Assignment 5: CS 754, Advanced Image Processing

Due: 15th April before 11:55 pm

1. Consider an inverse problem of the form $\boldsymbol{y} = \mathcal{H}(\boldsymbol{x}) + \boldsymbol{\eta}$ where $\boldsymbol{y}$ is the observed degraded and noisy image, $\boldsymbol{x}$ is the underlying image to be estimated, $\boldsymbol{\eta}$ is a noise vector, and $\mathcal{H}$ represents a transformation operator. In case of denoising, this operator is represented by the identity matrix. In case of compressed sensing, it is the sensing matrix, and in case of deblurring, it represents a convolution. The aim is to estimate $\boldsymbol{x}$ given $\boldsymbol{y}$ and $\mathcal{H}$ as well as the noise model. This is often framed as a Bayesian problem to maximize $p(\boldsymbol{x}|\boldsymbol{y}, \mathcal{H}) \propto p(\boldsymbol{y}|\boldsymbol{x}, \mathcal{H})p(\boldsymbol{x})$. In this relation, the first term in the product on the right hand side is the likelihood term, and the second term represents a prior probability imposed on $\boldsymbol{x}$.
   With this in mind, we refer to the paper 'User assisted separation of reflections from a single image using a sparsity prior' by Anat Levin, IEEE Transactions on Pattern Analysis and Machine Intelligence. Answer the following questions:

   - In Eqn. (7), explain what $A_{j\rightarrow}$ and b_j represent, for each of the four terms in Eqn. (6).
   - In Eqn. (6), which terms are obtained from the prior and which terms are obtained from the likelihood? What is the prior used in the paper? What is the likelihood used in the paper?
   - Why does the paper use a likelihood term that is different from the Gaussian prior? [7+12+6=25 points]

   Sol:

   (a) The Equation (7) in the paper is shown below:

   $$J_3(\nu) = \sum_j \rho_j(A_{j\rightarrow}\nu - b_j) \tag{1}$$

   Here $A_{j\rightarrow}$ is the j^{th} row of $\mathbf{A}$ which represents the forward filter matrix for computing the derivative of the vectorized image I_1 i.e. ν. Applying $A_{j\rightarrow}$ on ν gives us the derivative centered on the j^{th} pixel. If pixel j belongs to the set S_1, b_j is taken to be the input image derivative. This corresponds to the second term of Equation (6). If pixel j belongs to the set S_2, b_j is taken to be 0. This corresponds to the last term of Equation (6). ρ_j can be modified to be $(\rho(1 + \lambda)$to account for the first two terms of Equation (6) in the modified matrix representation of Equation (6) as shown in Equation (7).

   (b) The terms obtained from the likelihood in Equation (6) are $\rho(f_{i,k}.I_1)$ and $\rho(f_{i,k}(I - I_1))$ where $\rho(.)$ represents the Laplacian Mixture Model. The terms obtained from the prior in Equation (6) are $f_{i,k}(I - I_1)$ and $f_{i,k}.I_1$. The Laplacian Mixture Model is used for computing the likelihood in the paper and the log-histogram of the derivative filters was used as the prior in the paper.

   (c) The Laplacian Mixture Model models the natural images sparsely by using it to maximize the likelihood of the output filter histogram i.e. our prior. Using a Gaussian prior does not work as well here because the Gaussian distribution is not sparser than the Laplacian distribution itself as shown in Figure 1. Therefore, the Laplacian model is used because the sparseness of the derivative filter output is a robust property of natural images and we are in fact interested in separating the reflected images in natural images only.

