ENGG 5781: Matrix Analysis and Computations	2018-19 First Term

Assignment 1 Solution

Instructor: Wing-Kin Ma — October 5, 2018

Answer Problems 1–3, and **either** Problem 4 or Problem 5.

Note:

1. By submitting this assignment, we are assumed to have read the homework guideline `http://www.ee.cuhk.edu.hk/~wkma/engg5781/hw/hw_guidelines.pdf` thereby understanding and respecting the guideline mentioned there.
2. You are allowed to use properties and theorems up to Lecture 2; unless specified, that includes all the results and proofs in the additional notes.

Problem 1 (30%) Are the following sets subspaces? Provide your answer with a proof.

(a) $\mathcal{S} = \{\mathbf{X} \in \mathbb{R}^{m\times n} \mid \mathbf{X}^T\mathbf{Y} + \mathbf{Y}^T\mathbf{X} = \mathbf{0}\}$, where $\mathbf{Y} \in \mathbb{R}^{m\times n}$ is given.

(b) $\mathcal{S} = \{\mathbf{X} \in \mathbb{R}^{m\times n} \mid \mathbf{X} = \mathbf{A}\mathbf{B}^T, \mathbf{A} \in \mathbb{R}^{m\times r}, \mathbf{B} \in \mathbb{R}^{n\times r}\}$, where $r < \min\{m, n\}$.

(c) $\mathcal{S} = \{\boldsymbol{\mathcal{X}} \in \mathbb{R}^{m\times n\times p} \mid x_{ijk} = \sum_{\ell=1}^r a_{i\ell}b_{j\ell}c_{k\ell},\ \forall i,j,k,\ c_{k\ell} \in \mathbb{R}\ \forall k,\ell\}$, where $a_{i\ell}, b_{j\ell} \in \mathbb{R}$ are given.

(d) $\mathcal{S} = \{\mathbf{y} \in \mathbb{C}^m \mid |\sum_{i=1}^m y_i(\alpha_j)^i| = 0,\ j = 1,\ldots,n\}$, where $\alpha_1,\ldots,\alpha_n \in \mathbb{C}$ are given.

Solution:

(a) Yes. Let $\mathbf{X}_1, \mathbf{X}_2 \in \mathcal{S}$. For any $\alpha, \beta \in \mathbb{R}$, we have

$$(\alpha\mathbf{X}_1 + \beta\mathbf{X}_2)^T\mathbf{Y} + \mathbf{Y}^T(\alpha\mathbf{X}_1 + \beta\mathbf{X}_2) = \alpha(\mathbf{X}_1^T\mathbf{Y} + \mathbf{Y}^T\mathbf{X}_1) + \beta(\mathbf{X}_2^T\mathbf{Y} + \mathbf{Y}^T\mathbf{X}_2) = 0.$$

This means that $\alpha\mathbf{X}_1 + \beta\mathbf{X}_2 \in \mathcal{S}$ for any α, β.

(b) No. As a counter example, consider $m = n$, with m being even, $r = m/2$. Let

$$\mathbf{A}_1 = [\ \mathbf{e}_1,\ldots,\mathbf{e}_r\ ],\ \mathbf{B}_1 = [\ \mathbf{e}_1,\ldots,\mathbf{e}_r\ ],\ \mathbf{A}_2 = [\ \mathbf{e}_{r+1},\ldots,\mathbf{e}_m\ ],\ \mathbf{B}_2 = [\ \mathbf{e}_{r+1},\ldots,\mathbf{e}_m\ ],$$

and let $\mathbf{X}_1 = \mathbf{A}_1\mathbf{B}_1^T$, $\mathbf{X}_2 = \mathbf{A}_2\mathbf{B}_2^T$. We have

$$\mathbf{X}_1 + \mathbf{X}_2 = \begin{bmatrix}\mathbf{A}_1 & \mathbf{A}_2\end{bmatrix}\begin{bmatrix}\mathbf{B}_1^T \\ \mathbf{B}_2^T\end{bmatrix} = (\mathbf{I})(\mathbf{I})^T = \mathbf{I}.$$

Since $\text{rank}(\mathbf{X}_1 + \mathbf{X}_2) = m > r$, $\mathbf{X}_1 + \mathbf{X}_2$ does not lie in $\mathcal{S}$.

