ECE 712 Assignment on Least Squares

James P. Reilly

1. Show that the rank of a projector is equal to its trace.

2. Data $\boldsymbol{b} \in \Re^m$ are observed according to the over–determined model

$$\boldsymbol{b} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{\nu}, \tag{1}$$

where $\boldsymbol{A} \in \Re^{m \times n}, m > n$. The quantity $\boldsymbol{\nu}$ is an error vector with zero mean and covariance matrix $\boldsymbol{\Sigma}$, which is an arbitrary symmetric positive definite matrix. In this case, the LS solution is derived by minimizing the quantity $||\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b})||_2^2$.

(a) Derive the LS estimate of $\boldsymbol{x}$.

(b) What is $\boldsymbol{P}$ so that $\boldsymbol{r}_{LS} = \boldsymbol{P}(\boldsymbol{\Sigma}^{-1/2}\boldsymbol{b})$? ($\boldsymbol{r}_{LS}$ is the LS residual $\boldsymbol{\Sigma}^{-1/2}\left(\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}_{LS}\right)$. Does $\boldsymbol{P}$ satisfy all the properties of a projector?

(c) Give a geometrical interpretation.

3. Consider the over–determined set of equations

$$\boldsymbol{b} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{n} \tag{2}$$

where $\text{cov}(\boldsymbol{n}) = \sigma^2\boldsymbol{I}$. Suppose in a given experiment, we had control over the values of the matrix $\boldsymbol{A}$, so that $||\boldsymbol{a}_i||_2^2 = k, i = 1, \ldots, n$, where k is an arbitrary constant > 0, and $\boldsymbol{a}_i$ is the ith column of $\boldsymbol{A}$. Explain how to choose $\boldsymbol{A}$ so that the variance of each element of the LS estimate $\boldsymbol{x}_{LS}$ of $\boldsymbol{x}$ is minimum.

Hint: use the Hadamard inequality: For a positive definite square symmetric matrix $\boldsymbol{X} \in \Re^{n \times n}$,

$$\det(\boldsymbol{X}) \leq \prod_{i=1}^{n} x_{ii}, \tag{3}$$

with equality iff $\boldsymbol{X}$ is diagonal.

4. **a.** Given $\boldsymbol{A} \in \Re^{m \times n}$, and a set $\boldsymbol{x}_i \in \Re^m, \boldsymbol{y}_i \in \Re^n$, $i = 1, \ldots, k$, find a set of coefficients a_i so that

$$\left|\left|\boldsymbol{A} - \sum_{i=1}^{k} a_i \boldsymbol{x}_i \boldsymbol{y}_i^T\right|\right|_F^2 \tag{4}$$

is minimized.

b. What are the set $\boldsymbol{x}_i$, $\boldsymbol{y}_i$ that minimize the minimum in (4)?

c. What constraint is there on k so that the solution is unique?

5. Consider a non–white noise process $x[k]$ of duration m with a known covariance matrix $\boldsymbol{\Sigma}$. The sequence $f[k]$ of duration $n \ll m$ operates on $x[k]$ to give an output sequence $y[k]$ according to

$$y[k] = \sum_i f[i]x[k-i].$$

a. Show that this operation (convolution) can be expressed as a matrix–vector multiplication.

b. Find $f[k]$ so that $\left|\left|y[k]\right|\right|_2$ is minimized, subject to $||\boldsymbol{f}||_2 = 1$.

6. (From *Applied Numerical Linear Algebra*, by James W. Demmel.)

Let $\boldsymbol{A}, \boldsymbol{B}$ and $\boldsymbol{C}$ be matrices with dimensions such that the product $\boldsymbol{A}^T\boldsymbol{C}\boldsymbol{B}^{\boldsymbol{T}}$ is defined. Let χ be the set of matrices $\boldsymbol{X}$ minimizing $||\boldsymbol{AXB} - \boldsymbol{C}||_F$, and let $\boldsymbol{X}_o$ be the unique member of χ minimizing $||\boldsymbol{X}||_F$. Show that $\boldsymbol{X}_o = \boldsymbol{A}^+\boldsymbol{C}\boldsymbol{B}^+$. *Hint:* use the SVDs of $\boldsymbol{A}$ and $\boldsymbol{B}$.