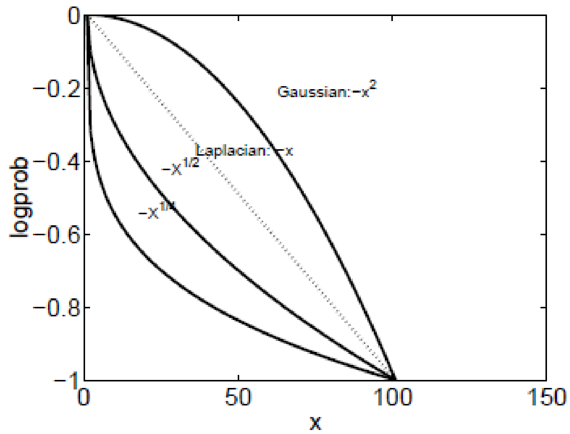

Figure 1: Sparseness of probability distributions

2. Consider compressive measurements of the form $\boldsymbol{y} = \boldsymbol{\Phi x} + \boldsymbol{\eta}$ under the usual notations with $\boldsymbol{y} \in \mathbb{R}^m, \boldsymbol{\Phi} \in \mathbb{R}^{m\times n}, m \ll n, \boldsymbol{x} \in \mathbb{R}^n, \boldsymbol{\eta} \sim \mathcal{N}(0, \sigma^2 \boldsymbol{I}_{m\times m})$. Instead of the usual model of assuming signal sparsity in an orthonormal basis, consider that $\boldsymbol{x}$ is a random draw from a zero-mean Gaussian distribution with known covariance matrix $\boldsymbol{\Sigma_x}$ (of size $n \times n$). Derive an expression for the maximum a posteriori (MAP) estimate of $\boldsymbol{x}$ given $\boldsymbol{y}, \boldsymbol{\Phi}, \boldsymbol{\Sigma_x}$. Also, run the following simulation: Generate $\boldsymbol{\Sigma_x} = \boldsymbol{U\Lambda U}^T$ of size 128×128 where $\boldsymbol{U}$ is a random orthonormal matrix, and $\boldsymbol{\Lambda}$ is a diagonal matrix of eigenvalues of the form $ci^{-\alpha}$ where $c = 1$ is a constant, i is an index for the eigenvalues with $1 \leq i \leq n$ and α is a decay factor for the eigenvalues. Generate 10 signals from $\mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma_x})$. For $m \in \{40, 50, 64, 80, 100, 120\}$, generate compressive measurements of the form $\boldsymbol{y} = \boldsymbol{\Phi x} + \boldsymbol{\eta}$ for each signal $\boldsymbol{x}$. In each case, $\boldsymbol{\Phi}$ should be a matrix of iid Gaussian entries with mean 0 and variance $1/m$, and $\sigma = 0.01\times$ the average absolute value in $\boldsymbol{\Phi x}$. Reconstruct $\boldsymbol{x}$ using the MAP formula, and plot the average RMSE versus m for the case $\alpha = 3$ and $\alpha = 0$. Comment on the results - is there any difference in the reconstruction performance when α is varied? If so, what could be the reason for the difference? [25 points]

Sol:

We have:

$$\mathbf{y} = \mathbf{\Phi x} + \eta \tag{2}$$

with $\boldsymbol{y} \in \mathbb{R}^m, \boldsymbol{\Phi} \in \mathbb{R}^{m\times n}, m \ll n, \boldsymbol{x} \in \mathbb{R}^n, \boldsymbol{\eta} \sim \mathcal{N}(0, \sigma^2 \boldsymbol{I}_{m\times m})$ and $\boldsymbol{\Phi} \sim \mathcal{N}\left(0, \frac{1}{m}\right)$, and $\sigma = 0.01\times$the average absolute value of $mathbf\Phi x$. $\mathbf{x}$ is smapled from a Gaussian distribution with mean 0 and covariance matrix($\mathbf{\Sigma_x}$).

The MAP estimate of $\mathbf{x}$ given $\mathbf{y}$, $\mathbf{\Phi}$, and $\mathbf{\Sigma_x}$:

$$\begin{aligned} \hat{\mathbf{x}} &= argmax_{\mathbf{x}}\ p(\mathbf{x}|\mathbf{y}) \\ \implies \hat{\mathbf{x}} &= argmax_{\mathbf{x}}\ p(\mathbf{y}|\mathbf{x})p(\mathbf{x}) \end{aligned}$$

We know that $p(\mathbf{x})$ is a joint Gaussian random vector and $p(\mathbf{y}|\mathbf{x})$ is the distribution of the noise η.

