College of Engineering

School of Aeronautical and Astronautical Engineering

AAE 567: Introduction To Applied Stochastic Processes

---

# HW4

---

*Professor:*
A. E. Frazho
Purdue AAE Professor

*Student:*
Tomoki Koike
Purdue AAE Senior

March 11, 2021

## Table of Contents

## Problem 1

Let $\mathbf{x}$ be a uniform random variable over the interval [0,4]. Moreover, $\mathbf{v}$ is a uniform random variable over the interval [-1,1]. Assume that $\mathbf{x}$ and $\mathbf{v}$ are independent. Let $\mathbf{y}$ be the random variable given by $\mathbf{y} = \mathbf{x} + \mathbf{v}$.

Let $\mathcal{H}$ be the space spanned by $\{1, \mathbf{y}, \mathbf{y^2}, \mathbf{y^3}\}$.Then compute

$$P_{\mathcal{H}}\mathbf{x} = a + b\mathbf{y} + c\mathbf{y^2} + d\mathbf{y^3} \quad \text{and} \quad d_4^2 = E|\mathbf{x} - P_{\mathcal{H}}\mathbf{x}|^2 \quad .$$

Compute the conditional expectation

$$\hat{g}(y) = E(\mathbf{x}|\mathbf{y} = y) \quad .$$

Plot $\hat{g}(y)$ and its approximation $a + by + cy^2 + dy^3$ on the same graph over the interval [-1,5].

---

**Solution:**
For convenience the heavy computations are done using MATLAB. (The code will be at the end of this problem.) Let

$$g = \begin{bmatrix} 1 \\ y \\ y^2 \\ y^3 \end{bmatrix}$$

$$f = \mathbf{x} \quad .$$

Also since $\mathbf{x}$ and $\mathbf{v}$ are uniform distributions

$$\begin{aligned} E\mathbf{x} &= 2 \\ E\mathbf{v} &= 0 \\ f_{\mathbf{x}(x)} &= \frac{1}{4} \\ f_{\mathbf{v}(v)} &= \frac{1}{2} \quad . \end{aligned}$$

Then

$$R_{fg} = R_{\mathbf{x}g} = Exg = Ex\begin{bmatrix}1 & y & y^2 & y^3\end{bmatrix} = \begin{bmatrix}Ex & Exy & Exy^2 & Exy^3\end{bmatrix} .$$

Here

$$\begin{aligned} Exy &= Ex(x+v) = Ex^2 + Exv = Ex^2 + ExEv \\ &= Ex^2 \\ &= \int_{-\infty}^{\infty} x^2 f_{\mathbf{x}}(x)dx = \frac{1}{4}\int_0^4 x^2 dx \\ &= \frac{16}{3} \quad . \end{aligned}$$

Furthermore,

$$\begin{aligned}
Exy^2 &= Ex(x^2+2xv+v^2) = Ex^3 + 2Ex^2Ev + ExEv^2 \\
&= \int_{-\infty}^{\infty} x^3 f_{\mathbf{x}}(x)dx + 2\int_{-\infty}^{\infty} v^2 f_{\mathbf{v}}(v)dv \\
&= \frac{1}{4}\int_0^4 x^3 dx + 2\int_{-1}^{1} v^2 dv \\
&= 16 + \frac{2}{3} = \frac{50}{3}
\end{aligned}$$

Similarly,

$$Exy^3 = \frac{1}{8}\int_{-1}^{1}\int_0^4 x(x+v)^3 dx dv = 56.5333$$

Which gives

$$\therefore R_{fg} = \left[\begin{array}{cccc} 2 & 5.3333 & 16.6667 & 56.5333 \end{array}\right].$$

Next,

$$R_g = Egg^* = E\begin{bmatrix} 1 \\ y \\ y^2 \\ y^3 \end{bmatrix}\begin{bmatrix} 1 & y & y^2 & y^3 \end{bmatrix} = \begin{bmatrix} E1 & Ey & Ey^2 & Ey^3 \\ Ey & Ey^2 & Ey^3 & Ey^4 \\ Ey^2 & Ey^3 & Ey^4 & Ey^5 \\ Ey^3 & Ey^4 & Ey^5 & Ey^6 \end{bmatrix}$$

$$= \left[\begin{array}{cccc} 1 & 2 & 5.6667 & 18 \\ 2 & 5.6667 & 18 & 62.0667 \\ 5.6667 & 18 & 62.0667 & 226 \\ 18 & 62.0667 & 226 & 857.2857 \end{array}\right].$$

Now the coefficients $a, b, c,$ and $d$ become

$$\begin{aligned}
\begin{bmatrix} a & b & c & d \end{bmatrix} &= R_{fg}R_g^{-1} \\
&= \left[\begin{array}{cccc} 2 & 5.3333 & 16.6667 & 56.5333 \end{array}\right]\left[\begin{array}{cccc} 1 & 2 & 5.6667 & 18 \\ 2 & 5.6667 & 18 & 62.0667 \\ 5.6667 & 18 & 62.0667 & 226 \\ 18 & 62.0667 & 226 & 857.2857 \end{array}\right]^{-1} \\
&= \left[\begin{array}{cccc} 0.4320 & 0.4286 & 0.2666 & -0.0444 \end{array}\right].
\end{aligned}$$

Hence,

$$P_{\mathcal{H}}\mathbf{x} = 0.4320 + 0.4286y + 0.2666y^2 - 0.0444y^3 \quad .$$

Then the error becomes

$$\begin{aligned}
d_4^2 &= R_{\mathbf{x}} - R_{fg}R_g^{-1}R_{gf} \\
&= 0.2525
\end{aligned}$$

Thus,

$$\therefore d_4 = 0.5024 \quad .$$

Now we will compute the conditional expectation

$$\hat{g}(y) = E(\mathbf{x}|\mathbf{y} = y) \quad .$$

Which is equivalent to

$$\hat{g}(y) = \int_{-\infty}^{\infty} x f_{\mathbf{x}|\mathbf{y}}(x|y)dx = \int_{-\infty}^{\infty} x \frac{f_{\mathbf{x},\mathbf{y}}(x,y)}{f_{\mathbf{y}}(y)} dx \quad .$$

