ECE 712 Take–home final exam 2011

James P. Reilly

Revision 1

Please hand in completed exam to Cheryl Gies in ITB A112 by Monday Dec. 19, 2011, by 4:00pm. If you have a valid reason why this deadline is inappropriate for you, then please contact me by email, even though you may have already talked to me in person.

This exam is based on the honour system. As such, please sign the following statement:

This exam is entirely my own work. I may consult any written material or text, but I have NOT received nor offered any help from any other person.

Name: Date:

Signature:

Given that individuals may be sitting closely together in the same lab, doing this exam at the same time, it may be difficult to avoid discussion of this exam. I therefore urge you avoid this situation and work e.g., in the library. It is your responsibility to work in an environment where you can honour the above oath. Students can feel free to contact me for clarification of any questions. Please note that any violations of this oath will be taken very seriously.

For each question, explain your method fully, but as precisely and concisely as possible. Include any matlab code if appropriate.

1. Given $\boldsymbol{Z} \in \mathbb{R}^{m \times n}$, and a set $\boldsymbol{x}_i \in \mathbb{R}^m, \boldsymbol{y}_i \in \mathbb{R}^n$, $i = 1, \ldots, k,\ k \leq \min(m, n)$

a. find a set of coefficients a_i so that

$$\left|\left|\boldsymbol{Z} - \sum_{i=1}^{k} a_i \boldsymbol{x}_i \boldsymbol{y}_i^T\right|\right|_F^2 \tag{1}$$

is minimized. *Hint:* Vectorize each outer product $\boldsymbol{x}_i \boldsymbol{y}_i^T$; i.e., for each outer product, stack each column on top of each other to form one long vector. This will lead to an expression of the form $\boldsymbol{U}\boldsymbol{a} - \boldsymbol{z}$.

b. What are the set $\boldsymbol{x}_i$, $\boldsymbol{y}_i$ that minimize the minimum in (1) for the optimal set of coefficients a_i? *Hint:* Check out the "Interesting Theorem" in the Lecture 3 course notes. It is the case where $k < \min(m, n)$ which is of specific interest here.

2. Consider the Gaussian elimination process $\boldsymbol{A} = \boldsymbol{LU}$, where $\boldsymbol{U}$ is obtained by applying successive Gauss transforms on $\boldsymbol{A}$. Note that the inverse $\boldsymbol{L}^{-1}$ of $\boldsymbol{L}$ is given as $\boldsymbol{L}^{-1} = \boldsymbol{L}_{n-1}^{-1} \dots \boldsymbol{L}_2^{-1}\boldsymbol{L}_1^{-1}$, where each $\boldsymbol{L}_i^{-1}$ is the Gauss transform eliminating the ith column.

a. Suggest a sequence of modified Gauss transform matrices $\boldsymbol{M}_i^{-1}$, to be applied to $\boldsymbol{U}$ that transforms it into a diagonal matrix. With this modified transform, the last (right-most) column of $\boldsymbol{U}$ is eliminated first, then the second, etc., down to the second column. These modified transform matrices pre-multiply the result in the previous stage.

b. Show how the matrices $\boldsymbol{L}^{-1}$ and $\boldsymbol{M}^{-1}$ can be used to calculate $\boldsymbol{A}^{-1}$ in an efficient manner. The matrix $\boldsymbol{M}^{-1}$ is defined in an analogous way to $\boldsymbol{L}^{-1}$.

Write a Matlab program to demonstrate your procedure. Use the matrix $\boldsymbol{A}$ associated with file exam2011 Q2.mat on the website. This program should first apply ordinary Gauss transforms on $\boldsymbol{A}$ to produce the upper triangular matrix $\boldsymbol{U}$. Then, it should apply your modified transforms on $\boldsymbol{U}$ to give a diagonal matrix $\boldsymbol{D}$. Finally, use the inherent matrices $\boldsymbol{L}^{-1}$ and $\boldsymbol{M}^{-1}$ and $\boldsymbol{D}$ to find $\boldsymbol{A}^{-1}$.

3. Suppose we have some process which has m_x input variables $\boldsymbol{x}_i, i = 1, \dots, n$, where n is the number of observations, and has m_y corresponding outputs $\boldsymbol{y}_i, i = 1, \dots, n$. This situation may occur for example, in a chemical plant, where the $\boldsymbol{x}_i$ are measurements of controllable quantities such as temperature, pressure, quantity of input reactants, etc. in a big reactor tank, and the $\boldsymbol{y}_i$ may be corresponding outputs such as concentration of various by-products of the reaction. The set of input and output variables may be assembled into matrices $\boldsymbol{X} \in \mathbb{R}^{n \times m_x}$ and $\boldsymbol{Y} \in \mathbb{R}^{n \times m_y}$, respectively. Typically, $n \gg m_x, m_y$.

