Numerical Methods 2023 — Final exam

Solutions

Problem 1 (10 pts). Consider a linear system

$$\begin{bmatrix} 3 & -1 & 2 \\ -1 & 3 & -1 \\ 2 & -1 & 3 \end{bmatrix} \boldsymbol{x} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}.$$

Please use CG method with

$$\epsilon = 0$$

to solve it, and show your calculations including $k, \boldsymbol{x}, \boldsymbol{r}, \rho, \boldsymbol{w}, \alpha$. Calculate $A\boldsymbol{x}$ to verify if it is equal to $\boldsymbol{b}$.

Solution.

We have

$$k = 0,\ \boldsymbol{x} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix},\ \boldsymbol{r} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix},\ \rho_0 = \boldsymbol{r}^T \boldsymbol{r} = 1$$

in the beginning. For $k = 1$, since

$$\sqrt{\rho_0} = 1 > 0$$

we calculate

$$\begin{aligned} \boldsymbol{p} =& \boldsymbol{r} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix},\ \boldsymbol{w} = A\boldsymbol{p} = \begin{bmatrix} -1 \\ 3 \\ -1 \end{bmatrix},\ \alpha = \frac{\rho_0}{\boldsymbol{p}^T \boldsymbol{w}} = \frac{1}{3}, \\ \boldsymbol{x} =& \boldsymbol{x} + \alpha \boldsymbol{p} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} + \frac{1}{3} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 1/3 \\ 0 \end{bmatrix}, \\ \boldsymbol{r} =& \boldsymbol{r} - \alpha \boldsymbol{w} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} - \frac{1}{3} \begin{bmatrix} -1 \\ 3 \\ -1 \end{bmatrix} = \begin{bmatrix} 1/3 \\ 0 \\ 1/3 \end{bmatrix}, \\ \rho_1 =& \boldsymbol{r}^T \boldsymbol{r} = \frac{2}{9}. \end{aligned}$$

Next, when $k = 2$,

$$\sqrt{\rho_1} = \frac{\sqrt{2}}{3} > 0.$$

Thus, we calculate

$$\begin{aligned}
\beta =& \frac{\rho_1}{\rho_0} = \frac{2}{9}, \\
\boldsymbol{p} =& \boldsymbol{r} + \beta \boldsymbol{p} = \begin{bmatrix} 1/3 \\ 0 \\ 1/3 \end{bmatrix} + \frac{2}{9} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1/3 \\ 2/9 \\ 1/3 \end{bmatrix} \\
\boldsymbol{w} =& A\boldsymbol{p} = \begin{bmatrix} 13/9 \\ 0 \\ 13/9 \end{bmatrix}, \ \alpha = \frac{\rho_1}{\boldsymbol{p}^T \boldsymbol{w}} = \frac{2/9}{26/27} = \frac{3}{13}, \\
\boldsymbol{x} =& \boldsymbol{x} + \alpha \boldsymbol{p} = \begin{bmatrix} 0 \\ 1/3 \\ 0 \end{bmatrix} + \frac{3}{13} \begin{bmatrix} 1/3 \\ 2/9 \\ 1/3 \end{bmatrix} = \begin{bmatrix} 1/13 \\ 5/13 \\ 1/13 \end{bmatrix}, \\
\boldsymbol{r} =& \boldsymbol{r} - \alpha \boldsymbol{w} = \begin{bmatrix} 1/3 \\ 0 \\ 1/3 \end{bmatrix} - \frac{3}{13} \begin{bmatrix} 13/9 \\ 0 \\ 13/9 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \\
\rho_2 =& \boldsymbol{r}^T \boldsymbol{r} = 0.
\end{aligned}$$

When $k = 3$, we have

$$\sqrt{\rho_2} = 0,$$

which satisfies the stopping condition. Therefore, the algorithm stops, and the solution is

$$\boldsymbol{x} = \begin{bmatrix} 1/13 \\ 5/13 \\ 1/13 \end{bmatrix}.$$

To verify the answer,

$$A\boldsymbol{x} = \begin{bmatrix} 3 & -1 & 2 \\ -1 & 3 & -1 \\ 2 & -1 & 3 \end{bmatrix} \begin{bmatrix} 1/13 \\ 5/13 \\ 1/13 \end{bmatrix} = \begin{bmatrix} 3/13 - 5/13 + 2/13 \\ -1/13 + 15/13 - 1/13 \\ 2/13 - 5/13 + 3/13 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}.$$

