

# CS 754 - Advanced Image Processing

## Assignment 4 - Report

Shaan ul Haque - 180070053  
Mantri Krishna Sri Ipsit - 180070032

April 16, 2021

### Question 1

#### 1.a

The equation (7) given in the paper is as follows:

$$J_3(v) = \sum_j \rho_j(A_{j \rightarrow} v - b_j)$$

Here  $A_{j \rightarrow}$  represents a matrix of derivative filters centered on pixel  $j$  to be applied on vectorized image  $I_1$  i.e,  $v$ . If the pixel  $j \in S_1$ , then as per equation (6), we have to subtract the gradient of original image  $I$  and then apply  $\rho$ . Hence when  $j \in S_1$ ,  $b_j$  has the derivative of  $I$  centered at  $j$ .

When  $j \in S_2$ , as per equation (6), we apply  $\rho$  directly on the derivative of  $I_1$  centered at  $j$ . Hence in this case the value of  $b_j = 0$  when  $j \in S_2$ .

#### 1.b

- The following terms are obtained from the likelihood:

$$\rho(f_{i,k} \cdot I_1) + \rho(f_{i,k} \cdot (I - I_1))$$

- The following terms are obtained from the prior:

$$f_{i,k} \cdot I_1, f_{i,k} \cdot (I - i_1)$$

The log-histogram of derivative filters was the prior used in the paper. The Laplacian mixture model was used as the likelihood in the paper.

#### 1.c

The paper exploits the statistics of natural images. It is known that the log-histograms of the derivative filters have a sparse distribution. They lie below the straight line connecting

the minimal and maximal values. The Gaussian distribution is not sparse because it is always above the straight line. The Laplacian distribution is exactly at the border between sparse and non-sparse distributions. This prior has been used because the sparse nature of derivative filter outputs is a robust property of natural images and this paper focuses on removing reflections from natural images.

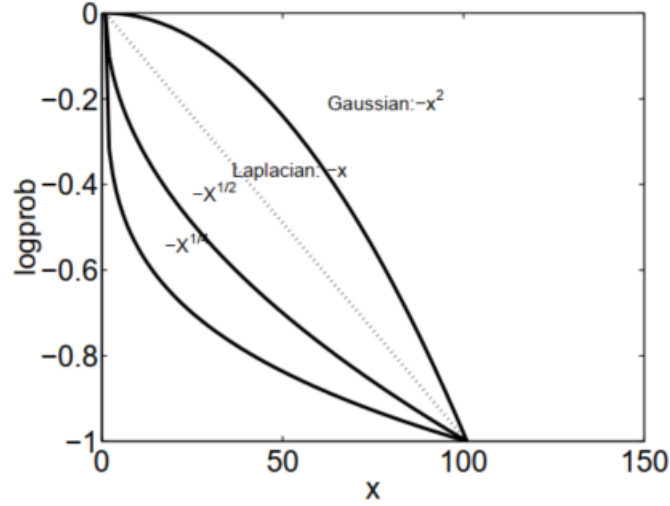


Figure 1: Sparseness of different distributions (taken from the paper provided). The Laplacian distribution is exactly at the border between sparse and non-sparse. Any distribution lying above Laplacian is non-sparse and any distribution lying below Laplacian is sparse.

## Question 2

Given model:

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

where  $\eta \sim N(0, \sigma^2 \mathbf{I}_{m \times m})$ ,  $\Phi \sim N(0, 1/m)$ ,  $\sigma = 0.01 \times$  the average absolute value of in  $\Phi \mathbf{x}$  and  $\mathbf{x}$  is also drawn from a Gaussian distribution whose mean is zero while the covariance matrix ( $\Sigma_{\mathbf{x}}$ ) has polynomial decrement in eigenvalues.

The MAP for estimating  $\mathbf{x}$  given  $\mathbf{y}$ ,  $\Phi$  and  $\Sigma_{\mathbf{x}}$  is:

$$\mathbf{x} = \operatorname{argmax}_{\mathbf{x}} p(\mathbf{x}|\mathbf{y})$$

Using Bayes rule, we get:

$$\operatorname{argmax}_{\mathbf{x}} p(\mathbf{x}|\mathbf{y}) = \operatorname{argmax}_{\mathbf{x}} p(\mathbf{y}|\mathbf{x})p(\mathbf{x})$$

where  $p(\mathbf{y})$  is dropped as it has no role to play in the maximization problem. Since  $p(\mathbf{y}|\mathbf{x})$  has the same distribution as noise but mean shifted to  $\Phi \mathbf{x}$  and  $p(\mathbf{x})$  is a joint Gaussian vector, we have:

$$p(\mathbf{y}|\mathbf{x})p(\mathbf{x}) = \exp(-\|\mathbf{y} - \Phi \mathbf{x}\|^2 / 2\sigma^2) \frac{\exp(-\frac{1}{2} \mathbf{x}^T \Sigma_{\mathbf{x}}^{-1} \mathbf{x})}{(2\pi)^{N/2} |\Sigma_{\mathbf{x}}|^{1/2}}$$

