CS754 Assignment 3 Report

Arpon Basu
Shashwat Garg

Spring 2022

Contents

Introduction

Welcome to our report on CS754 Assignment 3. We have tried to make this report comprehensive and self-contained. We hope reading this would give you a proper flowing description of our work, methods used and the results obtained.

Hope you enjoy reading the report. Here we go!

1 Problem 1

2 Problem 2

(a) The restricted eigenvalue condition is defined as follows:
The matrix $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ is said to satisfy the Restricted Eigenvalue Property over a constraint set $\mathcal{C}$ with coefficient γ if

$$\frac{1}{N\cdot\|\nu\|_2^2}\nu^T X^T X\nu \geq \gamma \;\forall\; \nu \in \mathcal{C}\setminus\{\mathbf{0}\}$$

$$\implies \frac{1}{N\cdot\|\nu\|_2^2}(X\nu)^T X\nu \geq \gamma \;\forall\; \nu \in \mathcal{C}\setminus\{\mathbf{0}\}$$

$$\implies \frac{1}{N}\|X\nu\|_2^2 \geq \gamma\|\nu\|_2^2 \;\forall\; \nu \in \mathcal{C}$$

Thus our definition condenses to: **The matrix $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ is said to satisfy the Restricted Eigenvalue Property over a constraint set $\mathcal{C}$ with coefficient γ if**

$$\boldsymbol{\frac{1}{N}\|X\nu\|_2^2 \geq \gamma\|\nu\|_2^2 \;\; \forall \nu \in \mathcal{C}}$$

(b) We have that

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

$$J(\boldsymbol{x}) := \frac{1}{2N}\|y - X\boldsymbol{x}\|_2^2 + \lambda_N\|\boldsymbol{x}\|_1$$

Thus

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

On the other hand, for $\widehat{\nu} := \widehat{\beta} - \beta^*$, we have

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

But $\widehat{\beta}$ was, **by construction** (as mentioned on Pg# 308, third last line), the optimizer of the LASSO function $J(\boldsymbol{x})$, and consequently $G(\widehat{\nu}) = J(\widehat{\beta}) \leq J(\beta^*) = G(0)$, as desired.
(c) We just have to piece together some equations and inequalities to derive equation 11.21. They go as follows:

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

$$y = X\beta^* + w$$
$$\widehat{\beta} = \beta^* + \widehat{\nu}$$

The first inequality $G(\widehat{\nu}) \leq G(0)$, combined with the fact that $\widehat{\beta} = \beta^* + \widehat{\nu}$ yields

$$G(\widehat{\nu}) \leq G(0)$$
$$\implies \frac{1}{2N}\|y - X\widehat{\beta}\|_2^2 + \lambda_N\|\widehat{\beta}\|_1 \leq \frac{1}{2N}\|y - X\beta^*\|_2^2 + \lambda_N\|\beta^*\|_1$$
$$\implies \frac{1}{2N}\|y - X(\beta^* + \widehat{\nu})\|_2^2 + \lambda_N\|\beta^* + \widehat{\nu}\|_1 \leq \frac{1}{2N}\|y - X\beta^*\|_2^2 + \lambda_N\|\beta^*\|_1$$

Now, using the fact that $y = X\beta^* + w$, we get

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

Now, $\|w - X\widehat{\nu}\|_2^2 = (w - X\widehat{\nu})^T \cdot (w - X\widehat{\nu}) = \|w\|_2^2 + \|X\widehat{\nu}\|_2^2 - 2w^T X\widehat{\nu}$. Substituting this in the equation above yields

