Assignment 5: CS 754, Advanced Image Processing

Mohit Kumar Meena - 213070021
Shashwat Pathak - 213070010

1. - In Eqn. (7), explain what $A_{j\to}$ 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?

Solution: (a)
Eqn (7) is given as

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

By notation, $A_{j\to}$ is the j^{th} row and it corresponds to the derivative filters centered on pixel j that needs to be applied on vectorized image $\mathbf{v}$, ρ is the scaler to scaler function and b_j either has a derivative of I centered at j or it's zero. Now eqn (6)

$$J_2(I_1) = \sum_{i,k} \rho(f_{i,k}\cdot I_1) + \rho(f_{i,k}\cdot(I - I_1)) + \lambda\sum_{i\in S_1,k}\rho(f_{i,k}\cdot I_1 - f_{i,k}.I) + \lambda\sum_{i\in S_2,k}\rho(f_{i,k}\cdot I_1) \tag{2}$$

Since the derivative filters are linear so we can write

$$f_{i,k}\cdot I = F_{i,k}vec(I) = p_{j,k} \quad (let)$$

(a) In the first term,

$$\rho(f_{i,k}\cdot I_1) = \rho(F_{i,k}vec(I_1) = \rho(F_{i,k}v)$$

Now comparing it with (1) $A_{j\to} = F_{i,k}, b_j = 0$ and $\rho_j(.) = \rho(.)$

(b) In the second term,

$$\rho(f_{i,k}\cdot(I - I_1)) = \rho(f_{i,k}\cdot(I) - f_{i,k}I_1 = \rho(F_{i,k}vec(I) - F_{i,k}v = \rho(p_{i,k} - F_{i,k}v)$$

comparing with (1), we get $A_{j\to} = -F_{i,k}, b_j = p_{j,k}$ and $\rho_j(.) = \rho(.)$

(c) In the third term,

$$\lambda\rho(f_{i,k}\cdot I_1 - f_{i,k}\cdot I) = \lambda\rho(F_{i,k}vec(I_1) - F_{i,k}vec(I)) = \lambda\rho(F_{i,k}v - p_{i,k})$$

comparing with (1), we get $A_{j\to} = F_{i,k}, b_j = p_{j,k}$ and $\rho_j(.) = \lambda\rho(.)$

(d) In the fourth term,

$$\lambda\rho(f_{i,k}\cdot I_1) = \lambda\rho(F_{i,k}vec(I_1)) = \lambda\rho(F_{i,k}v)$$

comparing with (1), we get $A_{j\to} = F_{i,k}, b_j = 0$ and $\rho_j(.) = \lambda\rho(.)$

(b). We can break (2) into two parts (let $\mathbf{A}$ and $\mathbf{B}$)

$$J_2(I_I) = \mathbf{A} + \mathbf{B}$$

$$\mathbf{A:}\sum_{i,k}\rho(f_{i,k}I_1) + \rho(f_{i,k}(I - I_1))$$

$$\textbf{B:} \sum_{i\in S_1,k} \rho(f_{i,k}I_1 - f_{i,k}\cdot I) + \sum_{i\in S_2,k} \rho(f_{i,k}\cdot I_1)$$

Here, **A** has derived from the prior function and **B** has been derived from the likelihood function.
Prior model: It's been mentioned in the paper, mixture of laplacian distribution is sparse in nature. Mixture of two laplacian distribution is given by

$$Pr(x) = \frac{\pi}{2s_1} e^{-|x|/s_1} + \frac{\pi}{2s_2} e^{-|x|/s_2}$$

So, the sparsity prior is used in the paper and this can be achieved by mixing a narrow laplacian distribution centered on zero and a broad laplacian distributions centered on zero.
Likelihood: It can be noted from (2) that expression depends upon the gradients of sets of image location. i.e. $\sum_{i\in S_1,k} \rho(f_{i,k}\cdot I_1 - f_{i,k}\cdot I)$ imposes gradients in location S_1 belong to layer 1 and $\sum_{i\in S_2,k} \rho(f_{i,k}\cdot I_1)$ imposes gradients in location S_2 belongs to layer 2.
Therefore, the likelihood used in the paper depends upon the parameters like the input image and the gradients in the sets of image locations.

