CS 754 - Assignment 3

Nihar Mehta, Ushasi Chaudhuri

March 18, 2018

Question 1

Tomography has a wide variety of applications like clinical tomography, structural biology/virology, nuclear medicine, positron emission tomography, structure and fault analysis, wood structure analysis, etc. One such important area of usage of tomography is very high resolution space borne SAR tomography in urban environment [1]. Synthetic aperture radar tomography (TomoSAR) extends the synthetic aperture principle into the elevation direction for 3-D imaging. It uses stacks of several acquisitions from slightly different viewing angles (the elevation aperture) to reconstruct the reflectivity function along the elevation direction by means of spectral analysis for every azimuth–range pixel.

Model- This paper demonstrates the potential of the new class of VHR space borne SAR systems for TomoSAR in urban environment. They introduce a new Wiener type regularization to the singular value decomposition (SVD) method for TomoSAR. Different model selection schemes for the estimation of the number of scatterers are evaluated and validated. Two parametric estimation algorithms of the scatterers' elevation and their velocities are evaluated. They also demonstrate first 3-D and 4-D reconstructions of an entire building from spaceborne VHR data by pixelwise TomoSAR. Volumetric objects, like trees, can be considered incoherent in X-band repeat pass and are hence treated as noise. A limited number (typically one to three) of scatterers is expected along every elevation profile that allows parametric estimation

Mathematical formulation- In VHR X-band data, the following component contributes,

- Weak diffuse scattering from mostly horizontal or vertical rough surfaces (roads and building walls).
- Strong returns from metallic structures or specular and dihedral or trihedral reflections.
- Returns from volumetric scatterers, e.g., from vegetation.

The noise sources are the following.

- Gaussian noise, which is caused by thermal noise and temporal decorrelation.
- Calibration errors in amplitude.
- Phase errors caused by atmospheric delay and unmodeled motion.

One SAR acquisition may be considered to be one tomographic projection of the complex reflectivity of the object along elevation.

$$g_n = \int_{\Delta s} \gamma(S) \exp(-j2\pi\xi_n s)\, ds \tag{1}$$

$n = 1, ...N$ and $\gamma(s)$ represents the reflectivity function along elevation s. $\xi = -2b_n/(\lambda r)$ is the spatial frequency. The continuous-space system model of this can be approximated by discretizing the continuous reflectivity function along s.

$$g_n = \delta_s \dot{\Sigma} \gamma(S_l) \exp(-j2\pi\xi_n s_l)\, ds \tag{2}$$

where L is the number of discrete elevation indices. After dropping the inconsequential leading constant δs, the system imaging model becomes

$$g = R\gamma \tag{3}$$

where g is the measurement vector with N elements.

Reconstruction For reconstruction, two frameworks have been suggested, one using SVD, and another using wiener regularization. The discrete reflectivity signal γ can be reconstructed from g through pseudo inversion of the imaging system matrix. However, due to the nonuniform track distribution, the solution may include significant noise propagation. The SVD is used for analyzing image quality and the amount of independent information about the unknowns that can be reliably retrieved from observations in the presence of noise. Due to the reciprocal of σn, noise propagation caused by small SVs will compromise this solution, and regularization tools are required.

Transforming the MAP estimator of to the SV space results in a soft thresholding, e.g., weighting the SVs according to their magnitudes, also referred to as a Tikhonov regularization.

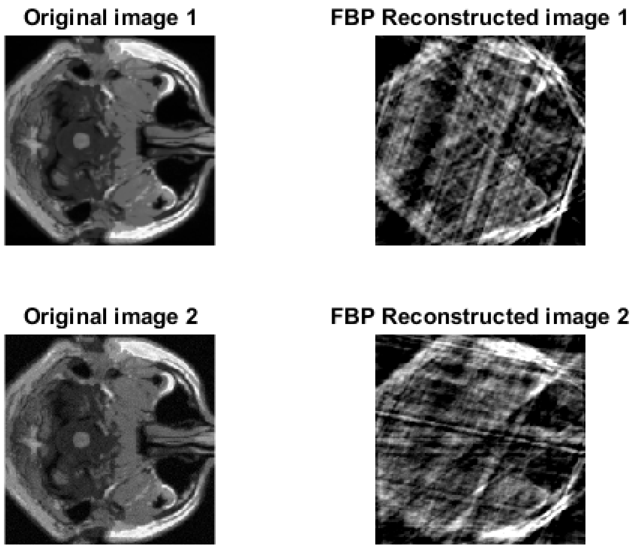
$$\vec{\gamma}_{MAP} = (\Sigma^T\Sigma + |\eta|^2 I)^{-1} V \Sigma^T U^T g = \Sigma \sigma_n^{-1}, Wiener(u_n^T g) v_n \tag{4}$$

