$\vec{y} = \vec{A}\cdot\vec{x} + b = A: \vec{y} = \begin{vmatrix} x_1 & 1 \\ x_2 & 1 \\ x_3 & 1 \end{vmatrix} \cdot \begin{bmatrix} a \\ b \end{bmatrix}$

$\hat{x}$ - estimat x - true

$\tilde{x}$ - linear estimator

$$E[\hat{\Theta}] = (F^T F)^{-1} F^T E(y) = (F^T F)^{-1} F^T \cdot E(F\Theta + \epsilon)$$

1) a) $y = F\Theta + \epsilon$ lsq: $\min_{\Theta} \epsilon^T\epsilon$ $\epsilon \sim \mathcal{N}(0, I)$

sol.: $\hat{\Theta} = (F^T F)^{-1} F^T y$

assuming Θ is deterministic, y is deterministic

$$E[\hat{\Theta}] = E\left[\Theta + (F^T F)^{-1} F^T \epsilon\right] = \Theta + (F^T F)^{-1} F^T \cdot E(\epsilon) = \underline{\Theta}$$

$$\mathrm{cov}(\hat{\Theta}) = E\left[(\hat{\Theta} - \Theta)(\hat{\Theta} - \Theta)^T\right] = E\left[(\Theta + (F^T F)^{-1} F^T \epsilon - \Theta)(\Theta + (F^T F)^{-1} F^T \epsilon - \Theta)^T\right]$$

$$= E\left[(F^T F)^{-1} F^T \epsilon \epsilon^T F (F^T F)^{-1}\right] = (F^T F)^{-1} F^T E(\epsilon\epsilon^T) F (F^T F)^{-1}$$

$E[\epsilon\epsilon^T] \rightarrow$ since $\epsilon \sim \mathcal{N}(0, I)$, the off-diagonal elements will be 0, and the diagonal values will be 1.

$\Downarrow$

$$E[\epsilon\epsilon^T] = I$$

$$\mathrm{cov}(\hat{\Theta}) = (F^T F)^{-1} F^T I F (F^T F)^{-1} = \underline{(F^T F)^{-1}}$$

b/ $(F^T F) = \begin{bmatrix} 6 & 3 \\ 3 & 2 \end{bmatrix}$ $(F^T F)^{-1} = \begin{bmatrix} \frac{2}{3} & -1 \\ -1 & 2 \end{bmatrix}$

$(F^T F)^{-1} F^T = \begin{bmatrix} -\frac{1}{3} & \frac{1}{3} & \frac{2}{3} \\ 1 & 0 & -1 \end{bmatrix}$ $\hat{\Theta} = (F^T F)^{-1} F^T y = \begin{bmatrix} \frac{1}{3} \\ 1 \end{bmatrix}$

$y - F\hat{\Theta} = \begin{bmatrix} -\frac{1}{3} \\ \frac{1}{3} \\ -\frac{1}{3} \end{bmatrix}$ $\|y - F\Theta\|_2^2 = \frac{1}{\sqrt{3}}$

c/ $\hat{\Theta} = [\hat{\Theta}_a \; \hat{\Theta}_b \; \hat{\Theta}_M]^T$ $\hat{\Theta}_a \in \mathbb{R}$ $\hat{\Theta}_b \in \mathbb{R}$ $\hat{\Theta}_M \in \mathbb{R}^{n-2}$

split up $\hat{\Theta}$ just as F

$\begin{bmatrix} 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1^T \\ 2 & 1 & 3\mathbb{1}^T \end{bmatrix} \cdot \hat{\Theta}$ $\hat{\Theta}_a + 2\hat{\Theta}_b = 3$ $\hat{\Theta}_a = 3 - 2\hat{\Theta}_b$

$\epsilon = y - F\hat{\Theta} = y - f_a \hat{\Theta}_a - f_b \hat{\Theta}_b - M\hat{\Theta}_M = y - M\hat{\Theta}_M - f_a(3 - 2\hat{\Theta}_b) -$