Linear least squares estimation may be used to predict a value of $\boldsymbol{y}$ from a known input value of $\boldsymbol{x}$. However, an alternative procedure that has been attracting attention of late is the so-called *partial least–square* procedure. The PLS procedure finds an orthonormal subspace $\boldsymbol{T}$ that is common to both $\boldsymbol{X}$ and $\boldsymbol{Y}$. Column vectors $\boldsymbol{t}_i$ in $\boldsymbol{T}$ are called *latent variables.*

The method works by finding latent variables that maximize correlations between $\boldsymbol{X}$ and $\boldsymbol{Y}$. Let m_t be the number of latent variables chosen for the problem. A schema for iteratively finding the latent vectors $\boldsymbol{t}_i$ is as follows:

for $i = 1, \dots, m_t$

- $\boldsymbol{A} = \boldsymbol{X}^T\boldsymbol{Y}\boldsymbol{Y}^T\boldsymbol{X}$
- $\boldsymbol{w}$ = max eigenvector of $\boldsymbol{A}$
- $\boldsymbol{t}_i = \boldsymbol{X}\boldsymbol{w}$
- $\boldsymbol{t}_i = \boldsymbol{t}_i/\text{norm}(\boldsymbol{t}_i)$
- COMMENT: deflate X by projecting it into the orthogonal complement subspace of $\boldsymbol{t}_i$
- $\boldsymbol{X} = \boldsymbol{X} - \boldsymbol{t}_i\boldsymbol{t}_i^T\boldsymbol{X}$

endfor

Questions:

a. Prove that $\boldsymbol{T} = [\boldsymbol{t}_1 \ldots, \boldsymbol{t}_{m_t}]$ is orthonormal.

b. Implement the PLS procedure in Matlab using the $\boldsymbol{X}$ and $\boldsymbol{Y}$ variables found in the file exam2011 Q3.mat on the website. Use $m_t = 2$. Predict the value $\boldsymbol{y}_o$ corresponding to the input values $\boldsymbol{x}_o$ found in the file.

Here are a few hints on how to do this question. We can write

$$\boldsymbol{X} = \boldsymbol{TP} + \boldsymbol{E}_X \tag{2}$$

$$\boldsymbol{Y} = \boldsymbol{TQ} + \boldsymbol{E}_Y, \tag{3}$$

and find $\boldsymbol{P}$ and $\boldsymbol{Q}$ using a least squares procedure. Then define

$$\boldsymbol{X}_m = \boldsymbol{TP} \tag{4}$$

$$\boldsymbol{Y}_m = \boldsymbol{TQ}. \tag{5}$$

Then we can assume that $\boldsymbol{Y}_m = \boldsymbol{X}_m\boldsymbol{B}+\boldsymbol{E}$, where again $\boldsymbol{B}$ is determined through an LS procedure. Knowledge of $\boldsymbol{B}$ is sufficient to predict $\boldsymbol{y}_o$ from $\boldsymbol{x}_o$. Be careful in this last LS procedure.

c. Now use a linear least-squares estimation (regression) procedure directly to predict $\boldsymbol{Y}$ from $\boldsymbol{X}$ and thus to predict $\boldsymbol{y}_o$ from $\boldsymbol{x}_o$. Be careful about what kind of LS procedure you use!

d. Provide a brief discussion comparing the effectiveness of the PLS and LS procedures, when both $\boldsymbol{X}$ and $\boldsymbol{Y}$ are rank deficient, or nearly so.

4. From the QR decomposition, a matrix $\boldsymbol{Z}$ can be written as

$$\boldsymbol{Z} = \boldsymbol{QR} \tag{6}$$

where $\boldsymbol{Q} \in \mathbb{R}^{m\times m}$ is orthonormal, and $\boldsymbol{R} \in \mathbb{R}^{m\times(n+k)}$ is upper triangular. The QR decomposition of $\boldsymbol{Z}$ can be partitioned as

$$\boldsymbol{Z} = \; m \underset{\quad n \quad\; m-n}{\begin{bmatrix} \boldsymbol{Q}_1 & \boldsymbol{Q}_2 \end{bmatrix}} \underset{n \quad k}{\begin{bmatrix} \boldsymbol{R}_{11} & \boldsymbol{R}_{12} \\ \boldsymbol{0} & \boldsymbol{R}_{22} \end{bmatrix}} \; m \tag{7}$$

The matrix $\boldsymbol{Z}$ itself can be partitioned as

$$\boldsymbol{Z} = \; m \underset{n \quad k}{\begin{bmatrix} \boldsymbol{Z}_1 & \boldsymbol{Z}_2 \end{bmatrix}}. \tag{8}$$

Express the projection of $\boldsymbol{Z}_2$ onto $\boldsymbol{Z}_1$ using only quantities shown in eq.(7).