where $\sigma_n^{-1}, Wiener$ denotes optimum weights.

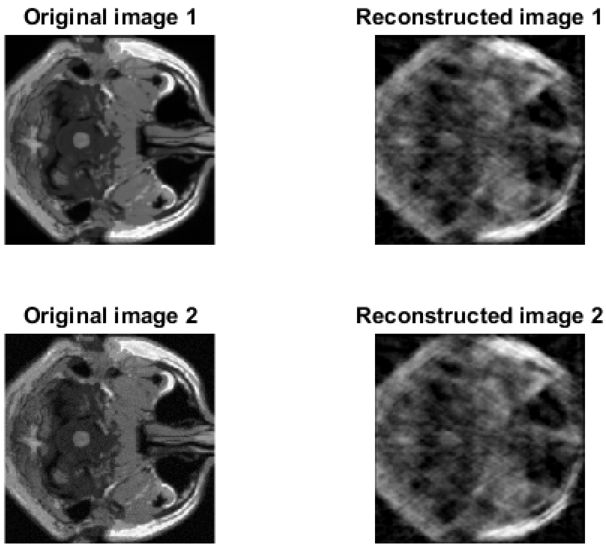
$$\sigma_n^{-1}, Wiener = \frac{\sigma_n}{|\sigma_n|^2 + |\eta|^2} \tag{5}$$

Question 2

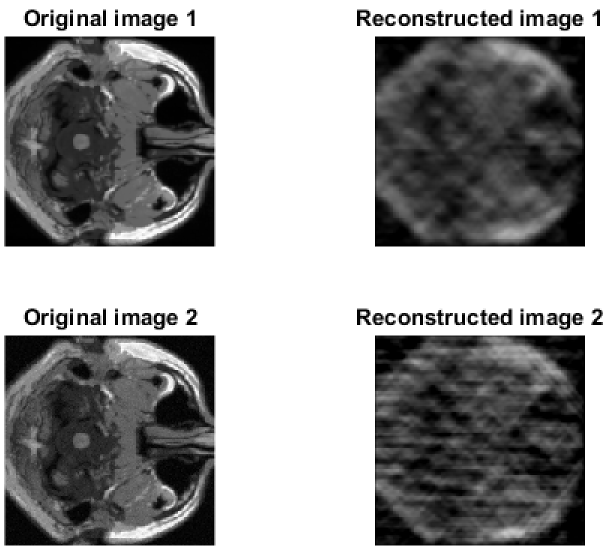
a) Reconstruction via Filtered Back Projection:



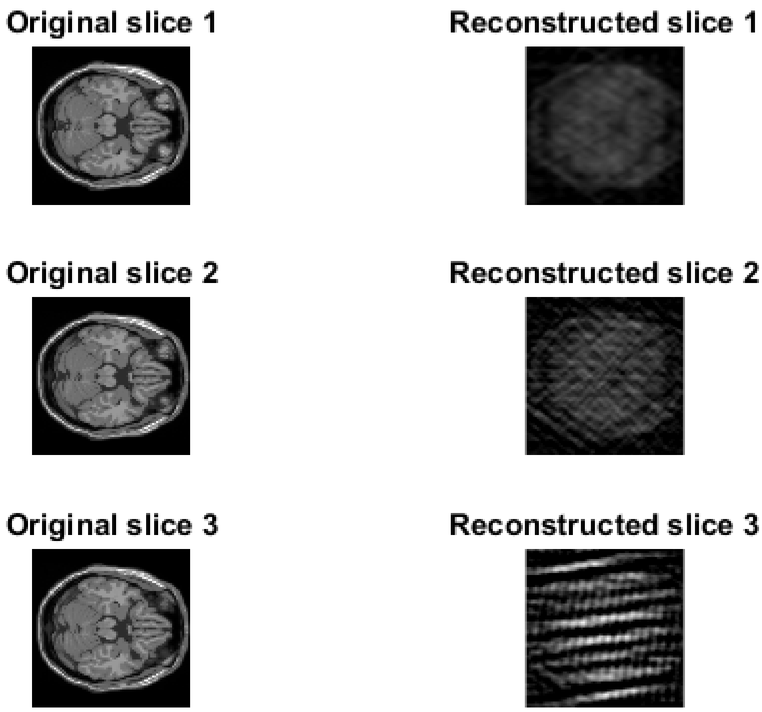
b) Reconstruction via compressed sensing:



