Assignment 1: CS 754

M.D. Kamran, Samiksha Das

Due 9 February 2021

Contents

Question 1

Part 1:
As s increases $s^{-1/2}$ and $\|\theta - \theta_s\|$ factors obviously decreasing with s. However δ_{2s} ;increases with an increase in 2s. Now C_0 and C_1 increase with an increase in δ_{2s} , the decrease caused by $s^{-1/2}$ and $\|\theta - \theta_s\|$ can be compensated with increase in C_0 and C_1.

Part 2:
The error term is not independent of m. This is because ϵ which is present in the upper bound on error is dependent on the number of samples. We generally take $\epsilon \leq \sigma\sqrt{m}$ whose upper bounds are clearly dependent on m. This implies that the upper bound of the error term is further bounded by $C_1 s^{-1/2}\|\theta - \theta_s\|_1 + C_2(\sigma\sqrt{m})$.

Part 3:
Clearly a given value of δ_{2s} corresponds to a value of C_0 and C_1 since these are dependent only on δ_{2s}. Now, theorem 3 is the same as theorem 3A for all values of $\delta_{2s} \leq 0.3$. In addition, theorem 3 can be applied to get an upper bound on $\|\theta - \theta*\|_2$ for $0.3 \leq \delta_{2s} \leq 0.41$ also. Hence theorem 3 is better than theorem 3A.

Part 4:
If the vector η has non-zero magnitude, then ϵ which is an upper bound on magnitude of η can not be set to value zero since ϵ is greater than equal to magnitude η . Hence the premise of the above claim cannot be satisfied and so the claim is not valid.

Question 2

MONDAY

Week - 46

Days - 312-054

Appointment-Meeting

Q2: (c) x is an unknown vector of dimensions $(H^*W^*T)\times 1$ which is vectorized form of the video sequence.

b is a known vector of dimensions $(H^*W)\times 1$ which is vectorized form of the coded snap snapshot after adding the gaussian noise to it.

$A = \Phi\Psi$, $b = Ax = \Phi\Psi x = \Phi f$

Here, Ψ is a 2D-DCT basis and f is sparse in Ψ. Φ is the measurement matrix with dimension $(H^*W)\times(H^*W^*T)$ which is constructed by concatenating T diag diagonal matrices of dimensions $(H^*W)\times(H^*W)$ where each diagonal has (H^*W) elements which represent the vectorized form of the randomly generated binary code C_t for all t from 1 to T.

Therefore, A is the product of the two matrices ie, measurement matrix and 2D DCT basis matrix. Also Ψ has the dimensions $(H^*W^*T)\times(H^*W^*T)$.

Priorities

Figure 1:

(d) $A = \Phi\Psi$, $b = Ax = \Phi\Psi x = \Phi f$

Similar to the previous case, the terms technically are the same but are over a smaller patch from the video.

x is the unknown vector of dimensions $(8*8*T)\times 1$ which is vectorized form of i^{th} patch of video.

b is of dimension $(8*8)\times 1$ vector which is vectorized form of any 8×8 patch of the coded snapshot.

Ψ is again 2D DCT basis matrix with dimensions $(8*8*T)\times(8*8*T)$. Φ is again measurement matrix with dimensions as $(8*8)\times(8*8*T)$ formed in similar manner for 8×8 patch instead of the previous $H*W$ patch from the random binary code. And A is the product of these two matrices.

Figure 2:

Meeting

(d) In the OMP algorithm, we have a gaussian noise with zero mean and sigma = 2.

$A = \Phi \Psi$ has dimension $64 \times 64T = m \times n$ because in our algorithm we are taking patches of size 8x8 and not the entire image. So, $m = 8$.

