Math 125 **Problem Set 4** **Due: Thursday 11/10 at 11:59 p.m. EST**

Instruction: Read the assignment policy. For problem 3, include a printout your code with your homework submission. You should submit your assignment on Gradescope.

1. Let $C[0,1]$ be the vector space of real-valued functions that are continuous on $[0,1]$. Prove that the following definition gives a valid inner product

$$\langle f\,, g\rangle = \int_0^1 f(x)\, g(x)\, dx,$$

where f and g are elements of $C[0,1]$.

Remark: An inner product $\langle\cdot,\cdot\rangle$ satisfies the following four properties. Let f, g, and h be vectors and c be a scalar.

1. $\langle f+g, h\rangle = \langle f, h\rangle + \langle g, h\rangle$.
2. $\langle cg, h\rangle = c\langle g, h\rangle$.
3. $\langle f, g\rangle = \langle g, f\rangle$.
4. $\langle f, f\rangle \geq 0$ and equal to 0 if and only if $f = 0$.

Solution: We check all the four properties.

$$\langle f+g, h\rangle = \int_0^1 (f(x)+g(x))h(x)\, dx = \int_0^1 f(x)h(x)\, dx + \int_0^1 g(x)h(x)\, dx = \langle f, h\rangle + \langle g, h\rangle,$$

where we have used the distributive property of function addition over function multiplication and the fact that the integral of a sum is the sum of the integrals. Next, we consider the second property.

$$\langle cg, h\rangle = \int_0^1 cg(x)h(x)\, dx = c\int_0^1 g(x)h(x)\, dx = c\langle g, h\rangle,$$

where we have used the fact that the integral of a constant multiple of a function is the product of the constant with the integral of the function. The third property simply follows from the fact that the product of two functions is commutative.

$$\langle f, g\rangle = \int_0^1 f(x)g(x)\, dx = \int_0^1 g(x)f(x)\, dx = \langle g, f\rangle$$

Finally, we consider the last property.

$$\langle f, f\rangle = \int_0^1 [f(x)]^2 \geq 0,$$

which follows from the non-negativity of $f(x)^2$. We note that $f(x)^2$ is continuous. Then, $\int_0^1 [f(x)]^2 = 0$ if and only if $f(x) = 0$. Having checked all the properties, we can conclude that the inner product is valid.

2. Let W be a subspace of functions on the interval $[0,1]$ with basis $\{f_1, f_2, f_3\}$. We define an inner product on this space as follows

$$\langle f\,, g\rangle = \int_0^1 f(x)\, g(x)\, dx,$$

where f and g are elements of W.

(a) Define the following matrix $\mathbf{G} = \begin{bmatrix} \langle f_1\,, f_1\rangle & \langle f_1\,, f_2\rangle & \langle f_1\,, f_3\rangle \\ \langle f_2\,, f_1\rangle & \langle f_2\,, f_2\rangle & \langle f_2\,, f_3\rangle \\ \langle f_3\,, f_1\rangle & \langle f_3\,, f_2\rangle & \langle f_3\,, f_3\rangle \end{bmatrix}$. Prove that G is symmetric and positive definite. [Hint: Show that $\mathbf{z}^T\mathbf{G}\mathbf{z} > 0$ for any nonzero vector $\mathbf{z}$].

(b) Any function g that is in W can be written as a linear combination of f_1, f_2 and f_3 i.e. $g = c_1 f_1 + c_2 f_2 + c_3 f_3$. Let $\mathbf{c}$ denote the vector $\begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix}$. Prove that $\mathbf{c}$ is a solution to the linear system $\mathbf{G}\mathbf{c} = \mathbf{y}$ where $\mathbf{y} = \begin{bmatrix} \langle g\,, f_1\rangle \\ \langle g\,, f_2\rangle \\ \langle g\,, f_3\rangle \end{bmatrix}$. Is there a unique solution?