Problem 2 (25 pts). In our slides "sparse_CG2.pdf", we have a Lemma tells that solving

$$\begin{aligned}
&\min_{\boldsymbol{p}} \|\boldsymbol{p} - \boldsymbol{r}_{k-1}\|_2 \\
&\text{s.t. } \boldsymbol{p} \in \operatorname{span}\{A\boldsymbol{p}_1, \ldots, A\boldsymbol{p}_{k-1}\}^\perp
\end{aligned} \tag{1}$$

is equivalent to solving

$$\min_{\boldsymbol{z}} \ \|\boldsymbol{r}_{k-1} - AP_{k-1}\boldsymbol{z}\|_2, \tag{2}$$

where

$$P_{k-1} = \begin{bmatrix} \boldsymbol{p}_1 & \cdots & \boldsymbol{p}_{k-1} \end{bmatrix}$$

Let us re-prove this Lemma in this problem. Without loss of generality, we assume

$$\boldsymbol{p} \in \mathbb{R}^n \text{ and } \boldsymbol{r}_{k-1} \in \mathbb{R}^n.$$

Thus, there exists some vectors

$$\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n \in \operatorname{span}\{A\boldsymbol{p}_1, \ldots, A\boldsymbol{p}_{k-1}\}^\perp$$

such that

$$\boldsymbol{r}_{k-1} = \sum_{i=1}^{k-1} a_i A\boldsymbol{p}_i + \sum_{j=1}^{n} b_j \boldsymbol{q}_j \tag{3}$$

$$\boldsymbol{p} = \sum_{j=1}^{n} c_j \boldsymbol{q}_j,$$

with some $\{a_i\}, \{b_j\}$ and $\{c_j\}$. Here $\{a_i\}, \{b_j\}$ are constants, but $\{c_j\}$ are variables to be decided. Then (1) can be rewritten to

$$\min_{c_1,\ldots,c_n} \left\| \sum_{j=1}^{n} c_j \boldsymbol{q}_j - \left(\sum_{i=1}^{k-1} a_i A\boldsymbol{p}_i + \sum_{j=1}^{n} b_j \boldsymbol{q}_j \right) \right\|_2,$$

which is further equivalent to

$$\min_{c_1,\ldots,c_n} \left\| \sum_{j=1}^{n} c_j \boldsymbol{q}_j - \left(\sum_{i=1}^{k-1} a_i A\boldsymbol{p}_i + \sum_{j=1}^{n} b_j \boldsymbol{q}_j \right) \right\|_2^2 \tag{4}$$

by taking the square.

To complete the proof, please help us to prove the following problems:

(a) (5 pts) Prove that

$$\left(\sum_{j=1}^{n} d_j \boldsymbol{q}_j \right)^T \left(\sum_{i=1}^{k-1} e_i A\boldsymbol{p}_i \right) = 0 \tag{5}$$

for any $\{d_j \mid j = 1, \ldots n\}$ and $\{e_i \mid i = 1, \ldots, k-1\}$.

(b) (10 pts) Use

$$\|\boldsymbol{x}\|_2^2 = \boldsymbol{x}^T \boldsymbol{x}.$$

to expand (4) and apply (5) to prove that

$$\boldsymbol{p} = \sum_{j=1}^{n} b_j \boldsymbol{q}_j$$

is an optimal solution for (4).

(c) (10 pts) To finish the proof, we show that there exists $\boldsymbol{z}$ such that it is an optimal solution of (2), and

$$\sum_{j=1}^{n} b_j \boldsymbol{q}_j = \boldsymbol{r}_{k-1} - AP_{k-1}\boldsymbol{z}.$$

To do this, you need to use (3) and (5). Then with (b), we have

$$\boldsymbol{p}_k = \boldsymbol{r}_{k-1} - AP_{k-1}\boldsymbol{z}_{k-1},$$

where $\boldsymbol{p}_k$ solves (1) and $\boldsymbol{z}_{k-1}$ solves (2).

Solution.