Therefore for each i from 1 to m, $\eta_i \sim N(0, \sigma^2)$. The squared magnitude of vector η is a chi squared random variable & hence, the magnitude of η will lie within 3 sigmas from the mean.

$$\varepsilon \geq 3\sigma\sqrt{m}$$
$$\varepsilon \geq 3 \times 2 \times \sqrt{64}$$
$$\varepsilon \geq 3 \times 2 \times 8$$
$$\varepsilon \geq 48$$

Therefore the error term bound is $\varepsilon \geq 48$ for OMP algorithm. When considering squared norm then $\varepsilon \geq 9\sigma^2 m = 2304$,

Figure 3:

T=3

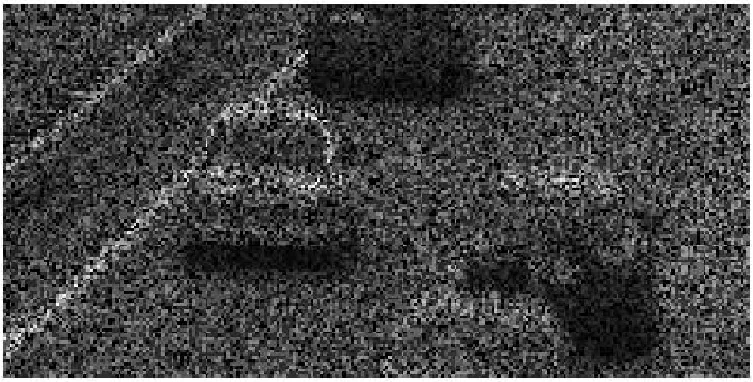

Figure 4: Coded Snapshot for T=3

Figure 5: Reconstruction Frame 1 for T=3

Figure 6: Reconstruction Frame 2 for T=3

Figure 7: Reconstruction Frame 3 for T=3

T=5

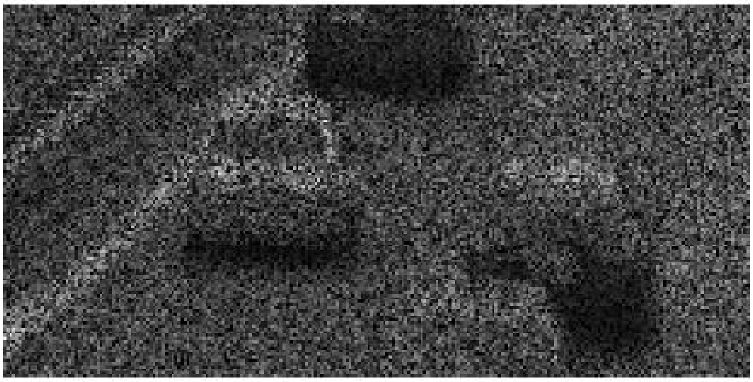

Figure 8: Coded Snapshot for T=5

Figure 9: Reconstruction Frame 1 for T=5

Figure 10: Reconstruction Frame 2 for T=5