(c) Yes. Let $\boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{Y}} \in \mathcal{S}$, for which we have $x_{ijk} = \sum_{\ell=1}^r a_{i\ell}b_{j\ell}c_{k\ell}$ and $y_{ijk} = \sum_{\ell=1}^r a_{i\ell}b_{j\ell}\tilde{c}_{k\ell}$ for some $\{c_{k\ell}\}$ and $\{\tilde{c}_{k\ell}\}$. Since, for any $\alpha, \beta \in \mathbb{R}$,

$$\alpha x_{ijk} + \beta y_{ijk} = \sum_{\ell=1}^r a_{i\ell}b_{j\ell}(\alpha c_{k\ell} + \beta\tilde{c}_{k\ell}), \quad \forall i,j,k,$$

it is true that $\alpha\boldsymbol{\mathcal{X}} + \beta\boldsymbol{\mathcal{Y}} \in \mathcal{S}$ for any α, β.

(d) Yes. Having $|\sum_{i=1}^m y_i(\alpha_j)^i| = 0$ is equivalent to $\sum_{i=1}^m y_i(\alpha_j)^i = 0$. By letting $\bar{\mathbf{a}}_j = [\, \alpha_j, \alpha_j^2, \ldots, \alpha_j^m \,]^T$ and

$$\mathbf{A} = \begin{bmatrix} \bar{\mathbf{a}}_1^T \\ \vdots \\ \bar{\mathbf{a}}_n \end{bmatrix},$$

we can equivalently rewrite $\mathcal{S} = \mathcal{N}(\mathbf{A})$. As a nullspace, $\mathcal{S}$ is a subspace.

Problem 2 (20%) A non-empty subset $\mathcal{S}$ of $\mathbb{R}^m$ is said to be an affine set if

$$\alpha \in \mathbb{R}, \mathbf{x}, \mathbf{y} \in \mathcal{S} \quad \Longrightarrow \quad \alpha\mathbf{x} + (1-\alpha)\mathbf{y} \in \mathcal{S}.$$

(a) Show that if $\mathcal{S}$ is affine, then any affine combination of $\mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathcal{S}$, i.e.,

$$\mathbf{y} = \sum_{i=1}^n \alpha_i \mathbf{a}_i, \quad \alpha_1, \ldots, \alpha_n \in \mathbb{R}, \ \sum_{i=1}^n \alpha_i = 1,$$

lies in $\mathcal{S}$.

(b) Show that an affine set $\mathcal{S}$ can always be represented by $\mathcal{S} = \mathcal{V} + \mathbf{b}$, where $\mathbf{b} \in \mathcal{S}$ and $\mathcal{V}$ is a subspace[1].

Solution:

(a) The proof can be done by induction. Let $k \geq 2$ be an integer. Suppose that it is true that any affine combination of $\mathbf{a}_1, \ldots, \mathbf{a}_{k-1}$ lies in $\mathcal{S}$. For $k = 2$ the above assumption is true (by definition). Let $\alpha_1, \ldots, \alpha_k \in \mathbb{R}$, with $\sum_{i=1}^k \alpha_i = 1$. Also, assume $\alpha_k \neq 1$; we will come back to the case of $\alpha_k = 1$ later. Then we can write

$$\begin{aligned} \sum_{i=1}^k \alpha_i \mathbf{a}_i &= (1-\alpha_k) \left(\sum_{i=1}^{k-1} \frac{\alpha_i}{1-\alpha_k} \mathbf{a}_i \right) + \alpha_k \mathbf{a}_k \\ &= (1-\alpha_k) \sum_{i=1}^{k-1} \left(\frac{\alpha_i}{\sum_{k=1}^{k-1} \alpha_j} \mathbf{a}_i \right) + \alpha_k \mathbf{a}_k. \end{aligned}$$

Since $\sum_{i=1}^{k-1} \left(\frac{\alpha_i}{\sum_{k=1}^{k-1} \alpha_j} \mathbf{a}_i \right) \in \mathcal{S}$, by the definition of affine sets it holds that $\sum_{i=1}^k \alpha_i \mathbf{a}_i \in \mathcal{S}$.

