

# HW1 - CS754

## 1

---

1. We know that  $\delta_{2s} \propto S$  (also the upperbound:  $\delta_{2s} \leq \mu(s-1)$ ). So, as  $S$  increases,  $\delta_{2s}$  will increase. Also as  $m$  is staying constant, the chances of any  $2S$  columns of measuring matrix being linearly independent decreases as the length of column is fixed and  $S$  is increasing.

The constant terms ( $C_1, C_2$ ) are increasing functions of  $\delta_{2s}$ , so they increase as  $S$  increases

So from the above statements:

- $\delta_{2s}$  increases as  $S$  increases
- $C_1, C_2$  increases as  $\delta_{2s}$  increases.

Also, if we change  $m$ ,  $\epsilon$  will have a tight lower bound thus resulting in larger  $C_2\epsilon$  term.

Overall they increase in such a way that the overall term increases if  $S$  increases

2.  $m$  might not be present in the error bound but it is hidden in various terms.
  - The second term contains  $\epsilon$  which depends on  $m$ . It depends on  $m$  and the distribution of the noise. For eg.,
    - For  $\eta_i \sim \text{Uniform}(-r, r)$ , i.e. uniform random noise with known error  $r$ ,  $\epsilon \geq r^2 m$
    - For  $\eta \sim N(0, \sigma^2)$  with known  $\sigma$ ,  $\epsilon \geq 9m\sigma^2$
  - Theorem 3 holds only when the matrix satisfies RIP of order  $2S$ . Now if we vary  $m$ ,  $\delta_{2s}$  will change and  $C_1, C_2$  will change accordingly.
  - As  $m$  decreases, since  $m > 2S$  is necessary for RIP, chances of matrix satisfying RIP will decrease.
  - For a sensing matrix  $\Phi$  constructed, the matrix  $A$  will obey the RIP of order  $S$  with overwhelming probability provided that the number of rows  $m \geq CS \log(n/S)$ .

From the above mentioned points we can conclude that even if the error bound appears to be independent of  $m$ , terms in the error bound indirectly depend on  $m$ .

3. Theorem 3 is more useful than 3A as it has more relaxed condition on  $\delta_{2s}$ .
  - When  $\delta_{2s} < 0.3$ , reconstruction will be similar in both the cases
  - When  $\delta_{2s} > 0.3$ :
    - If  $\delta_{2s} > 0.41$ : In this case both the theorems will not be applicable.
    - If  $0.3 < \delta_{2s} < 0.41$ : In this case, theorem 3A will not be applicable. But we can use theorem 3. The  $C_1, C_2$  error bound terms are increasing functions of  $\delta_{2s}$ , so the error might be more but we can reconstruct the image. Also since we can apply theorem 3, we can manipulate other terms (decreasing  $m$  will increase  $\delta_{2s}$ ) and can compress the signal more and reconstruct from it.
4. If we set  $\epsilon = 0$  in BP, even if the noise vector  $\eta$  has non-zero magnitude. The reconstruction won't give result as expected:
  - As discussed in the 1st part of this ques, if we increase  $S$ ,  $1/\sqrt{2} * \|y - \Phi\Psi\theta\|_2^2$  term will decrease. therefore

we need the  $C_2\epsilon$  term to avoid any such discrepancy in our theorem.

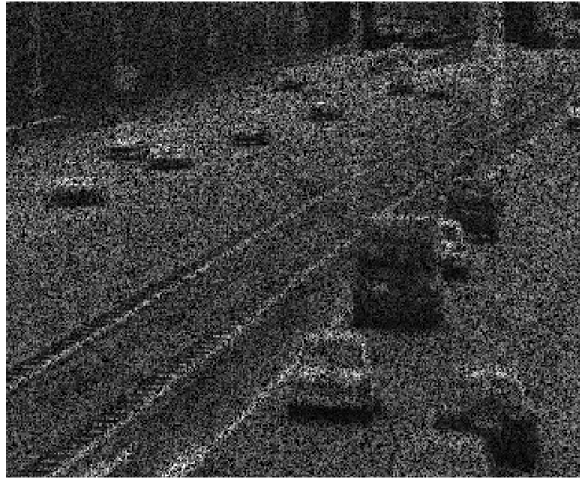
- To get a better reconstruction, we must choose  $\epsilon > \kappa$  where,  $\|\eta\|_2^2 \leq \kappa$ . And  $\eta$  has non-zero magnitude therefore choosing  $\epsilon = 0$  will give poor reconstruction result
- If we take  $\epsilon = 0$ , reconstruction algorithms will take more time to converge (or might even fail to converge).

