in the right hand side the following inequality is also being used :

$$||\hat{\nu_S}||_1 \leq \sqrt{k}||\hat{\nu}||_2$$

(h).

We know that the least squares objective function is always convex.

$$f_N(\beta) = \frac{1}{2N}||y - X\beta||_2^2$$

where, $\beta \in R^p$ and X $\in R^{N\times p}$
But, we want to check the conditions under which it is strongly convex in nature. One way is to find $\nabla^2 f(\beta) = \frac{X^T X}{N}$ we know that matrix of this form $X^T X$ has rank of atmost $\min\{N, p\}$.So, it will always be rank deficient and hence not strongly convex and we know that for least square loss to be strongly convex, the eigenvalues of matrix $X^T X$ should be uniformly bounded away from zero.The convex loss function will not be strongly convex rather it will be flat in some directions and curved in others. and hence the constraint of strong convexity is applied only to some subset. Now these sets are selected based on some constraint.If β^* is some sparse vector with support set as $S(\beta^*)$.Then the lasso error is given by

$$\hat{\nu} = \hat{\beta} - \beta^*$$

and for the support set we have $\hat{\nu_S}$ and for $\hat{\nu_{S^c}}$ for the set C. then for choosing the l_1 ball of appropriate radius and regularization parameter.The lasso error turns out to be satisfying the cone constraint of the following form:

$$||\hat{\nu_{S^c}}||_1 \leq \alpha||\hat{\nu_S}||_1$$

and this is why the cone constraint is required.

(i).

Advantage of this theorem over the theorem 3 is that in theorem 3 we used the upper bound over the noise e i.e.

$$||e||_2 \leq \epsilon$$

but there is no such assumption for noise w in this theorem only thing that has been taken into consideration is that $w \in R^N$ and is gaussian with i.i.d. entries having zero mean and variance σ^2.
This algorithm does not requires RIP to be followed but theorem 3 requires that. Advantage of theorem 3 over this theorem is that this theorem requires normalized columns of matrix X whereas it not at all a necessary condition in Theorem 3 to have a orthonormal bases.

(j).

The common thread between the bounds on 'Dantzig selector' and the LASSO are: for the measurement of the type

$$y = \beta X + w$$

Dantzig selector:

$$min_\beta ||X^T(y - X\beta)||_{l_\infty} \quad subject \quad to \quad ||\beta||_{l_1} \leq s$$

Lasso:

$$min_\beta ||y - X\beta||_{l_2} \quad subject \quad to \quad ||\beta||_{l_1} \leq s$$

with the bound on $|\beta|_1$, Dantzig selector minimizes the maximum component of the gradient of the squared error function while LASSO minimizes the squared error.If s is large then both of these are same but for other values of s they both are different.

(k).

The advantages of the square-root LASSO over the LASSO are:

1. square-root LASSO does not need to pre-estimate σ i.e. standard deviation.
2. Square-root LASSO does not have to rely on normality or sub Gaussianity of noise.
3. The problem has global convexity(despite of having a square root over the least square function), which makes it computationally attractive.
4. The constraints take the form of second order cone and this allows the use of efficient algorithmic methods from linear programming literature.

Answer 2.

For tomographic reconstruction using three similar slices we consider the equation,

$$\begin{aligned} E(\beta_1, \beta_2, \beta_3) = \|y_1 - R_1U\beta_1\|^2 + \|y_2 - R_2U(\beta_1 + \beta_2)\|^2 + \|y_3 - R_3U(\beta_1 + \beta_3)\|^2 + \\ \lambda\|\beta_1\|_1 + \lambda\|\beta_1 + \beta_2\|_1 + \|\beta_1 + \beta_3\|_1 \end{aligned}$$

Here U is 2D-DCT matrix and $R1, R2, R3$ are radon transform matrix for the three images respectively. which can be written as,

$$= \left\| \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} - \begin{pmatrix} R_1U & 0 & 0 \\ R_2U & R_2U & 0 \\ R_3U & 0 & R_3U \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} \right\|^2 + \lambda \left\| \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} \right\|_1$$

Answer 3.

(a.)Shifting Property:

To Prove:

$$R(g(x-x_0,y-y_0))(\rho,\theta) = R(g(x,y))(\rho - x_0cos\theta - y_0sin\theta,\theta)$$

We know that,

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

where R denotes the Radon transform
we now want to find $R(g(x-x_0,y-y_0))$

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

$$Rg(x-x_0,y-y_0) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x-x_0,y-y_0)\delta(xcos\theta-x_0cos\theta+ysin\theta-y_0sin\theta-\rho)dx'dy'$$

let's take

$$x' = x - x_0$$

$$y' = y - y_0$$

$$\rho' = xcos\theta + ysin\theta - \rho$$

then,

$$Rg(x-x_0,y-y_0) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x',y')\delta(\rho' - x_0cos\theta - y_0sin\theta)dx'dy'$$

the above written equation is just a change of variable and can be rewritten as:

$$= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x,y)\delta(\rho - x_0cos\theta - y_0sin\theta)dxdy$$
$$= R(g(x,y))(\rho - x_0cos\theta - y_0\sin\theta,\theta)$$

(b.)<u>Rotation:</u>

To Prove:

$$R(g')(\rho,\theta) = R(g)(\rho,\theta-\psi_0)$$

The 2D radon transform in polar coordinates (r,ψ) is given by:

$$Rg(\rho,\theta) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(r,\psi)\delta(\rho - rcos\psi cos\theta - rsin\psi sin\theta)|r|drd\psi$$

$$= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(r,\psi)\delta(\rho - rcos(\psi-\theta))|r|drd\psi$$

