Optional homework

1. ***The Wiener and Kalman filters***

We have the stable linear state-space system

$$\begin{aligned} x_{t+1} &= Ax_t + v_t \\ y_t &= Cx_t + w_t \end{aligned}$$

where we measure the output sequence y and would like to estimate the state x. Here v_t is a Gaussian disturbance acting on the system, and w_t is Gaussian sensor noise, and

$$v_t \sim \mathcal{N}(0, \Sigma_v) \qquad w_t \sim \mathcal{N}(0, \Sigma_w)$$

and the initial state is

$$x_0 \sim \mathcal{N}(0, Q_0)$$

All of the random vectors $v_0, v_1, \ldots, w_0, w_1, \ldots$ and x_0 are independent.

We would like to find an estimator of the form

$$\hat{x}_t = F(t) \begin{bmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-r+1} \end{bmatrix}$$

where $F(t)$ is an appropriate matrix. Here we allow the estimator access to the past r measurements of y, and $\hat{x}_t$ is an estimate of x_t that minimizes the mean square error. In particular, we would like to find the *steady-state* estimator, that is, the matrix F given by

$$F = \lim_{t\to\infty} F(t)$$

so that in steady-state we can write the estimate as

$$\hat{x}_t = \begin{bmatrix} F_0 & F_1 & \ldots & F_{r-1} \end{bmatrix} \begin{bmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-r+1} \end{bmatrix}$$

where

$$F = \begin{bmatrix} F_0 & F_1 & \ldots & F_{r-1} \end{bmatrix}$$

Then the matrices F_i are the *impulse response* of a filter which gives the state estimate. The estimator F is called the *FIR Wiener filter.*

(a) Define the vector $y_{\text{past}}(t)$ to be

$$y_{\text{past}}(t) = \begin{bmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-r+1} \end{bmatrix}$$

Show that this has a steady-state covariance given by

$$\lim_{t\to\infty} \operatorname{cov}(y_{\text{past}}) = R$$

where R is the symmetric Toeplitz matrix

$$\begin{bmatrix} R_0 & R_1 & R_2 & \ldots & R_{r-1} \\ R_1^T & R_0 & R_1 & \ldots & \\ R_2^T & R_1^T & R_0 & & \\ \vdots & & & \ddots & \\ R_{r-1}^T & & & R_1^T & R_0 \end{bmatrix}$$

where $R_0 = CQC^T + \Sigma_w$ and

$$R_k = CA^kQC^T \qquad \text{for all } k = 1, \dots, r-1$$

and the matrix Q is the unique solution to the Lyapunov equation

$$Q = AQA^T + \Sigma_v$$

(b) Define the matrix S to be

$$S = \lim_{t\to\infty} \mathrm{E}\Big(x_t \begin{bmatrix} y_t^T & y_{t-1}^T & \dots & y_{t-r+1}^T \end{bmatrix}\Big)$$

Show that

$$S = \begin{bmatrix} QC^T & AQC^T & \dots & A^{r-1}QC^T \end{bmatrix}$$

(c) Show that the optimal F is given by

$$F = SR^{-1}$$

(d) Suppose X is the unique positive-definite solution to the Riccati equation

$$X = AXA^T - AXC^T(CXC^T + \Sigma_w)^{-1}CXA^T + \Sigma_v$$

Consider the Kalman filter, which gives the MMSE estimate $\check{x}_t$ of x_t given $y_0, y_1, \dots, y_t$. In steady-state, this may be written as

$$\check{x}_{t+1} = \check{A}\check{x}_t + Ky_{t+1}$$

(Note that this is *not* the same as the MMSE estimate of x_t conditioned on $y_0, y_1, \dots, y_{t-1}$.) Find the matrices $\check{A}$ and K in terms of X. You don't need to rederive the Kalman filter, and can quote any formulae you need from the notes.

(e) Hence write the Kalman filter estimate as

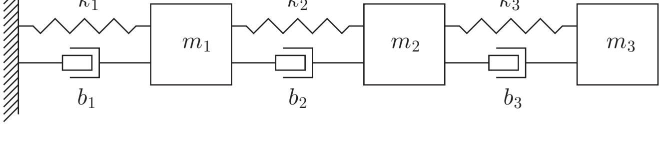
$$\check{x}_t = \sum_{k=0}^{t} G_k y_{t-k}$$

which, if we only consider the first r terms, is approximately

$$\check{x}_t \approx \begin{bmatrix} G_0 & G_1 & \dots G_{r-1} \end{bmatrix} \begin{bmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-r+1} \end{bmatrix}$$

Find G in terms of X.

(f) Now we have constructed an MMSE estimator for $\hat{x}_t$ based on $y_t, y_{t-1}, \dots, y_{t-r+1}$ in two different ways. The estimator F is optimal if we only have the past r measurements of y; the estimator G is optimal if we have all the past data. For large r, these should be indistinguishable. Let's check this. Consider the mass-spring system



where the masses $m_i = 1$, springs $k_i = 2$, dampers $b_i = 0.1$, $\Sigma_w = 0.1I$, $\Sigma_v = 0.2I$. We measure the position of the first mass, and would like to estimate the state. The sampling rate is $h = 0.1$. The first three states are the positions of masses 1 to 3; the second three states are the corresponding velocities.