$- f_b \hat{\Theta}_b = y - 3f_a - (f_b - 2f_a)\hat{\Theta}_b - M\hat{\Theta}_M$

$2 = 3 - 2\hat{\Theta}_b + \hat{\Theta}_b + \mathbb{1}^T \hat{\Theta}_M$

$\hat{\Theta}_b = \mathbb{1}^T \hat{\Theta}_M + 1$

$\epsilon = (y - 3f_a) - (f_b - 2f_a) \cdot (\mathbb{1}^T \hat{\Theta}_M + 1) - M\hat{\Theta}_M =$

$= \underbrace{(y - 3f_a - f_b + 2f_a)}_{y_{aug}} - \underbrace{(f_b \mathbb{1}^T - 2f_a \mathbb{1}^T + M)}_{F_{aug}} \hat{\Theta}_M$ → final least-squares problem

$\hat{\Theta}_M = (F_{aug}^T F_{aug})^{-1} F_{aug}^T y_{aug}$

② a)
$$p(\Theta) = \frac{1}{\sqrt{(2\pi)^n |P_\Theta|}} \cdot \exp\left(-\frac{1}{2}(\Theta-\mu_\Theta)^T P_\Theta^{-1}(\Theta-\mu_\Theta)\right) \quad n=\dim(\Theta)$$

$$p(y|\Theta) = \frac{1}{\sqrt{(2\pi)^n |P_{y|\Theta}|}} \cdot \exp\left(-\frac{1}{2}(y-\mu_{y|\Theta})^T P_{y|\Theta}^{-1}(y-\mu_{y|\Theta})\right) \quad n=\dim(y)$$

$$p(\Theta|y) = \frac{1}{\sqrt{(2\pi)^n |P_{y|\Theta}|}} \exp\left(-\frac{1}{2}(\Theta-\mu_{\Theta|y})^T P_{\Theta|y}^{-1}(\Theta-\mu_{\Theta|y})\right)$$

$$P_y^{-1} = (LL^T)^{-1} = W = W^T \quad P_\Theta > 0 \rightarrow \quad P_{\hat\Theta} = (P_\Theta^{-1} + F^T W F),$$

$$\mu_{y|\Theta} = F\Theta \quad \mu_{y|\Theta}^T = \Theta^T F^T$$

$$p(y|\Theta)\cdot p(\Theta) = \alpha \cdot \exp\left(-\frac{1}{2}(y^T - \Theta^T F^T) W (y - F\Theta) - \frac{1}{2}(\Theta-\mu_\Theta)^T P_\Theta^{-1}(\Theta-\mu_\Theta)\right) = \alpha \cdot \exp\left(-\frac{1}{2}y^T W y - \frac{1}{2}(\Theta^T F^T W F \Theta) + \frac{1}{2}\Theta^T F^T W y + \frac{1}{2} y^T W F\Theta - \frac{1}{2}\Theta^T P_\Theta^{-1}\Theta - \frac{1}{2}\mu_\Theta^T P_\Theta^{-1}\mu_\Theta + \frac{1}{2}\mu_\Theta^T P_\Theta^{-1}\Theta + \frac{1}{2}\Theta^T P_\Theta^{-1}\mu_\Theta\right) = \underline{\alpha \cdot \exp\left(-\frac{1}{2}\Theta^T(F^T W F + P_\Theta^{-1})\Theta + \frac{1}{2}(\Theta^T F^T W y + y^T W F\Theta + \mu_\Theta^T P_\Theta^{-1}\Theta + \Theta^T P_\Theta^{-1}\mu_\Theta) - \frac{1}{2}y^T W y - \frac{1}{2}\mu_\Theta^T P_\Theta^{-1}\mu_\Theta\right)}$$

$$\downarrow$$

2nd part

$$p_{\Theta|y} = \beta \cdot \exp\left(-\frac{1}{2}\Theta^T P_{\Theta|y}^{-1}\Theta + \underline{\frac{1}{2}(\Theta^T P_{\Theta|y}^{-1}\mu_{\Theta|y} + \mu_{\Theta|y}^T P_{\Theta|y}^{-1}\Theta)} - \underline{\frac{1}{2}(\mu_{\Theta|y} P_{\Theta|y}^{-1}\mu_{\Theta|y})}\right)$$

