ENGG 5781: Matrix Analysis and Computations	**2018-19 First Term**
Assignment 2	
Instructor: Wing-Kin Ma	Due: November 7, 2018

Answer Problems 1–2, and **either** Problem 3 or Problem 4.

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 4; unless specified, that includes all the results and proofs in the additional notes.

Problem 1 (30%)

(a) Let

$$\mathbf{T} = \begin{bmatrix} t_{11} & t_{12} & t_{13} \\ 0 & t_{22} & t_{23} \\ 0 & 0 & t_{33} \end{bmatrix}.$$

Write down $\mathbf{T}^2$. Conclude that $\mathbf{T}^2$ is upper triangular with diagonal elements $t_{11}^2, t_{22}^2, t_{33}^2$.

(b) Let

$$\mathbf{W} = \begin{bmatrix} w_{11} & w_{12} & w_{13} \\ 0 & w_{22} & w_{23} \\ 0 & 0 & w_{33} \end{bmatrix}.$$

Can we always decompose $\mathbf{W}$ as $\mathbf{W} = \mathbf{T}^2$, where $\mathbf{T}$ is upper triangular? If your answer is yes, provide the procedure of finding such a $\mathbf{T}$. If your answer is no, give a counter-example and provide reasons for the impossibility.

(c) Let $\mathbf{A} \in \mathbb{C}^{n\times n}$. Suppose that there exists a matrix $\mathbf{B} \in \mathbb{C}^{n\times n}$ such that $\mathbf{B}^2 = \mathbf{A}$. Show that $\lambda_i(\mathbf{A}) = \lambda_i(\mathbf{B})^2$ for $i = 1, \ldots, n$, where $\lambda_1(\mathbf{X}), \ldots, \lambda_n(\mathbf{X})$ denote the eigenvalues of $\mathbf{X}$. You may use (without proof) that the result in (a) holds for upper triangular matrices of any dimension.

Problem 2 (40%) This problem requires some prerequisite with random processes. A summary of random process properties is provided in the Appendix.

The problem setup is as follows. Let $\mathbf{x} \in \mathbb{R}^n$, with $\|\mathbf{x}\|_2 = 1$, be a deterministic unknown. Let $\mathbf{a} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ be a random vector of dimension n, and let $y = |\mathbf{a}^T\mathbf{x}|^2$ be a measurement. The problem of interest is to prove that we can recover $\mathbf{x}$ from $\mathbb{E}[y\mathbf{a}\mathbf{a}^T]$ up to a scale factor.

(a) Show that there exists an orthogonal matrix $\mathbf{Q}$ such that $\mathbf{Q}\mathbf{x} = \mathbf{e}_1$.

(b) Let $\mathbf{b} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. Show that

$$\mathbb{E}[b_1^2\mathbf{b}\mathbf{b}^T] = 2\mathbf{e}_1\mathbf{e}_1^T + \mathbf{I}.$$

(c) Let $\mathbf{R} = \mathbb{E}[y\mathbf{a}\mathbf{a}^T]$. Show that

$$\mathbf{R} = 2\mathbf{x}\mathbf{x}^T + \mathbf{I}.$$

(d) Show that the (real) eigenvector associated with the largest eigenvalue of $\mathbf{R}$ must be either $\mathbf{x}$ or $-\mathbf{x}$.

Problem 3 (30%) Let $\mathbf{A} \in \mathbb{R}^{n\times n}$. Let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of $\mathbf{A}$, arranged as $|\lambda_1| \geq |\lambda_2| \geq \ldots \geq |\lambda_n|$. The objective is to find an eigenvector $\mathbf{v}_1$ associated with λ_1 via the power method. We assume that i) the eigendecomposition of $\mathbf{A}$ exists, that ii) the starting point $\mathbf{v}^{(0)}$ is generated to lie in $\mathbb{R}^n$, and that iii) $[\mathbf{V}^{-1}\mathbf{v}^{(0)}]_1 \neq 0$.

(a) Do you agree that the power method will find the desired eigenvector under the assumption $|\lambda_1| > |\lambda_2|$? A very concise answer suffices.

(b) Suppose that any eigenvector $\mathbf{v}_1$ associated with λ_1 is complex-valued and can never be real-valued; i.e., $\mathbf{v}_1 \in \mathbb{C}^n, \mathbf{v}_1 \notin \mathbb{R}^n$. Do you think the power method will converge to one such $\mathbf{v}_1$? A concise justification for your answer suffices.

(c) Do you find that the answers in (a) and (b) (assuming that you answer them right) seem to contradict each other? What goes wrong? Can you solve the paradox through analysis? We expect a clear explanation.

Hint: It is only my personal advice. If you would randomly generate real $\mathbf{A}$'s and observe their eigenvalues (e.g., `A= randn(5,5); [V,D] = eig(A);` then display `diag(D)`), you would see certain patterns. That may give you insight into solving the problem.

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.

Problem 4 (30%) This is a MATLAB problem. Let $\mathbf{x} \in \mathbb{R}^n$. It is observed in a hidden manner. Specifically, we only have amplitude-only measurements

$$z_i = |\mathbf{a}_i^T\mathbf{x}|, \quad i = 1, \ldots, m,$$

where $\mathbf{a}_i \in \mathbb{R}^n$, $i = 1, \ldots, m$, are known. The challenge is to try to recover $\mathbf{x}$ from $z_1, \ldots, z_m$.