(Courtesy to an unnamed student who inspired us to get the following proof, which is more concise and beautiful) The case of $\alpha_k = 1$ is handled as follows. If $\alpha_1 = \cdots = \alpha_{k-1} = 0$, the result trivially follows. If not, we note from $\sum_{i=1}^{k-1} \alpha_i = 1 - \alpha_k = 0$ that some of the $\alpha_1, \ldots, \alpha_{k-1}$ must be positive, and some negative. Let $j \in \{1, \ldots, k-1\}$ be such that $\alpha_j > 0$. Let

$$\beta = \alpha_j + \alpha_k > 1,$$

and note that

$$\sum_{i=1, i \neq j}^{k-1} \alpha_i = 1 - \beta < 0.$$

[1] The notation $\mathbf{b} + \mathcal{V}$ means that $\mathbf{b} + \mathcal{V} = \{\mathbf{y} = \mathbf{b} + \mathbf{v} \mid \mathbf{v} \in \mathcal{V}\}$.

This lead us to

$$\sum_{i=1}^{k} \alpha_i \mathbf{a}_i = (1-\beta) \underbrace{\left(\sum_{i=1, i\neq j}^{k-1} \frac{\alpha_i}{1-\beta} \mathbf{a}_i \right)}_{\in S} + \beta \underbrace{\left(\frac{\alpha_j}{\beta} \mathbf{a}_j + \frac{\alpha_k}{\beta} \mathbf{a}_k \right)}_{\in S} \in \mathcal{S},$$

which completes the proof.

(b) Let $\mathcal{V} = \mathcal{S} - \mathbf{b} = \{\mathbf{v} = \mathbf{x} - \mathbf{b} \mid \mathbf{x} \in \mathcal{S}\}$. Let $\mathbf{v}_1, \mathbf{v}_2 \in \mathcal{V}$, which satisfy $\mathbf{v}_1 = \mathbf{x}_1 - \mathbf{b}, \mathbf{v}_2 = \mathbf{x}_2 - \mathbf{b}$ for some $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{S}$. For any $\alpha, \beta \in \mathbb{R}$ we have

$$\begin{aligned} \alpha \mathbf{v}_1 + \beta \mathbf{v}_2 &= \alpha(\mathbf{x}_1 - \mathbf{b}) + \beta(\mathbf{x}_2 - \mathbf{b}) \\ &= \alpha \mathbf{x}_1 + \beta \mathbf{x}_2 + (1 - \alpha - \beta)\mathbf{b} - \mathbf{b}. \end{aligned}$$

Since $\alpha \mathbf{x}_1 + \beta \mathbf{x}_2 + (1 - \alpha - \beta)\mathbf{b} \in \mathcal{S}$, which is implied by (a), it follows from the definition of $\mathcal{V}$ that $\alpha \mathbf{v}_1 + \beta \mathbf{v}_2 \in \mathcal{V}$. Thus, $\mathcal{V}$ is a subspace.

Problem 3 (20%) Let $\mathcal{S} = \{\mathbf{y} \in \mathbb{R}^m \mid \mathbf{y} = \mathbf{A}\mathbf{x} - \mathbf{b}, \mathbf{x} \in \mathbb{R}^n, \|\mathbf{x}\|_2 \leq 1\}$. Suppose $\mathbf{b} \notin \mathcal{R}(\mathbf{A})$.

(a) Is $\mathcal{S} \cap \mathcal{S}^\perp = \{\mathbf{0}\}$, or $\mathcal{S} \cap \mathcal{S}^\perp = \emptyset$?

(b) Show that $\mathcal{S}^\perp = \mathcal{N}([\ \mathbf{A}\ \mathbf{b}\ ]^T)$.

Solution:

(a) We have $\mathcal{S} \cap \mathcal{S}^\perp = \emptyset$ in this case. As written in the course note, we have either $\mathcal{S} \cap \mathcal{S}^\perp = \{\mathbf{0}\}$ or $\mathcal{S} \cap \mathcal{S}^\perp = \emptyset$. The condition $\mathbf{b} \notin \mathcal{R}(\mathbf{A})$ implies $\mathbf{A}\mathbf{x} - \mathbf{b} \neq \mathbf{0}$ for any $\mathbf{x}$. This implies that $\mathbf{0} \notin \mathcal{S}$, and thus we cannot have $\mathcal{S} \cap \mathcal{S}^\perp = \{\mathbf{0}\}$.