TD

2

*Cars.avi* (T=3)

Coded Snapshot with noise



**Coded Snapshot**



**Original**



**Reconstructed**



**Coded Snapshot**



**Original**



**Reconstructed**



**Coded Snapshot**



**Original**



**Reconstructed**



*Cars.avi* ( $T=5$ )

Coded Snapshot with noise



**Coded Snapshot**



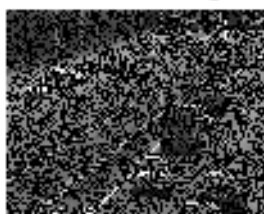
**Original**



**Reconstructed**



**Coded Snapshot**



**Original**



**Reconstructed**



**Coded Snapshot**



**Original**



**Reconstructed**



**Coded Snapshot**



**Original**



**Reconstructed**



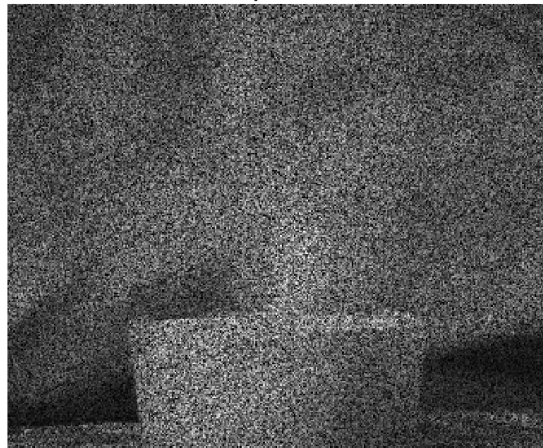


**Coded Snapshot    Original    Reconstructed**

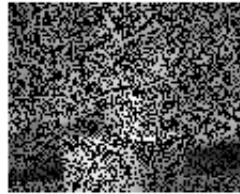


*Flames.avi* ( $T=5$ )

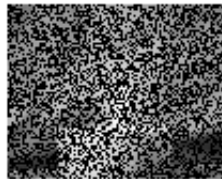
Coded Snapshot with noise



**Coded Snapshot    Original    Reconstructed**



**Coded Snapshot    Original    Reconstructed**



**Coded Snapshot    Original    Reconstructed**



**Coded Snapshot    Original    Reconstructed**



**Coded Snapshot    Original    Reconstructed**



# *Cars.avi* ( $T=7$ )

Coded Snapshot with noise



Coded Snapshot



Original



Reconstructed



Coded Snapshot



Original



Reconstructed



Coded Snapshot

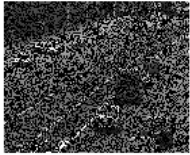


Original



Reconstructed



**Coded Snapshot****Original****Reconstructed****Coded Snapshot****Original****Reconstructed****Coded Snapshot****Original****Reconstructed****Coded Snapshot****Original****Reconstructed**

c)

We used the equation provided in the question to get another equation  $\mathbf{Ax} = \mathbf{b}$

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

$E$  is the coded image ( $I$  for our matlab code). We divided it into  $8 \times 8 \times T$  patches.

For each patch of  $8 \times 8$ , we define:

- $E^{i,j}$  be the patch of size  $8 \times 8 \times T$  with top left corner at  $i, j$ .
- $F^{i,j} = (\psi \otimes \psi) * \theta$ , where  $\theta$  is sparse.
- $C^{i,j}$  be the patch from the binary code  $C_t$  with top left corner at  $i, j$ .

$$E^{i,j} = \sum_{t=1}^T F_t^{i,j} \cdot C_t^{i,j}$$

Now, we can represent this as  $\mathbf{Ax} = \mathbf{b}$  form:

- $b = E_{i,j}$
- $A = [\text{diag}(C_1^{i,j}) \text{diag}(C_2^{i,j}) \text{diag}(C_3^{i,j})] \times (I_3 \otimes \psi \otimes \psi)$  ..... for  $T = 3$ 
  - $I_3$  will combine all the matrices to generate a single matrix which can process all the frames of a patch in a



single run

- $\psi \otimes \psi$  will take our  $\theta$  vector the the 2d-DCT basis

d)