If the two random variables are statistically independent we know that

$$f_{\mathbf{x},\mathbf{y}}(x,y) = f_{\mathbf{x}}(x) f_{\mathbf{v}}(y-x)$$
$$f_{\mathbf{y}}(y) = \int_{-\infty}^{\infty} f_{\mathbf{x}}(x) f_{\mathbf{v}}(y-x)dx$$

and if the range is defined to be $y \in [-1, 5]$, we can solve the first one to be

$$\begin{aligned} f_{\mathbf{x},\mathbf{y}}(x,y) &= f_{\mathbf{x},\mathbf{v}}(x,v) \cdot \left| det\Big(\nabla(x, x+v)\Big)\right|^{-1} \\ &= \frac{1}{4} \times \frac{1}{2} \begin{vmatrix} 1 & 0 \\ 1 & 1 \end{vmatrix}^{-1} \\ &= \frac{1}{8} \end{aligned}$$

Thus,

$$f_{\mathbf{x},\mathbf{y}}(x,y) = \frac{1}{8} \qquad \text{for} \quad -1 \le y \le 5 \quad .$$

Now for the marginal probability density function we use the range relations of

$$0 \le x \le 4$$
$$-1 \le y - x \le 1$$

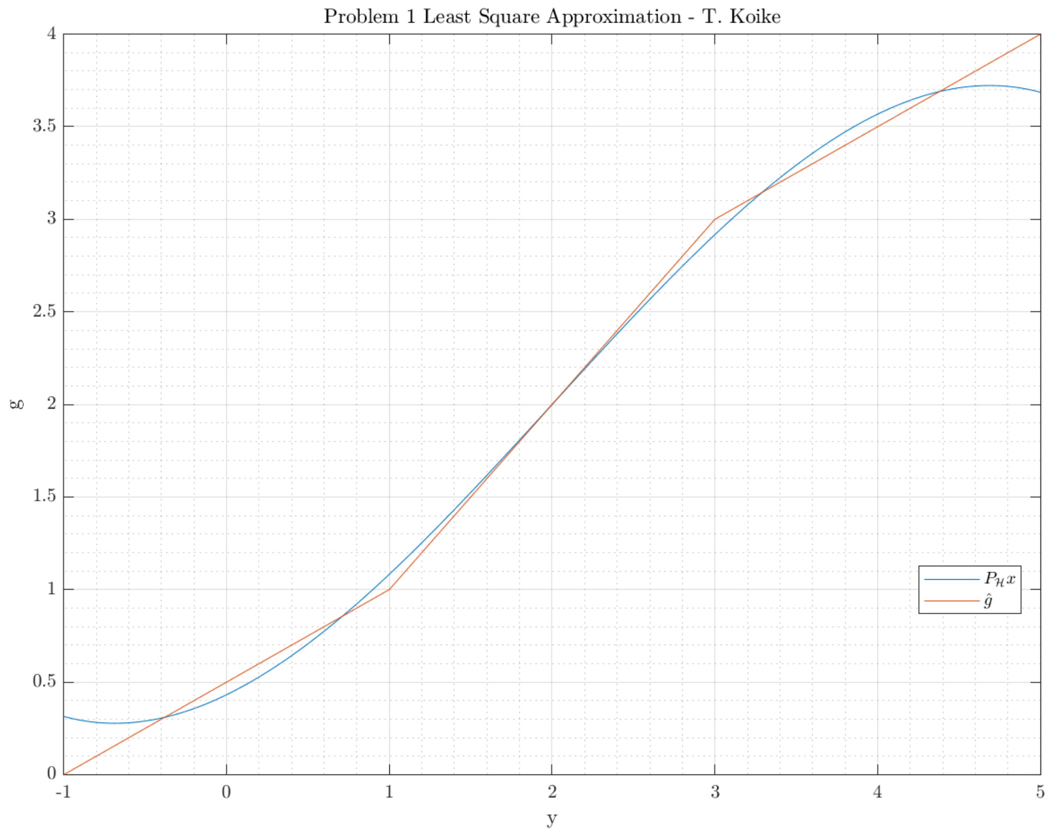
and we obtain

$$f_{\mathbf{y}}(y) = \begin{cases} \frac{1}{8}\int_0^{y+1} dx = \frac{y+1}{8} & \text{if} \quad -1 \le y \le 1 \\ \frac{1}{8}\int_{y-1}^{y+1} dx = \frac{1}{4} & \text{if} \quad 1 \le y \le 3 \\ \frac{1}{8}\int_{y-1}^{4} dx = \frac{5-y}{8} & \text{if} \quad 3 \le y \le 5 \end{cases}$$

Therefore

$$\hat{g}(y) = \begin{cases} \int_0^{y+1} x \left(\frac{1}{8}\right)\left(\frac{y+1}{8}\right)^{-1} dx = \frac{1}{2}y + \frac{1}{2} & \text{if} \quad -1 \le y \le 1 \\ \int_{y-1}^{y+1} x \left(\frac{1}{8}\right)\left(\frac{1}{4}\right)^{-1} dx = y & \text{if} \quad 1 \le y \le 3 \\ \int_{y-1}^{4} x \left(\frac{1}{8}\right)\left(\frac{5-y}{8}\right)^{-1} dx = \frac{1}{2}y + \frac{3}{2} & \text{if} \quad 3 \le y \le 5 \end{cases}$$

Finally we plot our results,




and the MATLAB code is as follows.

```
%% AAE 567 HW4 Problem1

% Housekeeping commands
clear all; close all; clc;
set(groot, 'defaulttextinterpreter','latex');
set(groot, 'defaultAxesTickLabelInterpreter','latex');
set(groot, 'defaultLegendInterpreter','latex');
outdir = pwd + "\output\hw4";
mdir = pwd + "\mfiles\hw4";
%%
% Define expectations
syms x v y g
y = x + v;
g = [1; y; y^2; y^3];
EX = @(A) int(1/4 * x.^A, 0, 4);
```

```
EV = @(A) int(1/2 * v.^A, -1, 1);
EY = @(A) 1/8 * int(int(y.^A, x, 0, 4), -1, 1);
EXY = @(A,B) 1/8 * int(int(x.^A * y.^B, x, 0, 4), -1, 1);

% P_Hx
Rfg = EXY(1, [0 1 2 3]);
A = [0:3; 1:4; 2:5; 3:6];
Rg = EY(A);
coef = Rfg * inv(Rg);
coef = eval(coef);
% Error d4
d4sq = EX(2) - Rfg*inv(Rg)*Rfg';
d4 = sqrt(d4sq);
%%
% Plotting
t = -1:0.01:5;
% ghat piecewise
t1 = t(-1 <= t & t < 1);
t2 = t(1 <= t & t < 3);
t3 = t(3 <= t & t <= 5);
ghat = [0.5*t1 + 0.5, t2, 0.5*t3 + 1.5];

Phx = coef(1) + coef(2)*t + coef(3)*t.^2 + coef(4)*t.^3;

fig = figure("Renderer","painters",'Position',[60 60 900 650]);
    plot(t, Phx)
    grid on; grid minor; box on; hold on;
    plot(t, ghat)
    hold off;
    title('Problem 1 Least Square Approximation - T. Koike')
    legend('$P_{\mathcal{H}}x$', '$\hat{g}$', "Location","best")
    xlabel('y')
    ylabel('g')
saveas(fig, fullfile(outdir, 'p1_lsqr_plot.png'));
%%
% Save file as .m
matlab.internal.liveeditor.openAndConvert('hw4_p1.mlx', ...
    convertStringsToChars(fullfile(mdir, 'hw4_p1.m')));
```