(b) To solve this sub-problem, it suffices to show that

$$(\mathbf{A}\mathbf{x} - \mathbf{b})^T \mathbf{y} = 0 \ \forall \|\mathbf{x}\|_2 \leq 1 \quad \Longleftrightarrow \quad \begin{bmatrix} \mathbf{A}^T \\ \mathbf{b}^T \end{bmatrix} \mathbf{y} = \mathbf{0}. \tag{$*$}$$

First, it is immediate that if the right-hand side of $(*)$ is true, the left-hand side of $(*)$ is also true. Let us show the converse. Suppose that the left-hand side of $(*)$ is true. Then, for $\mathbf{x} = \mathbf{0}$, we are led to $\mathbf{b}^T \mathbf{y} = 0$. The left-hand side of $(*)$ is thus reduced to

$$\mathbf{x}^T \mathbf{A}^T \mathbf{y} = 0, \ \forall \|\mathbf{x}\|_2 \leq 1.$$

Suppose that $\mathbf{A}^T \mathbf{y} \neq \mathbf{0}$. Then, by choosing $\mathbf{x} = \mathbf{A}^T \mathbf{y} / \|\mathbf{A}^T \mathbf{y}\|_2$ (which satisfies $\|\mathbf{x}\|_2 \leq 1$), we have $\mathbf{x}^T \mathbf{A}^T \mathbf{y} = \|\mathbf{A}^T \mathbf{y}\|_2 > 0$. This implies that $\mathbf{A}^T \mathbf{y} = \mathbf{0}$ must hold, and thus we have the right-hand side of $(*)$ to be true.

Problem 4 (30%) Let $\{\mathbf{a}_1, \ldots, \mathbf{a}_n\} \subset \mathbb{R}^m$ be a given set of linearly independent vectors. Let $\mathbf{q}_1, \ldots, \mathbf{q}_n \in \mathbb{R}^m$ whose construction will be specified, and let $\mathcal{S}_i = \text{span}\{\mathbf{q}_1, \ldots, \mathbf{q}_i\}$ for the sake of notational convenience. Consider the following procedure.

Algorithm 1:

$\mathbf{q}_1 = \mathbf{a}_1/\|\mathbf{a}_1\|_2$;
for $i = 2, \ldots, n$ **do**
  $\tilde{\mathbf{q}}_i = \Pi_{\mathcal{S}_{i-1}^\perp}(\mathbf{a}_i)$;
  $\mathbf{q}_i = \tilde{\mathbf{q}}_i/\|\tilde{\mathbf{q}}_i\|_2$;
end

(a) Show that, for any $i \in \{1, \ldots, n\}$,

$$\text{span}\{\mathbf{a}_1, \ldots, \mathbf{a}_i\} = \mathcal{S}_i.$$

DO NOT use the proof in Lecture 1, page 47–50. Consider the problem as if you did not know what is Gram-Schmidt, which you can easily find in textbooks or in the world-wide web. Use ONLY the basic notions of subspace, with projection onto subspaces included, to rediscover the result.

(b) Use the basic notions of subspaces, as well as those of orthogonality and LS, to show that $\tilde{\mathbf{q}}_i = \mathbf{a}_i - \mathbf{Q}_{i-1}\mathbf{Q}_{i-1}^T\mathbf{a}_i$, where $\mathbf{Q}_i = [\ \mathbf{q}_1, \ldots, \mathbf{q}_i\ ]$.

(c) Suggest how we may modify the algorithm when $\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$ is linearly dependent. Note that the ultimate goal is to find an orthogonal basis for $\text{span}\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$.

Note: Since this problem requires some explanation, we consider it as a semi-essay problem. English fluency, clarity of presentation, originality of the presentation (relative to others), etc., will be heavily taken into account.