(c). The main reason is that the gaussian distributions is not sparse because the term e^{-x^2} is always above the straight line, but the laplacian distribution is exactly at the borders between sparse and non-sparse as shown in figure 1. Moreover, the mixture of laplacian model is sparse in nature as discussed in part (b).

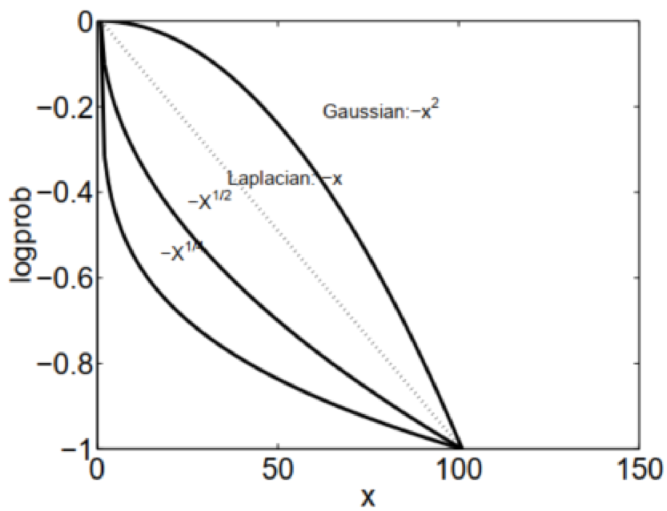


Figure 1: log probabilities plot of Gaussian and Laplacian distributions [paper]

In terms of results, the laplacian gives the decomposition for the image while the gaussian prior causes the decomposition to split edges into two low contrast edges, rather than putting the entire contrast in one of the layers as shown in figure 2. Since gaussian likelihood gives poor results in separation of reflections so the likelihood other than gaussian is used in paper.

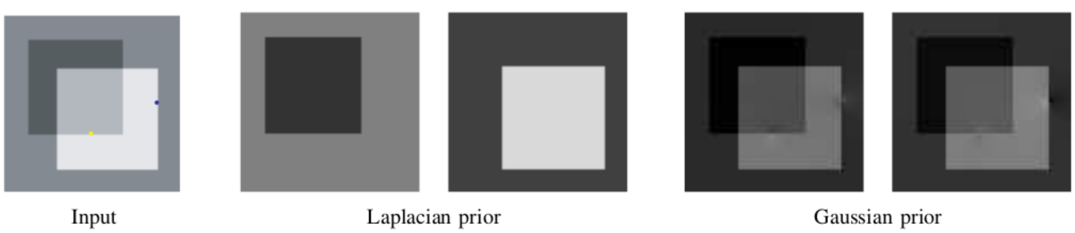
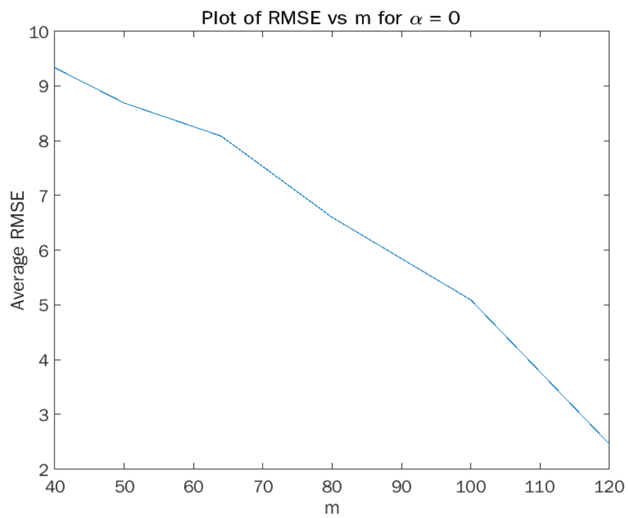


Figure 2: input (left) results of laplacian prior (middle) and gaussian prior (right) [paper]

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 $1 \le i \le n$ 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?