c) Reconstruction via 2-coupled compressed sensing



d) Reconstruction via 3-coupled compressed sensing:



Question 3

(a)
In a high-dimensional setting, in which the number of parameters p is larger than N,the least-squares objective function $f_N(\beta) = \frac{1}{2N}||y - X\beta||^2{}_2$ is always convex. The least-squares loss is strongly convex if and only if the eigenvalues of the $p \times p$ positive semidefinite matrix $X^T X$ are uniformly bounded away from zero. However, it is easy to see that any matrix of the form $X^T X$ has rank at most minN, p, so it is always rank-deficient and hence not strongly convex whenever $N < p$. A convex loss function in high-dimensional settings (with $p \geq\geq N$) cannot be strongly convex; rather, it will be curved in some directions but flat in others. For this reason, we need to relax our notion of strong convexity.
We say that a function f satisfies the restricted strong convexity at $\beta*$ with respect to C if there is a constant $\gamma \geq 0$, such that

$$\frac{v^T \delta^2 f(\beta) v}{||v||^2{}_2} \geq \gamma \tag{6}$$

for all nonzero $v \in C$
In the specific case of linear regression, this notion is equivalent to lower bounding the restricted eigenvalues of the model matrix in particular, requiring that

$$\frac{\frac{1}{N} v X^T X v}{||v||^2{}_2} \geq \gamma \tag{7}$$

for all nonzero $v \in C$
This is the restricted eigenvalue condition.

(b-g)
Please find the attached solution of part (b) to (g) in Fig. 1 to Fig. 5.

(h)
Relevance of the cone constraint-
Suppose that the parameter vector β^* is sparse, supported on the subset $S = S(\beta^*)$. Defining the lasso error $\vec{v} = \vec{\beta} - \beta^*$, let $\vec{v}_s \in R^{|S|}$ denote the sub vector indexed by elements of S, with $\vec{v}_s^c$ defined in an analogous manner.

For appropriate choices of the l_1-ball radius or equivalently, of the regularization parameter λ_N it turns out that the lasso error satisfies a cone constraint of the form $||\vec{v}_{s^c}||_1 \leq \alpha||\vec{v}_s||$, for some constant $\alpha \geq 1$. This fact is easiest to see for the lasso in its constrained version. If we solve the constrained lasso ie, minimize $||y - X\beta||^2{}_2$, for $||\beta||_1 \leq R^p$ with ball radius $R = ||\beta||_1$, then since $\vec{\beta}$ is feasible for the program, we have $R = ||\beta_s{}^*||_1 \geq ||\beta_s{}^*||_1 - ||\vec{v}_s||_1 + ||\vec{v}_{s^c}||_1$.

Rearranging this inequality, we see that the bound $||\vec{v}_{s^c}||_1 \leq \alpha||\vec{v}_s||$ holds with $\alpha = 1$. If we instead solve the regularized version (minimize $\frac{1}{2N}||y - X\beta||^2{}_2 + \lambda_N||\beta||_1$, for $\beta \in R^p$) of the lasso with a "suitable" choice of λ_N , then it turns out that the error satisfies the constraint $||\vec{v}_{s^c}||_1 \leq 3||\vec{v}_s||_1$.