- Consider  $y = \Phi f + \eta = \Phi \Psi \theta + \eta$
- If for each  $i = 1$  to  $m$ ,  $\eta_i$  is  $N(0, \sigma^2)$ , with known  $\sigma$ .
- The squared magnitude of the vector  $\eta$  is a chi-square random variable.
- Hence with very high probability, the magnitude of  $\eta$  will lie within 3 SD from the mean, i.e.  $\epsilon \geq 9m\sigma^2$ .
- Therefore, we set  $\epsilon \geq 9 * 64 * 2 * 2 = 2304$ .

e) Root mean squared error between reconstructed and original data:

- RMSE for T=5 (flames) is 0.033017
- RMSE for T=3 (cars) is 0.11745
- RMSE for T=5 (cars) is 0.15479
- RMSE for T=7 (cars) is 0.19776

### 3

---

$$\mu(\Phi, \Psi) = \sqrt{n} \max_{i,j \in \{0,1,\dots,n-1\}} |\Phi^{i^t} \Psi_j|$$

Any  $\mathbf{g} \in \mathbb{R}^n$  can be expressed as  $\mathbf{g} = \sum_{k=1}^n \alpha_k \Psi_k$  as  $\Psi$  is an orthonormal basis.

$$\mu(\mathbf{g}, \Psi) = \sqrt{n} \max_{i \in \{0,1,\dots,n-1\}} |\mathbf{g}^T \Psi_i|$$

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

$$\mu(\mathbf{g}, \Psi) = \sqrt{n} \max_{i \in \{0,1,\dots,n-1\}} \left| \sum_{k=1}^{n-1} \alpha_k \Psi_k^T \Psi_i \right|$$

$$\Psi_j^T \Psi_i = 0 \quad \forall i \neq j \quad \text{else} = 1$$

$$\mu(\mathbf{g}, \Psi) = \sqrt{n} \max_{i \in \{0,1,\dots,n-1\}} |\alpha_i|$$

$$\text{As } \mathbf{g} \text{ is a unit vector, } \|\mathbf{g}\| = 1$$

$$\mathbf{g} \mathbf{g}^T = \mathbf{1}$$

$$\sum_{k=0}^{n-1} \alpha_k \Psi_k \sum_{k=0}^{n-1} \alpha_k \Psi_k^T = \mathbf{1}$$

$$\sum_{k=0}^{n-1} \sum_{l=0}^{n-1} \alpha_k \alpha_l \Psi_k^T \Psi_l = \mathbf{1}$$

$$\sum_{k=0}^{n-1} \alpha_k^2 = 1$$

$$\text{So, } \mu(\mathbf{g}, \Psi) = \sqrt{n} \max_{i \in \{0,1,\dots,n-1\}} \frac{|\alpha_i|}{\sum_{k=0}^{n-1} \alpha_k^2}$$

Claim:  $\max |\alpha_i|$  is minimum when all  $|\alpha_i|$ s are equal and  $|\alpha_i| = \sqrt{k}/\sqrt{n}$  if  $\sum_{i=0}^{n-1} \alpha_i^2 = k$

Base Case:  $n=1$ , trivially true as  $\alpha_0^2 = 1$

Induction Hypothesis: Let this hold true for  $n-1$

Proof: w.l.o.g let's say  $|\alpha_0|$  is maximum among the  $\alpha_i$ s (It doesn't matter which index is chosen because addition is associative).

Lets say  $\sum_{i=1}^{n-1} \alpha_i^2 = k$

So  $\alpha_0^2 = 1 - k$  we want to minimize  $\alpha_0^2$

$\alpha_0^2$  should be greater than other elements so in order to ensure that we want  $\max_{i=1}^{n-1} \alpha_i$  to be minimum which will be when  $\alpha_1^2 = \alpha_2^2 = \dots = \alpha_{n-1}^2 = k/(n-1)$  [By induction Hypothesis]

We want to maximise  $k$  to minimise  $\alpha_0^2$

$$\begin{aligned} \alpha_0^2 &\geq k/(n-1) \\ 1 - k &\geq k/(n-1) \\ kn/(n-1) &\leq 1 \\ k &\leq (n-1)/n \\ \alpha_0^2 = 1 - k &\geq 1 - (n-1)/n \\ \alpha_0^2 &\geq 1/n \end{aligned}$$

So  $\min \alpha_0^2 = 1/n$ ,  $\alpha_1^2 = \alpha_2^2 = \dots = \alpha_{n-1}^2 = k/(n-1) = (1 - 1/n)/(n-1) = 1/n$

Therefore our claim holds.

