Advanced Image Processing HW2

Anilesh Bansal (22b0928) Parth Pujari (210100106)

1 Question 1

(a) Let Φ_i represent the i-th column of Φ. Then
trace$(D1) = \sum_{i=1}^{n} \Phi_i^t \Phi_i = n$ since all columns are unit-normalized.

(b) Consider the m^2 dimension vector $\mathbf{u}$ with it's terms as the entries of matrix $D2$. Let $\mathbf{v}$ be the m^2 dimensional vector corresponding to identity matrix I_m. Notice that

$$\langle u, u \rangle = trace(D2D2^T)$$

$$\langle v, v \rangle = m$$

$$\langle u, v \rangle = trace(D2)$$

Using Cauchy Shwarz Inequality, we get

$$\langle u, v \rangle^2 \leq \langle u, u \rangle \langle u, v \rangle$$

$$\implies trace(D2) \leq \sqrt{m}\sqrt{trace(D2D2^T)}$$

(c)

$$D2 = \Phi\Phi^t = \sum_{i=1}^{n} \Phi_i \Phi_i^t$$

Since $D2^t = D2$, $D2D2^t = \left(\sum_{i=1}^{n} \Phi_i \Phi_i^t\right)\left(\sum_{j=1}^{n} \Phi_j \Phi_j^t\right)$

$$D2D2^t = \sum_{i,j} (\Phi_i^t \Phi_j)(\Phi_i \Phi_j^t)$$

$$trace(D2D2^t) = \sum_{i,j} (\Phi_i^t \Phi_j) trace(\Phi_i \Phi_j^t) = \sum_{i,j} (\Phi_i^t, \Phi_j)^2$$

Since trace$(\Phi_i, \Phi_j^t) = \Phi_i^t \Phi_j$. Now as $\Phi_i^t \Phi_i = 1$ for all i, we get that

$$trace(D2D2^t) = n + \sum_{i,j;i\neq j} (\Phi_i^t, \Phi_j)^2$$

(d)

$$\begin{aligned} trace(D2) &= trace\left(\sum_{i=1}^{n} \Phi_i \Phi_i^t\right) \\ &= \sum_{i=1}^{n} trace\left(\Phi_i \Phi_i^t\right) \\ &= n \end{aligned}$$

Again since trace$(\Phi_i, \Phi_j^t) = \Phi_i^t \Phi_j = 1$ if $i = j$ Substituting we get the required result.

(e) Since mutual coherence is the maximum value of $\Phi_i^t \Phi_j$ for all $i, j; i \neq j$, we get that

$$n^2 \leq m \left(n + \sum_{i,j;i\neq j} (\Phi_i^t, \Phi_j)^2\right) \leq m(n + (n^2 - n)\mu^2)$$

$$n^2 \leq mn(1 + (n-1)\mu^2) \implies \mu \geq \sqrt{\frac{n-m}{m(n-1)}}$$

(f) Equality is achieved in Cauchy-Shwarz when vectorized D2 is proportional to I_m. Since the i,jth term of D2 is dot product of row i and row j of Φ, we get that the rows of Φ must be orthogonal to each other and have same L2 norm.

Further for the mutual coherence, equality is attained when all columns have $\Phi_i^t \Phi_j = \mu$. Thus we want both the conditions to exist simultaneously for the bound to be strict.

Specifically if n doesn't divide m, vectors corresponding to vertices of a n-regular polygon in m-dimensional space achieves the bound, eg. Tetrahedron in 3D.

2 Question 2

We consider Theorem 1 and its proof in Appendix A. We use some standard inequalities, namely:

$$\text{Triangle inequality 1) } ||x+y|| \leq ||x|| + ||y|| \tag{1}$$

$$\text{Triangle inequality 2) } ||x-y|| \geq |\,||x|| - ||y||\,| \tag{2}$$

$$\text{Caucy Schwarz' ineq } |<x,y>| \leq ||x|| \cdot ||y|| \tag{3}$$

We begin with $\mathbf{h} = \hat{\mathbf{x}} - \mathbf{x}$. $\mathbf{h}_0 = \mathbf{P}_{\mathcal{X}}\mathbf{h}$ with $\mathcal{X} = \text{supp}_{n_x}\mathbf{x}$.