Plot the 1,1 entry of F and that of G as a function of time when $r = 20$ and when $r = 200$. These are the impulse responses mapping the measurements y to the estimate of the position of the first mass.

Solution.

(a) Define $Q_t = \mathrm{cov}(x_t)$. Consider the random vector

$$\begin{bmatrix} x_t \\ x_{t+1} \end{bmatrix} = \begin{bmatrix} I & 0 \\ A & I \end{bmatrix} \begin{bmatrix} x_t \\ v_t \end{bmatrix}$$

This vector is Gaussian with zero mean and joint covariance

$$\begin{bmatrix} Q_t & Q_{t,t+1} \\ Q_{t+1,t} & Q_{t+1} \end{bmatrix} = \begin{bmatrix} Q_t & Q_t A^T \\ AQ_t & AQ_t A^T + \Sigma_v \end{bmatrix}$$

Define Q to be

$$Q = \lim_{t\to\infty} Q_t$$

In steady-state, we must have

$$Q = \lim_{t\to\infty} Q_t = \lim_{t\to\infty} Q_{t+1}$$

Hence, Q satisfies the Lyapunov equation

$$Q = AQA^T + \Sigma_v$$

Also, by recursion we can find that

$$Q_{s,t} = A^{s-t} Q_t \qquad s \geq t$$

Define the vectors $x_{\text{past}}(t)$ and $w_{\text{past}}(t)$ to be

$$x_{\text{past}}(t) = \begin{bmatrix} x_t \\ x_{t-1} \\ \vdots \\ x_{t-r+1} \end{bmatrix} \qquad w_{\text{past}}(t) = \begin{bmatrix} w_t \\ w_{t-1} \\ \vdots \\ w_{t-r+1} \end{bmatrix}$$

Using the above formulae for the covariance of x_t, we can define $\bar{Q}(t)$ as the joint covariance of $x_{\text{past}}(t)$

$$\bar{Q}(t) = \begin{bmatrix} Q_t & AQ_{t-1} & A^2 Q_{t-2} & \cdots & A^{r-1} Q_{t-r+1} \\ Q_{t-1} A^T & Q_{t-1} & AQ_{t-2} & & \\ {A^T}^2 Q_{t-2} & A^T Q_{t-2} & Q_{t-2} & & \\ \vdots & & & \ddots & \\ {A^T}^{r-1} Q_{t-r+1} & & & & Q_{t-r+1} \end{bmatrix}$$

In steady-state, this becomes

$$\bar{Q} = \lim_{t\to\infty} \bar{Q}(t) = \begin{bmatrix} Q & AQ & A^2 Q & \cdots & A^{r-1} Q \\ QA^T & Q & AQ & & \\ {A^T}^2 Q & A^T Q & Q & & \\ \vdots & & & \ddots & \\ {A^T}^{r-1} Q & & & & Q \end{bmatrix}$$

Now, consider the vector

$$\begin{bmatrix} y_{\text{past}}(t) \\ x_t \end{bmatrix} = \begin{bmatrix} C_d & I \\ e_1^T & 0 \end{bmatrix} \begin{bmatrix} x_{\text{past}}(t) \\ w_{\text{past}}(t) \end{bmatrix}$$

where

$$C_d = \begin{bmatrix} C & & 0 \\ & \ddots & \\ 0 & & C \end{bmatrix} \qquad e_1 = \begin{bmatrix} I \\ 0 \\ \vdots \end{bmatrix}$$

This vector is Gaussian with zero mean and joint covariance

$$\begin{bmatrix} \Sigma_y & \Sigma_{yx} \\ \Sigma_{xy} & \Sigma_x \end{bmatrix} = \begin{bmatrix} C_d \bar{Q}(t) C_d^T + \bar{\Sigma}_w & C_d \bar{Q}(t) e_1 \\ e_1^T \bar{Q}(t) C_d^T & e_1^T \bar{Q}(t) e_1 \end{bmatrix}$$

where

$$\bar{\Sigma}_w = \begin{bmatrix} \Sigma_w & & 0 \\ & \ddots & \\ 0 & & \Sigma_w \end{bmatrix}$$

Thus,

$$\mathrm{cov}(y_{\mathrm{past}}(t)) = C_d \bar{Q}(t) C_d^T + \bar{\Sigma}_w$$

In steady-state, this becomes

$$\lim_{t \to \infty} \mathrm{cov}(y_{\mathrm{past}}(t)) = C_d \bar{Q} C_d^T + \bar{\Sigma}_w = R$$

as desired.

(b) Using the results from part (a), we get

$$\mathrm{E}(x_t y_{\mathrm{past}}^T) = e_1^T \bar{Q}(t) C_d^T$$

In steady-state, this becomes

$$\lim_{t \to \infty} \mathrm{E}(x_t y_{\mathrm{past}}^T) = e_1^T \bar{Q} C_d^T = S$$

as desired.