Thus, in either its constrained or regularized form, the lasso error is restricted to a set of the form $C(S;\alpha) := v \in R^p|||v_{s^c}||_1 \leq \alpha||v_s||_1$ for some $\alpha \geq 1$.

(i)
For a noise vector w and a zero-mean Gaussian with standard deviation σ, the lasso-error satisfies the bound

$$||\vec{\beta} - \beta^*||_2 \leq \frac{c\sigma}{\gamma}\sqrt{\frac{\tau\kappa \log p}{N}} \tag{8}$$

with probability of at least $1 - 2\exp -\frac{1}{2}(\tau - 2)\log p$. On the other hand, theorem 3 states that to solve the optimization problem $\min||\theta||_1$ such that $||\theta^* - \theta||_2 \leq \frac{C_0}{\sqrt{S}}||\theta - \theta_s||_1 + C_1\epsilon$. The advantages of the example 11.1 are as follows:

- Example 11.1 is better since when N tends to infinity, the error tends to zero, giving the exactly correct estimate. Whereas for theorem 3, the error is bounded away from zero (not achievable).
- The error covariance in case of the given example asymptotically goes to zero at the rate of $1/sqrtN$, which is the best possible convergence rate one can achieve for any possible statistical estimator. Hence, this is the statistically an efficient estimator.

While we have these advantages of the example 11.1, we have a few observations which are in favor of theorem 3, making it better for certain cases. These are explained as follows:

- For a small sample size, the error bound of theorem 3 would be much tighter, hence the estimator is better in this case.
- In example 11.1, the error bound depends on the square root of log of sparsity. Whereas for theorem 3, the bound is inversely proportional to the square root of sparsity (S).
- Example 11.1 comes with an probability condition that it holds true with a probability of at least $1 - 2\exp -\frac{1}{2}(\tau - 2)\log p$. Whereas the theorem 3 holds true always.