2.1 A.1.1. Cone Constraint

Let $e_0 = 2||\mathbf{x} - \mathbf{x}_{\mathcal{X}}||_1$ with $\mathbf{x}_{\mathcal{X}} = \mathbf{P}_{\mathcal{X}}\mathbf{x}$; then

$||\mathbf{h} - \mathbf{h}_0||_1 \leq ||\mathbf{h}_0||_1 + e_0$. This is because:
$||\mathbf{x}||_1 \geq ||\hat{\mathbf{x}}||_1$ because both $\mathbf{x}$ and $\hat{\mathbf{x}}$ satisfy $||\mathbf{z} - \mathbf{A}\tilde{\mathbf{x}}||_2 \leq \eta$ but $\hat{\mathbf{x}}$ minimizes L1 norm.
$||\mathbf{x}||_1 \geq ||\hat{\mathbf{x}}||_1 = ||\hat{\mathbf{x}}_{\mathcal{X}}||_1 + ||\hat{\mathbf{x}}_{\mathcal{X}^c}||_1 = ||\mathbf{x}_{\mathcal{X}} + \mathbf{h}_0||_1 + ||\mathbf{h} - \mathbf{h}_0 + \mathbf{x}_{\mathcal{X}^c}||_1$
Using (2) on both terms (for eg. $||\mathbf{x}_{\mathcal{X}} + \mathbf{h}_0||_1 = ||\mathbf{x}_{\mathcal{X}} - (-\mathbf{h}_0)||_1 \geq ||\mathbf{x}_{\mathcal{X}}||_1 - ||\mathbf{h}_0||_1$) we have:
$||\mathbf{x}||_1 \geq ||\mathbf{x}_{\mathcal{X}}||_1 - ||\mathbf{h}_0||_1 + ||\mathbf{h} - \mathbf{h}_0||_1 - ||\mathbf{x}_{\mathcal{X}^c}||_1$ and on rearranging,

$||\mathbf{x}||_1 - ||\mathbf{x}_{\mathcal{X}}||_1 + ||\mathbf{x}_{\mathcal{X}^c}||_1 + ||\mathbf{h}_0||_1 \geq +||\mathbf{h} - \mathbf{h}_0||_1$
But by (2), $||\mathbf{x}||_1 - ||\mathbf{x}_{\mathcal{X}}||_1 \leq ||\mathbf{x} - \mathbf{x}_{\mathcal{X}}||_1$. Further, $\mathbf{x} - \mathbf{x}_{\mathcal{X}} = \mathbf{x}_{\mathcal{X}^c}$. Thus we have:

$2||\mathbf{x} - \mathbf{x}_{\mathcal{X}}||_1 + ||\mathbf{h}_0||_1 \geq ||\mathbf{h} - \mathbf{h}_0||_1$ and since $2||\mathbf{x} - \mathbf{x}_{\mathcal{X}}||_1 = e_0$, we have

$$||\mathbf{h} - \mathbf{h}_0||_1 \leq ||\mathbf{h}_0||_1 + e_0$$

Further, applying (2) to LHS we have $||\mathbf{h} - \mathbf{h}_0||_1 \geq ||\mathbf{h}||_1 - ||\mathbf{h}_0||_1$ and hence

$$||\mathbf{h}||_1 \leq 2||\mathbf{h}_0||_1 + e_0 \tag{4}$$

2.2 A.1.2 Tube Constraint

Now, given that the signal is noisy, under the constraint that $||\mathbf{A}\mathbf{x} - \mathbf{y}|| \leq \varepsilon$ and that $\hat{\mathbf{x}}$ follows the constraint for η we have:

$$||\mathbf{A}\mathbf{h}||_2 \leq ||\mathbf{A}\hat{\mathbf{x}} - \mathbf{y}||_2 + ||\mathbf{A}\mathbf{x} - \mathbf{y}||_2 \leq \eta + \varepsilon$$

2.3 A.1.3 Coherence based RIP

For any s-sparse vector x,

$$Ax = A_S x_S$$

where S is the set of indices with non zero value and $|S| \leq s$

Consider the matrix $B = A_S^T A_S$, where $b_{ij} = < a_i, a_j >$ and $b_{ii} = ||a_i^2|| = 1$ due to normalization of A.