(a) Because

$$\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n \in \operatorname{span}\{A\boldsymbol{p}_1, \ldots, A\boldsymbol{p}_{k-1}\}^{\perp},$$

we know that

$$\boldsymbol{q}_j^T \left(\sum_{i=1}^{k-1} e_i A\boldsymbol{p}_i \right) = 0$$

for all $j = 1, \ldots, n$ with given any $\{e_i \mid i = 1, \ldots, k-1\}$. Therefore,

$$\left(\sum_{j=1}^{n} d_j \boldsymbol{q}_j \right)^T \left(\sum_{i=1}^{k-1} e_i A\boldsymbol{p}_i \right) = \sum_{j=1}^{n} d_j \boldsymbol{q}_j^T \left(\sum_{i=1}^{k-1} e_i A\boldsymbol{p}_i \right) = 0.$$

(b) Because

$$\|\boldsymbol{x}\|_2^2 = \boldsymbol{x}^T \boldsymbol{x},$$

we have that is the function minimized in (4)

$$\begin{aligned}
&\left(\sum_{j=1}^{n} c_j \boldsymbol{q}_j - \left(\sum_{i=1}^{k-1} a_i A\boldsymbol{p}_i + \sum_{j=1}^{n} b_j \boldsymbol{q}_j \right) \right)^T \left(\sum_{j=1}^{n} c_j \boldsymbol{q}_j - \left(\sum_{i=1}^{k-1} a_i A\boldsymbol{p}_i + \sum_{j=1}^{n} b_j \boldsymbol{q}_j \right) \right) \\
&= \left(\sum_{j=1}^{n} c_j \boldsymbol{q}_j \right)^T \left(\sum_{j=1}^{n} c_j \boldsymbol{q}_j \right) - 2 \left(\sum_{j=1}^{n} c_j \boldsymbol{q}_j \right)^T \left(\sum_{i=1}^{k-1} a_i A\boldsymbol{p}_i - \sum_{j=1}^{n} b_j \boldsymbol{q}_j \right) \\
&\quad + \left(\sum_{i=1}^{k-1} a_i A\boldsymbol{p}_i - \sum_{j=1}^{n} b_j \boldsymbol{q}_j \right)^T \left(\sum_{i=1}^{k-1} a_i A\boldsymbol{p}_i - \sum_{j=1}^{n} b_j \boldsymbol{q}_j \right). \qquad (6)
\end{aligned}$$

By the result of (a), (**??**) can be derived as

$$\begin{aligned}
&\left(\sum_{j=1}^{n} c_j \boldsymbol{q}_j \right)^T \left(\sum_{j=1}^{n} c_j \boldsymbol{q}_j \right) + 2 \left(\sum_{j=1}^{n} c_j \boldsymbol{q}_j \right)^T \left(\sum_{j=1}^{n} b_j \boldsymbol{q}_j \right) \\
&+ \left(\sum_{i=1}^{k-1} a_i A\boldsymbol{p}_i \right)^T \left(\sum_{i=1}^{k-1} a_i A\boldsymbol{p}_i \right) + \left(\sum_{j=1}^{n} b_j \boldsymbol{q}_j \right)^T \left(\sum_{j=1}^{n} b_j \boldsymbol{q}_j \right) \\
&= \left\| \sum_{j=1}^{n} c_j \boldsymbol{q}_j - \sum_{j=1}^{n} b_j \boldsymbol{q}_j \right\|_2^2 + \left(\sum_{i=1}^{k-1} a_i A\boldsymbol{p}_i \right)^T \left(\sum_{i=1}^{k-1} a_i A\boldsymbol{p}_i \right)
\end{aligned}$$

so the minimization problem (4) is then equivalent to

$$\min_{c_1,\ldots,c_n} \left\| \sum_{j=1}^{n} c_j \boldsymbol{q}_j - \sum_{j=1}^{n} b_j \boldsymbol{q}_j \right\|_2^2 + \left(\sum_{i=1}^{k-1} a_i A\boldsymbol{p}_i \right)^T \left(\sum_{i=1}^{k-1} a_i A\boldsymbol{p}_i \right) \equiv \min_{c_1,\ldots,c_n} \left\| \sum_{j=1}^{n} c_j \boldsymbol{q}_j - \sum_{j=1}^{n} b_j \boldsymbol{q}_j \right\|_2^2. \qquad (7)$$

Since we can take

$$c_j = b_j$$

for all $j = 1, \ldots, n$ as the solution of (**??**), and it implies that

$$\boldsymbol{p} = \sum_{j=1}^{n} c_j \boldsymbol{q}_j = \sum_{j=1}^{n} b_j \boldsymbol{q}_j$$

is an optimal solution.

