Assignment 5: CS 754, Advanced Image Processing

Shreyas Narahari-203050037
Nitish Gangwar-203050069

April 20, 2021

Answer 1:

(a.)

The function is

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

we know,

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

$$J_3(v) = \rho_1(A_{1->v} - b_1) + \rho_2(A_{2->v} - b_2) + \rho_3(A_{3->v} - b_3) + \rho_4(A_{4->v} - b_4)$$

we know that,
Matrix **A** contains rows that denotes the derivative filters. **v** is the vectorized version of the image I_1.Now the above given terms denotes the convolution operation of f_k with I_k.

$$J(I_1) = \sum_{i,k} \rho((f_k * I_1)_i) + \rho((f_k * (I - I_1))_i) + \lambda \sum_{i \in S_1,k} \rho((f_k * I - f_k * I_1)_i) + \lambda \sum_{i \in S_2,k} \rho((f_k * I_1)_i)$$

and this equation is expressed as

$$J(v) = \sum_{j=1}^{4} \sum_k \rho(A_{j,k})$$

where $A_{j,k}$ is a matrix corresponding to filter f_k.

(b.)

The used function is:

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

Here,

$$\sum_{i,k} \rho(f_{i,k}.I_1) + \rho(f_{i,k}.(I - I_1)) \quad is \quad the \quad prior.$$

$$\sum_{i \in S_1,k} \rho(f_{i,k}.I_1 - f_{i,k}.I) \quad is \quad the \quad likelihood \quad term.$$

$$\sum_{i \in S_2,k} \rho(f_{i,k}.I_1) \quad is \quad the \quad likelihood \quad term.$$

first two terms are priors and last two terms are likelihood.
Prior over images is shown below:

$$Pr(I) \approx \prod_{i,k} Pr(f_{i,k}.I)$$

where,
f denotes the linear filter.
I denotes the image.
$f_{i,k}$ denotes the k'th derivative filter centered at pixel i.
Likelihood used in the paper is shown below:

$$logPr(f_{i,k}.I) \approx \rho(f_{i,k}.I)$$

$$\rho(x) = log(\frac{\pi_1}{2s_1}e^{-|x|/s_1} + \frac{\pi_2}{2s_2}e^{-|x|/s_2})$$

(c.)

Likelihood used in the paper is:

$$logPr(f_{i,k}.I) \approx \rho(f_{i,k}.I)$$

$$\rho(x) = log(\frac{\pi_1}{2s_1}e^{-|x|/s_1} + \frac{\pi_2}{2s_2}e^{-|x|/s_2})$$

which is an approximation to Laplacian mixture model and is different from Gaussian likelihood. Authors demonstrated the results by using different types of likelihood over the image patches selected at random from natural images.They marked gradients at random over the images and tried decomposing the image.They measured the absolute difference between the recovered layers and ground truth layers.From the result it can be clearly seen that Gaussian prior, results in a bad decomposition whereas the sparse prior and Laplacian outperforms the Gaussian prior.

Answer 2:

We consider compressive measurements of the form $y = \phi x + \eta$ where x is randomly drawn from a zero-mean Gaussian distribution with co-variance matrix Σ_x.
Given y, ϕ, Σ_x the maximum a posterior (MAP) estimate of x can be calculated as,

$$\hat{x} = argmax_x \; p(x \mid y, \phi)$$

From Bayes rule we get,

$$\hat{x} = argmax_x \; \frac{p(y \mid x, \phi) p(x)}{p(y)}$$

As $p(y)$ is not dependent on x we can rewrite the above equation as,

$$\hat{x} = argmax_x \; p(y \mid x, \phi) p(x)$$

From definition we know that,

$$p(x) = \frac{1}{\mid \Sigma_x \mid^{0.5} \, (2\pi)^{(n/2)}} exp(-x^t \Sigma_x x/2)$$

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

Substituting we get,

$$\hat{x} = argmax_x \; \frac{1}{\sigma^{m/2}(2\pi)^{m/2}} exp\left(-\frac{\|y - \phi x\|_2^2}{2\sigma^2}\right) \frac{1}{\mid \Sigma_x \mid^{0.5} \, (2\pi)^{(n/2)}} exp(-x^t \Sigma_x x/2)$$

