Problem 3:

We have $\mathcal{A}: \mathbb{R}^{m \times n} \to \mathbb{R}^p$ as a linear map.
X_0 is rank r matrix. $b := \mathcal{A}(X_0)$
We obtain estimate X^* as:

$$X^* := \underset{X}{\text{argmin}} \|X\|_* \quad \text{s.t.} \quad \mathcal{A}(X) = b.$$

Q1) Since X^* has minimum $\|X\|_*$ among all X satisfying $\mathcal{A}(X) = b$, we can say
$\|X\|_* \geq \|X^*\|_*$ $\forall$ X satisfying $\mathcal{A}(X) = b$.
Since X_0 satisfies $\mathcal{A}(X_0) = b$, we can conclude:

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

Q2) Lemma 2.3 states that for matrices A, B of same dimension satisfying $AB' = 0$ and $A'B = 0$, we have $\|A+B\|_* = \|A\|_* + \|B\|_*$
In this case we have decomposed R as
$R = R_0 + R_c$ with $\text{rank}(R_0) \leq 2\,\text{rank}(X_0)$, and $X_0 R_c' = 0$ and $X_0' R_c = 0$. So by lemma 2.3,
$\Rightarrow \|X_0 + R_c\|_* = \|X_0\|_* + \|R_c\|_*$, which explains the last equality in (3.3).

Q3) here we have split R_c into sum of $R_1, R_2, \ldots$
as $R_i := U_{I_i} \, \text{diag}(\sigma_{I_i}) V_{I_i}'$. $I_i = \{3r(i-1)+1, \ldots, 3ri\}$
now for some $k \in I_{i+1}$, we know
$k > 3ri \Rightarrow \sigma_k \leq \sigma_{3ri}$ (since singular values are arranged in decreasing order)
so

$$\sigma_k \leq \sigma_{3ri} = \frac{1}{3r} \sum_{m=1}^{3r} \sigma_{3ri} \leq \frac{1}{3r} \sum_{m=1}^{3r} \sigma_{3r(i-1)+m}$$

$$\Rightarrow \sigma_k \leq \frac{1}{3r} \sum_{m=1}^{3r} \sigma_{3r(i-1)+m}$$

$(\because 3ri \geq (3r(i-1)+m)$ $\forall 1 \leq m \leq 3r)$
$\hookrightarrow$ so $\sigma_{3ri} \leq \sigma_{3r(i-1)+m}$

$$= \frac{1}{3r} \sum_{j \in I_i} \sigma_j \quad (j = 3r(i-1) + m)$$

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

Q4)

$$\|R_{i+1}\|_F^2 = \sum_{k \in I_{i+1}} \sigma_k^2$$ (squared sum of singular values)

$$\leq \sum_{k \in I_{i+1}} \left(\frac{1}{3r} \sum_{j \in I_i} \sigma_j \right)^2$$ (from result of Q3))

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

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

$$= \frac{1}{(3r)^2} \times 3r \times \|R_i\|_*^2$$ (I_{i+1} has $3r$ elements)

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

Q5) From the previous result: $\left(\|R_{i+1}\|_F \leq \frac{1}{\sqrt{3r}} \|R_i\|_* \right)$

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

and $$\sum_{j \geq 1} \|R_j\|_* = \sum_{j \geq 1} \sum_{k \in I_j} \sigma_k = \sum_{k \in \bigcup_{j \geq 1} I_j} \sigma_k$$

$$= \sum_{k \geq 1} \sigma_k = \|R_c\|_*$$ (since I_j's are disjoint and $\bigcup_{j \geq 1} I_j$ is complete set of indices)

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

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

Q6) In (3.3) of the paper, we already showed that:

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

$$\Rightarrow \boxed{\|R_0\|_* \geq \|R_c\|_*}$$

So from this, we can directly say:

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

Q7) In equation (2.1) of the paper, it is stated that

$\|X\|_* \leq \sqrt{r}\, \|X\|_F$ where r is the rank of matrix X.