(c) Let h be a function that is not in W. From the best approximation theorem, we know that the closest element in W to h is the orthogonal projection of h onto W denoted by $\hat{h}$. Find an expression for $\hat{h}$. [Hint: Use orthogonality relations and the results in (b)].

Remark: The set $\{f_1, f_2, f_3\}$ is a basis but not necessarily **orthogonal**. Note that $\mathbf{z}^T\mathbf{G}\mathbf{z}$ can be written as

$$\mathbf{z}^T\mathbf{G}\mathbf{z} = \sum_{i=1}^{3}\sum_{j=1}^{3} G_{i,j} z_i z_j$$

Solution:

(a) We check if $\mathbf{z}^T\mathbf{G}\mathbf{z} > 0$ for any nonzero vector $\mathbf{z}$.

$$\mathbf{z}^T\mathbf{G}\mathbf{z} = \sum_{i=1}^{3}\sum_{j=1}^{3} G_{i,j} z_i z_j = \sum_{i=1}^{3}\sum_{j=1}^{3} \langle f_i\,, f_j\rangle z_i z_j = \sum_{i=1}^{3}\sum_{j=1}^{3} \langle z_i f_i\,, z_j f_j\rangle = \langle \sum_{i=1}^{3} z_i f_i, \sum_{i=1}^{3} z_i f_i\rangle \geq 0$$

Above, the third equality uses the inner product property $\langle cg, h\rangle = c\langle g, h\rangle$ where c is any scalar. The fourth equality uses the fact that the indices i and j are dummy. The final equality uses the fact that $\langle f, f\rangle \geq 0$. It now remains to show that $\langle \sum_{i=1}^{3} z_i f_i, \sum_{i=1}^{3} z_i f_i\rangle > 0$. $\langle \sum_{i=1}^{3} z_i f_i, \sum_{i=1}^{3} z_i f_i\rangle = 0$ if and only if $\sum_{i=1}^{3} z_i f_i = 0$. Since $\{f_1, f_2, f_3\}$ is basis, it is a linearly independent set. Therefore, $\sum_{i=1}^{3} z_i f_i = 0$ will only hold if $z_1 = z_2 = z_3 = 0$ i.e. z is a zero vector. For all nonzero vector $\mathbf{z}$, $\mathbf{z}^T\mathbf{G}\mathbf{z} > 0$. Therefore, $\mathbf{G}$ is positive definite.

(b) Consider $\langle g\,, f_1\rangle$. Using the properties of the inner product, we have

$$\langle g\,, f_1\rangle = c_1\langle f_1\,, f_1\rangle + c_2\langle f_1\,, f_2\rangle + c_1\langle f_1\,, f_3\rangle$$

In a similar manner, we obtain

$$\langle g\,, f_2\rangle = c_1\langle f_2\,, f_1\rangle + c_2\langle f_2\,, f_2\rangle + c_1\langle f_2\,, f_3\rangle$$
$$\langle g\,, f_3\rangle = c_1\langle f_3\,, f_1\rangle + c_2\langle f_3\,, f_2\rangle + c_1\langle f_3\,, f_3\rangle$$

The above three equations can be written as $\mathbf{G}\mathbf{c} = \mathbf{y}$ where $\mathbf{y} = \begin{bmatrix} \langle g\,,f_1\rangle \\ \langle g\,,f_2\rangle \\ \langle g\,,f_3\rangle \end{bmatrix}$ and $\mathbf{c} = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix}$. From (a), we know that $\mathbf{G}$ is positive definite. This implies that $\mathbf{G}$ is invertible. Therefore, there is a unique solution to the linear system.