(a) This is a simplified version of the problem for helping you kick-start. Generate $\mathbf{x}, \mathbf{A}, \mathbf{z}$ by

```
>> n= 30; m= 10000;
>> x= cos(2*pi*.075*(1:n).');
>> A= randn(m,n);
>> z= abs(A*x);
```

Here, note that $\mathbf{A} = [\ \mathbf{a}_1, \ldots, \mathbf{a}_m\ ]^T$ (which does not follow our conventional notation). The problem we studied in Problem 2 reveals an algorithmic way to estimate $\mathbf{x}$. Let

$$\mathbf{R} = \frac{1}{m}\sum_{i=1}^{m} z_i^2 \mathbf{a}_i\mathbf{a}_i^T,$$

which is a sampled version of $\mathbb{E}[z^2 \mathbf{a}\mathbf{a}^T]$ (assuming that $\mathbf{a}_i$'s do follow the i.i.d. Gaussian distribution). As seen in Problem 2(c)–(d), we may estimate $\mathbf{x}$, up to a scaling and sign, by computing the eigenvector associated with the largest eigenvalue of $\mathbf{R}$.

Plot the estimated $\mathbf{x}$ and see if it is consistent with the true $\mathbf{x}$. Note that the eigen method is subjected to sign and scaling ambiguities. Hence, if you get the right shape, it should be fine.

(b) Download `hw2-prob4.mat` from the course website, which stores $\mathbf{z}$ and $\mathbf{A}$ (the size of $\mathbf{A}$ is not small, so make sure your computer has enough memory). This time, $\mathbf{x}$ is a gray-level image with size 30×30. Try the eigen method described in (a), and plot the result. A recommended way to show the result is as follows. Let `hx` be the estimated $\mathbf{x}$ on MATLAB. Do this:

```
>> hx= sign(sum(hx))*hx;
>> hx= hx./abs(max(hx));
>> imshow(reshape(hx,[30,30]));
```

(c) Consider a recovery alternative where the recovery problem is formulated as

$$\min_{\mathbf{x}\in\mathbb{R}^n} \sum_{i=1}^{m} (z_i - |\mathbf{a}_i^T \mathbf{x}|)^2.$$

Argue why the above problem can be equivalently rewritten as

$$\min_{\mathbf{s}\in\{-1,1\}^m,\ \mathbf{x}\in\mathbb{R}^n} \quad \sum_{i=1}^{m} (s_i z_i - \mathbf{a}_i^T \mathbf{x})^2 = \|\mathbf{s} \odot \mathbf{z} - \mathbf{A}\mathbf{x}\|_2^2,$$

where $\odot$ denotes the Hadamard, or element-wise, product; i.e., $\mathbf{x} \odot \mathbf{y} = [\ x_1 y_1, \ldots, x_n y_n\ ]^T$. Then, build an alternating minimization algorithm for the above problem; i.e., like alternating LS for matrix factorization in Lecture 2 (or AM-RR in the last assignment), you minimize $\mathbf{s}$ fixing $\mathbf{x}$, minimize $\mathbf{x}$ fixing $\mathbf{s}$, and keep doing that until you feel you can stop the algorithm.

Write down the pseudo code of the alternating minimization algorithm; pseudo codes mean the algorithm descriptions you see in the course notes, not MATLAB or language-dependent codes. In particular, how you algorithmically perform the two minimization steps should be clearly shown. Plot the recovered image.

Note: Again, you will need to submit your MATLAB code online via Blackboard. Some of the previously mentioned guidelines for MATLAB problems also apply.

Appendix: Some Basic Notions of Random Processes

1. The notation $\mathbb{E}[\cdot]$ denotes expectation.

2. The notation $\xi \sim \mathcal{N}(0,1)$ means that ξ is a Gaussian-distributed random variable with mean zero and unit variance. It is known that, for $\xi \sim \mathcal{N}(0,1)$,

$$\mathbb{E}[\xi] = 0, \quad \mathbb{E}[\xi^2] = 1, \quad \mathbb{E}[\xi^3] = 0, \quad \mathbb{E}[\xi^4] = 3.$$

3. If ξ, ζ are independently distributed random variables, it is true that $\mathbb{E}[f(\xi)g(\zeta)] = \mathbb{E}[f(\xi)]\mathbb{E}[g(\zeta)]$ for any functions f, g.

4. The notation $\boldsymbol{\xi} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ means that $\xi_i \sim \mathcal{N}(0, 1)$ for all i, and any ξ_i, ξ_j, $i \neq j$, are independently distributed.

5. If $\boldsymbol{\xi} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ and $\mathbf{Q}$ is orthogonal, the transformed random vector $\boldsymbol{\zeta} = \mathbf{Q}\boldsymbol{\xi}$ satisfies $\boldsymbol{\zeta} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$.

6. Let $\mathbf{Z}$ be a random matrix. The notation $\mathbb{E}[\mathbf{Z}]$ denotes the element-wise expectation of $\mathbf{Z}$, i.e.,

$$\mathbb{E}[\mathbf{Z}] = \begin{bmatrix} \mathbb{E}[z_{11}] & \mathbb{E}[z_{12}] & \dots & \mathbb{E}[z_{1n}] \\ \mathbb{E}[z_{21}] & \mathbb{E}[z_{22}] & \dots & \mathbb{E}[z_{2n}] \\ \vdots & & & \vdots \\ \mathbb{E}[z_{m1}] & \mathbb{E}[z_{22}] & \dots & \mathbb{E}[z_{mn}] \end{bmatrix}.$$

7. Let $\mathbf{Z}$ be a random matrix, and let $\mathbf{A}, \mathbf{B}$ be deterministic matrices. Then

$$\mathbb{E}[\mathbf{AZB}] = \mathbf{A}\mathbb{E}[\mathbf{Z}]\mathbf{B}.$$