## Problem 2

Let $\mathbf{x}$ and $\mathbf{y}$ be two uniform random variables over the interval [0,1]. Let $\mathbf{a}$ be the random variable defined by the area $\mathbf{a}$=$\mathbf{xy}$. Clearly, the area is $0 \le \mathbf{a} \le 1$. Our problem is given the area $\mathbf{a}$ find the best estimate of $\hat{\mathbf{x}}$ of $\mathbf{x}$.

Let $\mathcal{H}$ be the space spanned by $\{1, \mathbf{a}, \mathbf{a^2}, \mathbf{a^3}\}$.Then compute

$$P_{\mathcal{H}}\mathbf{x} = \alpha + \beta\mathbf{a} + \gamma\mathbf{a^2} + \delta\mathbf{a^3} \quad \text{and} \quad d_4^2 = E|\mathbf{x} - P_{\mathcal{H}}\mathbf{x}|^2 \quad .$$

Compute the conditional expectation

$$\hat{g}(a) = E(\mathbf{x}|\mathbf{a} = a) \quad \text{and} \quad d_\infty^2 = E|\mathbf{x} - \hat{g}(\mathbf{a})|^2 \quad .$$

Plot $\hat{g}(a)$ and its approximation $\alpha + \beta a + \gamma a^2 + \delta a^3$ on the same graph over the interval [0,1]. Is $d_\infty < d_4$? Explain why or why not.

---

**Solution:**
For convenience the heavy computations are done using MATLAB. (The code will be at the end of this problem.) Let

$$g = \begin{bmatrix} 1 \\ a \\ a^2 \\ a^3 \end{bmatrix}$$

$$f = \mathbf{x} \quad .$$

Also since $\mathbf{x}$ and $\mathbf{v}$ are uniform distributions

$$\begin{aligned} E\mathbf{x} &= 0.5 \\ E\mathbf{y} &= 0.5 \\ f_{\mathbf{x}(x)} &= 1 \\ f_{\mathbf{y}(y)} &= 1 \quad . \end{aligned}$$

Then

$$R_{fg} = R_{\mathbf{x}g} = Exg = \begin{bmatrix} 1 & a & a^2 & a^3 \end{bmatrix} = \begin{bmatrix} Ex & Exa & Exa^2 & Exa^3 \end{bmatrix} .$$

Here

$$\begin{aligned} Exa &= Ex(xy) = Ex^2Ey \\ &= \int_{-\infty}^{\infty} x^2 f_{\mathbf{x}}(x)dx = \int_0^1 x^2 dx \\ &= 1.6667 \quad . \end{aligned}$$

Furthermore,

$$\begin{aligned}
Exa^2 &= Ex(x^2y^2) = Ex^3Ey^2 \\
&= \Big(\int_{-\infty}^{\infty} x^3 f_{\mathbf{x}}(x)dx\Big)\Big(\int_{-\infty}^{\infty} y^2 f_{\mathbf{y}}(y)dy\Big) \\
&= \Big(\int_0^1 x^3 dx\Big)\Big(\int_0^1 y^2 dy\Big) \\
&= 0.0833
\end{aligned}$$

Similarly,

$$Exa^3 = \int_0^1 \int_0^1 x(xy)^3 dxdy = 0.5$$

Which gives

$$\therefore R_{fg} = \begin{bmatrix} 0.5 & 0.1667 & 0.0833 & 0.5 \end{bmatrix}.$$

Next,

$$R_g = Egg^* = E\begin{bmatrix} 1 \\ a \\ a^2 \\ a^3 \end{bmatrix}\begin{bmatrix} 1 & a & a^2 & a^3 \end{bmatrix} = \begin{bmatrix} E1 & Ea & Ea^2 & Ea^3 \\ Ea & Ea^2 & Ea^3 & Ea^4 \\ Ea^2 & Ea^3 & Ea^4 & Ea^5 \\ Ea^3 & Ea^4 & Ea^5 & Ea^6 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0.2500 & 0.1111 & 0.0625 \\ 0.2500 & 0.1111 & 0.0625 & 0.0400 \\ 0.1111 & 0.0625 & 0.0400 & 0.0278 \\ 0.0625 & 0.0400 & 0.0278 & 0.0204 \end{bmatrix}.$$

Now the coefficients $\alpha, \beta, \gamma,$ and $\delta$ become

$$\begin{aligned}
\begin{bmatrix} \alpha & \beta & \gamma & \delta \end{bmatrix} &= R_{fg}R_g^{-1} \\
&= \begin{bmatrix} 0.5 & 0.1667 & 0.0833 & 0.5 \end{bmatrix} \begin{bmatrix} 1 & 0.2500 & 0.1111 & 0.0625 \\ 0.2500 & 0.1111 & 0.0625 & 0.0400 \\ 0.1111 & 0.0625 & 0.0400 & 0.0278 \\ 0.0625 & 0.0400 & 0.0278 & 0.0204 \end{bmatrix}^{-1} \\
&= \begin{bmatrix} 0.2140 & 1.8416 & -2.3412 & 1.3717 \end{bmatrix}.
\end{aligned}$$

Hence,

$$P_{\mathcal{H}}\mathbf{x} = 0.2140 + 1.8416a - 2.3412a^2 - 1.3717a^3 \quad .$$

Then the error becomes

$$\begin{aligned}
d_4^2 &= R_{\mathbf{x}} - R_{fg}R_g^{-1}R_{gf} \\
&= 0.0459
\end{aligned}$$