$$p(\mathbf{y}|\mathbf{x})p(\mathbf{x}) = \exp\left(-\frac{||\mathbf{y} - \mathbf{\Phi x}||^2}{2\sigma^2}\right)\left(\frac{\exp\left(-\frac{1}{2}\mathbf{x^T \Sigma_x^{-1} x}\right)}{(2\pi)^{\frac{n}{2}}\sqrt{|\mathbf{\Sigma_x}|}}\right) \tag{3}$$

Taking log of the posterior probability since it is a non-decreasing function retains the same optimization problem:

$$\hat{\mathbf{x}} = argmax_{\mathbf{x}}\left(\frac{(\mathbf{y} - \mathbf{\Phi x})^T(\mathbf{y} - \mathbf{\Phi x})}{2\sigma^2} + \frac{1}{2}\mathbf{x^T \Sigma_x^{-1} x}\right) \tag{4}$$

Taking derivative of the LHS gives us:

$$\begin{aligned} 2\left(\frac{\mathbf{\Phi^T \Phi \hat{x}} - \mathbf{y \Phi^T}}{2\sigma^2}\right) + \frac{1}{2}.2.\mathbf{\Sigma_x}^{-1}\hat{\mathbf{x}} &= 0 \\ \implies \hat{\mathbf{x}} &= (\mathbf{\Phi^T \Phi} + \sigma^2 \mathbf{\Sigma_x^{-1}})^{-1}\mathbf{\Phi^T y} \end{aligned} \tag{5}$$

The RMSE vs. m is obtained by computing the RMSE over 10 signal samples randomly drawn from a Gaussian distribution using the covariance matrix. This is carried out for different values of α. Log of the RMSE is taken for better visualization. Moreover, the RMSE decreases as α increases. The RMSE vs. m plots for different values of α is shown in Figure 2.