Taking log of the expression thus maximization problem becomes minimization problem as given below:

$$\mathbf{x} = \operatorname{argmax}_{\mathbf{x}} (\|\mathbf{y} - \Phi\mathbf{x}\|^2/2\sigma^2 + \frac{1}{2}\mathbf{x}^T \Sigma_{\mathbf{x}}^{-1} \mathbf{x})$$

Opening the square terms, we get:

$$\mathbf{x} = \operatorname{argmax}_{\mathbf{x}} ((\mathbf{y} - \Phi\mathbf{x})^T (\mathbf{y} - \Phi\mathbf{x})/2\sigma^2 + \frac{1}{2}\mathbf{x}^T \Sigma_{\mathbf{x}}^{-1} \mathbf{x})$$

Taking derivative w.r.t.  $\mathbf{x}$  and making it zero, we get:

$$2\Phi^T \Phi \mathbf{x} - 2\mathbf{y} \Phi^T / 2\sigma^2 + \frac{1}{2}(2\Sigma_{\mathbf{x}}^{-1} \mathbf{x}) = 0$$

$$(\Phi^T \Phi / \sigma^2 + \Sigma_{\mathbf{x}}^{-1} \mathbf{x}) = \Phi^T \mathbf{y} / \sigma^2$$

From the above equation we get the closed form solution as:

$$\mathbf{x} = (\Phi^T \Phi + \sigma^2 \Sigma_{\mathbf{x}}^{-1})^{-1} \Phi^T \mathbf{y}$$

The following figure shows the  $\log(\text{average RMSE} + 0.01)$  obtained over 10 signal samples randomly drawn from Gaussian distribution for different values of  $\alpha$ . Log was taken for better visualization as for large values of  $\alpha (\geq 2)$  error was significantly small.

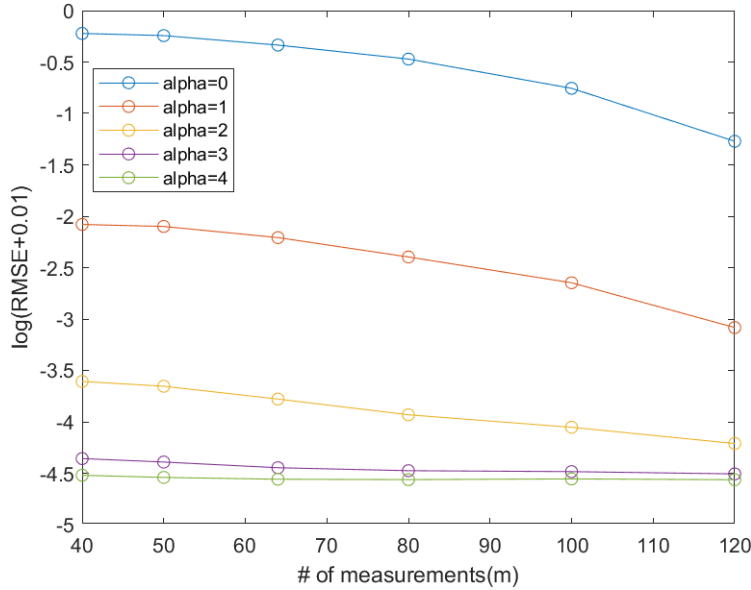


Figure 2: RMSE vs m

We observe that for the same  $m$ , RMSE decreases as  $\alpha$  increases. The magnitude of the eigenvalues of the covariance matrix tells us how much the random vector is spread along that direction about the mean. If the eigenvalue along that direction is very small then essentially it means that the random vector is almost constant along that direction. In our case since mean was 0, it means values of  $\mathbf{x}$  along that direction (or if  $i$ th eigenvector is small then  $i$ th element) is almost always 0.

The structure of the covariance matrix  $\Sigma_{\mathbf{x}}$  is that its eigenvalues decreases with increase in the

row number which is at faster rate for larger  $\alpha$ . Due to this most of the values in vector  $\mathbf{x}$  is zero because of negligibly small eigenvalues. Since most of the values are zero we need lesser number of measurements to estimate the non-zero values in  $\mathbf{x}$  as  $\alpha$  increases.