Figure 11: Reconstruction Frame 3 for T=5

Figure 12: Reconstruction Frame 4 for T=5

Figure 13: Reconstruction Frame 5 for T=5

T=7

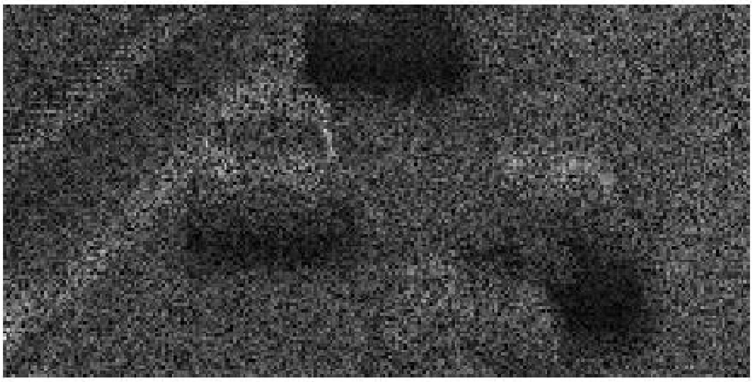

Figure 14: Coded Snapshot for T=7

Figure 15: Reconstruction Frame 1 for T=7

Figure 16: Reconstruction Frame 2 for T=7

Figure 17: Reconstruction Frame 3 for T=7

Figure 18: Reconstruction Frame 4 for T=7

Figure 19: Reconstruction Frame 5 for T=7

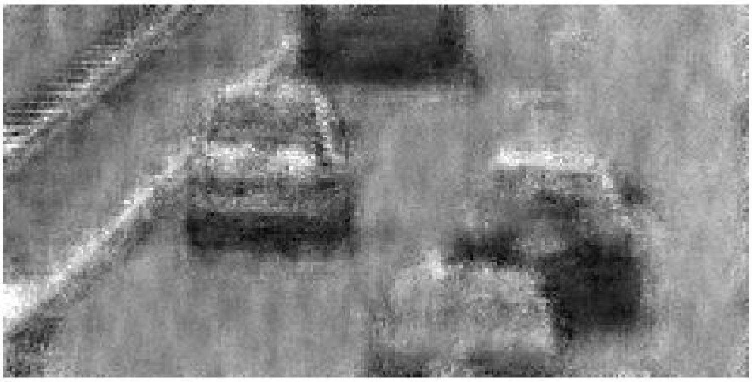

Figure 20: Reconstruction Frame 6 for T=7

Figure 21: Reconstruction Frame 7 for T=7

Flames

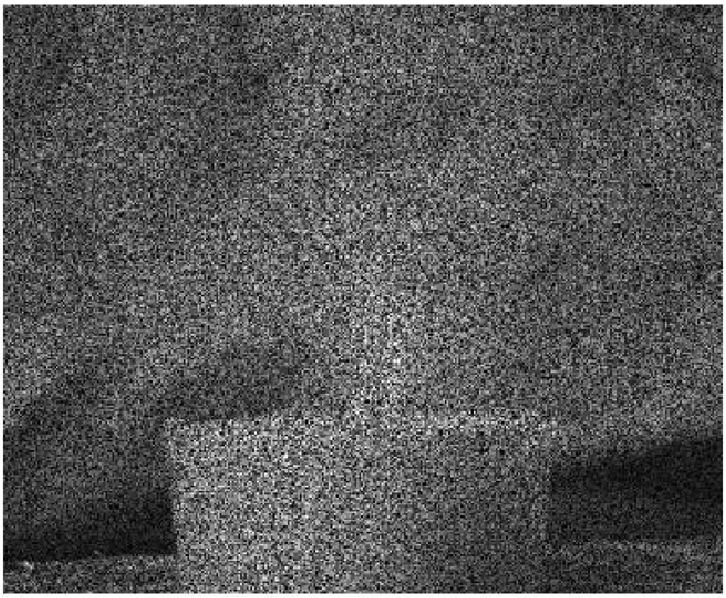

Figure 22: Coded Snapshot for flames

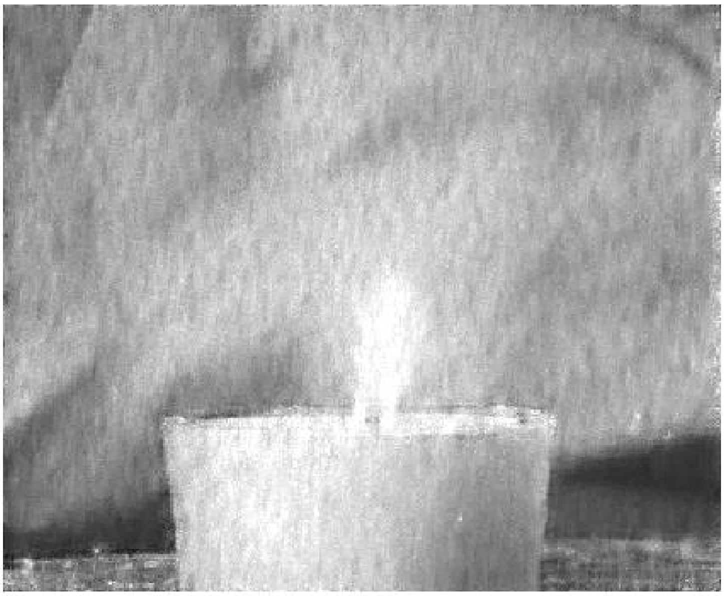

Figure 23: Reconstruction Frame 1 for flames

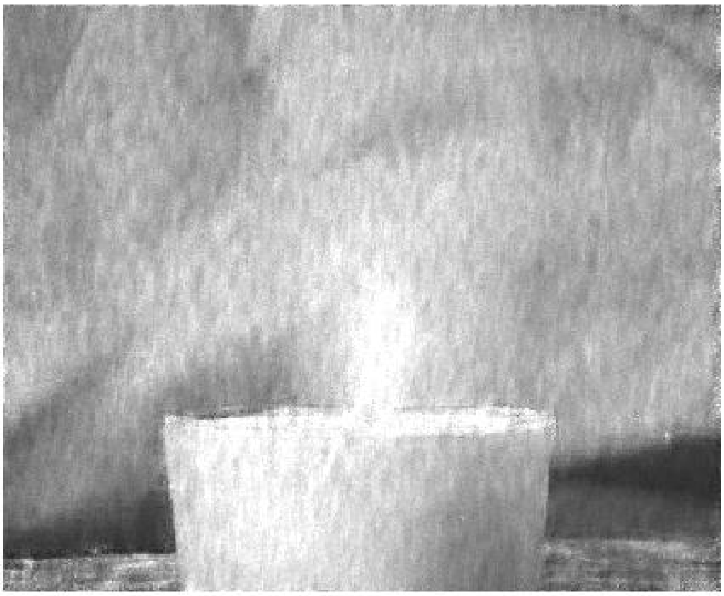

Figure 24: Reconstruction Frame 2 for flames

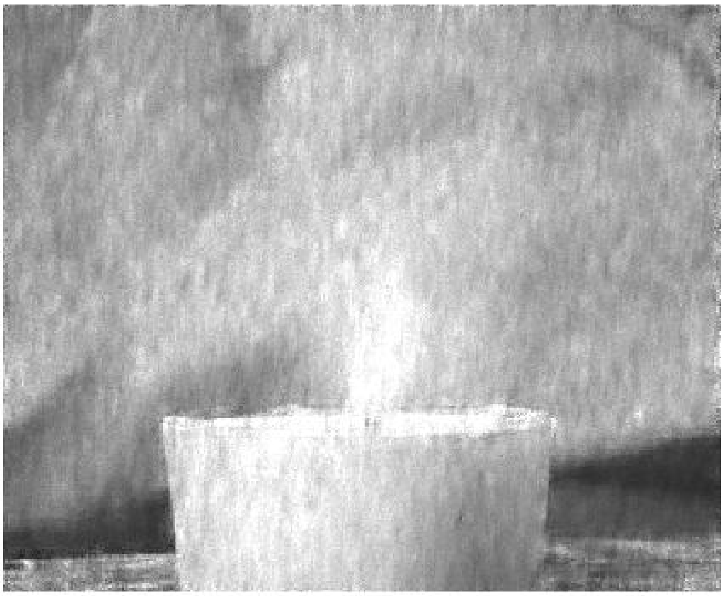