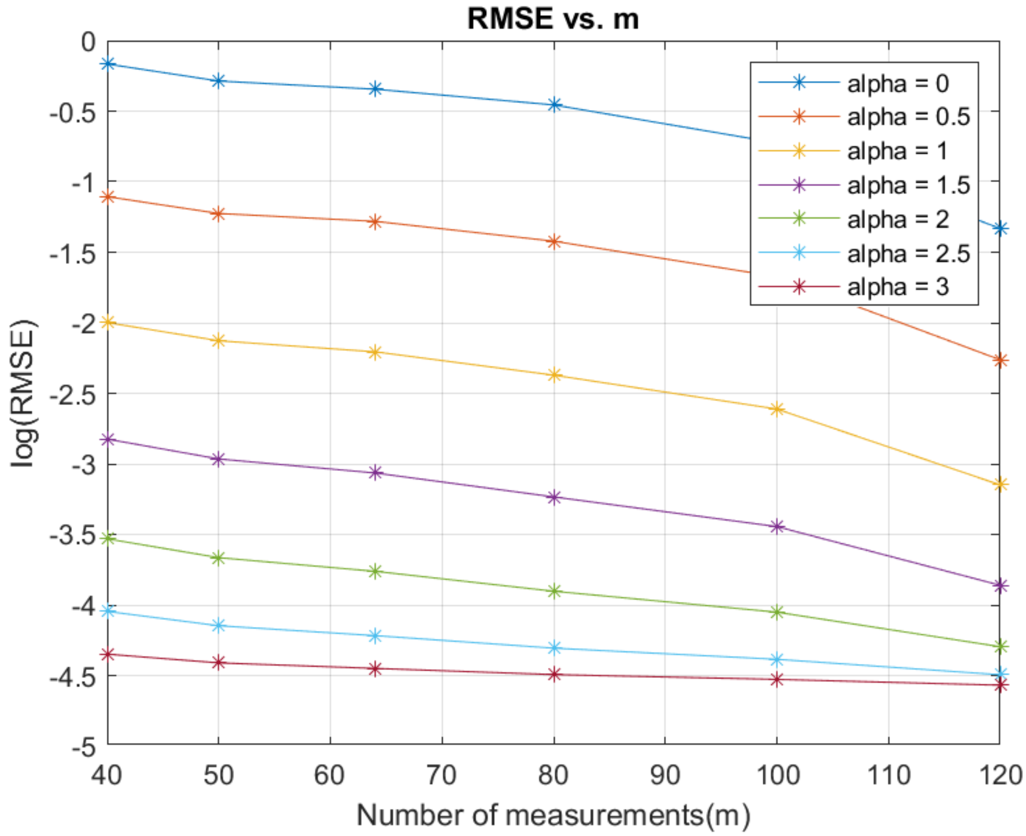

Figure 2: RMSE vs. m plot

Magnitude of the eigenvalues of a covariance matrix gives us a measure of the extent of spread of a random vector along the mean of the distribution. Since we are sampling $\mathbf{x}$ from a Gaussian distribution with mean 0, increasing α increases sparsity of $\mathbf{x}$. Therefore, for a given number of measurements m, we get a lower error as α increases as the number of actual number of non-zero values in $\mathbf{x}$ decreases. The other point that increasing the number of measurements m decreases RMSE is obvious enough.

3. Read through the proof of Theorem 3.3 from the paper 'Guaranteed Minimum-Rank Solutions of Linear Matrix Equations via Nuclear Norm Minimization' from the homework folder. This theorem refers to the optimization problem in Eqn. 3.1 of the same paper. Answer all the questions highlighted within the proof. You may directly use linear algebra results quoted in the paper without proving them from scratch, but mention very clearly which result you used and where. [12 × 2 +1 = 25 points]

Sol:

(a) **Justify that $\|X_0\|_* \geq \|X^*\|_*$**
We know that X^* is defined as the solution of the convex optimization problem

$$X^* := argmin_X \ \ \|X\|_* \qquad \text{s.t.} \qquad A(X) = b$$

and X_0 is a matrix of rank r such that $A(X_0) = b$.
Hence by definition X^* is matrix with minimum nuclear norm among all the matrices that satisfy the condition $A(X) = b$ and we know that the condition is satisfied by X_0. Hence the nuclear norm of X_0 should be greater than or equal to the nuclear norm of X^*. i.e.,

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

(b) **Explain how lemma 2.3 leads to the equality $\|X_0 + R_c\|_* - \|R_0\|_* = \|X_0\|_* + \|R_c\|_* - \|R_0\|_*$**
Lemma 2.3 : Let A and B be matrices of the same dimensions. If AB^T = 0 and A^TB = 0 then $\|A + B\|_* = \|A\|_* + \|B\|_*$.
We have R := $X^* - X_0$, Using Lemma 3.4, R is divided into R_0 and R_c such that R = $R_0 + R_c$, $rank(R_0) \leq 2rank(X_0)$ and $X_0 R_c^T = 0$ and $X_0^T R_c = 0$ and by construction X_0 and R_c have same dimensions. Hence X_0 and R_c satisfy all the conditions for lemma 2.3. Hence,

$$\|X_0 + R_c\|_* = \|X_0\|_* + \|R_c\|_* \tag{6}$$

Using this on part a we get,

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

Last part used reverse triangle inequality

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

Here we used the result form Lemma 2.3

(c) **Justify** $\sigma_k \leq \sum_{j \in I_i} \sigma_j \quad \forall k \in I_{i+1}$
We know that in singular value decomposition (SVD) the singular values are arranged in the descending order along the diagonal of matrix σ. Now the index sets I_i are each 3r length and disjoint by construction as they all 3r indices are consecutive.
Now consider a singular value σ_k such that $k \in I_{i+1}$. Then due to the descending order for every $j \in I_i$, we have

$$\sigma_k \leq \sigma_j \forall j \in I_i$$

Applying summation along the 3r values of j

$$\sum_{j \in I_i} \sigma_k \leq \sum_{j \in I_i} \sigma_j$$

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

$$\sigma_k \leq \frac{1}{3r} \sum_{j \in I_i} \sigma_j \qquad \forall k \in I_{i+1}$$

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

We know that the nuclear norm is sum of the singular values, i.e.,

$$\|R_i\|_* = \sum_{j \in I_i} \sigma_j$$

Now using the inequality from part (c), for any $k \in I_{i+1}$ we have

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

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

squaring on both sides

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

Summing over all 3r values of k

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

But we know that the Frobenius norm $\|R_{i+1}\|_F^2 = \sum_{k \in I_{i+1}} \sigma_k^2$. Therefore,