$$\frac{1}{2N}(\|w\|_2^2 + \|X\widehat{\nu}\|_2^2 - 2w^T X\widehat{\nu}) \leq \frac{1}{2N}\|w\|_2^2 + \lambda_N\{\|\beta^*\|_1 - \|\beta^* + \widehat{\nu}\|_1\}$$
$$\implies \frac{1}{2N}(\|X\widehat{\nu}\|_2^2 - 2w^T X\widehat{\nu}) \leq \lambda_N\{\|\beta^*\|_1 - \|\beta^* + \widehat{\nu}\|_1\}$$
$$\implies \frac{1}{2N}\|X\widehat{\nu}\|_2^2 - \frac{1}{N}w^T X\widehat{\nu} \leq \lambda_N\{\|\beta^*\|_1 - \|\beta^* + \widehat{\nu}\|_1\}$$
$$\implies \frac{\|X\widehat{\nu}\|_2^2}{2N} \leq \frac{w^T X\widehat{\nu}}{N} + \lambda_N\{\|\beta^*\|_1 - \|\beta^* + \widehat{\nu}\|_1\}$$

as desired.
(d) We use a version of Hölder's inequality for vectors as follows:
Let $p, q \in [1, \infty]$ be two real numbers such that $\frac{1}{p} + \frac{1}{q} = 1$. Let $\boldsymbol{x} = (x_1, x_2, ..., x_n)$ and $\boldsymbol{y} = (y_1, y_2, ..., y_n)$ be two vectors. Then

$$\boldsymbol{\sum_{k=1}^{n} |x_k y_k| \leq (\sum_{k=1}^{n} |x_k|^p)^{\frac{1}{p}} (\sum_{k=1}^{n} |y_k|^q)^{\frac{1}{q}} = \|x\|_p \|y\|_q}$$

We augment this inequality for our purposes: Note that $\boldsymbol{x}\cdot\boldsymbol{y} = \sum_{k=1}^{n} x_k y_k \leq \sum_{k=1}^{n} |x_k y_k| \leq (\sum_{k=1}^{n} |x_k|^p)^{\frac{1}{p}} (\sum_{k=1}^{n} |y_k|^q)^{\frac{1}{q}}$, thus showing that $\boldsymbol{x} \cdot \boldsymbol{y} \leq (\sum_{k=1}^{n} |x_k|^p)^{\frac{1}{p}} (\sum_{k=1}^{n} |y_k|^q)^{\frac{1}{q}} = \|\boldsymbol{x}\|_p \|\boldsymbol{y}\|_q$.
In our context let $\boldsymbol{x} = X^T w$ and $\boldsymbol{y} = \nu$, and let $p = \infty$, $q = 1$. Then, $(X^T w) \cdot \nu = (X^T w)^T \nu = w^T X\nu$. Then we have that

$$(X^T w) \cdot \nu \leq \|X^T w\|_\infty \|\nu\|_1$$
$$\implies w^T X\nu \leq \|X^T w\|_\infty \|\nu\|_1$$

Thus,

$$\frac{w^TX\widehat{\nu}}{N} + \lambda_N\{\|\widehat{\nu_S}\|_1 - \|\widehat{\nu_{S^c}}\|_1\} \leq \frac{\|X^Tw\|_\infty}{N}\|\nu\|_1 + \lambda_N\{\|\widehat{\nu_S}\|_1 - \|\widehat{\nu_{S^c}}\|_1\}$$

as desired.
(e) Let S be the set of indices where β^* is non-zero, and let $S^c := [n] \setminus S$. Now, we've already shown that

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

Now,

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

since $\beta^*_{S^c} = \mathbf{0}$, and applying the triangle inequality.
From the previous part, we also know that

$$\frac{w^TX\widehat{\nu}}{N} + \lambda_N\{\|\widehat{\nu_S}\|_1 - \|\widehat{\nu_{S^c}}\|_1\} \leq \frac{\|X^Tw\|_\infty}{N}\|\nu\|_1 + \lambda_N\{\|\widehat{\nu_S}\|_1 - \|\widehat{\nu_{S^c}}\|_1\}$$

$$\implies \frac{w^TX\widehat{\nu}}{N} + \lambda_N\{\|\widehat{\nu_S}\|_1 - \|\widehat{\nu_{S^c}}\|_1\} \leq \frac{\lambda_N}{2}\|\nu\|_1 + \lambda_N\{\|\widehat{\nu_S}\|_1 - \|\widehat{\nu_{S^c}}\|_1\}$$

since $\lambda_N \geq 2\frac{\|X^Tw\|_\infty}{N}$ by the premise of the theorem. Finally, note that by the additivity of the l_1-norm,