1st part

$\Theta^T P_{\Theta|y}^{-1}\mu_{\Theta|y}$ may not be symmetric b, $E\left[(\theta-\mu_\theta)\epsilon^T\right]=0$ $\qquad y = F\theta + L\epsilon$

unbiased : $E[\mu_{\theta|y}] = \mu_\theta$

minimum-variance : $E\left[(\theta-\mu_{\theta|y})(\theta-\mu_{\theta|y})^T\right]$ is minimized

$$\mu_{\theta|y} = [M\ N]\begin{bmatrix} y \\ \mu_\theta \end{bmatrix}$$

μ_θ is deterministic

$\rightarrow E[\mu_{\theta|y}] = E[My] + N\mu_\theta = E[MF\theta + M\epsilon] + N\mu_\theta = MF\,E[\theta] + M\,E[\epsilon] +$

$+ N\mu_\theta = MF\mu_\theta + N\mu_\theta = \mu_\theta \Rightarrow MF + N = I$ $\qquad E[\epsilon]=0$

$N = (I - MF)$

$\rightarrow (\theta - \mu_{\theta|y}) = \theta - MF\theta - ML\epsilon - N\mu_\theta = (I-MF)\theta - (I-MF)\mu_\theta - ML\epsilon =$

$= (I-MF)(\theta-\mu_\theta) - ML\epsilon$

$\rightarrow E\left[(\theta-\mu_{\theta|y})(\theta-\mu_{\theta|y})^T\right] = E\left[\left((I-MF)(\theta-\mu_\theta)-ML\epsilon\right)\left((I-MF)(\theta-\mu_\theta)-ML\epsilon\right)^T\right] =$

$= E\Big[(I-MF)(\theta-\mu_\theta)(\theta-\mu_\theta)^T(I-MF)^T - ML\epsilon(\theta-\mu_\theta)^T(I-MF) -$

$- (I-MF)(\theta-\mu_\theta)\epsilon^T L^T M^T + ML\epsilon\epsilon^T L^T M\Big] = (I-MF)\underbrace{E\left[(\theta-\mu_\theta)(\theta-\mu_\theta)^T\right]}_{P}$

$(I-MF)^T - ML\underbrace{E\left[\epsilon(\theta-\mu_0)^T\right]}_{(E[(\theta-\mu_0)\epsilon^T])^T = 0^T = 0}(I-MF) - (I-MF)\underbrace{E\left[(\theta-\mu_\theta)\epsilon^T\right]}_{0} + ML\underbrace{E\left[\epsilon\epsilon^T\right]}_{I}L^T M^T$

$(LL^T) = W^{-1}$

$\rightarrow E\left[(\theta-\mu_{\theta|y})(\theta-\mu_{\theta|y})^T\right] = (I-MF)P(I-MF)^T + MW^{-1}M^T =$

$$= [I\ \ -M]\begin{bmatrix} P & PF^T \\ FP & FPF^T \end{bmatrix}\begin{bmatrix} I \\ -M^T \end{bmatrix} + MW^{-1}M^T = [I\ \ -M]\underbrace{\begin{bmatrix} P & PF^T \\ FP & FPF^T + W^{-1} \end{bmatrix}}_{Q}\begin{bmatrix} I \\ -M^T \end{bmatrix}$$

$W, L > 0 \Rightarrow W^{-1} > 0$ $\qquad P > 0 \Rightarrow FPF^T > 0 \Rightarrow FPF^T + W^{-1} > 0$

since the bottom right entry of Q is positive definite the schur-decomposition can be applied.

$$\rightarrow \quad Q = \begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} I & BD^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} A-BD^{-1}C & 0 \\ 0 & D \end{bmatrix} \begin{bmatrix} I & 0 \\ D^{-1}C & I \end{bmatrix}$$

$$\rightarrow \quad Q = \begin{bmatrix} I & PF^T(FPF^T+W^{-1})^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} P-PF^T(FPF^T+W^{-1})^{-1}FP & 0 \\ 0 & FPF^T+W^{-1} \end{bmatrix} \begin{bmatrix} I & 0 \\ (FPF^T+W^{-1})^{-1}FP & I \end{bmatrix}$$