Whereas, for same  $\alpha$  RMSE decreases with increase in number of measurements is obvious as more observations would lead to more accurate estimation of  $\mathbf{x}$ .

**Note:** We used a ready made function to generate orthogonal matrix for this question which is uploaded with the file.

## Question 3

### 3.a

$X_0$  is a matrix of rank  $r$  such that  $\mathcal{A}(X_0) = b$ . We also have  $X^*$  such that

$$X^* := \arg \min_X \|X\|_* \quad \text{subject to } \mathcal{A}(X) = b$$

As  $X^*$  is a matrix with minimum nuclear norm subject to the condition  $\mathcal{A}(X) = b$  and  $X_0$  also satisfies  $\mathcal{A}(X_0) = b$ , the nuclear norm of  $X_0$  should be more than that of  $X^*$  i.e.,

$$\|X_0\|_* \geq \|X^*\|_*$$

### 3.b

$X_0$  and  $R_c$  have same dimension by construction. And by theorem (3.3) of the paper,  $X_0^T R_c = 0$  and  $X_0 R_c^T = 0$ . Hence,  $X_0$  and  $R_c$  satisfy the conditions of Lemma (2.3) of the paper and hence

$$\|X_0 + R_c\|_* = \|X_0\|_* + \|R_c\|_*$$

Hence

$$\|X_0\|_* \geq \|X_0 + R\|_* \geq \|X_0 + R_c\|_* - \|R_0\|_* = \|X_0\|_* + \|R_c\|_* - \|R_0\|_*$$

Using the relationship between the leftmost and the rightmost terms, we get

$$\|X_0\|_* \geq \|X_0\|_* + \|R_c\|_* - \|R_0\|_*$$

Hence

$$\|R_0\|_* \geq \|R_c\|_*$$

### 3.c

Without loss of generality, assume that the SVD has been performed in such a way that the singular values are arranged in decreasing order along the diagonal and the corresponding left and right singular vectors are re-ordered. The partitions  $I_i$  are disjoint by construction and are  $3r$  in cardinality.

Consider a singular value  $\sigma_k$  of  $R_c$  such that  $k \in I_{i+1}$ . By our assumption above, for every  $j \in I_i$ , we have

$$\sigma_k \leq \sigma_j$$

Repeating the above inequality for all  $j \in I_i$  and then adding those  $3r$  inequalities, we get

$$\sigma_k \times 3r \leq \sum_{j \in I_i} \sigma_j$$

hence

$$\sigma_k \leq \frac{1}{3r} \sum_{j \in I_i} \sigma_j$$

### 3.d

Note that  $\|R_i\|_* = \sum_{j \in I_i} \sigma_j$ . Hence, the above inequality becomes

$$\sigma_k \leq \frac{1}{3r} \|R_i\|_*$$

Squaring on both sides (as both the sides are positive quantities, the sign of the inequality doesn't change)

$$\sigma_k^2 \leq \frac{1}{9r^2} \|R_i\|_*^2 \quad \forall k \in I_{i+1}$$

Adding up the above  $3r$  inequalities, we get

$$\sum_{k \in I_{i+1}} \sigma_k^2 \leq \frac{3r}{9r^2} \|R_i\|_*^2$$

But  $\|R_{i+1}\|_F^2 = \sum_{k \in I_{i+1}} \sigma_k^2$ . Hence we have

$$\|R_{i+1}\|_F^2 \leq \frac{1}{3r} \|R_i\|_*^2$$

### 3.e

Using the result derived in 3.d, we have

$$\|R_{i+1}\|_F \leq \frac{1}{\sqrt{3r}} \|R_i\|_*$$

Repeating the above inequality and adding all of them, we have

$$\sum_{i \geq 1} \|R_{i+1}\|_F \leq \frac{1}{\sqrt{3r}} \sum_{i \geq 1} \|R_i\|_*$$

By change of variable, we have

$$\sum_{j \geq 2} \|R_j\|_F \leq \frac{1}{\sqrt{3r}} \sum_{j \geq 1} \|R_j\|_*$$

### 3.f

As all the  $R_j$  have orthogonal row space and column space, by the sub-additivity property of nuclear norm, we have

$$\sum_{j \geq 1} \|R_j\|_* = \|R_c\|_*$$

But by equation (3.4) of the paper (which is same as the result in 3.b above),  $\|R_0\|_* \geq \|R_c\|_*$ . Hence

$$\sum_{j \geq 2} \|R_j\|_F \leq \frac{1}{\sqrt{3r}} \sum_{j \geq 1} \|R_j\|_* = \frac{1}{\sqrt{3r}} \|R_c\|_* \leq \frac{1}{\sqrt{3r}} \|R_0\|_*$$