Figure 25: Reconstruction Frame 3 for flames

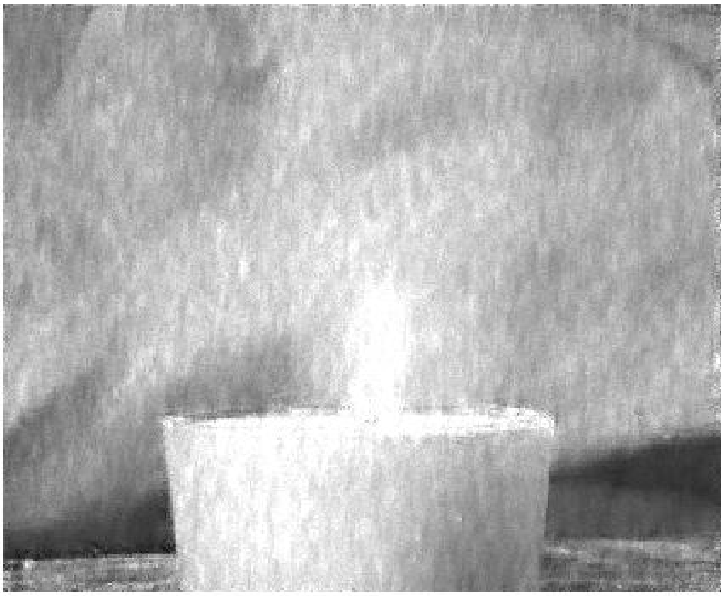

Figure 26: Reconstruction Frame 4 for flames

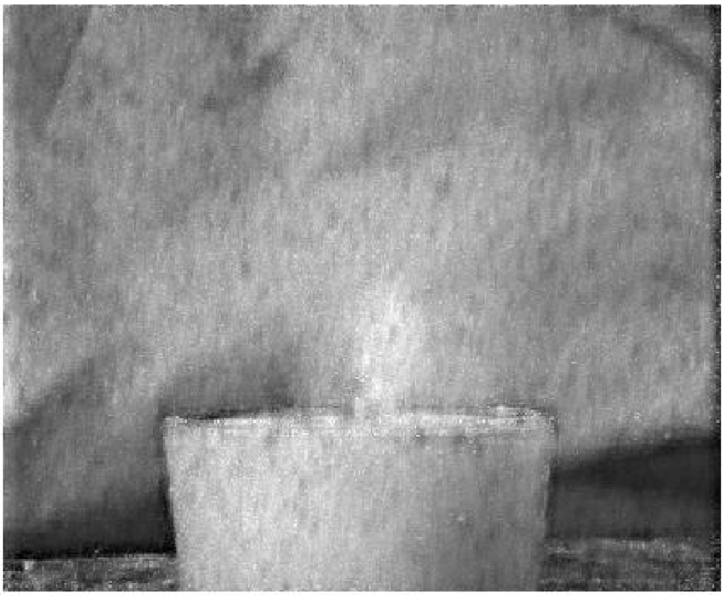

Figure 27: Reconstruction Frame 5 for flames

RMSE Values

- RMSE value for T=3 is 0.0250
- RMSE value for T=5 is 0.0458
- RMSE value for T=7 is 0.0632
- RMSE value for Flames is 0.0109

Question 3

given

$$g = \sum_{k=1}^{n} \alpha_k \Psi_k \implies \mu(g, \Psi) = \sqrt{n} max_{j \in \{1,..,n\}} |\langle g, \Psi_j \rangle| \implies \mu(g, \Psi) = \sqrt{n} max_{j \in \{1,..,n\}} |\langle \sum_{i=1}^{n} \alpha_i \Psi_i, \Psi_j \rangle|$$

since Ψ is orthonormal, $|\langle \Psi_i, \Psi_j \rangle| = 0$ when $i \neq j$ and $|\langle \Psi_i, \Psi_j \rangle| = 1$ when $i = j$

$$\implies \mu(g, \Psi) = \sqrt{n} max_{j \in \{1,..,n\}} |\langle \alpha_j \Psi_j, \Psi_j \rangle|$$