(c) Since $\hat{h}$ is in W, it can be written as a linear combination of f_1, f_2 and f_3 i.e. $\hat{h} = c_1 f_1 + c_2 f_2 + c_3 f_3$ where the constants c_1, c_2, c_3 are to be determined. Following the same procedure as in (b), we obtain $\mathbf{G}\mathbf{c} = \mathbf{y}$ where $\mathbf{y} = \begin{bmatrix} \langle \hat{h}\,,f_1\rangle \\ \langle \hat{h}\,,f_2\rangle \\ \langle \hat{h}\,,f_3\rangle \end{bmatrix}$. Note that, in contrast to the case in (b), we do not know the right hand side vector $\mathbf{y}$ as $\hat{h}$ is unknown (in fact that is the goal of this problem). In what remains, we will show that $\mathbf{y}$ can be determined using orthogonality relations. Using best approximation theorem, $\hat{h}$ is the orthogonal projection of h onto W. From this, it follows that $(\hat{h} - h) \perp z$ for any $z \in W$. In particular, using $z = f_1, z = f_2$ and $z = f_3$, we obtain

$$(\hat{h} - h) \cdot f_1 = 0 \rightarrow \hat{h} \cdot f_1 = h \cdot f_1$$
$$(\hat{h} - h) \cdot f_2 = 0 \rightarrow \hat{h} \cdot f_2 = h \cdot f_2$$
$$(\hat{h} - h) \cdot f_3 = 0 \rightarrow \hat{h} \cdot f_3 = h \cdot f_3$$

Therefore, $\mathbf{y}$ can also be written as $\mathbf{y} = \begin{bmatrix} \langle h\,,f_1\rangle \\ \langle h\,,f_2\rangle \\ \langle h\,,f_3\rangle \end{bmatrix}$. Since h is known, the vector $\mathbf{y}$ is explicit. To find $\mathbf{c}$, we can solve the linear system $\mathbf{G}\mathbf{c} = \mathbf{y}$. From (a), we know that $\mathbf{G}$ is positive definite. This implies that $\mathbf{G}$ is invertible. Therefore, there is a unique solution to the linear system.

3. Download the dataset `least_square_data.mat` from HW3 folder. If you use Python for programming, download `least_square_data.csv`.

(a) Find the best linear squares fit to the data. Plot the linear fit.

(b) Find the best quadratic squares fit to the data. Plot the quadratic fit.

[Remark: Note that the best least squares fit is determined from the normal equations $(\mathbf{A}^T\mathbf{A})\mathbf{c} = \mathbf{A}^T\mathbf{y}$. For the above problems, it suffices to define $\mathbf{A}$ and set up a linear system. You can use any numerical solver to find the coefficients $\mathbf{c}$ that obtain the least squares fit].

Solution: For details of implementation, refer to `least_squares.m` script in the `HW4_Soln` folder. For both cases, it suffices to construct the matrix $\mathbf{A}$ and the vector $\mathbf{y}$. Let the data be denoted as $(t_i, y_i)_{i=1}^{401}$. Then, the vector $\mathbf{y}$ is constituted from all observations y_i. Let $\mathbf{t}$ denote the vector of all t_i. For linear fit, $\mathbf{A} = [\mathbf{1}\ \mathbf{t}]$ where $\mathbf{1} \in \mathcal{R}^{401}$ denotes a vector of ones. For linear fit, $\mathbf{A} = [\mathbf{1}\ \mathbf{t}\ \mathbf{t}^2]$ where $\mathbf{t}^2$ denotes element wise squaring of $\mathbf{t}$.

(a) The best linear fit is $p(t) = 1.5666 + 7.1154t$. The figure is shown below.

(b) The best quadratic fit is $p(t) = 0.9341 + 3.3945t + 1.8605t^2$. The figure is shown below.

