3.1 Summary

For the last 2 courses, we learned about vector space.
We first introducerd the concept of vector space. A vector space is a collection of things that obey certain abstact. We also introduced the concept of vector addition and scalar multiplication which define the vector space. Then we learned concepts like lineraly dependent and basis.
After that, we leared about norms. A norm on a linear space S is a mapping form S to R. Then we introduced the concept of inner product, which is a mapping form S x S to C. After that we leared about the induced norm which is the norms defined by inner products. Then we learned about the completeness of a linear space.
Then we leared linear approximation in Hilbert space and orthogonal projection. Then we introduced Gram-Schemidt algorithm. After that, some important orthobasis is introduced.
In those orthobasis, the DCT is further discussed because it can apply on 2D image.

3.2

3.2. (Completeness, 2 pts) Completeness is a critical property when we are finding the best approximation of a given vector in a subspace. For example, the closest point of $\sqrt{2}$ does not exist in the set of rational numbers $\mathbb{Q} \subset \mathbb{R}$, which is incomplete with respect to absolute value distance. Now consider the vector space $\mathrm{C}([0,1])$ of all continual functions on $[0,1]$, please show that:

(a) $\mathrm{C}([0,1])$ is not complete with respect to $\|f\|_2$.

(b) (**OPTIONAL**) $\mathrm{C}([0,1])$ is complete with respect to $\|f\|_\infty$.

(a)

To simplify the discussion, I will work on [-1,1] rather than [0, 1]. It is very obvious that $C([0,1]) \in C([-1,1])$.
Let's assume that:

$$f_n(x) = \begin{cases} -1, x \in [-1, -\frac{1}{n}] \\ nx, x \in [-\frac{1}{n}, \frac{1}{n}] \\ 1, x \in [\frac{1}{n}, 1] \end{cases}$$

Let $n > m, t \in [0,1]$ we have:

$$|f_n(t)-f_m(t)| = \begin{cases} (nt-mt), t\in[0,\frac{1}{n}] \\ (1-mt), x\in[\frac{1}{n},\frac{1}{m}] \\ 0, x\in\frac{1}{m},1] \end{cases}$$

So we have:

$$\begin{aligned} ||f_n-f_m||_2 &= (\int_0^1 |f_n(t)-f_m(t)|^2 dt)^{1/2} \\ &= (\int_0^{\frac{1}{n}} (nt-mt)^2 dt + \int_{\frac{1}{n}}^{\frac{1}{m}} (1-mt)^2 dt)^{1/2} \\ &= \frac{m}{3n^2}+\frac{1}{3m}-\frac{2}{3n} \\ &= \frac{2}{3}(\frac{2m}{n^2}+\frac{2}{m}-\frac{1}{n}) \\ &< \frac{2m}{n^2}+\frac{2}{m}-\frac{1}{n} \\ &< \frac{2}{n}+\frac{2}{m}-\frac{1}{n} \\ &= \frac{1}{n}+\frac{2}{m} \\ &< \frac{3}{n} \end{aligned}$$

$\forall\epsilon, N > 3 over \epsilon, N\in\mathbb{N}$, we have:

$$\frac{3}{n} < \frac{3}{\frac{3}{\epsilon}} = \epsilon$$

So the f_n I defined before is a Cauchy sequence. However, consider $n\to\inf$,

$$f_n(x) = \begin{cases} -1, x\in[-1,-0) \\ 0, x=0 \\ 1, x\in(0,1] \end{cases}$$

we find that $f_n(x)\notin C[(0,1)]$.

(b)

Let $\{f_n(x)\}_{n=1}^{\inf}$ is a Cauchy sequence in $C([0,1])$. By defination, for every $\epsilon>0$, there is an N such that $||f_n-f_m||<\epsilon$ for all $n,m\geq N$. For any fixed $t\in[0,1]$, this implies that

$$|fn(t)-fm(t)|<\epsilon, \forall m,n\geq N.$$

So, $\{f_n(x)\}_{n=1}^{\inf}$ is a Cauchy sequence of real number.

3.3

3.3. (Gram-Schmidt, 3 pts) Consider a subspace $S_k \subset \mathrm{L}_2([-1,1])$ spanned by polynomial basis $\{1, x, x^2, \cdots, x^k\}$. The inner product is defined as $\langle f, g\rangle = \int_{-1}^{1} f(x)g(x)\mathrm{d}x$. Let $k = 3$. Apply Gram-Schmidt algorithm on this basis set to obtain an orthogonal basis of S_3.