$$\implies \mu(g, \Psi) = \sqrt{n} max_{j \in \{1,..,n\}} |\alpha_j| \tag{1}$$

$|g| = | < g, g > | = \sum_{i=1}^{n} |\alpha_i^2|$. since g is a unit vector $1 = \sum_{i=1}^{n} |\alpha_i^2|$. Now, since $\sum_{i=1}^{n} |\alpha_i^2| = 1$, $|\alpha_i^2| \leq 1$, then $max_{i=1ton} |\alpha_i^2| \leq 1$, then $max_{i=1ton} |\alpha_i| \leq 1$. Hence $\mu(g, \Psi) \leq (\sqrt{n}$.Now, let $max_{j \in \{1,..,n\}} |\alpha_j|$ be x,

$$\mu(g, \Psi) = \sqrt{n} x$$

also,

$$\sum_{i=1}^{n} |\alpha_i^2| \leq \sum_{i=1}^{n} |x^2| \implies 1 \leq n x^2 \implies |x| \geq \frac{1}{\sqrt{n}}$$

$$\implies \mu(g, \Psi) \geq 1$$

Now if $|x| = \frac{1}{\sqrt{n}}$, we have $|\alpha_i| \leq \frac{1}{\sqrt{n}}$ for i=1 to n which implies $\sum_{i=1}^{n} |\alpha_i^2| \leq 1$. But $\sum_{i=1}^{n} |\alpha_i^2| = 1$ hence for i=1 to n, $|\alpha_i| = \frac{1}{\sqrt{n}}$. In this case, we have $\mu(g, \Psi) = 1$.

