CS 754 - Advanced Image Processing
Assignment 1 - Report

Shaan ul Haque - 180070053
Mantri Krishna Sri Ipsit - 180070032

February 7, 2021

Question 1

1.a

We know that δ_{2s} is an increasing function of s(sparsity of signal). If we increase s, we are increasing δ_{2s} which in-turn increases C_1 and C_2 because they are increasing functions of δ_{2s} as well. Thus, both have conflicting effects so we can't assume the error bound to arbitrarily decrease as s increases.

1.b

Though, m(number of measurements) has no direct effect on the bound but it does have few implications on it. The measurement matrix has only m rows. If the sparsity of signal is s which is greater than m then we have:

$$\boldsymbol{y} = \sum_{i \epsilon \boldsymbol{S}} \boldsymbol{A_i} x_i$$

where $\boldsymbol{S}$ is the set of index where $\boldsymbol{x}$ has non-zero elements and $\boldsymbol{A_i}$ is the corresponding column. Now since s(cardinality of $\boldsymbol{S}$) is greater than m thus it is linear combination of s m-dimensional vectors which is clearly zero for some $\boldsymbol{x}$ as infinite solutions exists for such under-determined problem. Thus, the null set property of the matrix and hence RIP is lost. Thus, we must have m¿s, as a necessary condition for RIP. Infact $O(slog(n/s)) \le m$, and thus the value of s cannot be increased beyond the number of rows in $\boldsymbol{A}$.

1.c

Theorem 3 is better than theorem 3A. This is because a matrix obeying RIP for smaller bound on δ_{2s} would work fine on signals having more sparsity(smaller values of s) than a matrix whose RIC is larger. Thus, for larger δ_{2s} we can work with comparatively less sparse signals also and satisfy the BP error limit.

1.d

If we make the error ϵ zero in BP, then we want exact relation between y and θ which is:

$$y = \mathbf{\Phi\Psi\theta}$$

But since the measurements are noisy in, the estimated θ from solving such problem would be incorrect. Thus, we allow some margin so that we do not over-fit our estimated θ with noisy measurements. Moreover, we see that ϵ is not some independent parameter but it depends upon the noise power, therefore it is quite unreasonable to take it as zero if noise vector is non-zero in magnitude.

Question 2

2.a

The first $T = 3$ frames of the 'cars' video are as follows:

Figure 1: $T = 1$ Figure 2: $T = 2$ Figure 3: $T = 3$

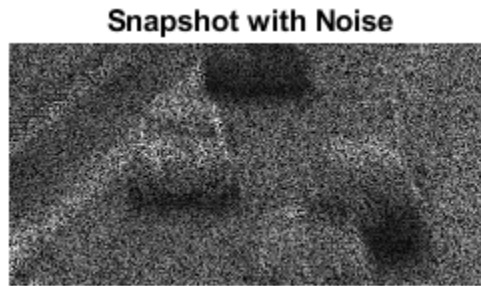## 2.b

The coded snapshot for $T = 3$ is as follows:

Figure 4: Coded Snapshot along with Gaussian Noise with standard deviation 2

2.c

We have

$$E_u = \sum_{t=1}^{T} C_t \cdot F_t$$

Rewriting it as a matrix product and vectorizing the coded snapshot and the image to be measured, we get

$$\mathbf{b} = \mathbf{A}\,\mathbf{x}$$

where

$$b = \text{Vectorized } E_u \in \mathbb{R}^{HW}$$

$$x = \text{Vectorized } F_u \in \mathbb{R}^{HWT}$$

$$A = [S_1|S_2|\cdots|S_T], S_t = \text{diag}(C_t(:)) \in \mathbb{R}^{HW \times HWT}$$