The following plots were obtained for different values of α.



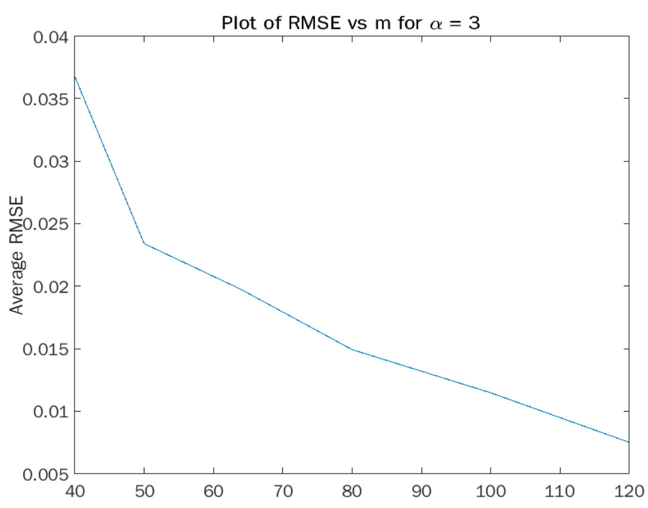


Figure 3: (Top) Average RMSE vs m for $\alpha = 0$, (Bottom) Average RMSE vs m for $\alpha = 3$

Comments : We observed that with increase in α there was a sharp decrease in the RMSE between the estimated and the original signal with increasing measurements. We believe this is intuitive as with increase in m, more number of measurements are there in the sensed signal which allows for better reconstruction with more information. Moreover, with a higher weighted diagonal matrix for $\alpha = 3$ we get better results as the map estimate is obtained from a better prior Gaussian representation than in the case of $\alpha = 0$.

Homework 5:

2) Derivation of MAP estimate of x given y, ϕ and Σ_x

Derivation:

We know that

$$y = \phi x + n$$

x → original signal R^n

n → Gaussian noise ~ $N(0, \sigma^2)$

ϕ → sensing matrix $R^{m \times n}$

y → measured signal (R^m)

Assumption: $x \in N(0, \Sigma_x)$

x → $n \times 1$; 0 → mean; Σ_x → $n \times n$

Now

$$x_{MAP} = \underset{x}{\operatorname{argmax}}\, P(y|x, \phi)\, p(x)$$

$p(x)$ → Prior probability

$P(y|x, \phi)$ → Posterori probability of y given x and ϕ.

Using the formula for multivariate gaussians, we know,

$$P(y|x,\phi) = \frac{1}{(\sigma 2\pi)^{m/2}} \exp\left(-\frac{\|y - \phi x\|_2^2}{2\sigma^2}\right)$$

and

$$P(x) = \frac{1}{|\Sigma|^{1/2}(2\pi)^{1/2}} \exp\left(-x^t \Sigma^{-1} x\right).$$

$L = \underset{x}{\text{argmax}}\ P(y|x, \phi)\, P(x)$.

$\Rightarrow \log L = \underset{x}{\text{argmax}}\ \log P(y|x, \phi) + \log P(x)$,

$\because$ log is a monotonically decreasing function.

thus,

$$\hat{x}_{MAP} = \underset{x}{\text{argmax}}\ \log P(y|x, \phi) + \log P(x).$$

$\Rightarrow$

$$\hat{x}_{MAP} = \underset{x}{\text{argmin}} \underbrace{\left(\frac{\|y - \phi x\|_2^2}{2\sigma^2}\right)}_{\text{Posterior term}} + \underbrace{x^t \Sigma^{-1} x}_{\text{Prior term}} + \underbrace{C}_{\text{constant}}.$$

thus, on solving we get.

$$\boxed{\hat{x}_{MAP} = \left(\frac{\phi^T \phi}{2\sigma^2} + \Sigma^{-1}\right)^{-1} \phi^T y.}$$

through Woodbury's decomposition of $(A+B)^{-1}$.
we get.

$$\boxed{\hat{x}_{MAP} = \left(\Sigma - \Sigma \phi^T [\sigma^2 I + \phi \Sigma \phi^T]^{-1}\right) \phi^T y}$$

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.