So  $\min \mu(\mathbf{g}, \Psi) = \min \sqrt{n} \max_{i \in \{0,1,\dots,n-1\}} |\alpha_i| = \sqrt{n} \times 1/\sqrt{n} = 1$

And it is attained when  $|\alpha_0| = |\alpha_1| = \dots = |\alpha_{n-1}| = 1/\sqrt{n}$

So minimal value of coherence is attained when  $\mathbf{g} = \sqrt{1/n} \sum_{k=0}^{n-1} \pm \Psi_k$

## 4

---

a) We can't estimate  $\mathbf{x}$  uniquely as we need to know the magnitude and position of the non-zero element but we only have one equation which is  $\alpha_i x_i = y$  where  $\alpha_i$  is the  $i^{th}$  element of  $\phi$ . So we don't know  $x$  as well as  $i$ . But if we are provided with  $i$  then we can determine  $\mathbf{x}$  uniquely by  $x_i = \frac{y}{\alpha_i}$  and  $x_j = 0, \forall j \neq i$ .

b) Yes we can determine  $\mathbf{x}$  uniquely now we have a measurement vector  $\mathbf{y}$  which is  $2 \times 1$  in dimensions. So  $\mathbf{y} = \alpha_i x_i$  where  $\alpha_i$  is the  $i^{th}$  element of  $\phi$ . If we have unique  $\frac{\alpha_{1,i}}{\alpha_{2,i}}$  for each  $i$  then we can determine  $i$  using  $\frac{\alpha_{1,i} x_i}{\alpha_{2,i} x_i}$ , i.e.,  $\frac{y_1}{y_2}$  which is unique for each  $i$ . Now  $x_i = y_1 / \alpha_{1,i}$  and  $x_j = 0 \forall j \neq i$ .

c) As we have  $m=3$ , so  $\phi = 3 \times n$ .

We have  $\mathbf{y} = \Phi \mathbf{x}$  where  $\|\mathbf{x}\|_0 = 2$

$$\begin{aligned}\Phi &= [\mathbf{V}_1 \dots \mathbf{V}_n]; \mathbf{V}_i = 3 \times 1 \\ \Phi \mathbf{x} &= x_1 \mathbf{V}_1 + \dots + x_n \mathbf{V}_n \\ \Phi \mathbf{x} &= x_i \mathbf{V}_i + x_j \mathbf{V}_j\end{aligned}$$

Suppose we have an  $\mathbf{x}'$  which also satisfies this equation. So  $\Phi(\mathbf{x} - \mathbf{x}') = \mathbf{0}$ .

We know  $\max(\|\mathbf{x} - \mathbf{x}'\|_0) = 4$ .

$$\Phi(\mathbf{x} - \mathbf{x}') = \sum_i \alpha_i \mathbf{V}_i$$

If these 4  $\mathbf{V}_i$ s are not linearly independent then there would be  $\alpha_i$ s other than zeros which satisfies this equation (which means we can't uniquely estimate  $\mathbf{x}$ ) so we have to make sure these are linearly independent to obtain a unique  $\mathbf{x}$ . But as  $\mathbf{V}_i \in \mathbb{R}^3$  so any 3 independent  $\mathbf{V}_i$ s can represent any vector in  $\mathbb{R}^3$  which means 4 such vectors can never be linearly independent.

A special instance of  $\Phi$  exists for which this problem doesn't arise, that is, when it doesn't have 4 vectors, i.e., we resort to  $\Phi : 2 \times 3$  and the three columns are linearly independent. In short we want  $n = 3$ , and at this instance of  $\Phi$  can uniquely determine  $\mathbf{x}$  as now  $\max(\|\mathbf{x} - \mathbf{x}'\|_0) = 3$ .

d) The same problem doesn't arise for  $m=4$ , so our matrix should be such that any 4 columns of  $\Phi$  are linearly independent.