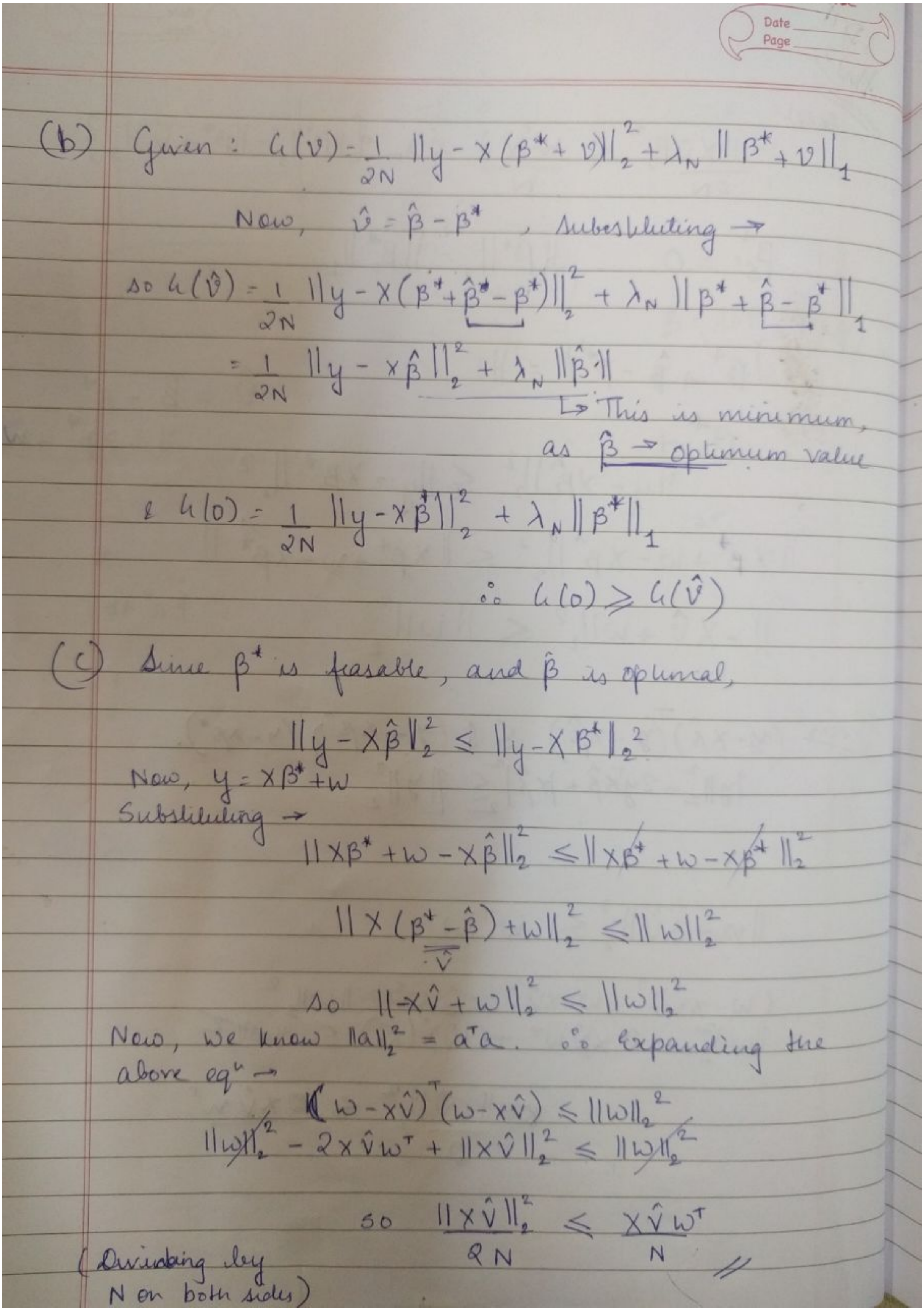
Question 4

a)

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

$= \int\int g(x,y)\delta(xcos(\theta) + ysin(\theta) - (\rho - x_0cos(\theta) - y_0sin(\theta)))dxdy$

$= \int\int g(x,y)\delta((x + x_0)cos(\theta) + (y + y_0)sin(\theta) - \rho)dxdy$


Date
Page

(b) Given: $h(v) = \frac{1}{2N} \|y - X(\beta^* + v)\|_2^2 + \lambda_N \|\beta^* + v\|_1$

Now, $\hat{v} = \hat{\beta} - \beta^*$, subestituting →

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

$= \frac{1}{2N} \|y - X\hat{\beta}\|_2^2 + \lambda_N \|\hat{\beta}\|$

↳ This is minimum, as $\hat{\beta}$ → optimum value

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

$\therefore h(0) \geq h(\hat{v})$

(c) Since β^* is feasable, and $\hat{\beta}$ is optimal,

$$\|y - X\hat{\beta}\|_2^2 \leq \|y - X\beta^*\|_2^2$$

Now, $y = X\beta^* + w$

Substituting →

$$\|X\beta^* + w - X\hat{\beta}\|_2^2 \leq \|X\beta^* + w - X\beta^*\|_2^2$$

$$\|X(\underbrace{\beta^* - \hat{\beta}}_{\hat{v}}) + w\|_2^2 \leq \|w\|_2^2$$

$$\text{so } \|-X\hat{v} + w\|_2^2 \leq \|w\|_2^2$$

Now, we know $\|a\|_2^2 = a^T a$. $\therefore$ expanding the above eq^n →

$$(w - X\hat{v})^T (w - X\hat{v}) \leq \|w\|_2^2$$

$$\|w\|_2^2 - 2X\hat{v}w^T + \|X\hat{v}\|_2^2 \leq \|w\|_2^2$$

$$\text{so } \frac{\|X\hat{v}\|_2^2}{2N} \leq \frac{X\hat{v}w^T}{N}$$

(Dividing by N on both sides)

Figure 1: Part b and c.

Now, since, $\frac{\|X\hat{v}\|_2^2}{2N} \leq \frac{w^T X \hat{v}}{N}$

Adding a term (non negative) to RHS would still preserve the inequality.

So $\frac{\|X\hat{v}\|_2^2}{2N} \leq \frac{w^T X \hat{v}}{N} + \lambda_N \{\|\beta^*\|_1 - \|\beta^* + \hat{v}\|_1\}$

(d) Since $\beta^*_{S^c} = 0$, $\|\beta^*\|_1 = \|\beta^*_S\|_1$.