Solution:

(a) We use induction. Suppose that $\mathcal{S}_{i-1} = \text{span}\{\mathbf{a}_1, \ldots, \mathbf{a}_{i-1}\}$; it is true for $i = 2$. We want to show that $\mathcal{S}_i = \text{span}\{\mathbf{a}_1, \ldots, \mathbf{a}_i\}$. Since $\mathbf{a}_1, \ldots, \mathbf{a}_i$ are linearly independent, the vector $\mathbf{a}_i$ cannot be written as $\mathbf{a}_i = \sum_{j=1}^{i-1} \alpha_j \mathbf{a}_j$ for any $\alpha_1, \ldots, \alpha_{i-1}$. Or, we have $\mathbf{a}_i \notin \text{span}\{\mathbf{a}_1, \ldots, \mathbf{a}_{i-1}\} = \mathcal{S}_{i-1}$. By the projection theorem we can always write

$$\mathbf{a}_i = \tilde{\mathbf{q}}_i + \mathbf{b}_i,$$

where $\tilde{\mathbf{q}}_i = \Pi_{\mathcal{S}_{i-1}^\perp}(\mathbf{a}_i)$, $\mathbf{b}_i = \Pi_{\mathcal{S}_{i-1}}(\mathbf{a}_i)$. As $\mathbf{a}_i \notin \mathcal{S}_{i-1}$, we must have $\tilde{\mathbf{q}}_i \neq \mathbf{0}$. It follows that

$$\begin{aligned}
\text{span}\{\mathbf{a}_1, \ldots, \mathbf{a}_i\} &\equiv \{\mathbf{y} = \textstyle\sum_{j=1}^{i-1} \alpha_j \mathbf{a}_j + \alpha_i \mathbf{q}_i \mid \boldsymbol{\alpha} \in \mathbb{R}^i\} \\
&= \text{span}\{\mathbf{a}_1, \ldots, \mathbf{a}_{i-1}\} + \text{span}\{\mathbf{q}_i\} \\
&= \text{span}\{\mathbf{q}_1, \ldots, \mathbf{q}_{i-1}\} + \text{span}\{\mathbf{q}_i\} \\
&= \text{span}\{\mathbf{q}_1, \ldots, \mathbf{q}_i\},
\end{aligned}$$

where the first inequality is due to the fact that $\mathbf{a}_i$ can be expressed as $\mathbf{a}_i = \|\tilde{\mathbf{q}}_i\|_2 \mathbf{q}_i + \sum_{j=1}^{i-1} \beta_j \mathbf{a}_j$ for some β_j's (note $\mathbf{b}_i \in \mathcal{S}_{i-1}$). Our proof is done.

(b) We know from Lectures 1–2 that for a full-column rank $\mathbf{A}$, we have $\Pi_{\mathcal{R}(\mathbf{A})}(\mathbf{y}) = \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{y}$ and $\Pi_{\mathcal{R}(\mathbf{A})^\perp}(\mathbf{y}) = \mathbf{y} - \Pi_{\mathcal{R}(\mathbf{A})}(\mathbf{y})$. By plugging $\mathbf{A} = \mathbf{Q}_{i-1}$, and using $\mathbf{Q}_{i-1}^T\mathbf{Q}_{i-1} = \mathbf{I}$, we get $\Pi_{\mathcal{S}_{i-1}^\perp}(\mathbf{a}_i) = \mathbf{a}_i - \mathbf{Q}_{i-1}\mathbf{Q}_{i-1}^T\mathbf{a}_i$.

(c) I would modify the algorithm this way:

Algorithm 2:

$j = 0$;
$\mathcal{A}_0 = \emptyset$;
for $i = 1, \ldots, n$ **do**
    if $\Pi_{\mathcal{S}_j^\perp}(\mathbf{a}_i) \neq \mathbf{0}$ (or $\|\Pi_{\mathcal{S}_j^\perp}(\mathbf{a}_i)\|_2 \leq \epsilon$ for some small $\epsilon > 0$), where $\mathcal{S}_j = \text{span}\,\mathcal{A}_j$, **then**
        $\tilde{\mathbf{q}}_{j+1} = \Pi_{\mathcal{S}_j^\perp}(\mathbf{a}_i)$;
        $\mathbf{q}_{j+1} = \tilde{\mathbf{q}}_{j+1}/\|\tilde{\mathbf{q}}_{j+1}\|_2$;
        $\mathcal{A}_{j+1} = \mathcal{A}_j \cup \{\mathbf{q}_{j+1}\}$;
        $j = j + 1$;
    end
end

