CS 754 Assignment 1 : Report

Mls Pragna - 180050064
Rahul Puli - 180050080

Question 1

(1)

We have,

$$||\theta - \theta^*||_2 <= C_1 s^{-1/2} ||\theta - \theta_s|| + C_2 \varepsilon$$

The constants C1 and C2 are increasing functions of δ_{2s}. δ_{2s} is an increasing function of s, this can be explained using the fact that a matrix A that obeys RIP of order s should have 2s linearly independent columns. With same A, the chances of having 2s linearly independent columns decreases with increase of s and therefore the chances of A satisfying RIP of order s also decreases which means δ_{2s} is more close to 1 when s is large. This implies the constants C1 and C2 are increasing with increase of s. Therefore, we cannot say the upper bound $||\theta - \theta^*||_2$ is decreasing with increase of s even though $s^{-1/2}$ is decreasing with increase of s.

(2)

$$||\theta - \theta^*||_2 <= C_1 s^{-1/2} ||\theta - \theta_s|| + C_2 \varepsilon$$

Error bound is indirectly dependent on m. The constants C1 and C2 are dependent on m because they are increasing functions of δ_{2s} and δ_{2s} is dependent on m because if the number of rows (m) in a matrix are high then there are high chances of having 2s linearly independent columns this means the matrix has high chances of obeying RIP order of s which implies δ_{2s} is less if m is high and $A = \Phi\Psi$ will obey RIP order of S with overwhelming probability if $m >= CSlog(n/S)$, therefore C1 and C2 are small when m is large . ε has a lower bound of L2-norm of noise and the noise is dependent on the measures (m) we take of the signal. Generally the noise is uniform random noise with known r then $\varepsilon > r^2 m$. Therefore ε is dependent on m.

(3)

Theorem 3 is more useful than Theorem 3A because to reconstruct the signal with high probability using theorem 3A, δ_{2s} should be less than 0.1 whereas in the theorem 3, δ_{2s} should be less than 0.414. If δ_{2s} is less than 0.414 but greater than 0.1 then according to theorem 3A signal cannot be reconstructed whereas theorem 3 can. More measurements (no.of rows=m) are required for the matrix $A = \Phi\Psi$ to obey the RIP of order s for smaller δ_{2s} this means compression of signal and reconstruction within the error limit is better in theorem 3. The number of ϕ measurement matrices having $\delta_{2s} < 0.414$ are higher than measurement matrices having $\delta_{2s} < 0.1$ therefore reconstruction can be done over more number of matrices by using theorem 3.

(4)

If we are setting ε to zero means that there is noise in the measurement as ε is the higher-bound of L2-norm of noise but practical measurements are always noisy. Obtained θ without considering the noise would give us incorrect reconstructed signal. By setting ε to zero we over-fit θ with noisy measurements and make our estimation vulnerable to noise. Therefore setting ε to zero when error vector has non zero magnitude would give us incorrect results.

Question 2

Upper bound:

As, all rows and columns of Φ and Ψ are unit normalized i.e., $|\Phi^i| = 1$ and $|\Psi_j| = 1$ for all $i\epsilon\{0, 1, 2, ...m-1\}, j\epsilon\{0, 1, 2, ...n-1\}$

By Cauchy–Schwarz inequality,

$$|\Phi^{i^t}\Psi_j| <= ||\Phi^{i^t}|| * ||\Psi_j||$$

$$\Rightarrow |\Phi^{i^t}\Psi_j| <= 1 * 1 = 1$$

$\mu(\Phi, \Psi) = \sqrt{n}.max_{i\epsilon\{0,1,...m-1\},j\epsilon\{0,1,...n-1\}}|\Phi^{i^t}\Psi_j|$
$\Rightarrow \mu(\Phi, \Psi) <= \sqrt{n} * 1 = \sqrt{n}$

Therefore, **upper bound** $= \sqrt{n}$

Lower bound:

Considering a unit vector $g\epsilon R^n$ such that $g = \sum_{k=1}^{n} \alpha_k \Psi_k$

As, Ψ is an orthonormal basis unit normalized matrix,

$$< \Psi^i, \Psi^j >= 0 \ \ \forall i, j\epsilon\{0, 1, 2, ...n-1\}, i \neq j < \Psi^i, \Psi^i >= 1 \ \ \forall i\epsilon\{0, 1, 2, ...n-1\} \tag{1}$$