Thus,

$$\therefore d_4 = 0.2143 \quad .$$

Now we will compute the conditional expectation

$$\hat{g}(y) = E(\mathbf{x}|\mathbf{y} = y) \quad .$$

Which is equivalent to

$$\hat{g}(a) = \int_{-\infty}^{\infty} x f_{\mathbf{x}|\mathbf{a}}(x|a)dx = \int_{-\infty}^{\infty} x \frac{f_{\mathbf{x},\mathbf{a}}(x,a)}{f_{\mathbf{a}}(a)} dx \quad .$$

and if the range is defined to be $y \in [0,1]$, we can solve the first one to be

$$\begin{aligned} f_{\mathbf{x},\mathbf{a}}(x,a) &= f_{\mathbf{x},\mathbf{y}}(x,y) \cdot \left| det\Big(\nabla(x,xy)\Big)\right|^{-1} \\ &= \begin{vmatrix} 1 & 0 \\ y & x \end{vmatrix}^{-1} \\ &= \frac{1}{x} \end{aligned}$$

Thus,

$$f_{\mathbf{x},\mathbf{a}}(x,a) = \frac{1}{x} \qquad \text{for} \quad 0 \le a \le 1 \quad .$$

Now for the marginal probability density function we use the range relations of

$$\begin{aligned} 0 \le xy \le 1 \\ 0 \le y \le 1 \\ 0 \le x \le 1 \end{aligned}$$

which leads to

$$0 \le a \le x \le 1$$

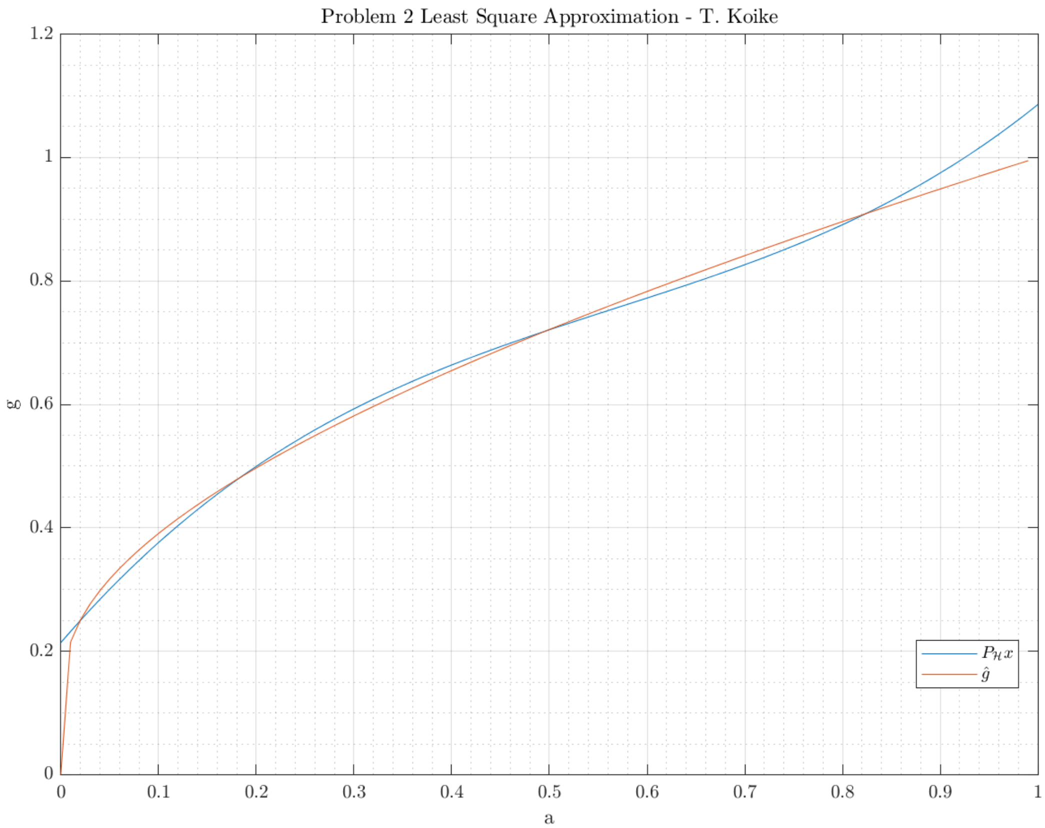
and we obtain

$$\begin{aligned} f_{\mathbf{a}}(a) &= \int_a^1 \frac{1}{x} dx \\ &= -\ln a \end{aligned}$$

Therefore

$$\begin{aligned} \hat{g}(y) &= \int_a^1 x \frac{1/x}{-\ln a} dx \\ &= \frac{1}{-\ln a} \Big\{ x \Big\}_a^1 \\ &= \frac{a-1}{\ln a} \quad . \end{aligned}$$

Finally we plot our results,




The error of $\hat{g}(y)$, which is also denoted as $d_\infty$ is equal to 0. This is because the $\hat{g}(y)$ is the exact value of the conditional expectation of products for 2 uniform distributions. This can be achieved by having $\mathcal{H}$ span infinite order which allows the rough estimation converge to $\hat{g}(t)$. This also means that $d_\infty < d_4$.

MATLAB code:

```
%% AAE 567 HW4 Problem2

% Housekeeping commands
clear all; close all; clc;
set(groot, 'defaulttextinterpreter','latex');
set(groot, 'defaultAxesTickLabelInterpreter','latex');
set(groot, 'defaultLegendInterpreter','latex');
outdir = pwd + "\output\hw4";
mdir = pwd + "\mfiles\hw4";
%%
m = 3;
% Define expectations
syms x v y g
y = x*v;
g = [];
for n = 0:m
    g = [g; y^n];
end
%g = [1; y; y^2; y^3; y^4; y^5; y^6];
xl = 0; xu = 1;  % range of x
vl = 0; vu = 1;  % range of v
EX = @(A) int(x.^A, xl, xu);
EV = @(A) int(v.^A, vl, vu);
EY = @(A) int(int(y.^A, x, xl, xu), vl, vu);
EXY = @(A,B) int(int(x.^A * y.^B, x, xl, xu), vl, vu);

% P_Hx
Rfg = EXY(1, 0:m);
%%
A = [];
for n = 0:m
    A = [A; n:n+m];
end
Rg = EY(A);
coef = Rfg * inv(Rg);
coef = eval(coef);
% Error d4
d4sq = EX(2) - Rfg*inv(Rg)*Rfg';
d4 = sqrt(d4sq);
%%
% Plotting
t = 0:0.01:1;
```

```
% ghat
Phx = 0;
for n = 1:length(coef)
    Phx = Phx + coef(n)*t.^(n-1);
end

fig = figure("Renderer","painters",'Position',[60 60 900 650]);
    plot(t, Phx)
    grid on; grid minor; box on; hold on;
    plot(t, (t-1)./log(t))
    hold off;
    title('Problem 2 Least Square Approximation - T. Koike')
    legend('$P_{\mathcal{H}}x$', '$\hat{g}$', "Location","best")
    xlabel('a')
    ylabel('g')
saveas(fig, fullfile(outdir, 'p2_lsqr_plot.png'));
%%
% Save file as .m
matlab.internal.liveeditor.openAndConvert('hw4_p2.mlx', ...
    convertStringsToChars(fullfile(mdir, 'hw4_p2.m')));
```