2.d

Rewriting the above equation patchwise, we have

$$\mathbf{A}\,\mathbf{x} = \mathbf{b}$$

where

$$A = \Phi\Psi \in \mathbb{R}^{64 \times 64T}$$

$$\Phi = [S_{i1}|S_{i2}|\cdots|S_{iT}], S_{it} = \text{diag}(C_{it}(:)) \in \mathbb{R}^{64 \times 64T}$$

$$\Psi = I_T \otimes \psi_{64,64} \otimes \psi_{64,64}$$

$$\psi_n = \text{2D DCT Basis Matrix of size } n \times n$$

$$x \in \mathbb{R}^{64T}, \text{ sparse}$$

$$b \in \mathbb{R}^{64}$$

Let θ^* be the optimal solution of the problem P0: $\min \|\theta\|_0$ such that $\|y - A\theta\|_2^2 \le \epsilon, A = \Phi\Psi$. Let the solution obtained using OMP be $\hat{\theta}$. Then we have the error as

$$
\begin{aligned}
E &= \left\|\theta^* - \hat{\theta}\right\|^2 \\
&= (\theta^* - \hat{\theta})^T (\theta^* - \hat{\theta}) \\
&= (A^T A)^{-1}(A\theta^* - A\hat{\theta})^T (A\theta^* - A\hat{\theta}) \\
&= (A^T A)^{-1}\left((y - A\hat{\theta}) - (y - A\theta^*)\right)^T \left((y - A\hat{\theta}) - (y - A\theta^*)\right) \\
&= (A^T A)^{-1}\, 4\epsilon \\
&= \frac{4\epsilon}{\|A\|^2}
\end{aligned}
$$

2.e

For $T = 3$, the relative mean squared error that we obtained is 0.0112. The reconstruction for $T = 3$ is as follows:

Figure 5: $t = 1$, Left - Original, Right - Reconstructed

Figure 6: $t = 2$, Left - Original, Right - Reconstructed

Figure 7: $t = 3$, Left - Original, Right - Reconstructed

2.f

The relative mean squared error value for $T = 5$ is 0.0194 and for $T = 7$ is 0.0316.

Please turn over

Snapshot with Noise

Figure 8: Coded snapshot with noise for $T = 5$

T = 1

Figure 9: $t = 1$

T = 2

Figure 10: $t = 2$

T = 3

Figure 11: $t = 3$

T = 4

Figure 12: $t = 4$

T = 5

Figure 13: $t = 5$

Figure 14: Left - Original, Right - Reconstructed

Please turn over

Snapshot with Noise

Figure 15: Coded snapshot with noise for $T = 7$

Figure 16: $t = 1$

Figure 17: $t = 2$

Figure 18: $t = 3$

Figure 19: $t = 4$

Figure 20: $t = 5$

Figure 21: $t = 6$

Figure 22: $t = 7$

Figure 23: Left - Original, Right - Reconstructed

2.g

We have extracted a portion of about 120×240 around the lower right corner of every frame to save time. Hence, all our result correspond to this part.

2.h

We have used the first 5 frames of the 'flame' video to do the experiment. We got an relative mean squared error of 0.0010.
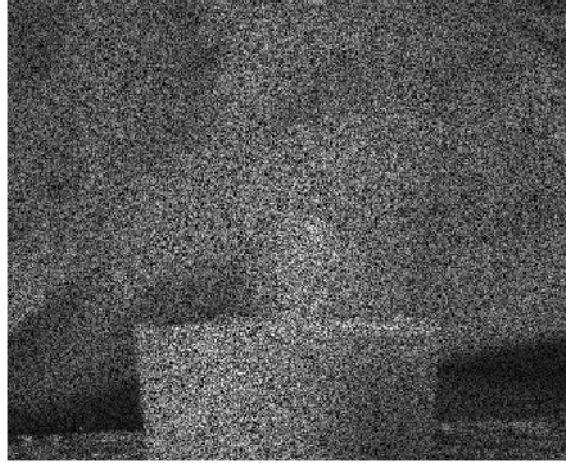