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

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

(e) **Show that** $\sum_{j \geq 2} \|R_j\|_F \leq \frac{1}{\sqrt{3r}} \sum_{j \geq 1} \|R_j\|_*$

From part (d), we have

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

applying square root on both sides

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

Summing the above inequality for all i

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

As nuclear norm is always non negative

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

Replacing i+1 by i in LHS

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

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

(f) **Justify** $\frac{1}{\sqrt{3r}}\|R_c\|_* \leq \frac{1}{\sqrt{3r}}\|R_0\|_*$

From part (b), we have

$$\|X_0\|_* \geq \|X_0\|_* + \|R_c\|_* - \|R_0\|_*$$
$$\|R_0\|_* \geq \|R_c\|_*$$
$$\text{Dividing both sides by } \sqrt{3r}$$
$$\frac{1}{\sqrt{3r}}\|R_c\|_* \leq \frac{1}{\sqrt{3r}}\|R_0\|_*$$

(g) **Justify** $\frac{1}{\sqrt{3r}}\|R_0\|_* \leq \frac{\sqrt{2r}}{\sqrt{3r}}\|R_0\|_F$
As seen in part (b), by construction, rank(R_0) $\leq$ 2rank(X_0) and rank(X_0) = r. Hence,

$$rank(R_0) \leq 2r$$

Now using the linear algebra result (2.1) from the paper, The relation between nuclear norm and frobenius norm.

$$\|X\|_* \leq \sqrt{rank(X)}\|X\|_F$$
$$\text{Note: above inequality follows from relation between L1 and L2 norm of a vector } \|x\|_1 \leq \sqrt{n}\|x\|_2$$
$$\implies \|R_0\|_* \leq \sqrt{2r}\|R_0\|_F$$
$$\text{Dividing both sides by } \sqrt{3r}$$
$$\frac{1}{\sqrt{3r}}\|R_0\|_* \leq \frac{\sqrt{2r}}{\sqrt{3r}}\|R_0\|_F$$

(h) **Justify that rank of $R_0 + R_1$ is atmost 5r**
R_c is partitioned into sum of R_1, R_2, ..., each of rank atmost 3r. Hence by construction, we have rank(R_1) $\leq$ 3r and as seen in part (g), rank(R_0) $\leq$ 2r. Now using the sub-additivity property of the rank (from linear algebra),i.e.,

$$rank(A + B) \leq rank(A) + rank(B)$$
$$\implies rank(R_0 + R_1) \leq 2r + 3r = 5r$$

Therefore rank of $R_0 + R_1$ is atmost 5r

(i) **Justify** $\|A(R)\| \geq \|A(R_0 + R_1)\| - \sum_{j\geq 2}\|A(R_j)\|$
We know that,

$$R = R_0 + R_c = R_0 + \sum_{j\geq 1} R_j = R_0 + R_1 + \sum_{j\geq 2} R_j$$

Hence we use triangle inequity twice as follows.

$$\text{First we use the reverse triangle inequality}$$
$$\|A(R)\| = \|A(R_0 + R_1 + \sum_{j\geq 2} R_j)\| \geq \|A(R_0 + R_1)\| - \|A(\sum_{j\geq 2} R_j))\|$$
$$\text{But using the triangle inequality on the second term}$$
$$\|A(\sum_{j\geq 2} R_j))\| \leq \sum_{j\geq 2}\|A(R_j)\|$$
$$\implies \|A(R)\| \geq \|A(R_0 + R_1)\| - \|A(\sum_{j\geq 2} R_j))\| \geq \|A(R_0 + R_1)\| - \sum_{j\geq 2}\|A(R_j)\|$$
$$\|A(R)\| \geq \|A(R_0 + R_1)\| - \sum_{j\geq 2}\|A(R_j)\|$$