## Problem 3

[Problem 1 from the Notes p.34.] Let $\mathbf{x}$ be a uniform random variable over the interval [0,10] and $\mathbf{v}$ a uniform random variable over [0,4]. Moreover, assume $\mathbf{x}$ and $\mathbf{v}$ are independent random variables. Now let $\mathbf{y}$ be the random variable defined by $\mathbf{y} = \mathbf{x} + \mathbf{v}$. Let $\mathcal{H}_3$ be the subspace spanned by $\{1, \mathbf{y}, \mathbf{y}^2\}$. Then compute the optimal estimate $\hat{\mathbf{x}} = P_{\mathcal{H}_3}\mathbf{x}$ and the error in estimation $E(\mathbf{x} - \hat{\mathbf{x}})^2$. Notice that in this case, $\mathbf{g}$=$[1, \mathbf{y}, \mathbf{y}^2]^{tr}$. Compute the conditional expectation $E(\mathbf{x}|\mathbf{y})$ and compare your answer $P_{\mathcal{H}_3}\mathbf{x}$ to the solution computed by the conditional expectation.

Hint: According to Lemma 12.3.1 in the Appendix, the joint probability density function $f_{\mathbf{x},\mathbf{y}}(x, y) = f_{\mathbf{x}}(x) f_{\mathbf{v}}(y - x)$. (The notation $f_{\mathbf{x}}(x)$ is the density function for the random variable $\mathbf{x}$ evaluated at the point $x$ on the real line.) Moreover, the density $f_{\mathbf{y}}$ for the random variable $\mathbf{y}$ is obtained by convolving $f_{\mathbf{x}}$ with $f_{\mathbf{v}}$. In other words, show that

$$\begin{aligned} f_{\mathbf{y}}(y) &= y/40 && \text{if} \quad 0 \le y \le 4 \\ &= 1/10 && \text{if} \quad 4 \le y \le 10 \\ &= (14 - y)/40 && \text{if} \quad 10 \le y \le 14 \end{aligned}$$

Verify that the conditional density $f_{\mathbf{x}|\mathbf{y}}$ is given by

$$\begin{aligned} f_{\mathbf{x}|\mathbf{y}} &= 1/y && \text{if} \quad 0 \le x \le y && \text{and} \quad 0 \le y \le 4 \\ &= 1/4 && \text{if} \quad y - 4 \le x \le y && \text{and} \quad 4 \le y \le 10 \\ &= 1/(14 - y) && \text{if} \quad y - 4 \le x \le 10 && \text{and} \quad 10 \le y \le 14 \end{aligned}$$

Finally, show that the conditional expectation is given by

$$\begin{aligned} E(\mathbf{x}|\mathbf{y} = y) &= y/2 && \text{if} \quad 0 \le y \le 4 \\ &= y - 2 && \text{if} \quad 4 \le y \le 10 \\ &= (y + 6)/2 && \text{if} \quad 10 \le y \le 14 \quad . \end{aligned}$$

Recall that $P_{\mathcal{H}_1}\mathbf{x} = 130\mathbf{y}/176$ was the best estimate of $\mathbf{x}$ corresponding to the one dimensional subspace $\mathcal{H}_1$ spanned by $\{\mathbf{y}\}$, and $P_{\mathcal{H}_2}\mathbf{x} = -30/29 + 25\mathbf{y}/29$ was the best estimate of $\mathbf{x}$ corresponding to the two dimensional subspace $\mathcal{H}_2$ spanned by $\{1,\mathbf{y}\}$; see Section 2.2.1. Plot 130y/176 and -30/29+25y/29 and the conditional expectation $E(\mathbf{x}|\mathbf{y} = y)$ along with your estimate $P_{\mathcal{H}_3}\mathbf{x}$ on the same graph. Finally, comment on the resulting graph.

---

**Solution:**
For convenience the heavy computations are done using MATLAB. (The code will be at the end of this problem.) Let

$$g = \begin{bmatrix} 1 \\ y \\ y^2 \end{bmatrix}$$

$$f = \mathbf{x} \quad .$$

Also since **x** and **v** are uniform distributions

$$\begin{aligned} E\mathbf{x} &= 5 \\ E\mathbf{v} &= 2 \\ f_{\mathbf{x}(x)} &= \frac{1}{10} \\ f_{\mathbf{v}(v)} &= \frac{1}{4} \quad . \end{aligned}$$

Then

$$R_{fg} = R_{\mathbf{x}g} = Exg = Ex \begin{bmatrix} 1 & y & y^2 \end{bmatrix} = \begin{bmatrix} Ex & Exy & Exy^2 \end{bmatrix}.$$

Here

$$\begin{aligned} Exy &= Ex(x+v) = Ex^2 + Exv = Ex^2 + ExEv \\ &= \int_{-\infty}^{\infty} x^2 f_{\mathbf{x}}(x)dx + 10 = \frac{1}{10}\int_0^{10} x^2 dx + 10 \\ &= 43.3333 \quad . \end{aligned}$$

Furthermore,

$$\begin{aligned} Exy^2 &= Ex(x^2 + 2xv + v^2) = Ex^3 + 2Ex^2Ev + ExEv^2 \\ &= \int_{-\infty}^{\infty} x^3 f_{\mathbf{x}}(x)dx + \frac{200}{3} \times 2 + 5\int_{-\infty}^{\infty} v^2 f_{\mathbf{v}}(v)dv \\ &= \frac{1}{10}\int_0^4 x^3 dx + \frac{400}{3} + 5\int_0^4 v^2 dv \\ &= 410 \end{aligned}$$

Which gives

$$\therefore R_{fg} = \left[ \begin{array}{ccc} 5 & 43.3333 & 410 \end{array} \right].$$