7. Let $\boldsymbol{A} \in \Re^{m\times n}$, and $\boldsymbol{b} \in \Re^n$. Find $\boldsymbol{x}$ so that $||\boldsymbol{A} - \boldsymbol{x}\boldsymbol{b}^T||_F$ is minimized. Hint: For matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ of compatible dimension, trace$(\boldsymbol{AB})$ = trace$(\boldsymbol{BA})$.

8. We have observations of data $\boldsymbol{z}(t) \in \Re^m, t = 1, \ldots, L$ which are generated according to

$$\boldsymbol{z}(t) = \boldsymbol{A}\boldsymbol{x}(t) + \boldsymbol{n}(t) \tag{5}$$

where $\boldsymbol{x}(t) \in \Re^n$ are known vectors, and $\boldsymbol{A} \in \Re^{m\times n}$ is unknown and to be estimated. The noise $\boldsymbol{n}(t)$ is a Gaussian–distributed random variable with zero mean and covariance matrix $\boldsymbol{Q}$, which is also to be estimated.

The matrices $\boldsymbol{A}$ and $\boldsymbol{Q}$ can be estimated by maximizing the logarithm of the probability distribution $p(\boldsymbol{z}\big|\boldsymbol{A}, \boldsymbol{Q}, \boldsymbol{x}(t))$ with respect to $\boldsymbol{A}$ and $\boldsymbol{Q}$. The log of this distribution is given through (5) as

$$\log p(\boldsymbol{z}(t)\big|\boldsymbol{A}, \boldsymbol{Q}, \boldsymbol{x}(t)) = \sum_{t=1}^{L} \big(\boldsymbol{z}(t) - \boldsymbol{A}\boldsymbol{x}(t)\big)^T \boldsymbol{Q}^{-1}\big(\boldsymbol{z}(t) - \boldsymbol{A}\boldsymbol{x}(t)\big) + \log\det \boldsymbol{Q} \tag{6}$$

Show that the maximization of (6) with respect to $\boldsymbol{A}$ and $\boldsymbol{Q}$ yields

$$\hat{\boldsymbol{A}} = \left(\sum_{t=1}^{L} \boldsymbol{z}(t)\boldsymbol{x}^T(t)\right)\left(\sum_{t=1}^{L} \boldsymbol{x}(t)\boldsymbol{x}^T(t)\right)^{-1} \tag{7}$$

$$\hat{\boldsymbol{Q}} = \left(\sum_{t=1}^{L} \boldsymbol{z}(t)\boldsymbol{z}^T(t)\right) - \left(\sum_{t=1}^{L} \hat{\boldsymbol{A}}\boldsymbol{x}(t)\boldsymbol{z}^T(t)\right) \tag{8}$$

respectively. The distribution $p(\boldsymbol{z}\big|\boldsymbol{A}, \boldsymbol{Q}, \boldsymbol{x}(t))$ is referred to as the *likelihood function* and the values given by (7) and (8) are *maximum likelihood* estimates.

9. Download the file q1.mat from the course website. The first column of the variable z is an array of x-variables and the second column contains the corresponding y-variables. The (x, y) pairs are points taken from a circle. Using methods taken in class, calculate the centre and radius of the circle. Explain your method fully. Include your matlab code.

10. Consider the projector $\boldsymbol{P} = \boldsymbol{X}\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T$ where $\boldsymbol{X}$ is a tall matrix, and let $\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-\frac{1}{2}}$ be an inverse square root factor of $\boldsymbol{X}^T\boldsymbol{X}$; i.e., $\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-\frac{T}{2}}\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-\frac{1}{2}} = \left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}$. What special property does the matrix $\boldsymbol{X}\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-\frac{1}{2}}$ possess? Use this property to explain how the projector eliminates the orthogonal component of a vector projected onto $R(\boldsymbol{X})$.

11. In file assig1Q4.mat on the course website, you will find three variables, $\boldsymbol{A} \in \mathbb{R}^{6\times 4}, \boldsymbol{b} \in \mathbb{R}^6$ and $\boldsymbol{B} \in \mathbb{R}^{6\times 4}$. (Do NOT use the pseudo–inverse to solve this problem.)