By Gershgorin's Disk Theorem, every eigenvalue of the matrix $A_S^T A_S$ lies in the the union of $D(b_{ii}, r_i)$ where r_i is sum of absolute values of off-diagonal elements of the i-th row.

$$r_i = \sum_{j=1, j\neq i}^{s} < a_i, a_j > \leq \mu(s-1)$$

Since the matrix $A_S^T A_S$ is positive semidefinite, it has positive real eigenvalues and they lie in the range $[1-\mu(s-1), 1+\mu(s-1)]$

Thus for every s-sparse vector x, we can say that $(1-\mu(s-1))||x||^2 \leq ||Ax||^2 \leq (1+\mu(s-1))||x||^2$

Applying to $n_{\mathcal{X}}$-sparse vector $\mathbf{h}_0$, we have

$$(1-\mu_a(n_{\mathcal{X}}-1))||\mathbf{h}_0||^2 \leq ||\mathbf{A}\mathbf{h}_0||^2 \leq (1+\mu_a(n_{\mathcal{X}}-1))||\mathbf{h}_0||^2 \tag{5}$$

2.4 A.2. Bounding the error $||\mathbf{h}_0||_2$ on support

$$\begin{aligned}
|\mathbf{h}^H \mathbf{A}^H \mathbf{A}\mathbf{h}_0| &= |(\mathbf{h} - \mathbf{h}_0 + \mathbf{h}_0)^H \mathbf{A}^H \mathbf{A}\mathbf{h}_0| = |(\mathbf{h} - \mathbf{h}_0)^H \mathbf{A}^H \mathbf{A}\mathbf{h}_0 + \mathbf{h}_0^H \mathbf{A}^H \mathbf{A}\mathbf{h}_0| \\
&\geq |\mathbf{h}_0^H \mathbf{A}^H \mathbf{A}\mathbf{h}_0| - |(\mathbf{h} - \mathbf{h}_0)^H \mathbf{A}^H \mathbf{A}\mathbf{h}_0| \qquad \text{...Using (2)} \\
&\geq (1-\mu_a(n_{\mathcal{X}}-1))||\mathbf{h}_0||^2 - \left| \sum_{k\in\mathcal{X}} \sum_{l\in\mathcal{X}^c} \left[\mathbf{h}_0^H\right]_k \mathbf{a}_k^H \mathbf{a}_l [\mathbf{h}]_l \right| \qquad \text{... Using (5) and expanding matrix product}
\end{aligned}$$

But we have $\mathbf{a}_k^H \mathbf{a}_l \leq \mu_a \ \forall k \neq l$. Hence:

$$\begin{aligned}
\left| \sum_{k\in\mathcal{X}} \sum_{l\in\mathcal{X}^c} \left[\mathbf{h}_0^H\right]_k \mathbf{a}_k^H \mathbf{a}_l [\mathbf{h}]_l \right| &\leq \mu_a \left| \sum_{k\in\mathcal{X}} \sum_{l\in\mathcal{X}^c} \left[\mathbf{h}_0^H\right]_k [\mathbf{h}]_l \right| \\
&= \mu_a \left| \sum_{k} \sum_{l} \left[\mathbf{h}_0^H\right]_k [\mathbf{h} - \mathbf{h}_0]_l \right| \\
&= \mu_a | < \mathbf{h}_0, \mathbf{h} - \mathbf{h}_0 > | \\
&\leq \mu_a ||\mathbf{h}_0||_1 \cdot ||\mathbf{h} - \mathbf{h}_0||_1 \qquad \text{...Cauchy Schwarz}
\end{aligned}$$

Plugging this into the main inequality we have:

$$
\begin{aligned}
|\mathbf{h}^H\mathbf{A}^H\mathbf{A}\mathbf{h}_0| &\geq (1-\mu_a(n_{\mathcal{X}}-1))||\mathbf{h}_0||^2 - \mu_a||\mathbf{h}_0||_1 \cdot ||\mathbf{h}-\mathbf{h}_0||_1 \\
&\geq (1-\mu_a(n_{\mathcal{X}}-1))||\mathbf{h}_0||^2 - \mu_a||\mathbf{h}_0||_1 \cdot (||\mathbf{h}_0||_1 + e_0) && \text{... Using (4)} \\
&\geq (1-\mu_a(n_{\mathcal{X}}-1))||\mathbf{h}_0||^2 - \mu_a n_{\mathcal{X}}||\mathbf{h}_0||_2^2 - \mu_a\sqrt{n_{\mathcal{X}}}||\mathbf{h}_0||_2 e_0 && \text{... Using } ||x||_1 \leq \sqrt{n_{\mathcal{X}}}||x||_2 \\
&= (1-\mu_a(2n_{\mathcal{X}}-1))||\mathbf{h}_0||^2 - \mu_a\sqrt{n_{\mathcal{X}}}||\mathbf{h}_0||_2 e_0
\end{aligned}
$$