As we only need the index of the maximum probability we can take logarithm on the R.H.S as logarithm is a increasing function,

$$\hat{x} = argmax_x \; log\left(\frac{1}{\sigma^{m/2}(2\pi)^{m/2}} exp\left(-\frac{\|y - \phi x\|_2^2}{2\sigma^2}\right) \frac{1}{\mid \Sigma_x \mid^{0.5} \, (2\pi)^{(n/2)}} exp(-x^t \Sigma_x x/2)\right)$$

Expanding,

$$\begin{aligned} \hat{x} = argmax_x \; log\left(\frac{1}{\sigma^{m/2}(2\pi)^{m/2}}\right) + log\left(exp\left(-\frac{\|y - \phi x\|_2^2}{2\sigma^2}\right)\right) + \\ log\left(\frac{1}{\mid \Sigma_x \mid^{0.5} \, (2\pi)^{(n/2)}}\right) + log\left(exp(-x^t \Sigma_x x/2)\right) \end{aligned} \tag{1}$$

We can take out the terms $log\left(\frac{1}{\sigma^{m/2}(2\pi)^{m/2}}\right)$ and $log\left(\frac{1}{|\Sigma_x|^{0.5}(2\pi)^{(n/2)}}\right)$ as **constants** as they will not affect the overall maximum index,

$$\hat{x} = argmax_x \; -\frac{\|y-\phi x\|_2^2}{2\sigma^2} - x^t\Sigma_x x/2 + \textbf{constants}$$

Inverting the sign inside $argmax_x$ we get $argmin_x$,

$$\hat{x} = argmin_x \; \frac{\|y-\phi x\|_2^2}{2\sigma^2} + x^t\Sigma_x x/2 + \textbf{constants}$$

Here we can see that we are finding out the index where the function is at its minimum. This can be done by taking the derivative of the right-hand side $\frac{\|y-\phi x\|_2^2}{2\sigma^2} + x^t\Sigma_x x/2 + \textbf{constants}$ and equating it to 0 we get,

$$-\frac{2\phi^T(y-\phi x)}{2\sigma^2} + \frac{2\Sigma^{-1}x}{2} = 0$$

Cancelling the constants,

$$-\frac{\phi^T(y-\phi x)}{\sigma^2} + \Sigma^{-1}x = 0$$

Rearranging,

$$\frac{\phi^T y - \phi^T\phi x}{\sigma^2} = \Sigma^{-1}x$$

Finally,

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

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

Answer 3:

(1).
Proof for:

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

Proof: As we know from theorem 3.2 of this paper that X_0 is the only matrix of rank at-most r which satisfies

$$A(X) = b$$

hence for other we say

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

proof for the optimality is given through contradiction:
Let us take a matrix of rank r which satisfies

$$A(X) = b$$

such that $X \neq X_0$.Then, Z has been taken as

$$Z = X_0 - X$$

where Z is a non zero matrix of rank at most 2r and we have A(Z)=0.but we know that

$$0 = ||A(Z)|| \geq (1 - \delta_{2r})||Z||_F$$

which is clearly a contradiction.This is the proof which is even given in the same paper.
(2).
Considering from triangle inequality we get, $\|X_0 + R_c\|_* - \|R_0\|$ and it is given that from the decomposition of the matrix R that $X_0R_c' = 0$ and $X_0'R_c = 0$. As this is true we can apply lemma 2.3 which states if $AB' = 0$ and $A'B = 0$ then $\|A + B\|_* = \|A\|_* + \|B\|_*$ applying this we get,

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

(3).
From the definition of the notation given in the paper we know that σ_i denotes the i^{th} largest singular value, here as the smallest element in the set I_{i+1} is $3ri+1$ which

is larger than all the indices in the set I_i so we can say that this inequality is trivially true. Thus,

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

(4).