Substituting this in the following term $\rightarrow$

$\|\beta^* + \hat{v}\|_1 = \|\beta^*_S + \hat{v}_S\|_1 + \|\hat{v}_{S^c}\|_1$ {$\hat{v}$ consists of both $\hat{v}_S$ & its compliment set $\hat{v}_{S^c}$}

By Summation inequality,

$\geq \|\beta^*_S\|_1 - \|\hat{v}_S\|_1 + \|\hat{v}_{S^c}\|_1$ $\{\|a+b\| \geq \|a\| - \|b\|\}$

From part (c),

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

Substituting,

$\leq \frac{w^T X \hat{v}}{N} + \lambda_N \{\cancel{\|\beta^*_S\|_1} - \cancel{\|\beta^*_S\|_1} + \|\hat{v}_S\|_1 - \|\hat{v}_{S^c}\|_1\}$

$\leq \frac{w^T X \hat{v}}{N} + \lambda_n \{\|\hat{v}_S\|_1 - \|\hat{v}_{S^c}\|_1\}$

By Holder's Inequality, $\|fg\|_1 \leq \|f\|_p \|g\|_q$ where $\frac{1}{p} + \frac{1}{q} = 1$.

Now in the term $\frac{w^T X \hat{v}}{N}$ $\rightarrow$ scalar part. so if $\boxed{p = 1, q = \infty}$

So Taking its L_1 norm doesn't change the eq^n.

$\leq \frac{\|w^T X\|_\infty}{N} \|\hat{v}\|_1 + \lambda_n \{\|\hat{v}_S\|_1 - \|\hat{v}_{S^c}\|_1\}$

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

Figure 2: Part d

(e) Till now, we have →

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

Since, $\frac{1}{N}\|X^T w\|_\infty \leq \frac{\lambda_N}{2}$ (by assumption)

Substituting,

$$\leq \frac{\lambda_N}{2}\|\hat{v}\|_1 + \lambda_N\{\|\hat{v}_S\|_1 - \|\hat{v}_{S^c}\|_1\}$$

$$\leq \frac{\lambda_N}{2}\{\|v_S\|_1 + \|v_{S^c}\|_1\} + \lambda_N\{\|\hat{v}_S\|_1 - \|\hat{v}_{S^c}\|_1\}$$

$$\leq \frac{3}{2}\lambda_N\|\hat{v}_S\|_1 - \frac{\lambda_N}{2}\underbrace{\|\hat{v}_{S^c}\|_1}_{\text{positive term (since } L_1 \text{ norm)}}$$

So a looser bound will give us →

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

Now, $\|v_S\|_1 \leq \|\hat{v}\|_1$

so

$$\leq \frac{3}{2}\lambda_N\|\hat{v}\|_1$$

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

(Norm change relation → $\|a\|_1 \leq \|a\|_2\sqrt{k}$)

where k is the sparsity.

(f) Assuming Lemma 11.1 to be true, we get →

$\lambda_N \geq 2\frac{\|X^T w\|_\infty}{N} > 0$, then $\hat{v} = \hat{\beta} - \beta^*$ belongs to cone set $C(S; 3)$

By the γ-RE condition from (11.10) →

$$\frac{1}{N}\|X\hat{v}\|_2^2 \geq \gamma\|\hat{v}\|_2^2$$

Substituting this in the equation →

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

Figure 3: Part e and f.

classmate

Date

Page

$$\frac{\gamma}{\cancel{2}}\|\hat{v}\|_2^2 \le \frac{3}{\cancel{2}}\lambda_N\sqrt{K}\|\hat{v}\|_2 \quad \text{———} \quad (*)$$

Substituting the result of Lema 11.14,

$$\boxed{0 < \frac{\|X^T w\|_\infty}{N} \le \frac{\lambda_N}{2}}$$

in the inequality of part (e) →

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

$$\underbrace{0}_{\text{Loose (also the lower bound of } \frac{\|X^T w\|_\infty}{N})} \le \frac{\lambda_N}{2}\|\hat{v}\|_1 + \lambda_N\{\|\hat{v}_S\|_1 - \|\hat{v}_{S^c}\|_1\}$$