QED

Question 4

02 WEDNESDAY
Week - 45 Days - 307-059

CS754

Appointment-Meeting

Q4. (a) If both the value and index is unknown, then we can not find either with $m=1$ because we need atleast $2S=2$ measurements to uniquely estimate 1 sparse signal.

If the location is known, then let the data value be α and we refer to that index in Φ and we can find α using the known value in y. Since $m=1$, $y_1 = \Phi_i \alpha$, y_1 & Φ_i is known because i is known therefore α is known.

(b) If $m=2$, then we can find both the value and the index of the unknown element. First we need to ensure that if we pick any 2 columns in Φ, then those 2 vectors must be linearly independent.

$$y_1 = \Phi_{1i}\alpha \qquad \frac{y_1}{y_2} = \frac{\Phi_{1i}}{\Phi_{2i}}$$
$$y_2 = \Phi_{2i}\alpha$$

From here we can check the ith column of Φ whose elements are in the ratio equal to the ratio of elements in y. From this we will know the index i of the element & from any of the 2 equations, we can get the value α.

Priorities

Figure 28:

SATURDAY

Days - 310-056

(c) For 2 non zero elements, α & β, both whose values & indices are unknown would require atleast $2S = 4$ measurements to uniquely estimate the $2S$ sparse signal. But in the case, $m = 3$, we have only atmost three columns which are linearly independent to each other. And when we takes any 4 columns of Φ, atleast one of the column vector would depend on other 3 column vectors and hence we can't find all the indices & values for both the elements.

And no, there can't be any special instance of Φ for which unique estimation can be done for both values & indices because for that we require 4 linearly independent vector which is possible only when we have atleast $4 \times n$ matrix which we'll have to design ourselves. Also in a 3D space we know that if we take 4 vectors, any 4th vector can be represented in terms of other 3 given the other 3 are linearly independent.

Figure 29:

(d) For $m=4$ case, it is possible to estimate the values and the indices for both the nonzero elements α and β because $2S=4$ condition is satisfied and also it should be taken care that any 4 columns when considered in the measurement matrix Φ should be linearly independent.

Now, the algorithm goes like this. First we pick any 2 columns, i & j, and from there we will have four equations.

$$y_1 = \Phi_{1i}\alpha + \Phi_{1j}\beta$$
$$y_2 = \Phi_{2i}\alpha + \Phi_{2j}\beta$$
$$y_3 = \Phi_{3i}\alpha + \Phi_{3j}\beta$$
$$y_4 = \Phi_{4i}\alpha + \Phi_{4j}\beta$$

From here we can easily check like we did in (b) part whether these i & j will give unique α & β or in case there is a contradiction in the values of α & β, then the indices we have pick are certainly wrong.

We initialize $i = 0$ & $j = i+1$, j goes till n and i goes till $n-1$, so this will be a quadratic time algorithm $O(n^2)$ but with this we can guarantee to estimate both the values & indices accurately.

Figure 30:

Question 5

Here Hitomi camera has been used to refer to the video compressed sensing architecture using coded snapshots.

(a)

Differences:-

The Hitomi camera uses an emulation imaging system with an LCoS device (which acts as a spatial light modulator) to achieve pixel-wise exposure control where as in the above paper coding is implemented by a chrome-on-glass binary transmission mask in an intermediate image plane.

As N increases, LCoS-driven temporal compression strategies (used in the Hitomi camera)must modulate N pixels C times per integration but translating a passive transmissive element (used in the above paper) attains C times temporal resolution without utilizing any additional bandwidth relative to conventional low frame rate capture.Translation of the coded aperture during exposure means that each temporal plane in the video stream is modulated by a shifted version of the code, thereby attaining per-pixel modulation using no additional sensor bandwidth. Signal detection and separation work by CDMA or code division multiple access. The object's temporal channels are isolated from the compressed data by inverting a highly-underdetermined system of equations.

In the Hitomi camera the reconstruction algorithm used is based on learned dictionary methods where as in the above paper the reconstruction algorithm is iterative. The above paper uses two reconstruction algorithms TWIST and GAP.TWIST applies a regularization function to penalize estimates of f that are unlikely or undesirable to occur in the estimated f_e while GAP takes ad- vantage of the structural sparsity of the subframes in transform domains such as wavelets and discrete cosine transform (DCT). However in Hitomi camera after learning the over complete dictionary, standard sparse reconstruction methods are applied.