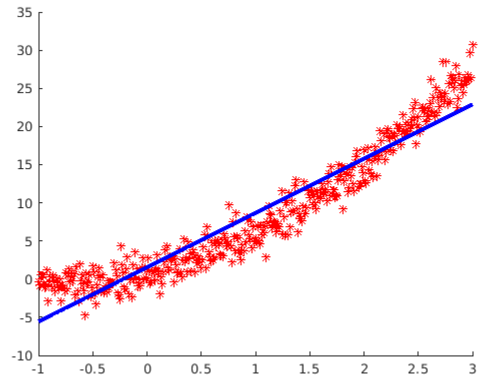


Figure 1: Original data: red, Linear fit: blue

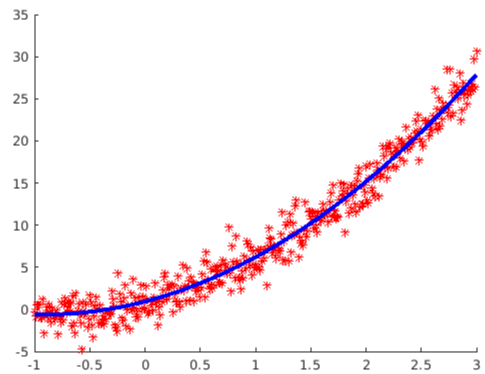


Figure 2: Original data: red, Quadratic fit: blue

4. We study the least problem for the case where all observations might not be equally reliable. The standard least square formulation assumes that each data point is equally reliable as the other. In certain applications, due to the nature of acquisition of data, this assumption is violated. In this setting, the observations can be assigned different weights. A larger weight indicate that a data point in consideration is more reliable. To integrate the weight in the least square formulation, define the following inner product: $\langle \boldsymbol{x} \,, \boldsymbol{y} \rangle = \boldsymbol{x}^T \boldsymbol{W}^T \boldsymbol{W} \boldsymbol{y}$ where $\boldsymbol{W}$ is a non-singular matrix of weights.

(a) Following the derivation of the normal equations for the least squares problem , show that the weighted least square solution is equivalent to solving the the weighted normal equation: $\boldsymbol{A}^T \boldsymbol{W}^T \boldsymbol{W} \boldsymbol{A} \mathbf{x} = \boldsymbol{A}^T \boldsymbol{W}^T \boldsymbol{W} \boldsymbol{b}$?

(b) Find the weighted least-squares solution to $\mathbf{A}\mathbf{x} = \boldsymbol{b}$:

$$\mathbf{A} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{pmatrix} \qquad \mathbf{b} = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \qquad \boldsymbol{W} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Plot the data and the least-squares fit on a graph.

(c) **Extra credit**: Why is it necessary to require invertibility of $\mathbf{W}$? [**Hint**: Check when the inner product is well-defined].

Solution

(a) Recall the orthogonality relation for the least squares problem: $\boldsymbol{b} - \mathbf{Ax} \perp \text{Range}(\mathbf{A})$. Let $\mathbf{a}_j$ denote the j-th column of $\mathbf{A}$. The orthogonality relation implies the following set of equations:

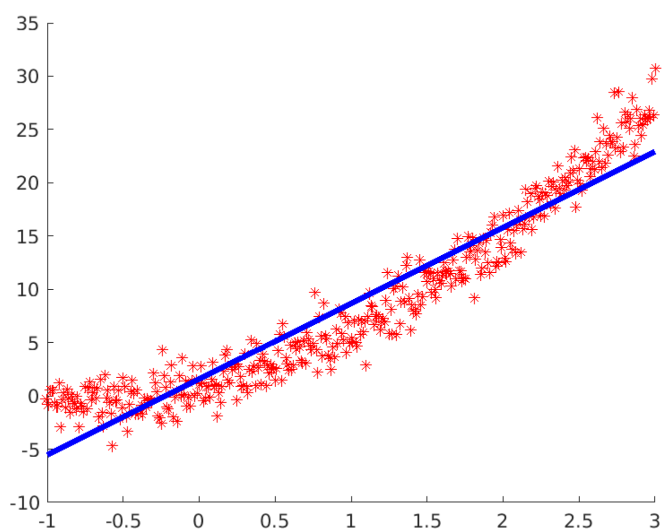
$$\begin{aligned} \boldsymbol{b} - \mathbf{Ax} \perp \mathbf{a}_1 &\rightarrow \langle \mathbf{a}_1 , \boldsymbol{b} - \mathbf{Ax} \rangle = 0 \\ \boldsymbol{b} - \mathbf{Ax} \perp \mathbf{a}_2 &\rightarrow \langle \mathbf{a}_2 , \boldsymbol{b} - \mathbf{Ax} \rangle = 0 \\ \vdots \qquad & \qquad\qquad\qquad \vdots \\ \boldsymbol{b} - \mathbf{Ax} \perp \mathbf{a}_n &\rightarrow \langle \mathbf{a}_3 , \boldsymbol{b} - \mathbf{Ax} \rangle = 0 \end{aligned}$$