Algorithm: Check for all combinations of  $i, j \in [n]$  and  $i < j$ . Such that  $y$  can be represented as a linear combination of  $\mathbf{V}_i$  and  $\mathbf{V}_j$  if it is possible then  $x_i = \text{coefficient of } \mathbf{V}_i$ ,  $x_j = \text{coefficient of } \mathbf{V}_j$  and rest of the elements of  $\mathbf{x}$  are zero. The  $\mathbf{x}$  obtained will be unique due to our choice of  $\Phi$ .

It is possible to determine the coefficients of  $\mathbf{V}_i$  and  $\mathbf{V}_j$  as it will give  $\geq 2$  equations for  $x_i$  and  $x_j$ . There are always  $\geq 2$  equations as if there were only 1 equation then that would mean  $\mathbf{V}_i$  and  $\mathbf{V}_j$  are not linearly independent which is not possible.

Existence of such a  $\Phi$ :

Four vectors in  $\mathbb{R}^4$  are linearly dependent iff they line in a hyperplane.

Consider the following process for building  $\Phi$  (we are considering it as a set of  $4 \times 1$  vectors). We can start with the empty set, and choose any three linearly independent vectors  $v_1, v_2, v_3 \in \mathbb{R}^4$  and add them to S. Then choose a fourth vector  $v_4$  to add to S, we must make sure it is not in the unique hyperplane containing (i.e. spanned by)  $v_1, v_2, v_3$ . Thus  $v_4$  can be any vector in  $\mathbb{R}^4 \setminus \text{span}(v_1, v_2, v_3)$ .

Similarly, if at some stage  $\Phi = \{v_1, \dots, v_k\}$ , we can add to  $\Phi$  any vector  $v_{k+1}$  in  $\mathbb{R}^4 \setminus \bigcup_{x_a, x_b, x_c} \text{span}(x_a, x_b, x_c)$ .  $\bigcup_{x_a, x_b, x_c} \text{span}(x_a, x_b, x_c)$  is a finite union of hyperplanes, so it can never be all of  $\mathbb{R}^4$ . In this way we can construct  $\Phi$  for any  $n$  with the desired property.

## 5

---

a)

Here we have used a passive coded aperture for low power space-time compressive measurement. Coding is implemented by a chrome-on-glass binary transmission mask in an intermediate image plane. Modulation of the image data stream is done by harmonic oscillations of this mask.

While in Hitomi camera we used an active coded aperture. Here an LCoS device is used to achieve pixel wise exposure control, this is our active aperture. In this device a CMOS chip controls the voltage on square reflective aluminium electrodes buried just below the chip surface, each controlling one pixel. Each pixel can only have a single bump (on-time) during one camera exposure. The start and end times of the bump are controlled.

Hardware designs used in both are also somewhat different. As for CACTI the mask is kept between the lens and image sensor. While for Hitomi, the LCoS is essentially a reflector, so first the light gets reflected by the device and then it reaches image sensor using a polarising beam splitter.

Both architectures make use of coded aperture which are implemented very differently (CACTI: Mechanical translation of passive aperture, Hitomi: Active aperture) and use CCD array for image sensing.

b)

$$C_f = \|\mathbf{g} - \mathbf{H}\mathbf{f}\|^2 + \lambda\Omega(\mathbf{f})$$

where  $\Omega(\mathbf{f})$  and  $\lambda$  are the regularizer and regularization weights. We employ a Total Variation (TV) regularizer given by

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

$C_f$  is our cost function for *TwIST* reconstruction.

$$\mathbf{f}_e = \arg \min_{\mathbf{f}} C_f$$

$f, f_e \in \mathbb{R}^{\sqrt{N} \times \sqrt{N} \times N_F}$ ;  $f$  is our data cube which contains  $N_F$  images of  $N$  pixels.

We have flattened out  $f$  as a column vector so  $f, f_e : NN_F \times 1$

$T \in \mathbb{R}^{\sqrt{N} \times \sqrt{N} \times N_F}$  is our time varying spatial transmission pattern which uniquely codes each of the  $N_F$  temporal channels of  $f$  prior to integrating into one detector image  $g \in \mathbb{R}^{\sqrt{N} \times \sqrt{N}}$  during acquisition.

$g$  is our detector image which is measured as

$$g_{i,j} = \sum_{k=1}^{N_F} T_{i,j,k} f_{i,j,k} + n_{i,j}$$

where  $n_{i,j}$  represents imaging noise at the  $(i,j)^{th}$  pixel.