We again use induction to prove why $S_j = \text{span}\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$ when the algorithm terminates. Suppose that at the ith iteration of the above algorithm, we have $\text{span}\{\mathbf{a}_1, \ldots, \mathbf{a}_{i-1}\} = \mathcal{S}_j$. By the projection theorem, we have (again) $\mathbf{a}_i = \tilde{\mathbf{q}}_i + \mathbf{b}_i$ where $\tilde{\mathbf{q}}_i = \Pi_{\mathcal{S}_j^\perp}(\mathbf{a}_i)$, $\mathbf{b}_i = \Pi_{\mathcal{S}_j}(\mathbf{a}_i)$. If $\tilde{\mathbf{q}}_i \neq \mathbf{0}$, we have the same result as in (a), i.e., $\text{span}\{\mathbf{a}_1, \ldots, \mathbf{a}_i\} = \mathcal{S}_j + \text{span}\{\mathbf{q}_{j+1}\} = \mathcal{S}_{j+1}$. If $\tilde{\mathbf{q}}_i = \mathbf{0}$, it implies that $\mathbf{a}_i \in \mathcal{S}_j$, and consequently, $\text{span}\{\mathbf{a}_1, \ldots, \mathbf{a}_i\} = \mathcal{S}_j$. It follows by induction that the result $S_j = \text{span}\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$ holds.

Problem 5 (30%) This is a MATLAB problem. The problem we deal with is a face recognition problem. Your main reference is [1], Chapter 9. I also recommend [2] as an additional reference for you to get better understanding of the context. Download from the course website the following three files: `X.mat`, `y_part_a.mat`, and `y_part_b.mat`. Loading `X.mat` on MATLAB, you will see five matrices, namely, `X_1`, `X_2`, `X_3`, `X_4`, `X_5`. I will sometimes use $\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathbf{X}_4, \mathbf{X}_5$, our standard matrix notations, to describe them. Each $\mathbf{X}_i$ is a collection of 63 images, with size 192×168, taken from the same person. Specifically, each column of $\mathbf{X}_i$ is an image stored in the vectorized form. You can see them, say, for $\mathbf{X}_1$, by calling

```
>> for i=1:63, subplot(8,8,i); imshow(reshape(X_1(:,i),192,168)); end;
```

I should mention that the data come from "Yale Face Database B."

(a) Load `y_part_a.mat`. You will find a vector `y`, or $\mathbf{y}$. You can see it by calling

```
>> imshow(reshape(y,192,168));
```

You will see that it is a noisy image. Form a multi-person data matrix $\mathbf{X} = [\ \mathbf{X}_1, \ldots, \mathbf{X}_5\ ]$, apply LS

$$\min_{\mathbf{w} \in \mathbb{R}^n} \ \|\mathbf{y} - \mathbf{X}\mathbf{w}\|_2^2,$$

and use the LS solution to construct an estimated, and hopefully noise-cleaned, image. Show us the image you recover. Then, from the LS solution, identify the person $\mathbf{y}$ is associated with. Note: You can certainly watch the images and *subjectively* identify who that person is, but I would like to see a quantitative way.

(b) Load `y_part_b.mat`, which contains a vector $\mathbf{y}$. Using `imshow`, you will see that a small part of the image is severely corrupted. Try LS and show the recovered image. Then, try Algorithm 11 in [1], AltMin for Robust Regression (AM-RR), with parameter $k = 4775$. You should also write a short description concerning what is the rationale of AM-RR, and how it works. Show the recovered image of AM-RR.

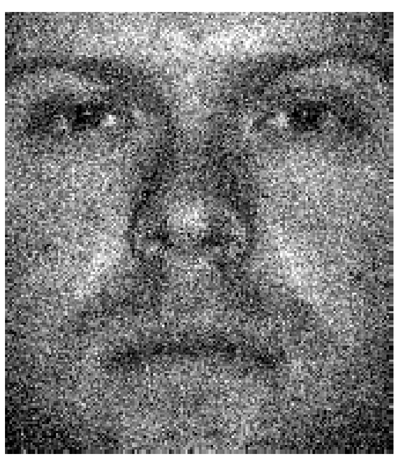
Note: You will need to submit your MATLAB code online via Blackboard. You also need to provide a description on your assignment. We do not do reverse engineering tasks such as guessing how a MATLAB code works in the absence of any description. If you are in doubt, talk to us to understand more.