   ******* Solutions are pasted below *******

Solution 3 (Homework - 5)

(a) We have X^* such that

$$X^* = \arg\min_X \|X\|_* \quad \text{s.t } A(X) = b \quad (\text{Eq } 3.1)$$

Here X^* is a matrix with minimum nuclear norm subject to the condition $A(X) = b$ and X_0 which is a matrix of rank r also satisfies $A(X_0) = b$, Hence

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

i.e the nuclear norm of X_0 should be more than that of the nuclear norm of X^*.

(b) In theorm (3.3) of paper, $X_0^T R_c = 0$ and $X_0 R_c^T = 0$ also X_0 and R_c have dimension by construction, Therefore X_0 and R_c satisfy the conditions of Lemma (2.3) of paper, i.e.

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

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

Note: $\|X_0 + R\|_* \leq \|X_0\|_*$

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

i.e $$\boxed{\|R_0\|_* \geq \|R_c\|_*}$$

(c) σ is obtained by singular value decomposition (SVD) of R_c.

Also σ_i's are the singular values of R_c with

$$\sigma_i \geq \sigma_j \quad \forall \; 1 \leq i \leq j \leq \text{rank}(R_c)$$

i.e $\quad \sigma_j \geq \sigma_K \quad \forall j \in I_i = \{3r(i-1)+1, \dots, 3ri\},$

$\quad \forall K \in I_{i+1} = \{3ri+1, \dots, 3r(i+1)\}$

This results,

$$\sum_{j \in I_i} \sigma_j \geq 3r\,\sigma_K$$

i.e $$\boxed{\sigma_K \leq \frac{1}{3r} \sum_{j \in I_i} \sigma_j} \quad \forall K \in I_{i+1} \quad \text{——— (1)}$$

(d) Using Eqn (1) and keeping in mind that $\|R_i\|_* = \sum_{j \in I_i} \sigma_j$, we get

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

Now squaring both the sides, we get

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

using the addition of $3r$ inequalities, we get

$$\sum_{K \in I_{i+1}} \sigma_K^2 \leq \frac{3r}{9r^2} \|R_i\|_*^2 \quad \text{——— (2)}$$

But we know that $\|R_{i+1}\|_F^2 = \sum_{K \in I_{i+1}} \sigma_K^2$, so we get from Eqn (1)

$$\boxed{\|R_{i+1}\|_F^2 \leq \frac{1}{3r} \|R_i\|_*^2} \quad \text{——— (3)}$$

(e) Using eqn (3), i.e

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

Now adding the equations for $i \in \{1, 2, 3, 4, \dots\}$ to get

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

Simplifying further,

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

(f) Using sub-additivity property of nuclear norm and exploiting the fact that all R_j have orthogonal row space and column space we get,

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

Using the result of part (B) i.e $||R_0||_* \geq ||R_c||_*$, we get

$$\sum_{j_{max} \geq j \geq 2} ||R_j||_F \leq \frac{1}{\sqrt{3r}} \sum_{j_{max} \geq j \geq 1} ||R_j||_* = \frac{1}{\sqrt{3r}} ||R_c||_* \leq \frac{1}{\sqrt{3r}} ||R_0||_* \quad (4)$$

(g) Using Lemma (3.4) of the paper and we know that rank of X_0 is r, we get

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

We know that, $||P||_* \leq \sqrt{\text{rank}(P)}\, ||P||_F$

using the above result on R_0, we get

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

Now using eqn (4), we get

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

$$\leq \frac{1}{\sqrt{3r}} ||R_0||_* \leq \frac{\sqrt{2r}}{\sqrt{3r}} ||R_0||_F$$