(j) **Justify** $\|A(R_0 + R_1)\| - \sum_{j\geq 2} \|A(R_j)\| \geq (1-\delta_{5r})\|R_0 + R_1\|_F - (1+\delta_{3r}) \sum_{j\geq 2} \|R_j\|_F$
From part (h), we have rank of $R_0 + R_1$ is atmost 5r and by construction, rank of R_j is atmost 3r for all j $\geq$ 2. Hence by using the restricted isometric property of A for order 5r and 3r,

$$\|A(R_0 + R_1)\| \geq (1-\delta_{5r})\|R_0 + R_1\|_F$$
$$\|A(R_j)\| \leq (1+\delta_{3r})\|R_j\|_F \qquad \forall j \geq 2$$

substituting these inequalities in (i), we have,

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

(k) **By what assumption can we say A(R) = A(X^* - X_0) = 0?**
The assumption that A is an affine transformation will give us the above equality. i.e,

$$A(R) = A(X^* - X_0) = A(X^*) - A(X_0)$$

Last step is possible because A is affine

$$\implies A(R) = b - b = 0$$

Since A(X^*) = b and A(X_0) = 0

(l) **Show that RHS of inequality (3.7) is positive when** $9\delta_{3r} + 11\delta_{5r} < 2$
The inequality (3.7) of the paper is

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

We have seen in part(k) that A(R) = 0. Hence if we make the factor in RHS to be strictly positive, then we will force R_0 to be zero which will imply $R_c = 0$ and hence $X^* = X_0$. So we want,

$$\left((1-\delta_{5r}) - \frac{9}{11}(1+\delta_{3r})\right) > 0$$
$$\frac{2 - 11\delta_{5r} - 9\delta_{3r}}{11} > 0$$
$$\implies 2 > 11\delta_{5r} + 9\delta_{3r}$$

Hence the required condition is

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

4. Read section 1 of the paper 'Exact Matrix Completion via Convex Optimization' from the homework folder. Answer the following questions: (1) Why do the theorems on low rank matrix completion require that the singular vectors be incoherent with the canonical basis (i.e. columns of the identity matrix)? (2) How would this coherence condition change if the sampling operator were changed to the one in Eqn. 1.13 of the paper? (3) The paper gives an example of a matrix which is low rank but cannot be recovered from its randomly sampled entries. What is that example and why cannot it not be recovered by the techniques in the paper? [5+5+5=15 points]
Sol:

(1) We are primarily interested in subspaces with low coherence as matrices whose column and row spaces have low coherence cannot really be in the null space of the sampling operator. That is incoherence of singular vectors with the canonical basis is required for matrix recovery to avoid the cases where the low rank matrix takes the form like all zeroes except 1 non zero element at random or very few non zero elements in the matrix. In this case the matrix M is highly likely to be in the null space of the sampling operator. And hence there is no way to recover this rank-1 matrix until we basically see all its entries. Hence when the singular vectors are sufficiently spread in the standard basis, we will require less number of observations to recover the low-rank matrix. We need both the left and the right singular vectors to be uncorrelated with the standard basis.

(2) When the sampling operation is changed as in equation 1.13 of the paper. i.e.,

$$f_i^T M g_j$$

Using the rank r representation of matrix M

$$f_i^T M g_j = \sum_{k=1}^{r} \sigma_k f_i^T u_k v_k^T g_j$$

Now we will again get low rank matrices with very few non zero elements if f_i are highly correlated with u_k or g_j are highly correlated with v_k. Then we will again be unable to recover the low rank matrix exactly with few measurements. Hence the new coherence condition is:
The right singular vectors (u_k) must be incoherent with basis f_i and the left singular vectors (v_k) must be incoherent with basis g_j.

(3) The example is a rank 2 symmetric matrix given by:

$$M = \sum_{k=1} 2\sigma_k u_k u_k^T, \qquad u_1 = (e_1 + e_2)/\sqrt{2}, u2 = (e_1 - e_2)/\sqrt(2)$$

e_1 and e_2 are standard basis vectors

And the singular values are arbitrary. This matrix vanishes everywhere except in the top-left corner and one would basically need to see all the entries of M to able to recover this matrix exactly by not only the techniques from this paper but by any method whatsoever.