Next,

$$\begin{aligned} R_g = Egg^* = E\begin{bmatrix} 1 \\ y \\ y^2 \end{bmatrix} \begin{bmatrix} 1 & y & y^2 \end{bmatrix} &= \begin{bmatrix} E1 & Ey & Ey^2 \\ Ey & Ey^2 & Ey^3 \\ Ey^2 & Ey^3 & Ey^4 \end{bmatrix} \\ &= \left[ \begin{array}{ccc} 1 & 7 & 58.6667 \\ 7 & 58.6667 & 546 \\ 58.6667 & 546 & 5.4379\text{e+}03 \end{array} \right]. \end{aligned}$$

Now the coefficients $a, b,$ and $c$ become

$$\begin{aligned}
\begin{bmatrix} a & b & c \end{bmatrix} &= R_{fg} R_g^{-1} \\
&= \begin{bmatrix} 5 & 43.3333 & 410 \end{bmatrix} \begin{bmatrix} 1 & 7 & 58.6667 \\ 7 & 58.6667 & 546 \\ 58.6667 & 546 & 5.4379\text{e}{+}03 \end{bmatrix}^{-1} \\
&= \begin{bmatrix} -1.0345 & 0.8621 & 0 \end{bmatrix}.
\end{aligned}$$

Hence,

$$P_{\mathcal{H}}\mathbf{x} = -1.0345 + 0.8621y \quad .$$

Then the error becomes

$$\begin{aligned}
d_4^2 &= R_{\mathbf{x}} - R_{fg} R_g^{-1} R_{gf} \\
&= 1.1494
\end{aligned}$$

Thus,

$$\therefore d_4 = 1.0721 \quad .$$

Now we will compute the conditional expectation

$$\hat{g}(y) = E(\mathbf{x}|\mathbf{y} = y) \quad .$$

Which is equivalent to

$$\hat{g}(y) = \int_{-\infty}^{\infty} x f_{\mathbf{x}|\mathbf{y}}(x|y) dx = \int_{-\infty}^{\infty} x \frac{f_{\mathbf{x},\mathbf{y}}(x,y)}{f_{\mathbf{y}}(y)} dx \quad .$$

If the two random variables are statistically independent we know that

$$\begin{aligned}
f_{\mathbf{x},\mathbf{y}}(x,y) &= f_{\mathbf{x}}(x) f_{\mathbf{v}}(y-x) \\
f_{\mathbf{y}}(y) &= \int_{-\infty}^{\infty} f_{\mathbf{x}}(x) f_{\mathbf{v}}(y-x) dx
\end{aligned}$$

and if the range is defined to be $y \in [0, 14]$, we can solve the first one to be

$$\begin{aligned}
f_{\mathbf{x},\mathbf{y}}(x,y) &= f_{\mathbf{x},\mathbf{v}}(x,v) \cdot \left| det\Big(\nabla(x, x+v)\Big)\right|^{-1} \\
&= \frac{1}{10} \times \frac{1}{4} \begin{vmatrix} 1 & 0 \\ 1 & 1 \end{vmatrix}^{-1} \\
&= \frac{1}{40}
\end{aligned}$$

Thus,

$$f_{\mathbf{x},\mathbf{y}}(x,y) = \frac{1}{40} \qquad \text{for} \quad 0 \le y \le 14 \quad .$$

Now for the marginal probability density function we use the range relations of

$$\begin{aligned} 0 \le x &\le 10 \\ 0 \le y - x &\le 4 \end{aligned}$$

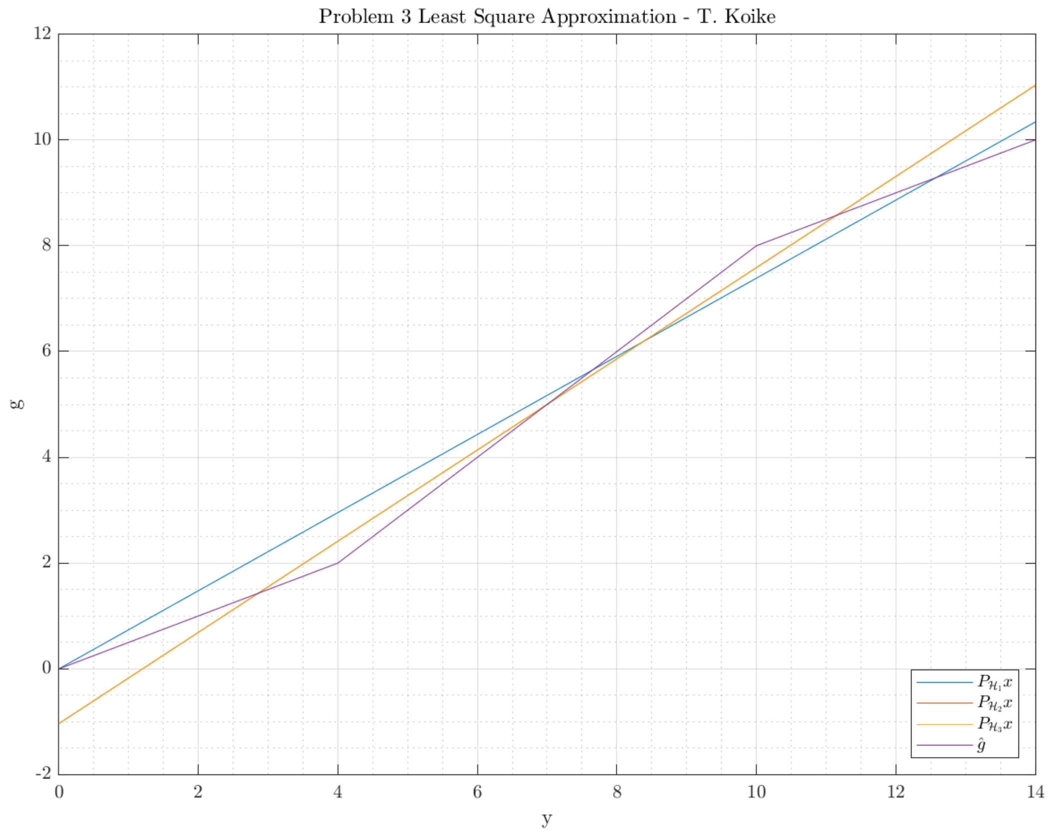
and we obtain

$$f_{\mathbf{y}}(y) = \begin{cases} \frac{1}{40}\int_0^y dx = \frac{y}{40} & \text{if} \quad 0 \le y \le 4 \\ \frac{1}{40}\int_{y-4}^y dx = \frac{1}{10} & \text{if} \quad 4 \le y \le 10 \\ \frac{1}{40}\int_{y-4}^{10} dx = \frac{14-y}{40} & \text{if} \quad 10 \le y \le 14 \end{cases}$$