(h) It is mentioned in the paper that for matrices the rank function is subadditive. so

$$\|R_0 + R_1\|_* \le \|R_0\|_* + \|R_1\|_*$$

From Lemma 3.4, we know that $\|R_0\|_* \le 2r$

also by construction $\operatorname{rank}(R_1) \le 3r$, so

$$\|R_0 + R_1\|_* \le 2r + 3r = 5r$$

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

(i) By construction of R_0 and R_c, we get

$$\|A(R)\| = \|A(R_0 + R_1 + \sum_{j\ge 2} R_j\|$$

since A is a linear map, so

$$\|A(R)\| = \|A(R_0 + R_1) + \sum_{j\ge 2} A(R_j)\|$$

Now using Triangle inequality, we get

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

i.e

$$\boxed{\|A(R)\| \ge \|A(R_0 + R_1)\| - \sum_{j\ge 2} \|A(R_j)\|} \quad \text{—— (5)}$$

(j) From the results of part (h) we know that rank of $R_0 + R_1$ is atmost $5r$ and by restricted isometry property, the rank of R_j is atmost $3r$ $\forall j \ge 2$. i.e

$$\|A(R_0 + R_1)\| \ge (1 - \delta_{5r}) \|R_0 + R_1\|_F \quad \text{—— (x)}$$

and

$$\|A(R_j)\| \le (1 + \delta_{3r}) \|R_j\|_F \quad \forall j \ge 2 \quad \text{—— (y)}$$

Substituting Eqn (x) and Eqn (y) in Eqn (5), we get

$$\|A(R)\| \geq \|A(R_0+R_1)\| - \sum_{j\geq 2} \|A(R_j)\| \qquad (\text{Eq}^n (5))$$

i.e

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

(k) We know that A is an affine transformation, also

$$A(X^*) = b \quad \text{and} \quad A(X_0) = b$$

therefore,

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

$$A(R) = b - b$$

$$\boxed{A(R) = 0}$$

(l) The inequality (3.7) is given as

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

Since $A(R) = 0$ so $\|A(R)\| = 0$ i.e

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

Here $\|R_0\|_F > 0$ (Frobenius norm is never negative)

so, simplifying further

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

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

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

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? ******* Solutions are pasted below *******

Solution 4 : (a)

We know that coherence of subspace U of $\mathbb{R}^n$ with respect to cannonical basis e_i is defined as:

$$\mu(U) = \frac{n}{r} \max_{1 \le i \le n} \|P_U e_i\|^2$$

Here, P_U is the orthogonal projection onto U, by this we mean that $P_U = U(U^T U)^{-1} U^T$.

Now suppose a sampling matrix M which has its SVD as $U \Sigma V^T$, then in this case we would be requiring a bound on coherence of both matrices U and V to be bounded by some constant. In simple words, there should exist a positive no. μ_0 such that

$$\max\{\mu(U), \mu(V)\} \le \mu_0 .$$

Incoherence of singular vector with cannonical basis is required for matrix recovery to avoid cases when the low rank matrix takes the form such as all zeros except 1 at some random position. Most of the time we would observe small variation among the sampled values when the random elements of such matrices are sampled.

It should be noted that we need the singular vectors to be uncorrelated with the standard basis to reduce the no. of observations needed to recover low-rank matrix. Also, the matrices whose column and row spaces have low coherence cannot really be in the null space of the sampling operator.