- **a.** Using the SVD, find an expression for $\boldsymbol{x}$ s.t. $||\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}||_2$ is minimized. Explain how this technique may be used for solving an overdetermined system of equations. Find an expression for the value of this 2-norm difference for the optimal value of $\boldsymbol{x}$. Using Matlab, provide specific values for these quantities.
- **b.** Is $\boldsymbol{b} \in R(\boldsymbol{A})$? If not, suggest a method of finding the closest $\boldsymbol{b}_o$ to $\boldsymbol{b}$ (in the 2-norm sense), which is. Repeat part **a** using $\boldsymbol{b}_o$ instead of $\boldsymbol{b}$ and comment.
- **c.** Repeat parts **a** and **b** above using the matrix $\boldsymbol{B}$ instead of $\boldsymbol{A}$.

12. Consider the configuration shown in Fig. 1a. The sequence $x[n]$ is of length m and $f_1[n]$ and $f_2[n]$ are sequences of length $n \ll m$. The outputs $y_1[n]$ and $y_2[n]$ are the convolution of $x[n]$ with $f_1[n]$ and $f_2[n]$ respectively; i.e.,

$$y_i[n] = \sum_k f_i[k]x[n-k], \quad i \in [1, 2].$$

We observe only the sequences $y_1[n]$ and $y_2[n]$.

- **a.** Show that the convolution operation can be expressed as a matrix–vector multiplication.
- **b.** Show how $f_1[n]$, $f_2[n]$ and $x[n]$ can all be determined, based ONLY on the observations $y_1[n]$ and $y_2[n]$. *Hint: Consider the configuration of Fig. 1b. What are w_1 and w_2 so that the output $z[n] = 0$?* .
- **c.** Are there any required conditions on $x[n]$, or on $f_1[n], f_2[n]$?

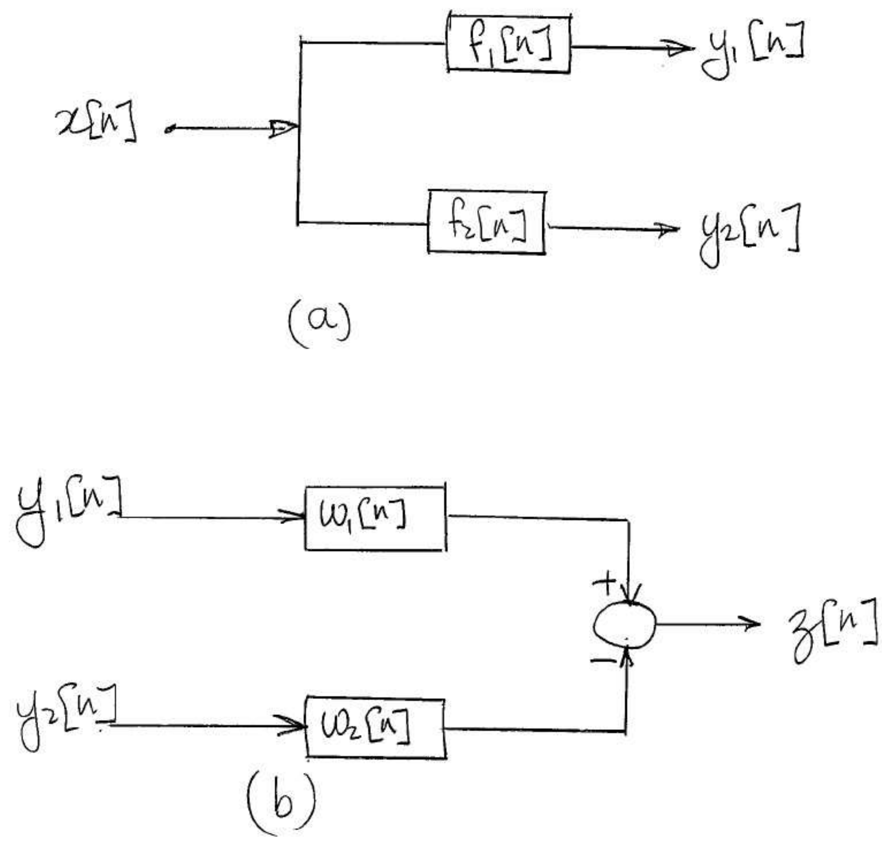

Figure 1: Configuration for Question 10.