(c) By (3), the square of the function minimized in (2) is equal to

$$\begin{aligned}\left\|\sum_{i=1}^{k-1} a_i A\boldsymbol{p}_i + \sum_{j=1}^{n} b_j \boldsymbol{q}_j - AP_{k-1}\boldsymbol{z}\right\|_2^2 &= \left\|\sum_{i=1}^{k-1} a_i A\boldsymbol{p}_i + \sum_{j=1}^{n} b_j \boldsymbol{q}_j - \sum_{i=1}^{k-1} z_i A\boldsymbol{p}_i\right\|_2^2 \\ &= \left\|\sum_{i=1}^{k-1}(a_i - z_i) A\boldsymbol{p}_i + \sum_{j=1}^{n} b_j \boldsymbol{q}_j\right\|_2^2 \\ &= \left\|\sum_{i=1}^{k-1}(a_i - z_i) A\boldsymbol{p}_i\right\|_2^2 + \left\|\sum_{j=1}^{n} b_j \boldsymbol{q}_j\right\|_2^2,\end{aligned}$$

where the last equality is from (5). We can take

$$z_i = a_i \tag{8}$$

for all $i = 1, \ldots, k-1$, to minimize (2). In this situation, (**??**) and (3) imply that

$$\boldsymbol{r}_{k-1} - AP_{k-1}\boldsymbol{z} = \sum_{j=1}^{n} b_j \boldsymbol{q}_j.$$

Problem 3 (15 pts). In our slides "equation_onevar1.pdf", we learned how to use Newton method to solve an one variable minimization problem

$$\min_x f(x)$$

by given an initial point $x^{(0)}$ and the update rule

$$x^{(k+1)} = x^{(k)} - \frac{f'(x^{(k)})}{f''(x^{(k)})}.$$

Now, we consider a two variables function

$$g(x_1, x_2) = x_1^2 - x_2^2,$$

and we also know that the update rule of two dimension Newton method is

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} - \left(\nabla^2 g(\boldsymbol{x}^{(k)})\right)^{-1} \nabla g(\boldsymbol{x}^{(k)}),$$

where

$$\nabla^2 g(\boldsymbol{x}) = \begin{bmatrix} \frac{\partial^2 g(\boldsymbol{x})}{\partial x_1^2} & \frac{\partial^2 g(\boldsymbol{x})}{\partial x_1 \partial x_2} \\ \frac{\partial^2 g(\boldsymbol{x})}{\partial x_2 \partial x_1} & \frac{\partial^2 g(\boldsymbol{x})}{\partial x_2^2} \end{bmatrix}.$$

(a) (10 pts) Run Newton method with an initial point

$$\boldsymbol{x}^{(0)} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

until reaching a point $\boldsymbol{x}^*$ that we cannot do further updates (i.e., $\nabla g(\boldsymbol{x}^*) = \boldsymbol{0}$.)

(b) (5 pts) Does the solution you find in (a) minimize the minimization problem

$$\min_{\boldsymbol{x}} g(\boldsymbol{x})?$$

If so, please prove it. Else, please give a counter example.

Solution.

(a) We can calculate

$$\nabla g(\boldsymbol{x}) = \begin{bmatrix} 2x_1 \\ -2x_2 \end{bmatrix}$$

and

$$\nabla^2 g(\boldsymbol{x}) = \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix}.$$

Therefore, when the update rule is applied,

$$\boldsymbol{x}^{(1)} = \begin{bmatrix} 2 \\ 1 \end{bmatrix} - \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix}^{-1} \begin{bmatrix} 4 \\ -2 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix} - \begin{bmatrix} 1/2 & 0 \\ 0 & -1/2 \end{bmatrix} \begin{bmatrix} 4 \\ -2 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix} - \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

and

$$\boldsymbol{x}^{(2)} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} - \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \boldsymbol{x}^{(1)}.$$