(b) Equation 1.13 is given for two orthonormal bases $f_1, f_2, \dots f_n$ and $g_1, g_2 \dots g_n$ of $\mathbb{R}^n$ in terms of rank minimization problem, i.e

$$\text{minimize} \quad \text{rank}(X) \quad \text{subject to} \quad f_i^* X g_j = f_i^* M g_j, \quad (i,j) \in \Omega.$$

When the sampling operator were changed to above optimization problem, it's not necessary that sampled matrix would now be incoherent with canonical basis. Now, we would be requiring that our rows and columns to be incoherent w.r.t to new bases. This can be done as follows:

- There exist unitary transformations F and G such that $e_j = F f_j$ and $e_j = G g_j$ for each $j = 1, \dots n$, Hence
$$f_i^* X g_j = e_i^* (F X G^*) e_j$$

- If the condition of theorem 1.3 (mentioned in paper) holds for the matrix FXG^*, then nuclear norm minimization finds the unique optimal solution of Eqn (1.13) when large enough random collection of inner products $f_i^* M g_j$ were provided.

- Hence, we need that columns and row spaces of M be respectively incoherent with the basis (f_i) and (g_i).

(c) As given in page 3, equation 1.1 of paper, i.e rank 1 matrix M

$$M = e_1 e_n^* = \begin{bmatrix} 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & & 0 & 0 \end{bmatrix}$$

Since there's only 1 in top-right corner and all other entries are 0 so clearly it would be a low rank matrix and this matrix cannot be recovered from a sampling of it's entries unless we pretty much see all the entries.

The reason is that for most sampling sets, we would only get to see zeroes so that we would have no way of guessing that the matrix is not zero.

One cannot hope to reconstruct any low-rank matrix M - even of rank 1 - if the sampling set avoids any column or row of M. i.e Suppose the M is of rank 1 and of the form xy^*, $x, y \in \mathbb{R}^n$ so that (i,j)th entry is given by

$$M_{ij} = x_i y_j$$

Then if we do not have samples from the first row for example, one could never guess the value of the first component x_1, by any method.

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. ******* Solutions are pasted below *******

5 Title: Finding Correspondence from Multiple Images via Sparse and Low-Rank Decomposition

Venue: ECCV (European Conference on Computer Vision) 2012

Problem Statement

The objective here was to find visual pattern correspondence between multiple images. This correspondence is done on the information that the rank of ordered patterns from a set of linearly correlated images should be lower than that of the disordered pattern. Also, the error among the reordered patterns are sparse.
The problem is formulated as the estimation of a partial permutation matrix that minimizes the rank of the ordered features/patches.

Low Rank Matrix Recovery.

Since the ordered patterns correspond to a low rank matrix. Thus, estimation of a optimal partial permutation matrix to minimize the rank of ordered features can be used. Additionally, a sparse error term is also introduced to improve the robustness to gross corruption and occlusion

Images N: $I_1, I_2, \cdots I_N$.
for each image $k \rightarrow$ features $f_{n,1}, f_{n,2} \cdots f_{n,k}$
at $\bar{k}$ land mark locations.

$$F_n = [f_1, \cdots, f_{\bar{k}}] \in \mathbb{R}^{d \times \bar{k}}$$

$$A = [vec(F_1) \mid vec(F_2) \mid \cdots \mid vec(F_N)] \in \mathbb{R}^{d\bar{k} \times N}$$

should be approximately low-rank.

$$P_n = \{P_n \mid P_n \in \{0,1\}^{k_n \times \bar{k}}, \ 1_{k_n}^T P_n = 1_{\bar{k}}^T, \ P_n 1_k \leq 1_k\}.$$

$$[vec(F_1P_1) \mid \cdots \mid vec(F_N P_N)] \in \mathbb{R}^{d\bar{k} \times N}$$

$$\min_{P_n \in P_n|_{n=1}^{N}} rank(L) \text{ such that } [vec(F_1P_1) \mid \cdots \mid vec(F_nP_N)] = L$$

$$\min_{L, E, \theta_1 \cdots \theta_N} \|L\|_* + \lambda \|E\|_1.$$

$\hookrightarrow$ error term.

4). Theorem on low rank matrix completion require singular vectors to be incoherent with the canonical basis (i.e Columns of the identity matrix)