Figure 24: Coded snapshot for flame with noise for $T = 5$

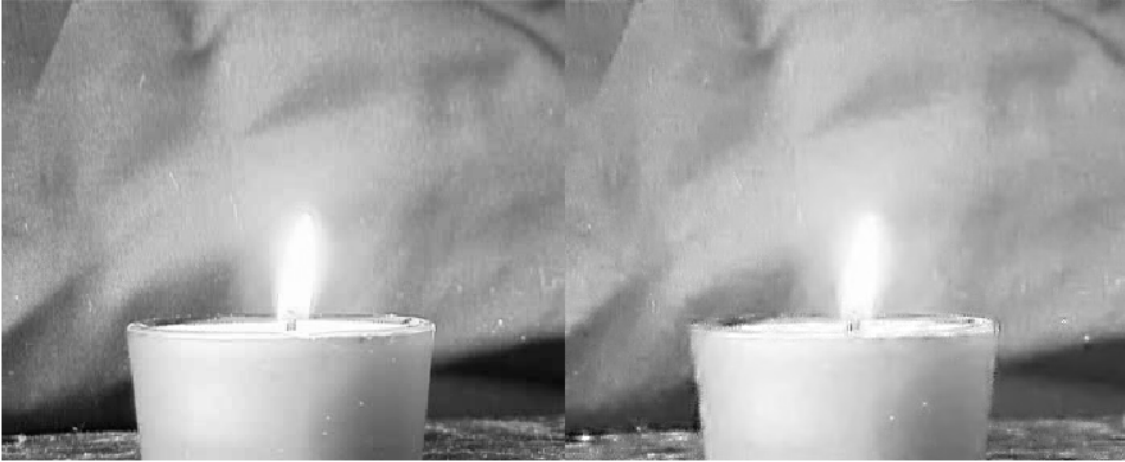Figure 25: $t = 1$, Left - Original, Right - Reconstructed

T = 2

Figure 26: $t = 2$, Left - Original, Right - Reconstructed

T = 3

Figure 27: $t = 3$, Left - Original, Right - Reconstructed

Figure 28: $t = 4$, Left - Original, Right - Reconstructed

Figure 29: $t = 5$, Left - Original, Right - Reconstructed

Question 3

For upper limit of the bound, we can use Cauchy-Schwartz inequality:

$$|\mathbf{\Phi^i \Psi_j}| \leq |\mathbf{\Phi^i}||\mathbf{\Psi_j}|$$

which is simply dot product of two vectors is always less than product of their magnitude. Since $\mathbf{\Psi}$ and $\mathbf{\Phi}$ are unit normalized along columns and rows respectively we have:

$$\max_{i,j \in \{1,2,3...,n\}} |\mathbf{\Phi^i \Psi_j}| \leq 1$$

And thus:

$$\max_{i,j \in \{1,2,3...,n\}} \sqrt{n}|\mathbf{\Phi^i \Psi_j}| \leq \sqrt{n}$$

For lower limit of the bound, we use the hint by creating a unit vector:

$$\mathbf{g} = \sum_{k=1}^{n} \alpha_k \mathbf{\Psi_k}$$

Since $\boldsymbol{g}$ is a unit vector we have:

$$\boldsymbol{g}^T\boldsymbol{g} = (\sum_{k=1}^{n} \alpha_k \boldsymbol{\Psi}_k^T)(\sum_{k=1}^{n} \alpha_k \boldsymbol{\Psi}_k) = 1$$

Now, since $\boldsymbol{\Psi}$ is orthonormal, dot product of columns of $\boldsymbol{\Psi}$ is zero when columns are different and equal to 1 if they are same. Thus we get the condition:

$$\sum_{k=1}^{n} \alpha_k^2 = 1$$

Now coherence between two matrices $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$ is defined as the maximum value of dot product between row vectors of $\boldsymbol{\Phi}$ and column vectors of $\boldsymbol{\Psi}$. If we take $\boldsymbol{\Phi}$ to be a vector we need maximum dot product of the vector with the columns of $\boldsymbol{\Psi}$ which in other words is:

$$\mu(\boldsymbol{g}, \boldsymbol{\Psi}) = \max_{j\epsilon\{1,2,3...,n\}} \sqrt{n} \frac{|\boldsymbol{g}^T\boldsymbol{\Psi}_j|}{||\boldsymbol{g}||_2 ||\boldsymbol{\Psi}_j||_2}$$