Now,

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

$$Rg(r,\psi-\psi_0) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(r,\psi-\psi_0)\delta(\rho - rcos(\psi-\theta))|r|drd\psi$$

let's take

$$\psi' = \psi - \psi_0$$

$$= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(r,\psi')\delta(\rho - rcos(\psi'+\psi_0-\theta))|r|drd\psi'$$

above given equation is terms of r and ψ'.Hence,

$$Rg(\rho,\theta-\psi_0)$$

(c.)<u>Convolution:</u>

Given: image f(x,y) and kernel k(x,y) To Prove:

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

two dimensional convolution is given by:

$$f*g = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x',y')k(x-x',y-y')dx'dy'$$

then radon transform of $f*k$ is given as:

$$R_\theta(f*k) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x',y')k(x-x',y-y')\delta(\rho - xcos\theta - ysin\theta)dx'dy'dxdy$$

Now by using shifting theorem we can say that radon transform of function f is shifted to the points (x',y') in integral of dx and dy:

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

Now let us include another integration in terms of ρ'.

$$=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}f(x',y')Rk(\rho-\rho',\theta)\delta(\rho'-x'cos\theta-y'sin\theta)dx'dy'd\rho'$$

Now we have radon transform in k for getting radon on f we can write:

$$=\int_{-\infty}^{+\infty}Rf(\rho',\theta)Rk(\rho-\rho',\theta)d\rho'$$

The above written equation denotes the convolution in terms of paramter ρ.

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

Answer 4.

We know from Ristricted Isometry Property for s-order:

$$(1-\delta_s)||x||_2^2\leq||Ax||_2^2\leq(1+\delta_s)||x||_2^2$$

we know that,

$$\delta_s=max_{S\subset[n],|S|\leq s}||A_S^*A_S-Id||_{2->2} \tag{3}$$

and

$$(1-\mu(s-1))||x||_2^2\leq||Ax||_2^2\leq(1+\mu(s-1))||x||_2^2$$

As $A_S^*A_S$ is a positive semi definite matrix having real and positive eigenvalues. It has also got orthogonal basis of eigen vectors.Due to the normalized $||a_i||_2=1$ for all i $\in$ [n].The eigen values of $A_S^*A_S$ are bounded by the union of the disks with center 1 as the diagonal entries of $A_S^*A_S$ are 1. By Gershgorin's disk theorem,

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

where λ denotes the eigen value.
B_{ii} denotes the diagonal entry.

r_i denotes the union/sum of non diagonal entries in the i-th row.
where

$$\begin{aligned} r_i &= \sum_{k \in S, k \neq i} |(A_S^*)_{i,k}| \\ &= \sum_{k \in S, k \neq i} |(a_k, a_i)| \leq \mu(s-1) \end{aligned} \tag{4}$$

and hence, using 3 and 4 we can write as

$$\delta_s \leq (s-1)\mu$$

Answer 5.

Title:

Inline discrete tomography system: Application to agricultural product inspection.

Link:

click here

Venue:

Computers and Electronics in Agriculture[ELSIVIER] Volume 138

Year of publication:

2017

Mathematical Problem:

Algebraic Reconstruction Methods are being used, which cast the reconstruction problem to system of linear equation like

$$p = Wv$$

where,
$\mathbf{p} \in R^M$ denotes the projected data of the scanned object.

$\mathbf{v} \in R^N$ this is what that needs to be reconstructed i.e. this is a vector associated with reconstructed image.
$\mathbf{W}_{M\times N}$ is a matrix which maps the vector v(vector associated with reconstructed image) with p(projected data).
for estimating $\mathbf{v}$, the following function is being minimized:

$$\chi = |Wv - p|^2$$

Method of optimization:

Authors want to improve the image reconstruction quality for the proposed inline scanning geometry under resource constrained environment like **limited angular range, limited number of projections and truncated projections**.The approaches they have used are as follows:

Scanning geometry with object rotation

Here, the object rotates around its own axis and translates linearly over the conveyor belt.Such rotation allows higher angular sampling about the scanned object and this is how the effects of limited angular range is reduced for the inline scanning setup.

Domain constrained discrete tomography

Due to prior knowledge(opticals sensors were employed in the algorithm,for determining the object contour) of outer structure the background pixels can be removed from the reconstruction domain. Hence, limited number of projections has to deal with only a subset of unknown pixels.
In this paper authors are using the extension of DART(Discrete Algebraic Reconstruction Technique) along with EOD(Expected Object Domain) and is referred to as EOD-DART which is being used,developed and evaluated.
Using these optimizations, authors were able to perform image reconstruction in small number of algorithmic iteration even in resource constrained environment.

References:

1. http://dsp-book.narod.ru/TAH/ch08.pdf

2. http://oftankonyv.reak.bme.hu/tiki-index.php?page=The+properties+of+the+Radon+transfor

3. https://link.springer.com/book/10.1007/978-0-8176-4948-7
4. http://dsp-book.narod.ru/TAH/ch08.pdf
5. https://en.wikipedia.org/wiki/H%C3%B6lder%27s$_inequality$
6. https://www.sciencedirect.com/science/article/abs/pii/S0168169916301909
7. https://www.clear.rice.edu/elec301/Projects00/tomography/code.htm
8. https://web.stanford.edu/ hastie/Papers/dantzig$_annals.pdf$
9. https://web.stanford.edu/ hastie/StatLearnSparsity$_files/SLS_corrected_1.4.16.pdf$
10. https://arxiv.org/pdf/math/0506081.pdf
11. https://mathoverflow.net/questions/265905/generalized-h%C3%B6lders-inequality-for-operator-subordinate-norms
12. https://deepblue.lib.umich.edu/bitstream/handle/2027.42/29625/0000714.pdf;sequence=1