we consider the result $\sigma_k \leq \frac{1}{3r} \sum_{j \in I_i} \sigma_j$ from the previous part. Form the definition of forbenius norm of a rank r matrix we consider,

$$\|A\|_F = \sqrt{\sigma_1^2 + \cdots + \sigma_r^2}$$

As we know that $\sigma_2, \cdots, \sigma_r \leq \sigma_1$ we get,

$$\sqrt{\sigma_1^2 + \cdots + \sigma_r^2} \leq \sqrt{\sigma_1^2 + \cdots + \sigma_1^2} = \sqrt{r}\sigma_1$$

Thus,

$$\|A\|_F = \sqrt{r}\sigma_1 \tag{2}$$

For converting the singular value to forbenius norm we can consider the index largest singular value of the set I_{i+1} ie., σ_{3r+1} which will hold for the inequality $\sigma_{3r+1} \leq \frac{1}{3r} \sum_{j \in I_i} \sigma_j$, squaring on both sides,

$$\sigma_{3r+1}^2 \leq \frac{1}{3r}^2 (\sum_{j \in I_i} \sigma_j)^2$$

Rearranging,

$$(3r)\sigma_{3r+1}^2 \leq \frac{1}{3r} (\sum_{j \in I_i} \sigma_j)^2$$

From equation 2 and the definition of nuclear norm we can get,

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

(5).

As it has been defined that $i \geq 1$ we sum up the previous on L.H.S and R.H.S accordingly to get,

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

(6).

As we already know $\|R_0\|_* \geq \|R_c\|_*$. Substituting it in the equation $\frac{1}{\sqrt{3r}}\|R_c\|_*$ we get,

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

(7).

From equation 2.1 in the paper we know that $\|X\|_* \leq \sqrt{r}\|X\|_F$ and from Theorem 3.2 statement in the paper we know that rank of X_0 is r so as $rank(R_0) \leq 2\ rank(X_0)$ we get that upper bound for $rank(R_0)$ is $2r$. Using this and the relation on the equation from the previous part we get,

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

(8).

As we already know from the properties of rank of matrix that $rank(A+B) \leq rank(A)+rank(B)$. From previous part we know that $rank(R_0) \leq 2r$ and $rank(R_1) \leq 3r$ from definition so we get,

$$rank(R_0 + R_1) \leq rank(R_0) + rank(R1) \leq 5r$$

(9).

It is already mentioned that $R = R_0 + R_c$ and R_c is the sum of matrices $R_1, R_2, \ldots,$ from this we can say,

$$R = R_0 + R_1 + R_2 \ldots$$

Applying linear transform on both sides,(The equality still holds from the definition of linear transform)

$$\mathcal{A}(R) = \mathcal{A}(R_0 + R_1 + R_2 + \ldots)$$

Applying operator norm on both sides we get,

$$\|\mathcal{A}(R)\| = \|\mathcal{A}(R_0 + R_1 + R_2 + \ldots)\|$$

Using following inequalities

$$||X|| \leq ||X||_F \leq ||X||_* \leq \sqrt{r}||X||_F \leq r||X||.$$

and using following fact that rank$(R_0) \leq$ 2r.

Using the property of triangle inequality of operator norm,

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

As all the terms in the above equation are positive we can tun this into a inequality by introducing a negative term in the second term,

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

(10).

We consider the inequality from the previous part. For the first term we need to consider the lower bound from "Definition 3.1" from the paper and rank of the $(R_0 + R_1)$ as $5r$,

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

For the second term we first consider the complete inequality expression from "Definition 3.1",

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

Multiplying by -1 reverses the inequalities giving,

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

Substituting the lower bound we get,

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

(11.)

From "Theorem 3.1" and "Theorem 3.2" we assume that, $\mathcal{A}(X^*) = b$ and $\mathcal{A}(X_0) = b$. And know that $R = X^* - X_0$ applying linear transformation we get,

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

(12).
As the right-hand side has to be positive,

$$\left((1-\delta_{5r})-\frac{9}{11}(1+\delta_3 r)\right)>0$$

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

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

Rearranging we get,

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

Answer 4:

(1)

Theorems on low rank matrix completion require that the singular vectors be incoherent with the canonical basis because the matrix that has to be recovered cannot lie in the null space of the sampling vectors of a matrix M.
Multiple examples of matrix have been used in the paper like top left corner of 2×2,top right corner entry is non zero rest all entries are zero.For such type of matrices we need to see all the entries for matrix recovery.Due to this they want that singular vectors have to be sufficiently spread i.e. incoherent with the canonical basis.