$\rightarrow$ substituting it back into $E((\Theta-\mu_\Theta)(\Theta-\mu_\Theta)^T)$

$$\rightarrow E\left[(\Theta-\mu_{\Theta|y})(\Theta-\mu_{\Theta|y})^T\right] = \begin{bmatrix} I & -M \end{bmatrix} \begin{bmatrix} I & PF \\ & \end{bmatrix} =$$

$$= \begin{bmatrix} I & PF^T(FPF^T+W^{-1})^{-1}-M \end{bmatrix} \begin{bmatrix} P-PF^T(FPF^T+W^{-1})^{-1}FP & 0 \\ 0 & (FPF^T+W^{-1}) \end{bmatrix} \begin{bmatrix} I \\ (FPF^T+W^{-1})^{-1}FP-M^T \end{bmatrix} =$$

$$= \begin{bmatrix} P-PF^T(FPF^T+W^{-1})^{-1}FP & PF^T-M(FPF^T+W^{-1}) \end{bmatrix} \begin{bmatrix} I \\ (FPF^T+W^{-1})^{-1}FP-M^T \end{bmatrix} = \left(P-PF^T(FPF^T+W^{-1})^{-1}FP\right) +$$

$$+ \left[PF^T(FPF^T+W^{-1})^{-1}-M\right)(FPF^T+W^{-1})\Big] \cdot \left((FPF^T+W^{-1})^{-1}FP-M^T\right)$$

since the goal is to minimize the above expression one can set either $(PF^T(FPF^T+W^{-1})^{-1}-M)$ or its transpose to 0

$$SO \Rightarrow M = PF^T(FPF^T + W^{-1})^{-1}$$

$$\rightarrow \mu_{\Theta|y} = PF^T(FPF^T+W^{-1})^{-1}y + \mu_\Theta - PF^T(FPF^T+W^{-1})^{-1}F\mu_\Theta =$$

$$= \underline{PF^T(FPF^T+W^{-1})^{-1}(y - F\mu_\Theta) + \mu_\Theta}$$

If $(PF^T(FPF^T+W^{-1})^{-1} - M)$ is set to 0 the remaining term

in the expression for $E\left[(\Theta-\mu_{\Theta|y})(\Theta-\mu_{\Theta|y})^T\right] =$

$$= \left(P - PF^T(FPF^T+W^{-1})^{-1}FP\right) = P_{\Theta|y}$$

c) since $P > 0$, and the Matrix inversion lemma
$\exists P^{-1}$

$$(A + BCD)^{-1} = A^{-1} - A^{-1}B(C^{-1} + DA^{-1}B)DA^{-1}$$

$\rightarrow (FPF^T + W^{-1})^{-1} = W - WF(P^{-1} + F^TWF)^{-1}F^TW$, $\mu_{\Theta|y}$ becomes

$$PF^T(W - WF(P^{-1} + F^TWF)^{-1}F^TW) = P(F^TW - F^TWF(P^{-1}+F^TWF)^{-1}F^TW) =$$

$$= P(I - F^TWF(P^{-1}+F^TWF)^{-1})F^TW = P((P^{-1}+F^TWF)\cdot(P^{-1}+F^TWF)^{-1} - F^TWF(P^{-1}+F^TWF))$$

$$F^TW = \cancel{P}((P^{-1} + \cancel{F^TWF} - \cancel{F^TWF})(P^{-1}+F^TWF))F^TW = \underline{(P^{-1}+F^TWF)F^TW}$$

sSC42025 Filtering & Identification

MATLAB EXERCISE HW1

Instructions: Fill in the live script with your code and answers. Then export the live script as a pdf (in the 'Live Editor tab', click on the arrow under 'Save' and then 'Export to pdf').

You are given measurements $y \in \mathbb{R}^{100\times 1}$. Assuming the measurement model

$$y = F \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \epsilon, \qquad \epsilon \sim \mathcal{N}(0, R),$$

with $F \in \mathbb{R}^{100\times 2}$ and $R = \sigma_y^2 I$, $\sigma_y = 0.5$, you would like to find an estimate for $[x_1 \quad x_2]^\top$. You also have a prior given by

$$\theta \sim \mathcal{N}\left([1.1 \quad 1.1]^\top, \begin{bmatrix} 0.1 & 0 \\ 0 & 0.1 \end{bmatrix}\right).$$

Note that this is a difficult estimation problem, since some measurements have a small signal-to-noise ratio, in other words, F is different in each row.