Acroding to the defination of inner product, the induced norm is:

$$
\begin{aligned}
||x|| &= (< x, x >)^{\frac{1}{2}} \\
&= \sqrt{\int_{-1}^{1} f(x)^2 dx}
\end{aligned}
$$

S_3 is $\{1, x, x^2, x^3\}$.
First, we calculate u_1:

$$
\begin{aligned}
w_1 &= 1 \\
u_1 &= \frac{1}{||1||} \\
&= \frac{1}{\sqrt{\int_{-1}^{1} 1 dx}} \\
&= \frac{\sqrt{2}}{2}
\end{aligned}
$$

Then, we calculate u_2:

$$
\begin{aligned}
w_2 &= v2 - < v2, u1 > u1 \\
&= x - \frac{\sqrt{2}}{2}\int_{-1}^{1} \frac{\sqrt{2}}{2} x dx \\
&= x \\
u_2 &= \frac{x}{||x||} \\
&= \frac{x}{\sqrt{\int_{-1}^{1} x^2 dx}} \\
&= \frac{\sqrt{6}}{2} x
\end{aligned}
$$

Then, we calculate u_3:

$$
\begin{aligned}
w_3 &= v3 - <v3, u2> u2 - <v3, u1> u1 \\
&= x^2 - \frac{\sqrt{6}}{2} x \int_{-1}^{1} \frac{\sqrt{6}}{2} x^3 dx - \frac{\sqrt{2}}{2} \int_{-1}^{1} \frac{\sqrt{2}}{2} x^2 dx \\
&= x^2 - \frac{1}{2} \int_{-1}^{1} x^2 dx \\
&= x^2 - \frac{1}{3} \\
u_2 &= \frac{x^2 - \frac{1}{3}}{||x^2 - \frac{1}{3}||} \\
&= \frac{x^2 - \frac{1}{3}}{\sqrt{\int_{-1}^{1} (x^2 - \frac{1}{3})^2 dx}} \\
&= \frac{3\sqrt{10}}{4} x^2 - \frac{\sqrt{10}}{4} \\
&= \frac{\sqrt{10}}{4} (3x^2 - 1)
\end{aligned}
$$

Then, we calculate u_4:

$$
\begin{aligned}
w_4 &= v3 - <v4, u3> u3 - <v4, u2> u2 - <v4, u1> u1 \\
&= x^3 - (\frac{3\sqrt{10}}{4} x^2 - \frac{\sqrt{10}}{4}) \int_{-1}^{1} (\frac{3\sqrt{10}}{4} x^2 - \frac{\sqrt{10}}{4}) x^3 dx - \frac{\sqrt{6}}{2} x \int_{-1}^{1} \frac{\sqrt{6}}{2} x^4 dx - \frac{\sqrt{2}}{2} \int_{-1}^{1} \frac{\sqrt{2}}{2} x^3 dx \\
&= x^3 - \frac{\sqrt{6}}{2} x \int_{-1}^{1} \frac{\sqrt{6}}{2} x^4 dx \\
&= x^3 - \frac{3}{5} x \\
u_2 &= \frac{x^3 - \frac{3}{5} x}{||x^3 - \frac{3}{5} x||} \\
&= \frac{x^3 - \frac{3}{5} x}{\sqrt{\int_{-1}^{1} (x^3 - \frac{3}{5} x)^2 dx}} \\
&= \frac{5\sqrt{14}}{4} x^3 - \frac{3\sqrt{14}}{4} x \\
&= \frac{\sqrt{14}}{4} (5x^3 - 3x)
\end{aligned}
$$

So the orthobasis is:

$$
< \frac{\sqrt{2}}{2}, \frac{\sqrt{6}}{2} x, \frac{\sqrt{10}}{4} (3x^2 - 1), \frac{\sqrt{14}}{4} (5x^3 - 3x) >
$$

3.4

3.4. (Linear Approximation, 3 pts) Determine the best quadratic approximation $\hat{x}(t) = a_1 + a_2 t + a_3 t^2$ to $x(t) = \log(1+t)$ over the interval $[0, 1]$, which minimizes the energy of the error

$$\|x(t) - \hat{x}(t)\|_{L_2([0,1])} = \sqrt{\int_0^1 |x(t) - \hat{x}(t)|^2 \, dt} \tag{1}$$

Note: This can be achieved by project $x(t)$ into the quadratic subspace. Following these steps may be useful:

(a) We set this up as a subspace approximation problem by assigning $\boldsymbol{T} = \{1, t, t^2\}$. Determine the Gram matrix G.