We now use the given inner product to rewrite the above set of equations in the following equivalent form:

$$\begin{aligned} \mathbf{a}_1^T \boldsymbol{W}^T \boldsymbol{W} (\boldsymbol{b} - \mathbf{Ax}) &= 0 \\ \mathbf{a}_2^T \boldsymbol{W}^T \boldsymbol{W} (\boldsymbol{b} - \mathbf{Ax}) &= 0 \\ \vdots \qquad & \quad \vdots \\ \mathbf{a}_n^T \boldsymbol{W}^T \boldsymbol{W} (\boldsymbol{b} - \mathbf{Ax}) &= 0 \end{aligned}$$

The above set of equations can be compactly represented as: $\mathbf{A}^T \boldsymbol{W}^T \boldsymbol{W} (\boldsymbol{b} - \mathbf{Ax}) = 0$. This leads to the weighted normal equation $\boldsymbol{A}^T \boldsymbol{W}^T \boldsymbol{W} \boldsymbol{A}\mathbf{x} = \boldsymbol{A}^T \boldsymbol{W}^T \boldsymbol{W} \boldsymbol{b}$ as desired.

(b) For implementation, see `weighted_ls.m` in `HW4_Soln` folder. Figure b shows the data points and the least square fit. Note that the second point, which is the highest weight, is the closest to the line as desired.



(c) To define the inner product $\langle \boldsymbol{x} , \boldsymbol{y} \rangle = \boldsymbol{x}^T \boldsymbol{W}^T \boldsymbol{W} \boldsymbol{y}$, we require that $\langle \mathbf{x} , \mathbf{x} \rangle = \mathbf{x}^T \boldsymbol{W}^T \boldsymbol{W} \mathbf{x} = 0$ if and only if $\mathbf{x} = \mathbf{0}$. Note that $\langle \mathbf{x} , \mathbf{x} \rangle = ||\mathbf{Wx}||_2^2$. This will be zero if and only if $\mathbf{Wx} = \mathbf{0}$. There will be a trivial solution if and only if $\mathbf{W}$ is invertible. Otherwise, there is a non-trivial solution to $\mathbf{Wx} = \mathbf{0}$ and $\langle \mathbf{x} , \mathbf{x} \rangle = 0$ could hold for non-trivial vectors (which violates the requirement of the inner product).

5. In this problem, we study the approximation of a function using Chebyshev polynomials. Recall the definition of a Chebyshev polynomial $T_n(x) = \cos(n \arccos(x))$ where $x \in [-1, 1]$. In lecture 8, we noted that $T_0(x) = 1, T_1(x) = x$ and $T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x)$.

(a) Prove that the set $\{T_0(x), T_1(x), T_2(x), ..., T_n(x)\}$ is orthogonal with the following inner product

$$\langle T_i(x) \, , T_j(x) \rangle = \int_{-1}^{1} \left(T_i(x) T_j(x) \frac{1}{\sqrt{1-x^2}} \right) dx.$$

(b) **Extra credit**: Find the least-squares approximation of the function $f(x) = \sin(x)$ on the interval $[-1, 1]$ using the Chebyshev basis $\{T_0(x), T_1(x), T_2(x)\}$. Plot the approximation and the function $f(x) = \sin(x)$ on the same plot in the interval $[-1, 1]$. Define the error of the approximation to be $E = \max_{-1 \le x \le 1} |f(x) - \hat{f}(x)|$ where $\hat{f}$ denotes the least squares approximation. What is E?