We can represent this using vector multiplication as

$$\mathbf{g} = \mathbf{H}\mathbf{f} + \mathbf{n}$$

$$\mathbf{g} : N \times 1 \quad \mathbf{H} : N \times NN_f \quad \mathbf{f} : NN_f \times 1 \quad \mathbf{n} : N \times 1$$

$\mathbf{H}$  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. The forward matrix is a 2-D representation of the 3-D transmission function  $\mathbf{T}$ :

$$\mathbf{H}_k := \text{diag}[T_{1,1,k} \ T_{2,1,k} \ \dots \ T_{\sqrt{N},\sqrt{N},k}], \ k = 1, \dots, N_F;$$

$$\mathbf{H} := [\mathbf{H}_1 \ \mathbf{H}_2 \ \dots \ \mathbf{H}_{N_F}],$$

where  $\mathbf{H}_k \in \mathbb{R}^{N \times N}$  is a matrix containing the entries of  $\mathbf{T}_k$  along its diagonals and  $\mathbf{H}$  is a concatenation of all  $\mathbf{H}_k$ ,  $k \in [N_F]$ .

## 6

---

a)

### High-speed flow microscopy using compressed sensing with ultrafast laser pulses

Received 15 Jan 2015; revised 29 Mar 2015; accepted 2 Apr 2015; published 15 Apr 2015

© OSA 2015 20 Apr 2015 | Vol. 23, No. 8 | DOI:10.1364/OE.23.010521 | OPTICS EXPRESS 10531

[Link to the paper](#)

b)

The principle of operation of the CHiRP-CS imaging system is to modulate pseudorandom patterns at an ultrahigh rate onto the optical spectra of broadband mode-locked laser pulses and then utilize these spectral patterns to create structured illumination of an object.

Light collected from the object is directed onto a single-pixel high-speed photodetector and the energy of each returned laser pulse is recorded continuously by a synchronized real-time ADC.

Broadband laser pulses are dispersed in optical fiber to accomplish spectrum-to-time mapping. Each pulse is modulated with a unique ultrahigh-rate pseudorandom binary pattern and then re-compressed in fiber (Dispersion compensation) to an ultrashort duration before passing through a 1D wavelength-to-space mapping diffraction grating and lens that focuses the spectral pattern onto the object plane, providing structured illumination of the object flow. The output pulse energy traveling back through the spatial disperser to the

photodiode and ADC represents an optically-computed inner product between the pseudorandom pattern and the object. The image is reconstructed via a sparsity-driven optimization from sub-Nyquist compressive measurements.

c)

The coefficient  $\alpha \in \mathbb{R}^{N_b \times M_b}$  of  $x$  under some sparsifying transform  $\tilde{\Psi}(\cdot)$  defined by

$$\alpha = \tilde{\Psi}(x)$$

should be sparse or compressible.

The recovery process estimates the set of sparse coefficients  $\{\alpha^k\}_{k=1}^p$  of the patch set  $\{x^k\}_{k=1}^p$  covering the entire image. Denoting  $\{\tilde{\alpha}^k\}_{k=1}^p$  as the sparse coefficients of the patches  $\{\tilde{x}^k\}_{k=1}^p$  extracted from original image  $\tilde{G} \in \mathbb{R}^{N \times M}$ , the 1D compressive measurement process can be written as

$$y_j = \Phi_j[P(\Psi(\tilde{\alpha}^k)_{k=1}^p)]_j, \forall j = 1, \dots, M$$

where  $\Psi(\cdot)$  is the inverse sparsifying transform of  $\tilde{\Psi}(\cdot)$  satisfying  $\tilde{x}^k = \Psi(\tilde{\alpha}^k), \forall k = 1, \dots, p$ ;  $P(\cdot)$  is the operator that combines the set of image patches  $\{\tilde{x}^k\}_{k=1}^p$  back to the original image, i.e.,  $\tilde{G} = P(\Psi(\tilde{\alpha}^k)_{k=1}^p)$ ;  $\Phi_j \in \mathbb{R}^{m \times N}, \forall j = 1, \dots, M$ , is the local pseudorandom sensing matrix used to measure row  $\tilde{g}_j$  of  $\tilde{G}$  and  $y_j$  is the corresponding measurement vector. Given the set of measurement vectors and sensing matrices  $\{(y_j, \Phi_j)\}_{j=1}^M$ , we propose to obtain the sparse coefficients from the following optimization problem