Exercise 1: Batch estimation

Compute the stochastic least squares estimate $[\hat{x}_1 \quad \hat{x}_2]^\top$ using the given matrices `y.mat` and `F.mat` provided in `data.mat`. Print the result.

```
clear; clc; close all;

load('data1.mat')
N       = 100;
invP0   = 10*eye(2);
x0      = [1.1; 1.1];
sigy    = .5;
R       = sigy^2*eye(size(F,1));

P_theta = [0.1 0; 0 0.1];
% P_theta is positive definite obviously but i already wrote the code so im
% not gonna change it
K = P_theta * F.' / (F* P_theta * F.' + R)
```

```
K = 2×100
    0.0076   -0.0197    0.0056    0.0156    0.0002   -0.0054   -0.0159    0.0079 ···
   -0.0102   -0.0026    0.0048    0.0081    0.0034   -0.0220   -0.0037    0.0015
```

```
theta_hat = K* y + (eye(2) - K*F) * x0
```

```
theta_hat = 2×1
    1.1397
    1.1721
```

Exercise 2: Recursive estimation

Compute the recursive least squares estimate, pretending that one measurement becomes available at a time. Print the result.

```
P_new = P_theta
```

```
P_new = 2×2
    0.1000         0
         0    0.1000
```

```
theta_new = [1.1,; 1.1]
```

```
theta_new = 2×1
    1.1000
    1.1000
```

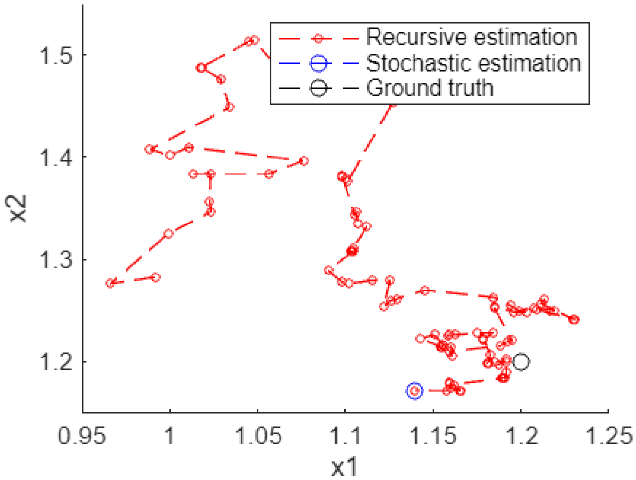
```
t_history = zeros(100, 2);
for i = (1:100)
    F_loop = F(i, :);
    y_loop = y(i, :);
    R_loop = R(i, i);
    K_new = P_new * F_loop.' / (F_loop* P_new * F_loop.' + R_loop);
    theta_new = K_new*y_loop + (eye(2) - K_new*F_loop) * theta_new;
    P_new = P_new - K_new * F_loop * P_new;
    t_history(i, :) = theta_new;
end

theta_new
```

```
theta_new = 2×1
    1.1397
    1.1721
```

Exercise 3: Visualization

Plot the results of a) and b) **into the same figure**. For b) plot the path connecting the estimates for each recursion step. Note that the recursive estimation should obtain the exact same result as the batch estimation. Also plot the ground truth, which is given by $[1.2 \quad 1.2]$. You could set the x-axis to represent estimate for x_1 and the y-axis to represent estimate for x_2 in the figure.

```
hold on
plot(t_history(:, 1), t_history(:, 2), '--ro', MarkerSize=3, DisplayName="Recursive estimation")
plot(theta_hat(1), theta_hat(2), '--bo', DisplayName="Stochastic estimation")
plot(1.2, 1.2, "--ko", DisplayName="Ground truth")
xlabel('x1')
ylabel('x2')
legend()
hold off
```



Recursive estimation
Stochastic estimation
Ground truth
1.5
1.4
1.3
1.2
x2
0.95
1
1.05
1.1
1.15
1.2
1.25
x1