now, we know

$$\text{rank}(R_0) \leq 2\,\text{rank}(X_0) = 2r$$

so,

$$\|R_0\|_* \leq \sqrt{\text{rank}(R_0)}\, \|R_0\|_F \leq \sqrt{2r}\, \|R_0\|_F$$

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

Q8) We know that for matrices A, B of same dimension:

$$\text{rank}(A+B) \leq \text{rank}(A) + \text{rank}(B)$$

This is a standard result in linear algebra. This is because if A has atmost r_1 linearly independent columns and B has atmost r_2 linearly independent columns then $A+B$ will have atmost r_1+r_2 linearly independant columns, as these columns are linear combination of columns of A and B.

So $\operatorname{rank}(R_0 + R_1) \leq \operatorname{rank}(R_0) + \operatorname{rank}(R_1)$
$\leq 2r + 3r$
($\operatorname{rank}(R_1) \leq 3r$ by construction and $\operatorname{rank}(R_0) \leq 2r$).

$\Rightarrow \operatorname{rank}(R_0 + R_1) \leq 5r.$

Q9) Triangular inequality: $\|x+y\| \geq \|x\| - \|y\|$
Using this, we can write:

$$\|A(R)\| = \|A(R_0 + R_c)\| \qquad (\because R_c = \sum_{j\geq 1} R_j)$$
$$= \|A(R_0 + R_1 + \sum_{j\geq 2} R_j)\|$$
$$= \|A(R_0+R_1) + \sum_{j\geq 2} A(R_j)\|$$
$$\geq \|A(R_0+R_1)\| - \|\sum_{j\geq 2} A(R_j)\| \qquad \text{(Triangular inequality)}$$

Also
$$\sum_{j\geq 2} \|A(R_j)\| \geq \|\sum_{j\geq 2} A(R_j)\| \qquad \text{(Triangular inequality)}.$$

So we get:
$$\|A(R)\| \geq \|A(R_0+R_1)\| - \|\sum_{j\geq 2} A(R_j)\|$$
$$\geq \|A(R_0+R_1)\| - \sum_{j\geq 2} \|A(R_j)\|$$

$\therefore$ Proved.

Q10) We know $\operatorname{rank}(R_0 + R_1) \leq 5r$. By RIP of A:
$$\|A(R_0+R_1)\| \geq (1-\delta_{5r}) \|R_0 + R_1\|_F$$
Also $\operatorname{rank}(R_j) \leq 3r \;\forall\, j \geq 2$. So by RIP of A:
$$\|A(R_j)\| \leq (1+\delta_{3r}) \|R_j\|_F$$
Therefore:
$$\|A(R_0+R_1)\| - \sum_{j\geq 2} \|A(R_j)\| \geq (1-\delta_{5r}) \|R_0 + R_1\|_F - (1+\delta_{3r}) \sum_{j\geq 2} \|R_j\|_F$$

Q11) We have assumed that $A(X_0) = b$ and since X^* also satisfies the constraint $A(X^*) = b$ in the optimisation problem, we have

$$A(R) = A(X^* - X_0) = A(X^*) - A(X_0)$$
$$= b - b = 0 \qquad (\because A \text{ is linear})$$
$$\Rightarrow A(R) = 0$$

Q12)
We need: $(1 - \delta_{5r}) - \frac{9}{11}(1 + \delta_{3r}) > 0$

for right hand side to be positive.

$$\Leftrightarrow 11(1 - \delta_{5r}) - 9(1 + \delta_{3r}) > 0$$
$$\Leftrightarrow (11 - 9) - 11\delta_{5r} - 9\delta_{3r} > 0$$
$$\Leftrightarrow 2 - 11\delta_{5r} - 9\delta_{3r} > 0$$
$$\Leftrightarrow 2 > 9\delta_{3r} + 11\delta_{5r} \Leftrightarrow \boxed{9\delta_{3r} + 11\delta_{5r} < 2}$$

Problem 1:

In Eqn. (6), we have the objective function:

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

now we can define ρ_j, $A_{j\rightarrow}$ and b_j as follows to get eqn. (7): $(1 \le j \le 4)$

$$A_{1\rightarrow} = F := \begin{bmatrix} F_1 \\ F_2 \\ \vdots \\ F_K \end{bmatrix}, \quad b_1 = 0.$$

and $\boxed{\rho_1(v) := \sum_i \rho(v_i)}$ for a vector v as input.

here $F_1, F_2, \ldots, F_K$ are the K derivative filter matrices (K is the total no. of filters applied). So $F_i v$ is the vectorised output of the ith derivative filter ~~on the~~ applied to the image I_1 corresponding to v. (that is, $F_i v = \text{vec}(f_i * I_1)$).

and we define F as: $F := \begin{bmatrix} F_1 \\ \vdots \\ F_K \end{bmatrix}$

F is concatenation of all filters.

So, similarly: $A_{2\rightarrow} = -F$, $b_2 = -F\,\text{vec}(I)$

and $\boxed{\rho_2(v) := \sum_i \rho(v_i)}$, for vector inputs v.

$$\boxed{\begin{gathered} A_{3\rightarrow} = F, \quad b_3 = F\,\text{vec}(I) \\ \rho_3(v) := \lambda \sum_i \rho(v_i) \end{gathered}},$$

$\boxed{A_{4\to} = F\ ,\ b_4 = 0,\ \ \rho_4(v) := \lambda \sum_i \rho(v_i)}$ for input vectors v.

with the $A_{j\to}$, b_j, ρ_j, we get the equivalent objective function:

$$J_3(v) = \sum_{j=1}^{4} \rho_j(A_{j\to} v - b_j), \quad v = \text{vec}(I_1)$$

- Now let us mention the prior terms and likelihood terms in eqn (6)

the first summation containing the first two terms is obtained from the sparse ~~prior~~ prior, i.e., the terms:

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

and the likelihood terms are:

likelihood terms $\rightarrow$ $\lambda \sum_{i \in S_1, k} \rho(f_{i,k} \cdot I_1 - f_{i,k} \cdot I)$

$$+ \lambda \sum_{i \in S_2, k} \rho(f_{i,k} \cdot I_1)$$

which are obtained by enforcing agreement between the labelled and image gradients.

The prior used in the paper is a sparse distribution obtained by a mixture of two Laplacian distributions:

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

This prior was applied to the entire image as:

~~log Pr~~ $\quad Pr(I) \approx \prod_{i,k} Pr(f_{i,k} \cdot I)$

We make use of the negative log of this probability and define:

$$\rho(x) = -\log(Pr(x))$$

The likelihood in this paper uses the same sparse distribution as the prior. But, the likehood is used to impose agreement of derivative of I with I_1 on set S_1 and with I_2 on set S_2.

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

$$Pr(I_1 | I, S_1, S_2) \approx \prod_{i \in S_2, k} Pr(f_{ik} \cdot I_1) \prod_{i \in S_1, k} Pr(f_{ik} \cdot I_1 - f_{ik} \cdot I)$$

there's also a factor λ that is different between the prior and likelihood.
We take negative log likelihood and add it to the prior log terms to get the objective.

- We can see that the paper uses a sparse distribution instead of a gaussian for its likelihood term. This is based on statistics of images in which the logarithm of the histogram lies below the straight line connecting the maximum and minimum ~~values~~ values. The gaussian distribution is, on the other hand, above this line, which means it does not show a "sparse distribution", in line with image statistics.
  A distribution like the ~~gaussian~~ gaussian that is above the line (non sparse) will ~~prefer~~ prefer to split an edge of unit contrast into two edges (one in each layer) with half the contrast, while sparse

distributions (below the line) will prefer decompositions with edge appearing in only one of the layers. So it is not a good idea to use a non sparse distribution (on or above the line) for this problem.