Rearranging terms we get:

$$
\begin{aligned}
||\mathbf{h}_0||_2 &\leq \frac{|\mathbf{h}^H\mathbf{A}^H\mathbf{A}\mathbf{h}_0| + \mu_a\sqrt{n_{\mathcal{X}}}||\mathbf{h}_0||_2 e_0}{(1-\mu_a(2n_{\mathcal{X}}-1))||\mathbf{h}_0||_2} \\
&\leq \frac{|\mathbf{A}\mathbf{h}||_2||\mathbf{A}\mathbf{h}_0||_2 + \mu_a\sqrt{n_{\mathcal{X}}}||\mathbf{h}_0||_2 e_0}{(1-\mu_a(2n_{\mathcal{X}}-1))||\mathbf{h}_0||_2} && \text{... Using Cauchy Schwarz} \\
&\leq \frac{(\eta+\varepsilon)\sqrt{1+\mu_a(n_{\mathcal{X}}-1))}||\mathbf{h}_0||_2 + \mu_a\sqrt{n_{\mathcal{X}}}||\mathbf{h}_0||_2 e_0}{(1-\mu_a(2n_{\mathcal{X}}-1))||\mathbf{h}_0||_2} && \text{... Using Tube constraint and (5)} \\
&\leq \frac{(\eta+\varepsilon)\sqrt{1+\mu_a(n_{\mathcal{X}}-1))} + \mu_a\sqrt{n_{\mathcal{X}}}e_0}{1-\mu_a(2n_{\mathcal{X}}-1)} && \text{... Since } ||\mathbf{h}_0||_2 \neq 0 \text{ else the bound is trivial}
\end{aligned} \tag{6}
$$

2.5 A.3. Bounding the recovery error $||\mathbf{h}||_2$

$$
\begin{aligned}
||\mathbf{A}\mathbf{h}||_2^2 &= \sum_{k,l} \left[\mathbf{h}^H\right]_k \mathbf{a}_k^H\mathbf{a}_l[\mathbf{h}]_l \\
&= \sum_k ||\mathbf{a}_k||_2^2|[\mathbf{h}]_k|^2 + \sum_{k,l,k\neq l}\left[\mathbf{h}^H\right]_k \mathbf{a}_k^H\mathbf{a}_l[\mathbf{h}]_l \\
&\geq ||\mathbf{h}||_2^2 - \mu_a \sum_{k,l,k\neq l}\left[\mathbf{h}^H\right]_k [\mathbf{h}]_l && \text{... Since } ||\mathbf{a}_k||_2 = 1\ \forall k \text{ and } |\mathbf{a}_k^H\mathbf{a}_l| \leq \mu_a\ \forall k\neq l \\
&= ||\mathbf{h}||_2^2 + \mu_a\sum_k |[\mathbf{h}]_k|^2 - \mu_a\sum_{k,l}\left[\mathbf{h}^H\right]_k[\mathbf{h}]_l && \text{... Adding and subtracting terms where } l=k \\
&= (1+\mu_a)||\mathbf{h}||_2^2 - \mu_a||\mathbf{h}||_1^2 && \text{... } \sum_k |[\mathbf{h}]_k|^2 = ||\mathbf{h}||_2^2\ ;\ \sum_{k,l}\left[\mathbf{h}^H\right]_k[\mathbf{h}]_l = \sum_k\left[\mathbf{h}^H\right]_k^2 + 2\sum_{k\neq l}\left[\mathbf{h}^H\right]_k[\mathbf{h}]_l = ||\mathbf{h}||_1^2
\end{aligned}
$$

Rearranging we get;

$$
\begin{aligned}
||\mathbf{h}||_2^2 &\leq \frac{||\mathbf{A}\mathbf{h}||_2^2 + \mu_a||\mathbf{h}||_1^2}{1+\mu_a} \\
&\leq \frac{(\varepsilon+\eta)^2 + \mu_a(2||\mathbf{h}_0||_1+e_0)^2}{1+\mu_a} && \text{... Using Tube constraint and (4)} \\
||\mathbf{h}||_2 &\leq \frac{\sqrt{(\varepsilon+\eta)^2 + \mu_a(2||\mathbf{h}_0||_1+e_0)^2}}{\sqrt{1+\mu_a}} \\
&\leq \frac{(\varepsilon+\eta) + \sqrt{\mu_a}(2||\mathbf{h}_0||_1+e_0)^2}{\sqrt{1+\mu_a}} && \text{... Cauchy Schwarz} \\
&\leq \frac{(\varepsilon+\eta) + \sqrt{\mu_a}(2\sqrt{n_{\mathcal{X}}}||\mathbf{h}_0||_1+e_0)^2}{\sqrt{1+\mu_a}} && \text{...} ||x||_1 \leq \sqrt{n}||x||_2
\end{aligned}
$$