As, g is a unit vector,

$$\begin{aligned}||g||^2 &= ||\sum_{k=1}^{n} \alpha_k \Psi_k||^2 \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j < \Psi_i, \Psi_j > \\ &= \sum_{j=1}^{n} \alpha_j^2 \qquad (from\ equation(1)) \\ ||g||^2 &= 1 \\ \sum_{j=1}^{n} \alpha_j^2 &= 1\end{aligned}$$

Coherence between g and Ψ,

$$\begin{aligned}
\mu(\Phi,\Psi) =& \sqrt{n}.max_{i\epsilon\{0,1,...m-1\},j\epsilon\{0,1,...n-1\}}|\Phi^{i^t}\Psi_j| \\
\Rightarrow \mu(g,\Psi) =& \sqrt{n}.max_{i\epsilon\{0,1,...m-1\},j\epsilon\{0,1,...n-1\}}|g^{i^t}\Psi_j| \\
=& \sqrt{n}.max_{j\epsilon\{0,1,...n-1\}}|(\sum_{k=1}^{n}\alpha_k\Psi_k^t)\Psi_j| \\
=& \sqrt{n}.max_{j\epsilon\{0,1,...n-1\}}|\sum_{k=1}^{n}\alpha_k < \Psi_k, \Psi_j > | \\
=& \sqrt{n}.max_{j\epsilon\{0,1,...n-1\}}|\alpha_j| \qquad (from\ equation(1)) \\
=& \sqrt{n}.max_{i\epsilon\{0,1,...n-1\}}\frac{|\alpha_i|}{\sum_{j=1}^{n}\alpha_j^2}
\end{aligned}$$

The minimum value of $\mu(g,\Psi)$ can be achieved when all the values of α_i are $\frac{1}{\sqrt{n}}$ and the minumum value is 1 because $\sum_{j=1}^{n}\alpha_j^2 = 1$. For contradiction, if α_i is less than $\frac{1}{\sqrt{n}}$, there exists a α_j such that it is greater than $\frac{1}{\sqrt{n}}$, then $\mu(g,\Psi)$ value will be greater than 1, else α_i is greater than $\frac{1}{\sqrt{n}}$, then $\mu(g,\Psi)$ value will be greater than 1.

Therefore, minimum is achieved when $|\alpha_i| = \frac{1}{\sqrt{n}}\ \forall i\epsilon\{1,2,...,n\}$. This minimum is attained for $g = \frac{1}{\sqrt{n}}\sum_{k=1}^{n}\Psi_k$. At minimum, $\mu(g,\Psi) = \sqrt{n}.max_{i\epsilon\{0,1,...n-1\}}\frac{|\alpha_i|}{\sum_{j=1}^{n}\alpha_j^2} = \frac{1}{\sqrt{n}} * \sqrt{n} = 1$

Therefore, **Lower bound = 1**

Question 3

a)

x cannot be uniquely estimated if m=1 and if it is known that x has only one-zero element because for suppose $y = [\phi_i * x_i]$ (let ith row element in Φ be Φ_i) but y can also be written as $y = [\phi_j * x_j]$ for some $j \neq i$. In first case x can be zero at all positions except at ith position and in second case x can be zero at all positions except jth position. Hence unique estimation is not possible. If the index (let it be i) of the non-zero element of x is also known then x can be uniquely estimated because $y = [\phi_i * x_i]$ is the only possibility therefore x can be uniquely estimated by $\phi_i = y/x_i$.

b)

Let Φ_i be the ith column in Φ and y^T= $[y_1\ y_2]$.
For suppose $y = \Phi_i x_i$(as x has only one non-zero element) then to have unique solution there should not exist j such that $y = \Phi_j x_j$ and $j \neq i$ which means $\Phi_i x_i \neq \Phi_j x_j$ this implies no two columns of measurement matrix should be linearly independent. Such a measurement matrix is possible, for example, $\Phi_i^T = [z_i\ 1]$, all $z_i's$ are different.

$$y_1 = z_i x_i$$
$$y_2 = x_i$$

To get x, first we find the index of non-zero element of x by checking which column of ϕ has elements with ratio$(z_i/1 = z_i)$ $\frac{y_1}{y_2}$. Let's say k is the index, then $x_k = y_2$ and all other $x_i's$ are zero. Multiples columns may have the same ratio if columns are linearly dependent therefore to have unique solution no two columns of measurement matrix should be linearly dependent.