Therefore

$$\hat{g}(y) = \begin{cases} \int_0^y x\left(\frac{1}{40}\right)\left(\frac{y}{40}\right)^{-1} dx = \frac{y}{2} & \text{if} \quad 0 \le y \le 4 \\ \int_{y-4}^y x\left(\frac{1}{40}\right)\left(\frac{1}{10}\right)^{-1} dx = y & \text{if} \quad 4 \le y \le 10 \\ \int_{y-4}^{10} x\left(\frac{1}{40}\right)\left(\frac{14-y}{40}\right)^{-1} dx = \frac{1}{2}y + 3 & \text{if} \quad 10 \le y \le 14 \end{cases}$$

Finally we plot our results,




From the graph we can see that for $\mathcal{H}_2$ and $\mathcal{H}_3$ there is no difference in the approximation. Thus, to achieve a better approximation we would have to span the space for more than $\mathcal{H}_4$ like we did in Problem 1 of this homework.

MATLAB code:

```matlab
%% A AE 567 HW4 Problem 3

% Housekeeping commands
clear all; close all; clc;
set(groot, 'defaulttextinterpreter','latex');
set(groot, 'defaultAxesTickLabelInterpreter','latex');
set(groot, 'defaultLegendInterpreter','latex');
outdir = pwd + "\output\hw4";
mdir = pwd + "\mfiles\hw4";
%%
m = 2;
% Define expectations
syms x v y g
y = x + v;
g = [];
for n = 0:m
    g = [g; y^n];
end
xl = 0; xu = 10;  % range of x
vl = 0; vu = 4;  % range of v
fx = 1 / (xu - xl);
fv = 1 / (vu - vl);

EX = @(A) int(fx * x.^A, xl, xu);
EV = @(A) int(fv * v.^A, vl, vu);
EY = @(A) fx * fv * int(int(y.^A, x, xl, xu), vl, vu);
EXY = @(A,B) fx * fv * int(int(x.^A * y.^B, x, xl, xu), vl, vu);

% P_Hx
Rfg = EXY(1, 0:m);
A = [];
for n = 0:m
    A = [A; n:n+m];
end
Rg = EY(A);
coef = Rfg * inv(Rg);
coef = eval(coef);
% Error d4
d4sq = EX(2) - Rfg*inv(Rg)*Rfg';
d4 = sqrt(d4sq);
%%
% Plotting
```

```
t = 0:0.01:14;

% ghat piecewise
t1 = t(0 <= t & t < 4);
t2 = t(4 <= t & t < 10);
t3 = t(10 <= t & t <= 14);
ghat = [t1 ./ 2, t2 - 2, (t3 + 6)/2];

% PHx
Ph3x = 0;
for n = 1:length(coef)
    Ph3x = Ph3x + coef(n)*t.^(n-1);
end

Ph1x = 130 * t / 176;
Ph2x = -30/29 + 25/29*t;

fig = figure("Renderer","painters",'Position',[60 60 900 650]);
    plot(t, Ph1x)
    grid on; grid minor; box on; hold on;
    plot(t, Ph2x)
    plot(t, Ph3x)
    plot(t, ghat)
    hold off;
    title('Problem 3 Least Square Approximation - T. Koike')
    legend('$P_{\mathcal{H}_1}x$','$P_{\mathcal{H}_2}x$',...
       '$P_{\mathcal{H}_3}x$', '$\hat{g}$', "Location","best")
    xlabel('y')
    ylabel('g')
saveas(fig, fullfile(outdir, 'p3_lsqr_plot.png'));
%%
% Save file as .m
matlab.internal.liveeditor.openAndConvert('hw4_p3.mlx', ...
    convertStringsToChars(fullfile(mdir, 'hw4_p3.m')));
```

## Problem 4

[Problem 2 from the Notes p. 35.] Let $\mathbf{y}$ be a uniform random variable over [0,1] and $\mathbf{x}$ the random variable defined by $\mathbf{x} = e^{\mathbf{y}}$. Let $\mathcal{H}$ be the subspace spanned by $\{1,\ \mathbf{y},\ \mathbf{y}^2\}$. Then compute the optimal estmiate $\hat{\mathbf{x}} = P_{\mathcal{H}}\mathbf{x}$ and the error in estimation $E(x - \hat{x})^2$. Show that $E(\mathbf{x}|\mathbf{y} = y) = e^y$. Plot your estimate $P_{\mathcal{H}}\mathbf{x}$ and the conditional expectation $E(\mathbf{x}|\mathbf{y} = y) = e^y$ on the same graph, and compare these two estimates.

---

**Solution:**
For convenience the heavy computations are done using MATLAB. (The code will be at the end of this problem.) Let

$$g = \begin{bmatrix} 1 \\ y \\ y^2 \end{bmatrix}$$

$$f = \mathbf{x} \quad .$$

Also since $\mathbf{x}$ and $\mathbf{v}$ are uniform distributions

$$\begin{aligned} E\mathbf{y} &= 0.5 \\ E\mathbf{x} &= \int_0^1 e^y dy = e - 1 \\ f_{\mathbf{y}}(y) &= 1 \end{aligned}$$

Then

$$R_{fg} = R_{\mathbf{x}g} = Exg = Ex\begin{bmatrix}1 & y & y^2\end{bmatrix} = \begin{bmatrix}Ex & Exy & Exy^2\end{bmatrix}.$$

Here

$$\begin{aligned} Exy &= E(e^y)y \\ &= \int_{-\infty}^{\infty} ye^y f_{\mathbf{y}}(y)dy = \int_0^1 ye^y dy \\ &= 1 \quad . \end{aligned}$$

Furthermore,

$$\begin{aligned} Exy^2 &= Ey^2e^y \\ &= \int_{-\infty}^{\infty} y^2e^y f_{\mathbf{y}}(y)dy \\ &= \int_0^1 y^2e^y dy \\ &= e - 2. \end{aligned}$$

Which gives

$$\therefore R_{fg} = \left[ \begin{array}{ccc} (e-1) & 1 & (e-2) \end{array} \right].$$

Next,

$$R_g = Egg^* = E \begin{bmatrix} 1 \\ y \\ y^2 \end{bmatrix} \begin{bmatrix} 1 & y & y^2 \end{bmatrix} = \begin{bmatrix} E1 & Ey & Ey^2 \\ Ey & Ey^2 & Ey^3 \\ Ey^2 & Ey^3 & Ey^4 \end{bmatrix}$$

$$= \left[ \begin{array}{ccc} 1 & 0.5000 & 0.3333 \\ 0.5000 & 0.3333 & 0.2500 \\ 0.3333 & 0.2500 & 0.2000 \end{array} \right].$$