(2)

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

Coherence condition when sampling operator were changed to the **f** and **g**.Then in this case the column and row spaces of matrix M should be incoherent with the basis **f** and **g**.

Answer 5:

(1)

Link to the paper: click here
Title: Low-rank matrix recovery with structural incoherence for robust face recognition.
Venue: June 2012 Proceedings / CVPR, IEEE Computer Society Conference on Computer Vision and Pattern Recognition.
DOI: 10.1109/CVPR.2012.6247981

(2)

The paper deals with the problem of robust face recognition, with training and test image data being corrupted due to occlusion and disguise.Most of the previous literature does not consider data to be contaminated during training. As the training data is assumed to be taken under well maintained environment. So, they discarded the corrupted training images which could lead to over fitting and small sample size.multiple techniques have been employed in the past like Eigenfaces,Laplacian faces etc but these techniques are not robust to outliers or special cases of disguise and occlusion.Low rank matrix recovery method is used for the following part which shows good results.This method identifies a set of representative basis for the corrupted training data.To be specific authors have used low-rank matrix approximation with structural incoherence.

(3)

Low-rank matrix recovery seeks to decompose a data matrix D into A + E where
A is low-rank matrix.
E is sparse error.
So, LR minimizes the rank of matrix A and reduces the $||E||_0$ in order to approximate to data matrix D.But this is an NP-hard problem and hence following optimization problem is being solved.

$$min_{A,E}||A||_* + \lambda||E||_1 \quad s.t. \quad D = A + E$$

Where terms used are:
$||A||_*$ is nuclear norm which denotes the sum of singular values which approximates

the rank of matrix A.
$||E||_1$ denotes the l1 norm which includes the sum of absolute values of E.
solving this optimization problem is equivalent to solving the l0 containing term which is NP-hard in nature.To solve the above written optimization problem authors have used augmented lagrange multipliers(ALM) for efficient computation.
Optimization function being used are:

$$min_{A,E} \sum ||A_i||_* + \lambda||E_i||_1 + \eta \sum_{j \neq i} ||A_j^T A_i||_F^2$$
$$s.t. \quad D_i = A_i + E_i$$

for class-wise optimization problem the above problem is manipulated to

$$min_{A_i,E_i} ||A_i||_* + \lambda||E_i||_1 + \eta \sum_{j \neq i} ||A_j^T A_i||_F^2$$
$$s.t. \quad D_i = A_i + E_i$$

further changes are done to make problem more tractable by using the following property

$$||A_j^T A_i||_F^2 \leq ||A_j||_F^2 ||A_i||_F^2$$

hence, the optimization problem becomes

$$min_{A_i,E_i} ||A_i||_* + \lambda||E_i||_1 + \eta' ||A_i||_F^2$$
$$s.t. \quad D_i = A_i + E_i$$

Now, let us see which variable denotes what in the above stated equations
A is low-rank matrix.
E is sparse error.
η is a parameter that balances the low-rank matrix approximation and matrix incoherence.
where,
$\sum_{j \neq i} ||A_j^T A_i||_F^2$ denotes the sum of frobenius norms between each pair of the low-rank matrices A_i and A_j.It is an regularization term added to the objective function to include the incoherence between low-rank matrices.

$$\eta' = \eta \sum_{j \neq i} ||A_j||_F^2$$

D denotes the data matrix.which is being decomposed into **A** and **E**.We know that A is low rank matrix and E is sparse error matrix.

References:

1. https://webee.technion.ac.il/people/anat.levin/papers/Assisted-Reflections-Levin-Weiss-PAMI.pdf
2. https://arxiv.org/abs/0706.4138
3. https://arxiv.org/abs/1401.3409
4. https://arxiv.org/abs/0805.4471
5. Low-rank_matrix_recovery_with_structural_incoherence_for_robust_face_recognition
6. http://web.stanford.edu/class/archive/ee/ee263/ee263.1082/hw/hw9sol.pdf
7. https://math.stackexchange.com/questions/1671911/squared-reverse-triangle-inequality