Solution:

(a)

```
>> load X.mat; load y_part_a.mat;
>> X= [ X_1 X_2 X_3 X_4 X_5 ];
>> w= X\y;
>> hy= X*w;
>> imshow(reshape(hy,192,168));
```

The result, together with the original image, are shown below.



To identify which person $\mathbf{y}$ should be, partition $\mathbf{w}$ as

$$\mathbf{w} = \begin{bmatrix} \mathbf{w}_1 \\ \mathbf{w}_2 \\ \vdots \\ \mathbf{w}_5 \end{bmatrix}$$

where $\mathbf{w}_i \in \mathbb{R}^{63}$ for all i. By noting that

$$\mathbf{X}\mathbf{w} = \mathbf{X}_1\mathbf{w}_1 + \mathbf{X}_2\mathbf{w}_2 + \cdots + \mathbf{X}_5\mathbf{w}_5,$$

we can believe that the $\mathbf{w}_i$ with the largest $\|\mathbf{w}_i\|_2^2$ should be an indication of which $\mathbf{X}_i$, or person, $\mathbf{y}$ belong to.

```
>> W= reshape(w,63,5);
>> sum(abs(W).^2)

ans =

   0.8608    0.3211    0.1037    0.4181    0.1220
```

It appears that the answer is person 1.

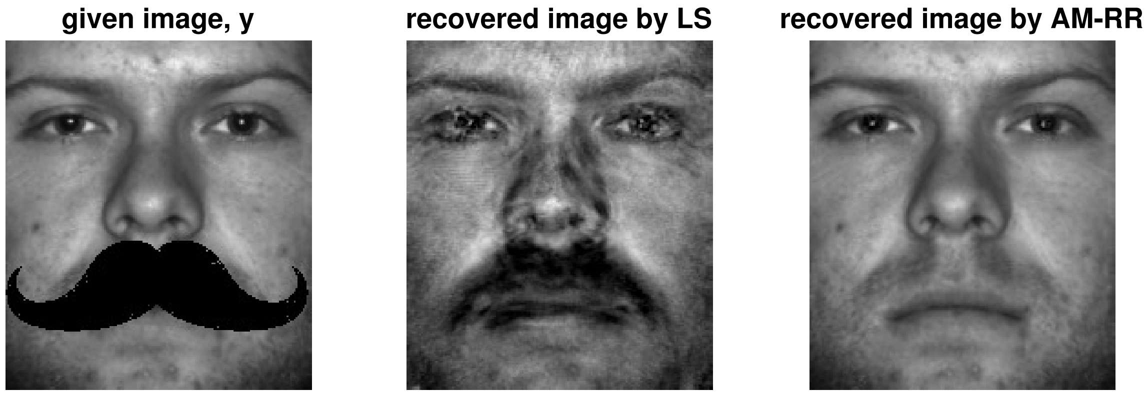
(b) In short, it is an algorithm that intends to solve a outlier-robust formulation

$$\begin{aligned} &\min_{\mathbf{w},\mathbf{b}} \|\mathbf{y} - \mathbf{X}\mathbf{w} - \mathbf{b}\|_2^2 \\ &\text{s.t. } \|\mathbf{b}\|_0 \leq k, \end{aligned}$$

where $\|\cdot\|_0$ denotes the number of nonzero elements of its argument. The idea is to ignore a number of k measurements in evaluating the loss between $\mathbf{y}$ and $\mathbf{X}\mathbf{w}$, thereby giving the algorithm some robustness against outliers. This problem is difficult to solve exactly, and AM-RR uses alternating minimization to deal with it; i.e., solve $\mathbf{w}$ fixing $\mathbf{b}$ at one time, solve $\mathbf{b}$ fixing $\mathbf{w}$ at another time, repeat the above until little progress is seen.

```
>> T= 50; k= 4775;
>> [m,n]= size(X);
>> S= 1:(m-k);
>> for t=1:T,
         w= X(S,:)\y(S);
         r= y- X*w;
         [notused,indx]= sort(abs(r),'ascend');
         S= indx(1:m-k);
   end;
>> hy= X*w;
```

The results are shown below.



References

[1] P. Jain and P. Kar. Non-convex optimization for machine learning. *Foundations and Trends® in Machine Learning*, 10(3–4):142–336, 2017.

[2] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma. Robust face recognition via sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(2):210–227, 2009.