The two laplacian mixture distribution used in this paper is sparse and so it is suitable for the problem.

classmate

Date

Page

Q2: $y = \phi x + \eta$, $x \in \mathbb{R}^n$, $\phi \in \mathbb{R}^{m \times n}$, $m <<< n$,

$\eta \sim \mathcal{N}(0, \sigma^2 I_{m\times m})$

The MAP estimator for x is,

$$\hat{x} = \arg\max_x p(x|y)$$

$$= \arg\max_x \frac{p(y|x)\, p(x)}{p(y)}$$

$$= \arg\max_x p(y|x)\, p(x)$$

$x \sim \mathcal{N}(0, \Sigma_x)$ and $\eta \sim \mathcal{N}(0, \sigma^2 I_{m\times m})$

$$\therefore \hat{x} = \arg\max_x \exp\left(\frac{-\|y-\phi x\|^2}{2\sigma^2}\right) \frac{\exp\left(\frac{-1}{2} x^T \Sigma_x^{-1} x\right)}{(2\pi)^{N/2} |\Sigma_x|^{1/2}}$$

$$= \arg\min_x \left\{ \frac{\|y-\phi x\|^2}{2\sigma^2} + \frac{1}{2} x^T \Sigma_x^{-1} x \right\}$$

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

$$\therefore \left.\frac{\partial S}{\partial x}\right|_{\hat{x}} = 0$$

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

$$\therefore \frac{\phi^T \phi x}{\sigma^2} + \Sigma_x^{-1} x = \frac{\phi^T y}{\sigma^2}$$

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

$$\therefore x = (\phi^T \phi + \sigma^2 \Sigma_x^{-1})^{-1} \phi^T y$$

See the graph in ~~as~~ graph-p2.png. Here, we can see that RMSE decreases as m increases which is obvious because more measurements lead to more accuracy. As σ is increased, RMSE decreases because the eigenvalues and hence the spread in x decreases. A lesser spread in x (quantified by standard deviation) implies lesser error in reconstruction.

Code can be found in codes/p2 and graph in images/p2. We have used a function we found online.

Q4: If the singular vectors are not incoherent with the canonical basis, there might arise cases like very sparse matrices or matrices with similar elements. (for example just one non-zero element). Rank-r matrices can be represented like this,

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

If it happens that the singular vectors are just standard orthonormal vectors then it may be possible that the matrix contains just r non-zero elements. This is a pathological case we want to avoid and hence, singular vectors need to be incoherent with canonical basis.

If our constraint problem changes to this,

$$\begin{aligned} &\text{minimize} \quad \text{rank}(X) \\ &\text{such that} \quad f_i^T X g_j = f_i^T M g_j \ , (i,j) \in \Omega \end{aligned}$$

we see that $f_i^T M g_j = \sum_{k=1}^{r} \sigma_k f_i^T u_k v_k^T g_j$
$(i,j) \in \Omega$

here if u_k and v_k are not incoherent with f_i and g_j then we may get a very sparse matrix. So, the column and row spaces of M should be incoherent with the bases f_i and g_j respectively.

A matrix like $e_1 v^T$ will have the first row as v^T and everything else zero. This matrix cannot be recovered unless we see all of its elements. And its a rank 1 matrix.

Q5: Title: Active Contours with Group Similarity

Venue: CVPR 2013 (Computer Vision and Pattern Recognition conference 2013)

Problem: Shape priors which are widely used in active contours (an image segmentation procedure) can constrain the contours well but they are not always feasible as they require a large amount of data. So, we have introduced group priors that are based on group similarity which is measured by the rank of contour matrix. This group similarity will be constrained in the paper.

Explanation: Shape contours of similar objects with different sizes and orientation are put as a matrix $X = [C_1, C_2, ---, C_n]$ where each contour is represented as a vector. Then, the following problem is solved,

$$\min_X \sum_i f_i(C_i) \quad \text{such that} \quad \text{rank}(X) \leq K$$

Here, the rank constraint will ensure that all the contours are similar and only differ in translation, scaling, rotation, etc.