### 3.g

As per the construction using the Lemma (3.4) of the paper and as rank of  $X_0$  is  $r$ , we have

$$\text{rank}(R_0) \leq 2r$$

Since the relation between the nuclear norm and Frobenius norm (from linear algebra) for some matrix  $P$  is

$$\|P\|_* \leq \sqrt{\text{rank}(P)} \|P\|_F$$

applying to  $R_0$ , we get

$$\|R_0\|_* \leq \sqrt{2r} \|R_0\|_F$$

Using this in the result of (3.f) above, we finally have

$$\sum_{j \geq 2} \|R_j\|_F \leq \frac{1}{\sqrt{3r}} \sum_{j \geq 1} \|R_j\|_* = \frac{1}{\sqrt{3r}} \|R_c\|_* \leq \frac{1}{\sqrt{3r}} \|R_0\|_* \leq \frac{\sqrt{2r}}{\sqrt{3r}} \|R_0\|_F$$

### 3.h

By construction,  $\text{rank}(R_1) \leq 3r$  and by Lemma (3.4) of the paper,  $\text{rank}(R_0) \leq 2r$ . Hence by subadditivity property of rank, we have

$$\text{rank}(R_0 + R_1) \leq 2r + 3r = 5r$$

### 3.i

We have

$$R = R_0 + R_c = R_0 + R_1 + \sum_{j \geq 2} R_j$$

Hence, by triangle inequality,

$$\|\mathcal{A}(R)\| = \left\| \mathcal{A}\left((R_0 + R_1) + \sum_{j \geq 2} R_j\right) \right\| \geq \|\mathcal{A}(R_0 + R_1)\| - \left\| \mathcal{A}\left(\sum_{j \geq 2} R_j\right) \right\|$$

But again using the triangle inequality, we have

$$\left\| \mathcal{A}\left(\sum_{j \geq 2} R_j\right) \right\| \leq \sum_{j \geq 2} \|\mathcal{A}(R_j)\|$$

Hence,

$$\|\mathcal{A}(R)\| \geq \|\mathcal{A}(R_0 + R_1)\| - \sum_{j \geq 2} \|\mathcal{A}(R_j)\|$$

### 3.j

As the rank of  $R_0 + R_1$  is atmost  $5r$  and the rank of  $R_j$  is atmost  $3r \forall j \geq 2$ , by using the restricted isometry property of  $\mathcal{A}$ , we have

$$\|\mathcal{A}(R_0 + R_1)\| \geq (1 - \delta_{5r}) \|R_0 + R_1\|_F$$

and

$$\|\mathcal{A}(R_j)\| \leq (1 + \delta_{3r}) \|R_j\|_F \quad \forall j \geq 2$$

Using the above two inequalities in the result of (3.j) above, we get

$$\begin{aligned} \|\mathcal{A}(R)\| &\geq \|\mathcal{A}(R_0 + R_1)\| - \sum_{j \geq 2} \|\mathcal{A}(R_j)\| \\ &\geq (1 - \delta_{5r}) \|R_0 + R_1\|_F - (1 + \delta_{3r}) \sum_{j \geq 2} \|R_j\|_F \end{aligned}$$

### 3.k

Since  $\mathcal{A}$  is an affine transformation and  $\mathcal{A}(X^*) = b$  and  $\mathcal{A}(X_0) = b$ , we have

$$\mathcal{A}(R) = \mathcal{A}(X^* - X_0) = \mathcal{A}(X^*) - \mathcal{A}(X_0) = b - b = 0$$

### 3.1

The inequality (3.7) of the paper is as follows:

$$\|\mathcal{A}(R)\| \geq \left( (1 - \delta_{5r}) - \frac{9}{11}(1 + \delta_{3r}) \right) \|R_0\|_F$$

Since  $\mathcal{A}(R) = 0$ ,  $\|\mathcal{A}(R)\| = 0$ . We know that the Frobenius norm of a matrix is always non-negative. Hence, the right hand side of the above inequality to be positive, we need

$$\left( (1 - \delta_{5r}) - \frac{9}{11}(1 + \delta_{3r}) \right) > 0$$

Taking LCM and then simplifying, we have

$$\frac{2 - (11\delta_{5r} + 9\delta_{3r})}{11} > 0$$

which is equivalent to numerator being positive or

$$11\delta_{5r} + 9\delta_{3r} < 2$$

## Question 4

Incoherence of singular vectors with the canonical basis is required for matrix recovery to avoid cases where the low rank matrix takes the form such as all zeros except 1 at the any random position or in other words there is not much variation among most of the elements of the matrix except for very few elements where value is significantly different. When random elements of such matrices are sampled most of the time we would observe small variation among the sampled values and we have no way of knowing what were the significantly different values. Matrices of rank less than  $r$  can be represented by:

$$\mathbf{M} = \sum_{k=1}^r \sigma_k \mathbf{u}_k \mathbf{v}_k^T$$

where all the symbols have there usual meanings. When singular vectors are uniformly sampled from the set of all families of  $r$  orthonormal vectors we are infact asking for incoherence with the canonical basis so as to avoid pathological cases as described above. This can be seen by taking an example where singular vectors are themselves are standard orthonormal vectors. Each summand in the sum would be a matrix with just one non-zero element making the net matrix have just  $r (< \min(m, n))$  making the net matrix almost all zeros with  $r$  non-zero elements giving us the pathological case we wanted to avoid.

When the constraint is changed to:

$$\text{minimize } \text{rank}(\mathbf{X})$$

$$\text{subject to : } \mathbf{f}_i^T \mathbf{X} \mathbf{g}_j = \mathbf{f}_i^T \mathbf{M} \mathbf{g}_j \quad (i, j) \in \Omega$$

Using the rank  $r$  representation of matrix  $\mathbf{M}$  we get:

$$\mathbf{f}_i^T \mathbf{M} \mathbf{g}_j = \sum_{k=1}^r \sigma_k \mathbf{f}_i^T \mathbf{u}_k \mathbf{v}_k^T \mathbf{g}_j \quad (i, j) \in \Omega$$

Again we see, that if the right( $\mathbf{v}_k$ ) and left( $\mathbf{u}_k$ ) singular vectors are highly coherent with  $\mathbf{g}_j$  and  $\mathbf{f}_i$  then again we will get almost all zero matrix with few non-zero elements. Thus, all that is needed is that the column and row spaces of  $\mathbf{M}$  be respectively incoherent with the basis ( $\mathbf{f}_i$ ) and ( $\mathbf{g}_i$ ).

## Question 5

- **Title:** Active Contours with Group Similarity
- **Venue:** Department of ECE, The Hong Kong University of Science and Technology; Image Analysis and Processing Group, Yale University
- **Problem Addressed:** Active contours are widely used in image segmentation. To cope with missing or misleading features in images, researchers have introduced various ways to model the prior of shapes and use the prior to constrain active contours. However, the shape prior is usually learnt from a large set of annotated data, which is not always accessible in practice. Moreover, it is often that the existing shapes in the training set are



not sufficient to model the new instance in the testing image. In this paper, the author proposes to use the group similarity of object shapes in multiple images as a prior to aid segmentation.

- **Method and Objective Function:** The paper proposes to represent each shape contour as a vector. Shape contours of similar objects(may have different size, rotations) are put into a matrix form. Intuitively, this matrix must have small rank because rank gives us an indication of the correlation among the vectors of a matrix. For example, the rank equals to 1 if the shapes are identical, and the rank may increase if some shapes change. Intrinsically, the rank of the shape matrix describes the degree of freedom of the shape change. The low-rank constraint will allow the global change of contours such as translation, scaling, rotation and principal deformation to fit the image data while truncating the local variation caused by image defects. The objective function they seek to minimize is:

$$\min_{\mathbf{X}} \sum_i f_i(C_i) \quad \text{subject to } \text{rank}(\mathbf{X}) \leq K$$

where  $C_i$  are the contours,  $\mathbf{X} = [\mathbf{C}_1 \ \mathbf{C}_2 \dots \mathbf{C}_n]$  and  $f_i(\cdot)$  is the energy of the active contour. For region-based energy it is given by:

$$f_i(C_i) = \int_{\Omega_1} (I_i(x) - u_1)^2 dx + \int_{\Omega_2} (I_i(x) - u_2)^2 dx + \beta \text{length}(C_i)$$

where  $I_i(x)$  is the intensity of  $i$ th image at pixel location  $x$ ,  $\Omega_1$  and  $\Omega_2$  are the regions inside and outside of the contour respectively and  $u_1$  and  $u_2$  are the mean intensity of pixels inside and outside contour respectively. Since the given objective function is difficult to optimize due to discrete nature a relaxed version of the optimization is given by:

$$\min_{\mathbf{X}} \left( \sum_i f_i(C_i) + \lambda \|\mathbf{X}\|_* \right)$$

where  $\|\mathbf{X}\|_*$  is the nuclear norm of  $\mathbf{X}$ .