Now, using (6), substituting $||\mathbf{h}_0||_2$ by the RHS we get;

$$
\begin{aligned}
||\mathbf{h}||_2 \leq (\varepsilon+\eta)&\frac{1-\mu_a(2n_{\mathcal{X}}-1)+2\sqrt{\mu_a n_{\mathcal{X}}}\sqrt{1+\mu_a(n_{\mathcal{X}}-1)}}{\sqrt{1+\mu_a}(1-\mu_a(2n_{\mathcal{X}}-1)} \\
&+ e_0\frac{2\mu_a\sqrt{\mu_a}n_{\mathcal{X}}+1-\mu_a(2n_{\mathcal{X}}-1)}{\sqrt{1+\mu_a}(1-\mu_a(2n_{\mathcal{X}}-1))} \\
&\leq (\varepsilon+\eta)\frac{1-\mu_a(2n_{\mathcal{X}}-1)+2\sqrt{\mu_a n_{\mathcal{X}}}\sqrt{1+\mu_a(n_{\mathcal{X}}-1)}}{\sqrt{1+\mu_a}(1-\mu_a(2n_{\mathcal{X}}-1)} + e_0\frac{\sqrt{\mu_a+\mu_a^2}}{(1-\mu_a(2n_{\mathcal{X}}-1))} \qquad \text{... Using } n_{\mathcal{X}} < (1+1/\mu_a)/2 \\
&= C_0(\eta+\varepsilon)+C_1||\mathbf{x}-\mathbf{x}_{\mathcal{X}}||_1
\end{aligned}
$$

The condition on $n_{\mathcal{X}}$ is imposed so that the solution to the minimizing equation has a unique solution.

3 Question 3

Original Images:

Figure 1: Original Barbara

Figure 2: Original Goldhill

Results for ISTA:
Barbara (RMSE = 0.0788 on only 60 iterations)
Goldhill (RMSE = 0.0803)

Figure 3: ISTA Barbara reconstruction

Figure 4: ISTA Goldhill reconstruction

Using**wavelet basis**:
We implemented a second function, using **function handles** for the discrete wavelet transform.

Barbara (RMSE = 0.1278 on 60 iterations)
Goldhill (RMSE = 0.2277)

Figure 5: ISTA (wavelet) Barbara reconstruction

Figure 6: ISTA (wavelet) Goldhill reconstruction

4 Question 4

Paper : **Compressed sensing for aperture synthesis imaging** (17th IEEE International Conference)

Compressive sensing has direct use in interferometric aperture synthesis in astronomy. Radio observations requires a large diameter telescope for adequate resolution, of the order of kms. Since building such a big telescope is impossible, instead Radio telescope arrays consist of several antennas that pick up electromagnetic waves from astronomical sources, and the correlation of the signals at each pair of antennas, termed visibility, cor responds to a single complex entry in the Fourier transform of the image of the source.

Here the sensing matrix depends on every pair of telescope arrays, and each entry corresponds to a entry in the u-v plane of Fourier transform called the baseline. As time passes, the orientation of telescope w.r.t source changes which results in newer entries. This is called **Point Spread Function**.

Sparse Basis is the **Fourier Basis**, which like many other practical applications results in sparse vectors for the kind of images to be reconstructed.

$$y = MSx$$

where M is the measurement matrix in Fourier domain and S is the matrix corresponding to Fourier Basis. The reconstruction is done by solving $\hat{x}$ **= argmin** $||x||_1$ **subject to the constraint** $||y - MSx||_2 \leq \eta$ where y is the measurements and η is a parameter based on the expected noise in signal accumulation and processing. This is solved using the **SparseRI algorithm**.

5 Question 5

Although theoretically an iid Gaussian random matrix can be used for constructing frames in an image using the theory discussed in class with m = number of pixels and $n = Nm$ where N = number of frames to be reconstructed, creating hardware that does this is not possible.

In video compressed sensing architecture of Hitomi, a pixel is switched on for some time Δt in between, which is a multiple of $1/N$. Pixel being 'on' corresponds to 1 in the sensing matrix and 0 otherwise. Thus the sensing matrix consists of entries from $\{0, 1\}$ only. Constructing hardware for a sensing matrix where entries are sampled from $\mathcal{N}(0, 1)$ in Hitomi's video CS architecture is not possible.