(b) Determine the right-hand vector $\boldsymbol{b}$. Then write down the system of equations that needs to be solved using G and $\boldsymbol{b}$.

(c) Solve the equation in (b) and derive your conclusion.

(d) Write a script to compare your approximation to the (truncated) Taylor expansion $\hat{x}_{\text{Taylor}}(t) = 0 + t - \frac{1}{2}t^2$, i.e., plot $x(t)$, $\hat{x}_{\text{Taylor}}(t)$ and $\hat{x}(t)$ in one figure.

(a)

Let's assume the inner product is:

$$< x(t), y(t) >= \int_0^1 x(t)y(t)dt$$

Gram matrix is:

$$G = \begin{bmatrix} <1,1> & <t,1> & <t^2,1> \\ <1,t> & <t,t> & <t^2,t> \\ <1,t^2> & <t,t^2> & <t^2,t^2> \end{bmatrix}$$

$$<1,1> = \int_0^1 1dt = 1$$

$$<1,t>=<t,1> = \int_0^1 tdt = \frac{1}{2}$$

$$<1,t^2>=<t,t>=<t^2,1> = \int_0^1 t^2dt = \frac{1}{3}$$

$$<t,t^2>=<t^2,t> = \int_0^1 t^3dt = \frac{1}{4}$$

$$<t^2,t^2>=<t^2,t> = \int_0^1 t^4dt = \frac{1}{5}$$

So,

$$G = \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{3} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} \\ \frac{1}{3} & \frac{1}{4} & \frac{1}{5} \end{bmatrix}$$

(b)

$$b = \begin{bmatrix} < \hat{x}, 1 > \\ < \hat{x}, t > \\ < \hat{x}, t^2 > \end{bmatrix}$$

$$\begin{aligned} < \hat{x}, 1 > &= \int_0^1 log(t+1)dt = ((t+1)log(t+1) - t)|_0^1 \\ &= 2log2 - 1 \\ < \hat{x}, t > &= \int_0^1 tlog(t+1)dt \\ &= [\frac{1}{4}(2(t^2-1)log(t+1) - (t-2)t)]|_0^1 \\ &= \frac{1}{4} \\ < \hat{x}, t^2 > &= \int_0^1 t^2 log(t+1)dt \\ &= [\frac{1}{18}(6(t^3+1)log(t+1) + (-2t^2+3t-6)t)]|_0^1 \\ &= \frac{12log2-5}{18} \end{aligned}$$

So,

$$b = \begin{bmatrix} 2log2 - 1 \\ \frac{1}{4} \\ \frac{12log2-5}{18} \end{bmatrix}$$

assume that

$$logx = log_2 x$$

So,

$$b = \begin{bmatrix} 0.386294 \\ 0.25 \\ 0.18432 \end{bmatrix}$$

We have,

$$a = G^{-1}b$$

(c)

$$G^{-1} = 3\begin{bmatrix} 3 & -12 & 10 \\ -12 & 64 & -60 \\ 10 & -60 & 60 \end{bmatrix}$$