5. Read section 5.9 of the paper 'Low-Rank Modeling and Its Applications in Image Analysis' from the homework folder. You will find numerous image analysis or computer vision applications of low rank matrix modelling and/or RPCA, which we did not cover in class. Your task is to glance through any one of the papers cited in this section and answer the following: (1) State the title and venue of the paper; (2) Briefly explain the problem being solved in the paper; (3) Explain how low rank matrix recovery/completion or RPCA is being used to solve that problem. Write down the objective function being optimized in the paper with meaning of all symbols clearly explained. [10 points]
Sol:

- **Title:** Parsing Façade with Rank-One Approximation
- **Venue:** 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA.
- **Problem: Façade parsing**. The purpose of façade parsing is to discover different regions of a façade with semantic meanings. Potential applications include reconstruction of buildings, navigation systems and location-based services, which require relatively simplified yet structurally faithful model representations.

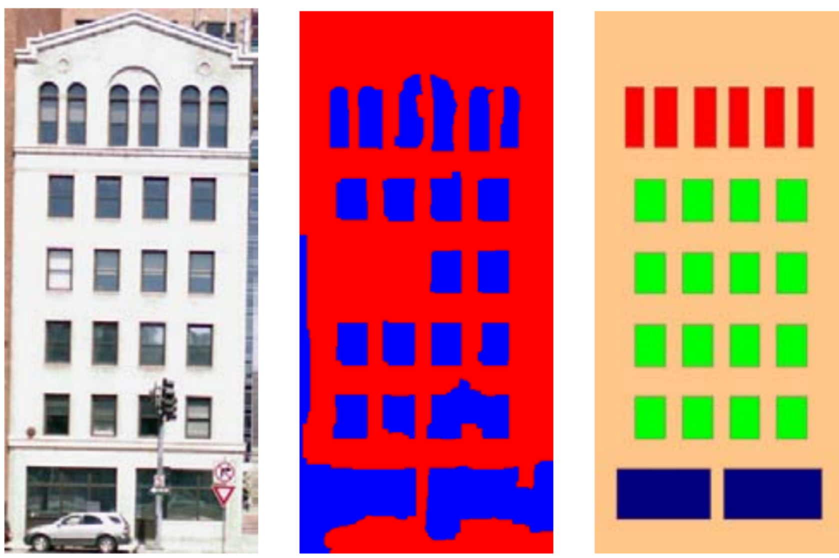

Figure 1. A façade is parsed into blocks of aligned and repetitive objects. Left: The input façade image. Middle: The initial wall/non-wall classification. Right: The parsed blocks with their repetitive objects, where a different color codes a different block.

- **How Low rank matrix approximation is being used:** paper follows three stages, described below:
  - **Façade rectification and low-rank extraction:** to globally rectify the input façade image and extract the low-rank texture that recovers the regularity of the façade.
  - **Wall/non-wall classification:** This step utilizes low-level color information and uses randomized forest as classifier to generate an initial classification for the input façade image. The output returns a pixel-wise classification of two labels: wall and non-wall.
  - **Rank-one approximation and block-wise partition** : In this step we seek to optimally partition the façade into blocks, and approximate each block with a rank-one matrix. The block-wise partition should best match with the classification generated at stage 2.

  The final step is where we use the Low rank (rank-1) approximation. This rank-1 approximation problem is similar to Robust PCA

(a) (b) (c) (d)

Figure 3. Three stages of our parsing procedure. (a) The input façade image. (b) The rectified façade with the low-rank part, which removes the traffic light and regularizes the shapes of the first rows of windows. (c) The initial classification gives two missing windows due to different window color. (d) The final parsing result displays three blocks with aligned patterns. Two missing windows in (c) are recovered by approximating the middle block with a rank-one matrix.

- **Objective function being minimized:**

$$\min_{M^0, E} \parallel E \parallel_1 \text{ subject to } M = M^0 + E; \text{ rank}(M^0) = 1,$$

Where M is the input matrix, M^0 is the rank 1 approximation of M and E is the error matrix which is optimized to be sparse. The low rank approximation problem is similar to Robust PCA and hence the authors use Augmented Lagrangian multiplier method which is shown to work for Robust PCA.

Remarks

- The function used for generating a random orthonormal matrix in Question 2 is taken from MathWorks File Exchange accessible here.
- The project folder for this assignment is shared in the github repository that can be accessed from here.