Rearranging the terms →

$$0 \le \frac{\lambda_N}{2}\{\|\hat{v}_S\|_1 + \|\hat{v}_{S^c}\|_1\} + \lambda_N\{\|\hat{v}_S\|_1 - \|\hat{v}_{S^c}\|_1\}$$

so $$0 \le \frac{3}{\cancel{2}}\lambda_N\|\hat{v}_S\|_1 - \frac{\lambda_N}{\cancel{2}}\|\hat{v}_{S^c}\|_1.$$

so $$\|\hat{v}_{S^c}\|_1 \le 3\|\hat{v}_S\|_1 \quad //$$ hence proved//

Substituting our findings in eqn (*) →

$$\|\hat{\beta} - \beta^*\|_2 \le \frac{3}{\gamma}\lambda_N\sqrt{K}$$

Dividing & multiplying by $\sqrt{N}$

$$\|\hat{\beta} - \hat{\beta}^*\|_2 \le \frac{3}{\gamma}\sqrt{\frac{K}{N}}\sqrt{N}\lambda_N \quad //$$ Theorem 11.1 (b)

Figure 4: continued.

Changing variables as $x_1 = x + x_0$ and $y_1 = y + y_0$, we get

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

$= \int\int g(x_1 - x_0, y_1 - y_0)\delta(x_1 cos(\theta) + y_1 sin(\theta) - \rho)dx_1 dy_1$

$= \int\int g(x - x_0, y - y_0)\delta(xcos(\theta) + ysin(\theta) - \rho)dxdy$

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

Hence proved.

b)

$R(g'(\rho, \theta))$

$= \int\int g'(x, y)\delta(xcos(\theta) + ysin(\theta) - \rho)dxdy$

$= \int\int g'(r, \psi)\delta((rcos(\psi)cos(\theta) + (rsin(\psi))sin(\theta) - \rho)rdrd\psi$

$= \int\int g(r, \psi - \psi_0)\delta((rcos(\psi)cos(\theta)) + (rsin(\psi)sin(\theta)) - \rho)rdrd\psi$

$= \int\int g(r, \psi_1)\delta((rcos(\psi_1 + \psi_0)cos(\theta)) + (rsin(\psi_1 + \psi_0)sin(\theta)) - \rho)rdrd\psi_1$

$= \int\int g(r, \psi)\delta((rcos(\psi + \psi_0)cos(\theta)) + (rsin(\psi + \psi_0)sin(\theta)) - \rho)rdrd\psi$

$= \int\int g(r, \psi)\delta((rcos(\psi - (\psi_0 - \theta)) - \rho)rdrd\psi$

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

Hence proved.

References

[1] X. X. Zhu and R. Bamler. Very high resolution spaceborne sar tomography in urban environment. *IEEE Transactions on Geoscience and Remote Sensing*, 48(12):4296–4308, Dec 2010.

(g) $\lambda_N \geq 2\frac{\|Xw\|_\infty}{N}$??

The Lemma 11.1 gives that if $\lambda_N \geq 2\|\frac{X^T w}{N}\|_\infty > 0$, then the error $\hat{v} := \hat{\beta} - \beta^*$ associated with any lasso solution $\hat{\beta}$ belongs to the cone set $C(S;3)$

This Lemma has been used in eq^n 11.23

i.e, $\frac{\|X\hat{v}\|_2^2}{2N} \leq \frac{3}{2}\sqrt{k}\,\lambda_N\|\hat{v}\|_2$

to yield the result → $\frac{\gamma}{2}\|\hat{v}\|_2^2 \leq \frac{3}{2}\lambda_N\sqrt{k}\,\|\hat{v}\|_2$

~~State~~ since it allows us to apply the γ-RE condition.

Using Lemma 11.1, further, its lower bound $2\frac{\|Xw\|_\infty}{N} > 0$ has been used

to prove the concluding part of theorem 11.1 as shown in part (f).

Figure 5: Part g.