Since $\boldsymbol{g}$ is a linear combination of columns of $\boldsymbol{\Psi}$ which is orthonormal also, we get:

$$\mu(\boldsymbol{g}, \boldsymbol{\Psi}) = \max_{j\epsilon\{1,2,3...,n\}} \sqrt{n} \frac{|\alpha_j|}{||\boldsymbol{g}||_2}$$

The equation finally reduces to:

$$\mu(\boldsymbol{g}, \boldsymbol{\Psi}) = \max_{j\epsilon\{1,2,3...,n\}} \sqrt{n} \frac{|\alpha_j|}{\sum_{k=1}^{n} \alpha_k^2} = \max_{j\epsilon\{1,2,3...,n\}} \sqrt{n}|\alpha_j|$$

The minimum value of the above expression can only be achieved when all the elements are equal which gives the lower limit of coherence:

$$n\alpha^2 = 1 \implies |\alpha| = 1/\sqrt{n}$$

This is very clear to see, let us assume if we decrease one of the parameters from $1/\sqrt{n}$ a little bit then other parameters will have to increase in order to satisfy the unit magnitude constraint and when max is taken across all elements then we will get some value greater than $1/\sqrt{n}$, whereas if we increase one of the parameters then others will have to decrease and max across all of them will again increase. Thus the minimum value is $\mu(\boldsymbol{g}, \boldsymbol{\Psi}) = 1$ when, $\boldsymbol{g} = 1/\sqrt{n} \sum_{k=1}^{n} \boldsymbol{\Psi}_k$

Question 4

4.a

If $\boldsymbol{x}$ is 1-sparse signal and only one measurement is taken then we won't be able to estimate $\boldsymbol{x}$ as we don't know which index element was non-zero. Whereas, if the index is known then the measurement can be divided by the value at the column of the measurement vector($n \times 1$, as only one measurement is taken) corresponding to given index to uniquely determine $\boldsymbol{x}$.

4.b

If two measurements are taken then also uniqueness is not guaranteed. Let us say some j-th index of x was non-zero while others were zero. Then we will get two measurements whose values are:

$$m_1 = \Phi_{1j} x_j$$
$$m_2 = \Phi_{2j} x_j$$

If we divide the measurements we can get ratio of Φ_{1j}/Φ_{2j} and check which column of Φ has elements with this ratio. The problem is, that there might be more than one column satisfying this relation which will lead to many estimations of x_j and hence compromising uniqueness.

4.c

When x is 2-sparse and three measurements are taken, then let i-th and j-th element of signal have non-zero elements:

$$m_1 = \Phi_{1i} x_i + \Phi_{1j} x_j$$
$$m_2 = \Phi_{2i} x_i + \Phi_{2j} x_j$$
$$m_3 = \Phi_{3i} x_i + \Phi_{3j} x_j$$

This equation can be re-written in more readable form as:

$$\begin{bmatrix} m_1 \\ m_2 \\ m_3 \end{bmatrix} = \begin{bmatrix} \Phi_{1i} \\ \Phi_{2i} \\ \Phi_{3i} \end{bmatrix} x_i + \begin{bmatrix} \Phi_{1j} \\ \Phi_{2j} \\ \Phi_{3j} \end{bmatrix} x_j$$

In other words, we need to check which two columns of Φ are co-planar with the measurement vector. Since Φ is randomly generated matrix the chance of more than two columns of Φ to satisfy the condition exactly is less but in some cases there might be more than one solution for x. Algorithm to determine x is:

1) Pick any two measurements say 1st and 2nd.
2) Loop over the columns of the matrix \Phi.
3) For any column p, pick any column q such that q>p.
4) Solve the two linear equations for x_i and x_j with coefficients from the 1st and 2nd row of \Phi.
5) Put the calculated values in the third equation, i.e. with coefficients from the third row and check if the calculated value is equal to m_3.

If more than two columns of Φ pass the algorithm then x is not uniquely determinable. Otherwise, calculated x_i and x_j are the values with the column indices as their location in x.