c)

Let Φ_i be the ith column in Φ and $y^T= [y_1\ y_2\ y_3]$.
For suppose x has two non-zero elements at ith and jth positions then $y = \Phi_i x_i + \Phi_j x_j$ $(i \neq j)$(as x has only two non-zero elements) and to have unique solution there should not exist k,l such that $y = \Phi_k x_k + \Phi_l x_l$ and not both $k = i$ and $j = l$ this implies for no i,j,k,l $\Phi_i x_i + \Phi_j x_j = \Phi_k x_k + \Phi_l x_l$. But there exists i,j,k,l such that $\Phi_i x_i + \Phi_j x_j - \Phi_k x_k - \Phi_l x_l = 0$ as Φ_i is a length 3 vector and any four vectors of length three are linearly dependent. Therefore x cannot be uniquely estimated for any measurement matrix as all sets of four vectors of length 3 are linearly dependent.

d)

Let Φ_i be the ith column in Φ and $y^T= [y_1\ y_2\ y_3\ y_4]$.
For suppose x has two non-zero elements at ith and jth positions then,

$$y = \Phi_i x_i + \Phi_j x_j$$

$$\begin{bmatrix} y1 \\ y2 \\ y3 \\ y4 \end{bmatrix} = \Phi_i x_i + \Phi_j x_j$$

$$\begin{bmatrix} y1 \\ y2 \\ y3 \\ y4 \end{bmatrix} = x_i \begin{bmatrix} \phi_{1i} \\ \phi_{2i} \\ \phi_{3i} \\ \phi_{4i} \end{bmatrix} + x_j \begin{bmatrix} \phi_{1j} \\ \phi_{2j} \\ \phi_{3j} \\ \phi_{4j} \end{bmatrix}$$

To have unique there should not exist k,l such that $y = \Phi_k x_k + \Phi_l x_l$ and not both $k = i$ and $j = l$ this implies for no i,j,k,l

$$\Phi_i x_i + \Phi_j x_j = \Phi_k x_k + \Phi_l x_l$$

$$x_i \begin{bmatrix} \phi_{1i} \\ \phi_{2i} \\ \phi_{3i} \\ \phi_{4i} \end{bmatrix} + x_j \begin{bmatrix} \phi_{1j} \\ \phi_{2j} \\ \phi_{3j} \\ \phi_{4j} \end{bmatrix} - x_k \begin{bmatrix} \phi_{1k} \\ \phi_{2k} \\ \phi_{3k} \\ \phi_{4k} \end{bmatrix} - x_l \begin{bmatrix} \phi_{1l} \\ \phi_{2l} \\ \phi_{3l} \\ \phi_{4l} \end{bmatrix} = 0$$

(there should not exist any such (i,j,k,l) for unique solution)

Hence unique solution is possible when no four columns of measurement are linearly dependent. Such a measurement matrix is possible since the number of rows in measurement matrix is four. For example, with $\Phi_i^T = [{z_i}^3\ {z_i}^2\ z_i\ 1]$, all $z_i's$ are different, unique solution is possible because no Φ_i, Φ_j, Φ_k, Φ_l are linearly dependent. For the sake of contradiction assume Φ_i, Φ_j, Φ_k, Φ_l are linearly dependent then the determinant of this matrix $[\Phi_i\ \Phi_j\ \Phi_k\ \Phi_l\ ]$ should be zero. That is,

$$\begin{vmatrix} \phi_{1i} & \phi_{1j} & \phi_{1k} & \phi_{1l} \\ \phi_{2i} & \phi_{2j} & \phi_{2k} & \phi_{2l} \\ \phi_{3i} & \phi_{3j} & \phi_{3k} & \phi_{3l} \\ \phi_{4i} & \phi_{4j} & \phi_{4k} & \phi_{4l} \end{vmatrix} = 0$$

(for linear dependency)

$$\Rightarrow \begin{vmatrix} {z_i}^3 & {z_j}^3 & {z_k}^3 & {z_l}^3 \\ {z_i}^2 & {z_j}^2 & {z_k}^2 & {z_l}^2 \\ z_i & z_j & z_k & z_l \\ 1 & 1 & 1 & 1 \end{vmatrix} = 0$$

$$\Rightarrow (z_i - z_j)(z_i - z_k)(z_i - z_l)(z_j - z_k)(z_j - z_l)(z_k - z_l) = 0$$

Hence contradiction as z_i, z_j, z_k, z_l have different values. Therefore, unique solution is possible with measurement matrix Φ such that $\Phi_i^T = [{z_i}^3 \ {z_i}^2 \ z_i \ 1]$, all $z_i's$ are different.