Similarities:-

Both the Hitomi camera and the architecture mentioned in the above paper employ some form of per-pixel modulation either through LCOS or translation of coded aperture. Both the architectures perform spatial and temporal resolution and sample and reconstruct discrete frames . Also both the architectures inherently estimate solutions to under determined linear systems.

(b)

For the TWIST reconstruction algorithm used in the given paper the cost function is :-

$$argmin_f(\left\|(g - Hf)\right\|^2 + \lambda\Omega(f))$$

In the above, $f \in \mathbb{R}^{NN_f \times 1}$ {after rasterisation}is the signal to be estimated. $H \in \mathbb{R}^{N \times 1}$ {after rasterisation} is the system's discrete forward matrix that accounts for sampling factors including the optical impulse response, pixel sampling function, and time-varying transmission function.g is the detector image $g = \sum_{k=1}^{N_F} T_{i,j,k} f_{i,j,k} + n_{i,j}$ rasterised where $n_{i,j}$ is noise in the (i,j)th pixel and time-varying spatial transmission pattern $T \in \mathbb{R}^{\sqrt{N} \times \sqrt{N} \times NF}$. The forward matrix is a a 2-dimensional representation of the 3-dimensional transmission function T. Lastly, $\Omega(f)$ is the Total Variation (TV) regularizer of dimension $NN_f \times 1$ given by:

$$\Omega(f) = \sum_{k}^{N_f} \sum i, j^N \sqrt{(f_{i+1,j,k} - f_{i,j,k})^2 + (f_{i,j+1,k} - f_{i,j,k})^2}$$

and λ is Total Variation (TV) regularization weight.

Question 6

(a)

GHz Optical Time-Stretch Microscopy by Compressive Sensing

It was published in IEEE photonics journal magazine on 1st March, 2017
(Volume 2, Issue 2, April 2017).

https://ieeexplore.ieee.org/abstract/document/7867817
https://www.researchgate.net/publication/314163187_GHz_Optical_Time-Stretch_Microscopy_by_Compressive_Sensing

(b)

A pulse laser is employed as the optical source. The pulses are first spatially dispersed by a diffraction grating, and then focused onto the target object by an objective lens. The 1-D cross-sectional profiles of the object in the lateral direction are encoded on the spectra of the incoming pulses.

The transmitted pulses through the object are collected by a second objective lens, modulated by a random binary mask, and recombined by a second diffraction grating. After that, the pulses are coupled into the temporal dispersive fiber spool and stretched in the time domain. Finally, the overlapped pulses are detected by a photodetector and digitized by a digitizer.

(c)

Consider an image with M rows and N columns. A uniform shift is introduced between each pair of successive rows, i.e, the k^{th} row of the image is shifted by $(k-1)^*T$ columns where, T is the number of sampling points during one pulse period. At higher pulse repetition rates, $T < N$ which leads to fewer measurements than the actual dimensionality of the image.

The reconstruction technique used is the **isotropic total-variation norm** defined by

$$\rho\left(I\right) = \sum_{x,y} \sqrt{\left[I\left(x+1,\ y\right) - I\left(x,y\right)\right]^2 + \left[I\left(x,y+1\right) - I\left(x,y\right)\right]^2} \tag{2}$$

to enforce sparse gradients in the recovered image. The image of the target is recovered by solving the following **convex optimization problem** given by

$$\hat{I} = \arg\ \min_I \left\{E - O(I)^2 + \beta \cdot \rho\left(I\right)\right\} \tag{3}$$

where, $E \in \mathbb{R}^K$ is the measured signal, $I \in \mathbb{R}^{M \times N}$ is the estimated original image during the iterations $K < (M \times N)$, $\hat{I} \in \mathbb{R}^{M \times N}$ is the recovered image, $O : \mathbb{R}^{M \times N} \rightarrow \mathbb{R}^K$ is the linear measurement operator which indicates the imaging process, and β is a non-negative scalar that provides a trade-off between data fidelity and gradient sparsity. The algorithm used is the **fast iterative shrinkage-thresholding algorithm (FISTA)** to solve the optimization problem. The value of the scalar β is tuned manually.