4.d

When x is 2-sparse and four measurements are taken, then let i-th and j-th element of signal have non-zero elements:

$$m_1 = \Phi_{1i}x_i + \Phi_{1j}x_j$$
$$m_2 = \Phi_{2i}x_i + \Phi_{2j}x_j$$
$$m_3 = \Phi_{3i}x_i + \Phi_{3j}x_j$$
$$m_4 = \Phi_{4i}x_i + \Phi_{4j}x_j$$

This equation can be re-written in more readable form as:

$$\begin{bmatrix} m_1 \\ m_2 \\ m_3 \\ m_4 \end{bmatrix} = \begin{bmatrix} \Phi_{1i} \\ \Phi_{2i} \\ \Phi_{3i} \\ \Phi_{4i} \end{bmatrix} x_i + \begin{bmatrix} \Phi_{1j} \\ \Phi_{2j} \\ \Phi_{3j} \\ \Phi_{4j} \end{bmatrix} x_j$$

In this case, the probability that more than two columns of Φ satisfy the above condition is even more less. But in some bizarre cases we can encounter the situation that x is not unique. We have the same algorithm as used the previous part with only one more step, that is we should put the calculated values in the fourth linear equation also and check whether the calculated values satisfy it or not.

Question 5

5.a

Both the papers use sensors to compute coded snapshots. Hitomi's paper changes the code for each frame by sending signals to sensors directly whereas CACTI uses mechanical translation to compute a coded snapshot for several frames. Making uses of mechanics helps in reducing the power consumption and operating bandwidth of the camera. CACTI captures an object's spatiotemporal-multiplexed information. This mechanical translation helps in compressing temporal sampling without substantially increasing system volume or power.

The main difference between Hitomi's and CACTI is that Hitomi's paper tries to achieve high spatial and spectral resolution whereas CACTI tries to achieve high spatial, spectral and temporal resolution by using the fact that the information present in natural images are highly correlated. In Hitomi's, the code/mask is fixed for an integration time whereas in CACTI, the mask changes continuously.
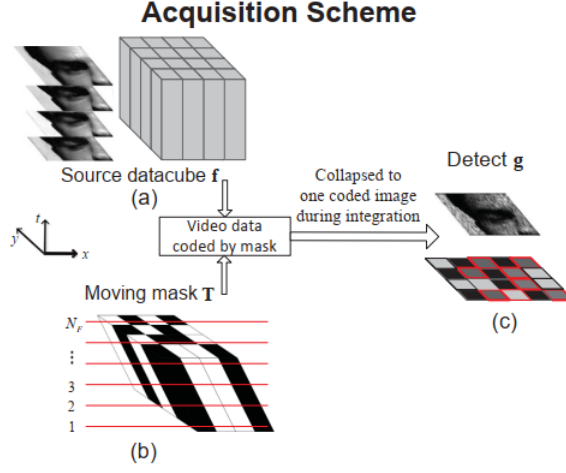
Acquisition Scheme

Fig. 1. CACTI image acquisition process. (a) A discrete space-time source datacube is (b) multiplied at each of N_F temporal channels with a shifted version of a coded aperture pattern. (c) Each detected frame **g** is the summation of the coded temporal channels and contains the object's spatiotemporal-multiplexed information. The dark grey (red-outlined) and black detected pixels in (c) pictorially depict the code's location at the beginning and the end of the camera's integration window, respectively.

Figure 30: CACTI

(a) Architecture of CMOS image sensor

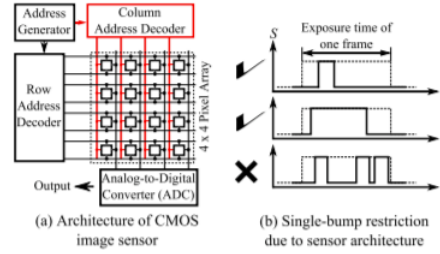(b) Single-bump restriction due to sensor architecture

Figure 2: **CMOS image sensor architecture and its limitations.** (a) Current CMOS image sensors have row addressing capability (black horizontal connections) which provides row-wise exposure control. Per-pixel exposure control can be implemented by adding column addressing (red vertical connections). (b) Most CMOS sensors do not have per-pixel frame-buffers on chip. Thus, each pixel can have only a single bump (on-time) during one camera exposure. The start and end times of the bump can be controlled.