$$\|\widehat{\nu}\|_1 = \|\widehat{\nu_S}\|_1 + \|\widehat{\nu_{S^c}}\|_1$$

yielding

$$\frac{w^TX\widehat{\nu}}{N} + \lambda_N\{\|\widehat{\nu_S}\|_1 - \|\widehat{\nu_{S^c}}\|_1\} \leq \frac{\lambda_N}{2}\{\|\widehat{\nu_S}\|_1 + \|\widehat{\nu_{S^c}}\|_1\} + \lambda_N\{\|\widehat{\nu_S}\|_1 - \|\widehat{\nu_{S^c}}\|_1\}$$

$$= \frac{3\lambda_N}{2}\|\widehat{\nu_S}\|_1 - \frac{\lambda_N}{2}\|\widehat{\nu_{S^c}}\|_1 \leq \frac{3\lambda_N}{2}\|\widehat{\nu_S}\|_1$$

For the final step, let $|S| = k$. Then by the RMS-AM inequality on vectors,

$$\frac{\|\boldsymbol{x}\|_1}{n} \leq \frac{\|\boldsymbol{x}\|_2}{\sqrt{n}}$$

where $\boldsymbol{x} \in \mathbb{R}^n$.
Thus $\|\widehat{\nu_S}\|_1 \leq \sqrt{|S|}\|\widehat{\nu_S}\|_2 \leq \sqrt{|S|}\|\widehat{\nu}\|_2 = \sqrt{k}\|\widehat{\nu}\|_2$. Combining this with the inequality above yields

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

as desired.
(f) We have

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

We also know that

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

since X satisfies the restricted eigenvalue property.
Thus

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

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

$$\implies \|\widehat{\nu}\|_2 \leq \frac{3\lambda_N}{\gamma}\sqrt{k} = \frac{3}{\gamma}\sqrt{\frac{k}{N}}\sqrt{N}\lambda_N$$

Finally, recalling that $\widehat{\nu} := \widehat{\beta} - \beta^*$, we get that

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

as desired.
(g) When we show $\frac{\|X\widehat{\nu}\|_2^2}{2N} \leq \frac{3}{2}\sqrt{k}\lambda_N\|\widehat{\nu}\|_2^2$, we need to use the fact $\frac{1}{N}\|X^T w\|_\infty \leq \frac{\lambda_N}{2}$. It's here that we need to utilise the fact that $\lambda_N \geq 2\frac{\|X^T w\|_\infty}{N}$.
(h) The cone constraint naturally arises when we study the **error vector** $\widehat{\nu}$. In fact, when we define $\widehat{\nu} := \widehat{\beta} - \beta^*$, and try to derive bounds on the l_2-norm of $\widehat{\nu}$ **with suitable constraints on or regularization parameter λ_N, the vector $\widehat{\nu}$ automatically gets constrained within a cone set $\mathcal{C}$.** In fact, as we just proved in the previous parts, when we assumed λ_N to be sufficiently large (as in $\lambda_N \geq 2\frac{\|X^T w\|_\infty}{N}$ for our sensing matrix $\boldsymbol{X}$ and a Gaussian noise $w \sim \mathcal{N}(\mathbf{0}_N, \Sigma)$, where $\Sigma \in \mathbb{R}^{N\times N}$; $\det(\Sigma) = \sigma^2$) we saw that $\widehat{\nu}$ automatically satisfied the cone constraint $\mathcal{C}(S;3)$, ie:-

$$\|\widehat{\nu_{S^c}}\|_1 \leq 3\|\widehat{\nu_S}\|_1 \implies \widehat{\nu} \in \mathcal{C}(S;3)$$

Finally, the cone constraints help in deriving the final forms for the bounds of $\|\nu\|_2$ by upper-bounding variable terms such as $\|X^T w\|$ in terms of expressions in λ_N.
(i)
(j)