Hence,

$$\boldsymbol{x}^* = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

(b) There is a point $(0, 1)$ such that

$$g(0, 1) = -1 < 0 = g(0, 0),$$

so the solution $\boldsymbol{x}^*$ in (a) does not minimize the minimization problem

$$\min_{\boldsymbol{x}} g(\boldsymbol{x}).$$

The reason is that Newton method solves the first-order condition

$$\nabla g(\boldsymbol{x}) = \mathbf{0} \tag{9}$$

of the second-order approximation

$$g(\boldsymbol{x}) + \nabla g(\boldsymbol{x})^T \boldsymbol{d} + \frac{1}{2} \boldsymbol{d}^T \nabla^2 g(\boldsymbol{x}) \boldsymbol{d},$$

and (**??**) only guarantees the solution is a stationary point, which contains local minimum, local maximum, global minimum, global maximum, saddle point and inflection point. In function g, we have

$$g(\boldsymbol{x}^* + \delta \begin{bmatrix} 1 \\ 0 \end{bmatrix}) = (0 + \delta)^2 - 0^2 = \delta^2 > 0 = g(\boldsymbol{x}^*)$$

and

$$g(\boldsymbol{x}^* + \delta \begin{bmatrix} 0 \\ 1 \end{bmatrix}) = 0^2 - (0 + \delta)^2 = -\delta^2 < 0 = g(\boldsymbol{x}^*)$$

for all $\delta > 0$, which means $\boldsymbol{x}^*$ is a minimal point in the $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ direction but a maximal point in the $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$ direction. Usually, we call this point as a "saddle point", because the shape of the simplest example is like a saddle.

Problem 4 (25 pts). Consider the function

$$f(x) = -(x-2)^2.$$

To approximate the function $f(x)$ using discrete Fourier transform, we will consider these $2m$ points

$$(x_0, f(x_0)), \cdots, (x_{2m-1}, f(x_{2m-1}))$$

where

$$x_j = \frac{2j}{m}, \; j = 0, \dots, 2m-1.$$

We wish to approximate the function using a Fourier series with $2n$ coefficients:

$$S_n(z) = \frac{a_0 + a_n \cos nz}{2} + \sum_{k=1}^{n-1} \left(a_k \cos kz + b_k \sin kz\right) \tag{10}$$

In this problem, we will consider the case where $m = n = 2$.

(a) (5 pts) Transform the coordinates x_j into z_j so that the new coordinates z_j are in the interval $[-\pi, \pi]$. Then, calculate and list the values

$$(z_0, f(x_0)), \cdots, (z_{2m-1}, f(x_{2m-1})).$$

(b) (10 pts) Give the matrix A_2, A_1 and P required by the fast Fourier transform algorithm. That is, the discrete Fourier transform matrix F can be decomposed into

$$F = A_2 A_1 P.$$

(c) (5 pts) Following (b), calculate the coefficient vector $\boldsymbol{c}$ given by

$$\boldsymbol{c} = A_2 A_1 P \boldsymbol{y}$$

where $y_j = f(x_j)$. Then, calculate the coefficients a_k and b_k required by equation (6). Finally, write down the obtained Fourier series in terms of z in the form of (6).

(d) (5 pts) The series we obtained in subproblem (c) approximate the shifted and scaled version of $f(x)$ in the interval $[-\pi, \pi]$. However, we are interested in approximating the original $f(x)$ in the interval $[0, 4]$. Given any $x \in [0, 4]$, show how to calculate the approximated value given by the Fourier series. That is, you need to rewrite the $S_n(z)$ obtained in subproblem (c) in terms of x.

Solution.

(a) The given x_j are in the interval $[0, 4]$. Therefore, to transform them into the interval $[-\pi, \pi]$, we can let

$$z_j = \frac{x_j - 2}{2}\pi.$$

Then, the function data are therefore

$$(z_0, f(x_0)) = (-\pi, -4)$$
$$(z_1, f(x_1)) = (-\frac{\pi}{2}, -1)$$
$$(z_2, f(x_2)) = (0, 0)$$
$$(z_3, f(x_3)) = (\frac{\pi}{2}, -1).$$

(b) The δ we will be using is

$$e^{-i\pi/m} = -i.$$

To calculate A_2, we have

$$L = 2^2 = 4, \quad r = \frac{2m}{L} = \frac{4}{4} = 1,$$

and so

$$\begin{aligned} A_2 &= I_r \otimes B_L \\ &= I_1 \otimes B_4 \\ &= \begin{bmatrix} 1 \end{bmatrix} \otimes \begin{bmatrix} I_2 & \Omega_2 \\ I_2 & -\Omega_2 \end{bmatrix}, \quad \text{where } \Omega_2 = \begin{bmatrix} 1 & 0 \\ 0 & \delta \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & \delta \\ 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -\delta \end{bmatrix} \end{aligned} \tag{11}$$