Solution

(a) The inner product can be written as follows

$$\begin{aligned} \langle T_i(x) \, , T_j(x) \rangle &= \int_{-1}^{1} \cos(i \arccos(x)) \, \cos(j \arccos(x)) \frac{1}{\sqrt{1-x^2}} \, d\theta \\ &= -\int_{\pi}^{0} \cos(i \, \theta) \, \cos(j \, \theta) \, d\theta, \end{aligned}$$

where we have substituted $\theta = \arccos(x)$ and used the fact that $d\theta = -\frac{1}{\sqrt{1-x^2}} \, dx$. Using the cosine product identity, $\cos(i \, \theta) \, \cos(j\theta) = \frac{1}{2} \left(\cos(i\theta - j \, \theta) + \cos(i\theta + j\theta) \right)$. Using this, we obtain

$$\begin{aligned} \langle T_i(x) \, , T_j(x) \rangle &= \frac{1}{2} \int_{0}^{\pi} \left(\cos(i \, \theta - j \, \theta) + \cos(i \, \theta + j \, \theta) \right) \, d\theta \\ &= \frac{1}{2} \frac{1}{i-j} \sin(i\theta - j\theta)|_{\theta=0}^{\theta=\pi} + \frac{1}{2} \frac{1}{i+j} \sin(i \, \theta + j \, \theta)|_{\theta=0}^{\theta=\pi} \\ &= \frac{1}{2} \frac{1}{i-j} \sin((i-j)\pi) + \frac{1}{2} \frac{1}{i+j} \sin((i+j)\pi) \end{aligned}$$

When $i \neq j$, $\langle T_i(x) \, , T_j(x) \rangle = 0$ since $\sin(k\pi) = 0$ for any integer k. Therefore, the set is orthogonal with the given product. For completeness, we consider the inner products when $i = j \neq 0$ and $i = j = 0$. When $i = j \neq 0$, we have

$$\langle T_i(x) \, , T_j(x) \rangle = \frac{1}{2} \int_{0}^{\pi} \left(1 + \cos(i \, \theta + j \, \theta) \right) \, d\theta = \frac{\pi}{2}$$

When $i = j = 0$, we have

$$\langle T_i(x) \, , T_j(x) \rangle = \frac{1}{2} \int_{0}^{\pi} 2 \, d\theta = \pi$$

(b) Having established orthogonality of the Chebyshev polynomials in (a), using best approximation theorem, the least-squares approximation is

$$\hat{f} = c_0 T_0 + c_1 T_1 + c_2 T_2,$$

where $c_i = \frac{\langle f\,, T_i \rangle}{\langle T_i\,, T_i \rangle}$. We start with c_0.

$$c_0 = \frac{\langle \sin(x)\,, 1 \rangle}{\langle T_0\,, T_0 \rangle} = \frac{1}{\pi} \int_{-1}^{1} \sin(x) \frac{1}{\sqrt{1-x^2}}\, dx = 0,$$

since the integrand is odd. We next consider c_1.

$$c_1 = \frac{\langle \sin(x)\,, x \rangle}{\langle T_1\,, T_1 \rangle} = \frac{2}{\pi} \int_{-1}^{1} x \sin(x) \frac{1}{\sqrt{1-x^2}}\, dx \approx 0.88,$$

where the last approximation can be obtained numerically or relating integrand to Bessel's function of first kind. Finally, we compute c_2.

$$c_2 = \frac{\langle \sin(x)\,, x \rangle}{\langle T_2\,, T_2 \rangle} = \frac{2}{\pi} \int_{-1}^{1} (2 * x^2 - 1) \sin(x) \frac{1}{\sqrt{1-x^2}}\, dx = 0,$$

since the integrand is odd. Therefore, the least square approximation is

$$\hat{f} = 0.88x$$

$E = 0.0394$ and a plot of the true function against the approximation is shown below