3 Problem 3

4 Problem 4

Paper Details	
Title of the Paper	Coastal Acoustic Tomography System and Its Field Application
Link of the paper	**Click Here**
Author List	Haruhiko Yamoaka, Arata Kaneko, Jae-Hun Park, Hong Zheng, Noriaki Gohda, Tadashi Takano, Xiao-Hua Zhu and Yoshio Takasugi
Publication Date	August 2002
Publication Venue	IEEE Journal of Oceanic Engineering, Volume 27, Issue 2

4.1 Introduction and Aim

This paper aims to map the structure of the "strongly nonlinear tidal currents in the coastal sea" by using multiple synchronised coastal acoustic tomography system (CATS). Using GPS clock signals and separate codes to distinguish between signals of individual systems, reconstruction of tidal process behaviour is done through an inverse analysis of the acoustic signals obtained by the sensors.

4.2 Mathematical Formulation

5 Problem 5

We know that Radon Transform is given by-

$$R_\theta(f) = g(\rho,\theta) = \int_{-\infty}^{+\infty} f(\rho\cos\theta - z\sin\theta, \rho\sin\theta + z\cos\theta)dz$$

We can write the same as-

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

Let the scaled image be denoted by $h(x,y) = f(ax,ay)$. This is the same image as original, but scaled by a factor of a, in both x and y directions.

We can write the same Radon Transform as-

$$R_\theta(h) = g'(\rho,\theta) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} h(x,y)\delta(x\cos\theta + y\sin\theta - \rho)dxdy$$

$$R_\theta(h) = g'(\rho,\theta) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(ax,ay)\delta(x\cos\theta + y\sin\theta - \rho)dxdy$$

$$R_\theta(h) = g'(\rho,\theta) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x',y')\delta\left(\frac{x'\cos\theta + y'\sin\theta - a\rho}{a}\right)\frac{dx'}{a}\frac{dy'}{a}$$

Since $\delta(ax) = \delta(x)/a$, we get-

$$R_\theta(h) = g'(\rho,\theta) = \frac{1}{a}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x',y')\delta(x'\cos\theta + y'\sin\theta - a\rho)dx'dy'$$

$$R_\theta(h) = g'(\rho,\theta) = \frac{1}{a}g(a\rho,\theta)$$

Thus, we can see that the Radon transform of the scaled image is also scaled by a factor of a in the size of projection, but the intensity of each projection has reduced by a as well.

6 Problem 6

We know that the Radon Transform is given by-

$$R_\theta(f) = g(\rho, \theta) = \int_{-\infty}^{+\infty} f(\rho\cos\theta - z\sin\theta, \rho\sin\theta + z\cos\theta)dz$$

We can write the same as-

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

Now, let $f(x, y) = \delta(x - x_0, y - y_0)$ for some given constants x_0, y_0. Also, for the sake of simplification, call $\delta(x\cos\theta + y\sin\theta - \rho)$ as $h(x, y, \rho, \theta)$.
Then the Radon transform of our function $f(x, y)$ becomes

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x, y)\delta(x\cos\theta + y\sin\theta - \rho)dxdy$$

$$= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \delta(x - x_0, y - y_0)h(x, y, \rho, \theta)dxdy$$

Now, by a well known property of delta functions, the integration of a function multiplied by the delta function over any space (including the delta function's singularity) yields the evaluation of the function at the singularity point. In the context of our problem, we can state the above property as

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \delta(x - x_0, y - y_0)h(x, y)dxdy = h(x_0, y_0)$$

Applying this property verbatim on our integral above yields

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \delta(x - x_0, y - y_0)h(x, y, \rho, \theta)dxdy$$

$$= h(x_0, y_0, \rho, \theta) = \delta(x_0\cos\theta + y_0\sin\theta - \rho)$$

since ρ, θ are constant parameters within the integration.
Thus the Radon transform of the unit impulse function is another impulse function, ie:-

$$\boldsymbol{R_\theta(\delta(x - x_0, y - y_0))) = g(\rho, \theta) = \delta(x_0\cos\theta + y_0\sin\theta - \rho)}$$

$$\boldsymbol{R_\theta(\delta(x, y))) = g(\rho, \theta) = \delta(-\rho) = \delta(\rho)}$$