To get A_1, we have

$$L = 2^1 = 2, \quad r = \frac{2m}{L} = 2.$$

and

$$\begin{aligned} A_1 &= I_r \otimes B_L \\ &= I_2 \otimes B_2 \\ &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \otimes \begin{bmatrix} I_1 & \Omega_1 \\ I_1 & -\Omega_1 \end{bmatrix}, \quad \text{where } \Omega_1 = \begin{bmatrix} 1 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & -1 \end{bmatrix} \end{aligned} \tag{12}$$

The permutation is calculated by reversing the binary representation.

$00 \to 00$	column 0 swapped to column 0
$01 \to 10$	column 1 swapped to column 2
$10 \to 01$	column 2 swapped to column 1
$11 \to 11$	column 3 swapped to column 3

Therefore, we have

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{13}$$

(c) First, the vector $\boldsymbol{y}$ is

$$\begin{bmatrix} f(-\pi) \\ f(-\frac{\pi}{2}) \\ f(0) \\ f(\frac{\pi}{2}) \end{bmatrix} = \begin{bmatrix} -4 \\ -1 \\ 0 \\ -1 \end{bmatrix}.$$

Therefore, by (??) we have

$$P\boldsymbol{y} = \begin{bmatrix} -4 \\ 0 \\ -1 \\ -1 \end{bmatrix}.$$

Then by (??) and (??) we can then calculate

$$A_1(P\boldsymbol{y}) = \begin{bmatrix} -4 \\ -4 \\ -2 \\ 0 \end{bmatrix}$$
$$\text{and } \boldsymbol{c} = A_2(A_1 P\boldsymbol{y}) = \begin{bmatrix} -6 \\ -4 \\ -2 \\ -4 \end{bmatrix}.$$

From $\boldsymbol{c}$ we can then calculate

$$a_0 = \frac{\operatorname{Re}(c_0)(-1)^0}{2} = -3$$
$$a_1 = \frac{\operatorname{Re}(c_1)(-1)^1}{2} = 2$$
$$a_2 = \frac{\operatorname{Re}(c_2)(-1)^2}{2} = -1$$
$$b_1 = \frac{\operatorname{Im}(c_1)(-1)^1}{2} = 0.$$

Therefore, the transformed series in terms of z is

$$S_n(z) = \frac{-3 - \cos(2z)}{2} + 2\cos(z). \tag{14}$$

(d) Because

$$z = \frac{x-2}{2}\pi,$$

for any x we can calculate the approximated value by simply substituting z in (??):

$$S_n(x) = \frac{-3 - \cos\left((x-2)\pi\right)}{2} + 2\cos\left(\frac{(x-2)}{2}\pi\right).$$

Problem 5 (25 pts). Consider the continuous least square problem on the interval $[a, b] = [0, 1]$.

(a) (5 pts) We would like to approximate

$$f(x) = x$$

using this list of polynomials:

$$\phi_1(x) = 1$$
$$\phi_2(x) = x^2$$

That is, we are solving the following minimization problem:

$$\min E = \int_0^1 (x - a_1\phi_1(x) - a_2\phi_2(x))^2 dx$$

Derive the linear system

$$A \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = b$$

we need to solve for this least square problem. You only need to calculate the values of A and b and are not required to solve the linear equations.

Note: Our functions ϕ_1 and ϕ_2 are 1 and x^2, instead of 1 and x. Thus, either you directly check $\frac{\partial E}{\partial a_j} = 0$, or you need to apply the equations on page 4 of "FFT_basic1.pdf" with care.

(b) (5 pts) Following the definition of orthogonality in lecture slide "FFT_basic1.pdf", are ϕ_1 and ϕ_2 orthogonal? Show your calculation.