Now the coefficients $a, b,$ and $c$ become

$$\begin{aligned} \begin{bmatrix} a & b & c \end{bmatrix} &= R_{fg} R_g^{-1} \\ &= \left[ \begin{array}{ccc} 1.7183 & 1 & 0.7183 \end{array} \right] \left[ \begin{array}{ccc} 1 & 0.5000 & 0.3333 \\ 0.5000 & 0.3333 & 0.2500 \\ 0.3333 & 0.2500 & 0.2000 \end{array} \right]^{-1} \\ &= \left[ \begin{array}{ccc} 1.0130 & 0.8511 & 0.8392 \end{array} \right]. \end{aligned}$$

Hence,

$$P_{\mathcal{H}}\mathbf{x} = 1.0130 + 0.8511y + 0.8392y^2 \quad .$$

Then the error becomes

$$\begin{aligned} d_4^2 &= R_{\mathbf{x}} - R_{fg} R_g^{-1} R_{gf} \\ &= 2.7835\text{e} - 5 \end{aligned}$$

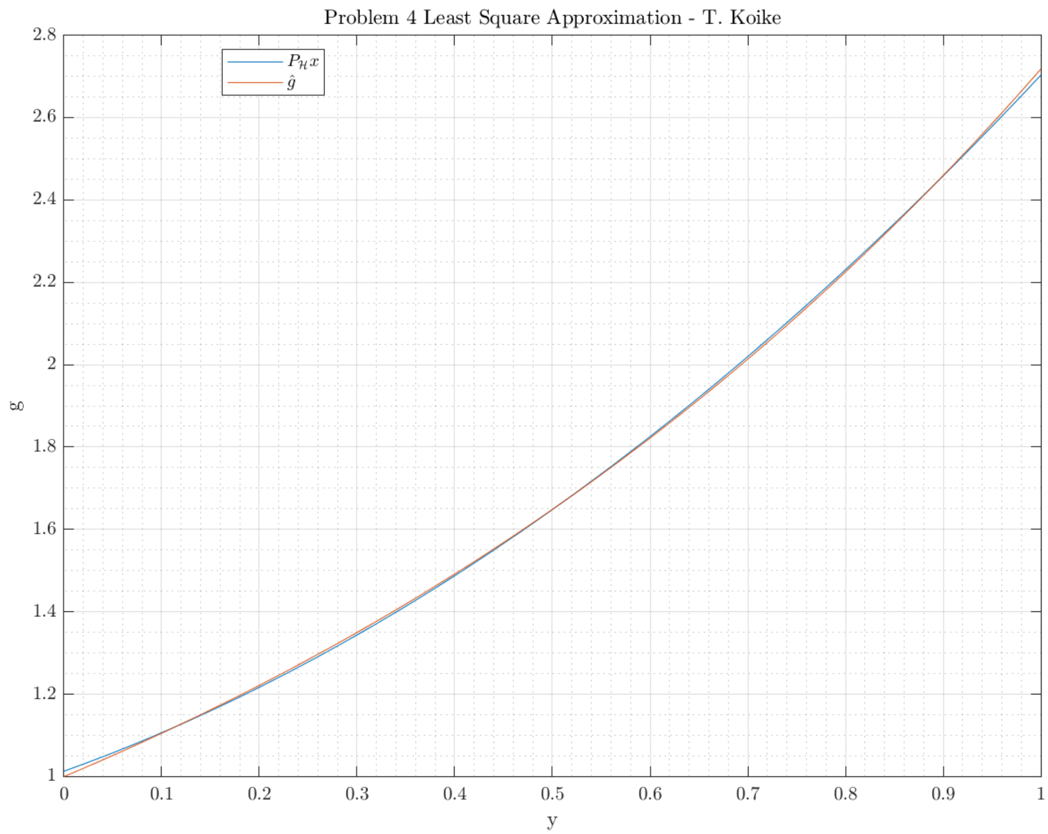
Thus,

$$\therefore d_4 = 0.0053 \quad .$$

It is known that the conditional expectation

$$\hat{g}(y) = E(\mathbf{x}|\mathbf{y} = y) = e^y \quad .$$

Thus, we plot the 2 on the same plot for comparison.




From the plot, we can see that the approximation is very close to the original exponential equation in the range of [0,1]. This visual result is coherent with the error value that we have calculated.

The MATLAB code is as follows.

```
%% AAE 567 HW4 Problem4

% Housekeeping commands
clear all; close all; clc;
set(groot, 'defaulttextinterpreter','latex');
set(groot, 'defaultAxesTickLabelInterpreter','latex');
set(groot, 'defaultLegendInterpreter','latex');
outdir = pwd + "\output\hw4";
mdir = pwd + "\mfiles\hw4";
%%
m = 2;
% Define expectations
```

```
syms x y
g = [];
for n = 0:m
    g = [g; y^n];
end
yl = 0; yu = 1;  % range of y
xl = 0; xu = 1; % range of x
fy = 1 ;

EY = @(A) int(y.^A, yl, yu);
EX = @(A) int(expm(A*y), yl, yu);
EXY = @(A,B) int(y.^B .* expm(A*y), yl, yu);

% P_Hx
Rfg = EXY(1, 0:m);
A = [];
for n = 0:m
    A = [A; n:n+m];
end
Rg = EY(A);

coef = Rfg * inv(Rg);
coef = eval(coef);
% Error d4
d4sq = EX(2) - Rfg*inv(Rg)*Rfg'
d4 = sqrt(d4sq);
%%
% Plotting
t = 0:0.01:1;
% ghat
Phx = 0;
for n = 1:length(coef)
    Phx = Phx + coef(n)*t.^(n-1);
end

fig = figure("Renderer","painters",'Position',[60 60 900 650]);
    plot(t, Phx)
    grid on; grid minor; box on; hold on;
    plot(t, exp(t))
    hold off;
    title('Problem 4 Least Square Approximation - T. Koike')
    legend('$P_{\mathcal{H}}x$', '$\hat{g}$', "Location","best")
    xlabel('y')
    ylabel('g')
saveas(fig, fullfile(outdir, 'p4_lsqr_plot.png'));
```

```
%%
% Save file as .m
matlab.internal.liveeditor.openAndConvert('hw4_p4.mlx', ...
    convertStringsToChars(fullfile(mdir, 'hw4_p4.m')));
```