Figure 31: Hitomi's

5.b

This paper solves the following unconstrained optimization problem:

$$\arg \min_{\mathbf{f}} \|\mathbf{g} - \mathbf{Hf}\|^2 + \lambda \Omega(\mathbf{f})$$

Here, for an N-pixel active sensor area,

- $\mathbf{g} \in \mathbb{R}^N$ is the sensed data
- $\mathbf{f} \in \mathbb{R}^{\sqrt{N} \times \sqrt{N} \times N_F}$ is a $(\sqrt{N} \times \sqrt{N}) \times N_F$ voxel spatiotemporal datacube i.e., the scene.

13

- **H** $\in \mathbb{R}^{N \times NN_F}$ s the system's discrete forward matrix that accounts for sampling factors including the optical impulse response, pixel sampling function, and time-varying transmission function. The forward matrix is a 2-dimensional representation of the 3-dimensional transmission function **T**

$$\mathbf{H}_k \overset{\text{def}}{=} \text{diag}\left[T_{1,1,k}, T_{2,1,k}, \cdots, T_{\sqrt{N},\sqrt{N},k}\right], \quad k = 1, \ldots, N_F$$

$$\mathbf{H} \overset{\text{def}}{=} [\mathbf{H}_1 | \mathbf{H}_2 | \cdots | \mathbf{H}_{N_F}]$$

- λ is a regularizer. This is chosen using experimental optimization.

- $\boldsymbol{\Omega}(\mathbf{f})$ is the total variation of **f**

$$\boldsymbol{\Omega}(\mathbf{f}) = \sum_k^{N_F} \sum_{i,j}^{N} \sqrt{(f_{i+1,j,k} - f_{i,j,k})^2 + (f_{i,j+1,k} - f_{i,j,k})^2}$$

Question 6

6.a

- Title - **Sparse imaging for fast electron microscopy**

- Authors - **Hyrum S. Anderson**, **Jovana Ilic-Helms**, **Brandon Rohrer**, **Jason Wheeler** and **Kurt Larson**

- Place - **Sandia National Laboratories**

- Link - https://cfwebprod.sandia.gov/cfdocs/CompResearch/docs/1210508c.pdf

6.b

The scanning electron microscope acquires image by raster sampling (images in old phones or video games using dot matrix structure) focused electron beam. At each location, electrons in the incident beam interacts with sample, producing various signals about the composition or topography of the sample's surface. These signals are detected and digitally assigned to the image pixel value at the corresponding sample location. The electron probe is then repositioned via electromagnetic or electrostatic deflection to the subsequent pixel location.

In order to produce high-quality SEM images, long (on the order of microseconds) integration times per pixel are required to reduce noise. In many non-trivial cases detector response times and other factors necessitate very long integration time. In sum, well-engineered systems are SNR-limited in their data acquisition speed, and can require months to collect millimeters or centimeters of data for some applications.

The degrees of freedom of typical electron microscope images are many fewer than the number of image pixels. Hence, they focused the electron beam at random image points and acquired some M measurements, thereby reconstructing the image using BP.

6.c

The reconstruction technique involves minimizing TVF norm with slight variation. The equations goes as:

$$\mathbf{y} = \Phi\mathbf{x} + \mathbf{n}$$

where $\mathbf{n}$ is noise vector Φ is is a subset of rows of identity $\mathbf{I}$ and $\mathbf{n}$ is noise with power σ^2. The image is obtained by solving regularized basis pursuit:

$$\min_{x} ||\mathbf{\Psi}^T \boldsymbol{x}||_1 + ||\boldsymbol{\nabla}\boldsymbol{x}||_1$$
$$s.t. \ ||\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x}|| \leq \boldsymbol{\sigma^2}$$

where $||\boldsymbol{\nabla}\boldsymbol{x}||_1 = \sum_i \sqrt{|\boldsymbol{\nabla}_h\boldsymbol{x_i}|^2 + |\boldsymbol{\nabla}_v\boldsymbol{x_i}|^2}$. By the low mutual coherence between the DCT basis and image-domain sampling, $\mathbf{\Psi}$ is chosen to be a block-DCT basis with 32×32 pixel blocks.