Question 4

Given,
P1 : Minimize $||x||_1$ w.r.t x such that $||y - Ax||_2 \leq e$.
Q1 : Minimize $||y - Ax||_2$ w.r.t x such that $||x||_1 \leq t$.

Proof:

To prove, if x is a unique minimizer of P1 for some value $e \geq 0$, then there exists some value $t \geq 0$ for which x is also a unique minimizer of Q1

Let, vector k be the unique solution of P1 and $t = ||k||_1$.
As, k is the unique solution to P1, for all x, $x \neq k$, $||k||_1 < ||x||_1$ such that $||y - Ax||_2 \leq e$ and $||y - Ak||_2 \leq e$. This implies by contrapositive, if $||k||_1 \geq ||x||_1$ and $x \neq k$ then $||y - Ax||_2 > e$. Therefore, Minimum value of $||y - Ax||_2$ is e when $||x||_1 \leq ||k||_1$ and the minimum is obtained only when $x = k$ as for all other $||x||_1 \geq ||k||_1$, $||y - Ax||_2$ is greater than e. Hence proved.

Question 5

a)

- **Title of the Paper :** Environmental Monitoring via Compressive Sensing
- **Published place :** SensorKDD '12: Proceedings of the Sixth International Workshop on Knowledge Discovery from Sensor Data
- **Published on :** 12 August 2012
- **Conference Venue :** KDD '12: The 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Beijing, China
- **Link :** `https://dl.acm.org/doi/pdf/10.1145/2350182.2350189`

b)

Networks of spatially distributed autonomous sensors are installed to monitor physical or environmental conditions such as noise pollution, gas emissions from vehicles etc. Expensive detection sensors are used to measure the target environmental substances over a monitored area in environmental monitoring stations. Large scale sensor networks suffer from transmission cost as it is impossible for all passive sensors to be kept alive and transmitting data to central node all the time. In this paper, only $M << N$ sensors are randomly deployed in the monitoring area to get the N dimensional environmental signal through compressive sensing reconstruction methods. Readings and location information of these sensors are collected to generate a sparse measurement vector $Y^T = [y_1, y_2, ...y_M]$ with their coordinates $C^T = [c_1, c_2, ...c_M]$.

c)

Three reconstruction methods are used and compared their performances. These three reconstruction methods are l_1-approch, Greedy algorithm and Bayesian Compressive sensing.

$$y = \phi x = \Phi\Psi w = \Theta w$$

where w is a S-sparse, basis vector Φ is Gaussian kernel. Based on the coordinate matrix C, ϕ is defined as $\phi_i = 1$ if i = c_j else 0 where i = 1,2,..,N and j =1,2,....,M.

l_1-approach:

$$\min \|w\|_1 : \text{y} = \Theta w$$

BP can be casted as Linear programming problem and splev in polynomial time.
Iterative thresholding algorithm, Lasso can be used,

$$\min \tfrac{1}{2}\|\Theta w - y\|_2^2 + \mu\|w\|_1$$

Greedy Algorithm:

In greedy algorithms, current sparse vector w is refined iteratively by selecting columns of Θ according to their correlation with y determined by inner product.

Bayesian Compressive Sensing:

BCS recovery method uses the knowledge of sparse bayesian compressive learning and RVM. It can provide a sparse solution even the measurement matrix doesnot obey RIP property. BCS converts the CS problem into a linear regression problem with a constraint that w is sparse. Under the common conditions of zero-mean Gaussian noise, Guassian likelihood model:

$$p(y|w,\sigma^2) = (2\pi\sigma^2)^{\frac{-M}{2}} exp(-\frac{1}{2\sigma^2}\|y - \Theta w\|^2)$$

It has been observed that recovery error rates using recovery methods l_1-norm and greedy algorithms are extremely large and these both methods fail to recover the signal. The reason is that the both methods depend on the assumption that CSmatrix satisies RIP. Whereas the BCS method guarantes the recovery without requiring CS matrix to satisfy RIP.