So,

$$a = \begin{bmatrix} 0.006246 \\ 0.915816 \\ -0.23358 \end{bmatrix}$$

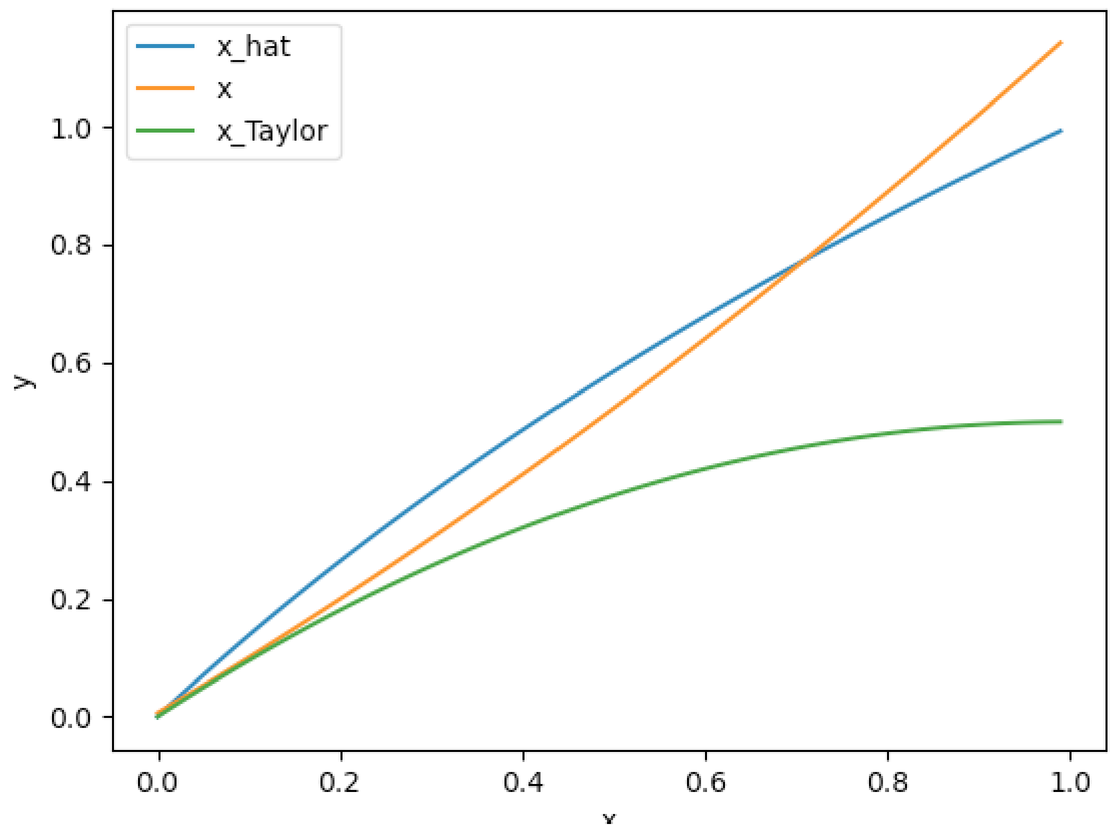
The Linear appoximation of $\hat{x}(t) = log(t+1)$ is

$$x(t) = 0.006246 + 0.915816t - -0.23358t^2$$

(d)

```
import math

import numpy as np

import matplotlib.pyplot as plt

x = np.arange(0, 1, 0.01)

y1 = []

y2 = []

y3 = []

for t in x:

y_hat = math.log2(t + 1)

y1.append(y_hat)

y = 0.006246 + 0.915816*t - -0.23358 * t**2

y2.append(y)

y_tay = t - 0.5*t**2

y3.append(y_tay)

plt.plot(x, y1, label="x_hat")

plt.plot(x, y2, label="x")

plt.plot(x, y3, label="x_Taylor")

plt.xlabel("x")

plt.ylabel("y")
```

```
plt.legend()

plt.show()
```



3.5

(RKHS, **OPTIONAL**) The Reproducing Kernel Hilbert Space (RKHS) is a Hilbert space H of functions, in which there exists a kernel function $k(\cdot, \cdot)$ such that evaluation of some function can be computed by its inner product with respect to this kernel:

$$f(x) = \langle f, k(\cdot, x) \rangle_H, \forall f \in H \tag{9}$$

(a) Consider the space of band-limited continual signals with cut-off frequency Ω_b, and equipped with inner product $\langle f, g \rangle_H = \int_{-\infty}^{\infty} f(x)\overline{g(x)}\mathrm{d}x$. Verify that the kernel

$$k(x, y) = \frac{\sin(\Omega_b(x - y))}{\pi(x - y)}$$

satisfies equation 9.

(b) RKHS is usually set as assumption space in many machine learning problems and it can lead to dramatic simplification. Consider a regression problem over a dataset $D = \{(x^{(i)}, y^{(i)})\}_{i=1}^N$, we want to learn a function f within an RKHS H that can minimize regularized empirical mean squared error:

$$\underset{f \in H}{\text{minimize}} \sum_{i=1}^{N} (f(x^{(i)}) - y^{(i)})^2 + \lambda \|f\|_H^2$$

Show that the minimizer f^* must be a linear combination of kernels over training points:

$$f^*(x) = \sum_{i=1}^{N} \alpha_i^* k(x, x^{(i)})$$

Hint: Decompose $f \in H$ into one component that lies in subspace spanned by $\{k(\cdot, x^{(i)})\}$, and another component that is a orthogonal complement. Then explain why the latter one must be zero.