(c) (10 pts) In order to solve the coefficient for each polynomials independently without solving a system of linear equations, we will need orthogonal polynomials. Identifying orthogonal polynomials are not easy. Fortunately, we can use the Gram-Schmidt process to orthogonalize a set of independent functions. Formally, given a set of linearly independent functions

$$\{v_1, v_2, \ldots, v_n\},$$

the Gram-Schmidt process goes as follows:

$$\begin{aligned} u_1 &= v_1 \\ u_2 &= v_2 - \text{proj}_{u_1}(v_2) \\ u_3 &= v_3 - \text{proj}_{u_1}(v_3) - \text{proj}_{u_2}(v_3) \\ &\vdots \\ u_n &= v_n - \sum_{i=1}^{n-1} \text{proj}_{u_i}(v_n) \\ \text{where } \text{proj}_u(v) &= \frac{\int_0^1 u(t)v(t)dt}{\int_0^1 u(t)^2 dt} u(x) \end{aligned}$$

The set of output functions

$$\{u_1, u_2, \ldots, u_n\}$$

will be orthogonal. Apply the Gram-Schmidt process to $\{\phi_1, \phi_2\}$ to obtain a set of orthogonal polynomials $\{u_1, u_2\}$. Check that u_1 and u_2 are indeed orthogonal after the process.

(d) (5 pts) Solve the continuous least square problem but with the new orthogonal polynomials:

$$\min E = \int_0^1 (x - a_1 u_1(x) - a_2 u_2(x))^2 dx$$

Show how the coefficients a_1 and a_2 can be calculated without solving a system of linear equations.

Solution.

(a) The minimizer of E will satisfy

$$\frac{\partial E}{\partial a_1} = \frac{\partial E}{\partial a_2} = 0.$$

For a_1 we have

$$\begin{aligned}\frac{\partial E}{\partial a_1} &= 2\int_0^1 (x - a_1 - a_2x^2)(-1)dx = 0 \\ \iff \frac{x^2}{2} - a_1x - a_2\frac{x^3}{3}\bigg|_0^1 &= \frac{1}{2} - a_1 - \frac{a_2}{3} = 0.\end{aligned}$$

Then, for a_2 we have

$$\begin{aligned}\frac{\partial E}{\partial a_2} &= 2\int_0^1 (x - a_1 - a_2x^2)(-x^2)dx = 0 \\ \iff \frac{x^4}{4} - a_1\frac{x^3}{3} - a_2\frac{x^5}{5}\bigg|_0^1 &= \frac{1}{4} - \frac{a_1}{3} - \frac{a_2}{5} = 0.\end{aligned}$$

Therefore, the system of linear equations can be represented as

$$A = \begin{bmatrix} 1 & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{5} \end{bmatrix} \text{ and } b = \begin{bmatrix} \frac{1}{2} \\ \frac{1}{4} \end{bmatrix}.$$

(b) Because

$$\int_0^1 1 \cdot x^2 dx = \frac{x^3}{3}\bigg|_0^1 = \frac{1}{3} \neq 0,$$

they are not orthogonal.

(c) Following the Gram-Schmidt process, we have

$$\begin{aligned} u_1(x) &= \phi_1(x) = 1 \\ u_2(x) &= \phi_2(x) - \text{proj}_{u_1}(\phi_2) \\ &= x^2 - \frac{\int_0^1 t^2 \cdot 1dt}{\int_0^1 1^2 dt} \cdot 1 \\ &= x^2 - \frac{1}{3}. \end{aligned}$$

Because

$$\int_0^1 1 \cdot (x^2 - \frac{1}{3})dx = \frac{x^3}{3} - \frac{x}{3}\bigg|_0^1 = 0,$$

they are indeed orthogonal.

(d) Because the polynomials are now orthogonal, according to lecture slide “FFT_basic1.pdf”, we can calculate the coefficients as follows:

$$
\begin{aligned}
a_1 &= \frac{\int_0^1 x \cdot 1 dx}{\int_0^1 1^2 dx} = \frac{1}{2} \\
a_2 &= \frac{\int_0^1 x \cdot (x^2 - \frac{1}{3}) dx}{\int_0^1 (x^2 - \frac{1}{3})^2 dx} \\
&= \frac{\frac{x^4}{4} - \frac{x^2}{6}\Big|_0^1}{\frac{x^5}{5} - \frac{2}{9}x^3 + \frac{1}{9}\Big|_0^1} \\
&= \frac{\frac{1}{12}}{\frac{4}{45}} = \frac{15}{16}.
\end{aligned}
$$