(c) Using the standard MMSE results, we have

$$\hat{x}_t = \Sigma_{xy} \Sigma_y^{-1} y_{\mathrm{past}}(t)$$

Using the results of parts (a) and (b), in steady-state we get

$$\lim_{t \to \infty} \Sigma_{xy} \Sigma_y^{-1} = S R^{-1} = F$$

(d) We want to find the estimate for x_{t+1} given $y_0, y_1, \ldots, y_{t+1}$. Using the Kalman filter notation in the notes, we want

$$\check{x}_{t+1} = \hat{x}(t+1|t+1) = \hat{x}(t+1|t) + \Sigma_{t+1|t} C^T (C \Sigma_{t+1|t} C^T + \Sigma_w)^{-1} (y_{t+1} - C \hat{x}(t+1|t))$$

Using $\hat{x}(t+1|t) = A\hat{x}(t|t)$, we get

$$\begin{aligned} \check{x}_{t+1} &= A\hat{x}(t|t) + \Sigma_{t+1|t} C^T (C \Sigma_{t+1|t} C^T + \Sigma_w)^{-1} (y_{t+1} - CA\hat{x}(t|t)) \\ &= A\check{x}_t + \Sigma_{t+1|t} C^T (C \Sigma_{t+1|t} C^T + \Sigma_w)^{-1} (y_{t+1} - CA\check{x}_t) \end{aligned}$$

For the covariance matrix $\Sigma_{t+1|t}$, we have

$$\begin{aligned} \Sigma_{t+1|t} &= A \Sigma_{t|t} A^T + \Sigma_v \\ &= A \Sigma_{t|t-1} A^T - A \Sigma_{t|t-1} C^T (C \Sigma_{t|t-1} C^T + \Sigma_w)^{-1} C \Sigma_{t|t-1} A^T + \Sigma_v \end{aligned}$$

Define X to be

$$X = \lim_{t \to \infty} \Sigma_{t|t-1}$$

In steady-state we must have

$$X = \lim_{t\to\infty} \Sigma_{t+1|t} = \lim_{t\to\infty} \Sigma_{t|t-1}$$

Hence, X satisfies the Riccati recursion

$$X = AXA^T - AXC^T(CXC^T + \Sigma_w)^{-1}CXA^T + \Sigma_v$$

Substituting back into $\check{x}_{t+1}$ yields

$$\begin{aligned} \check{x}_{t+1} &= A\check{x}_t + XC^T(CXC^T + \Sigma_w)^{-1}(y_{t+1} - CA\check{x}_t) \\ &= \check{A}\check{x}_t + Ky_{t+1} \end{aligned}$$

where

$$K = XC^T(CXC^T + \Sigma_w)^{-1} \qquad \check{A} = (I - KC)A$$

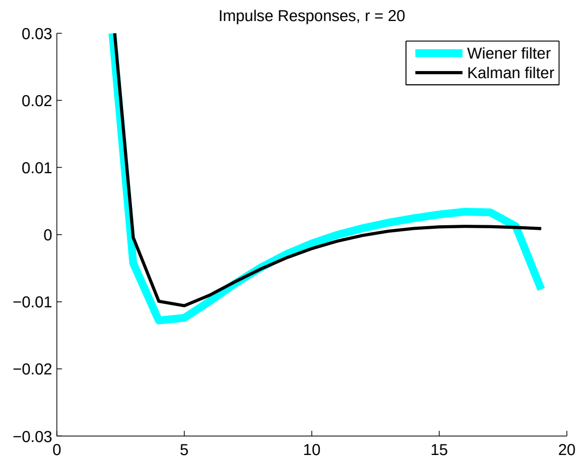
(e) Performing the recursion in part (d) produces

$$\check{x}_t = Ky_t + \check{A}Ky_{t-1} + \check{A}^2Ky_{t-2} + \dots$$

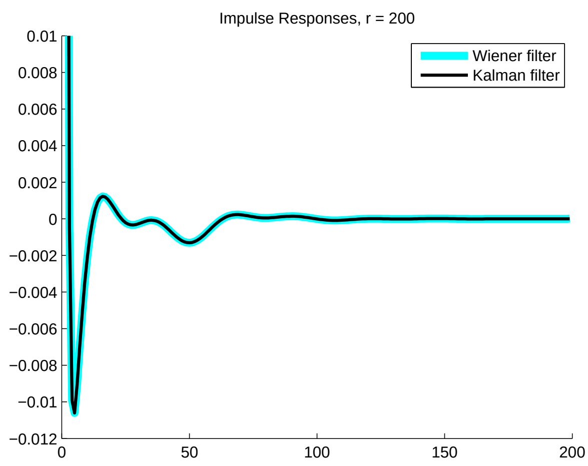
Hence, we see

$$G_k = \check{A}^k K$$

(f) Using this mass-spring system, the impulse responses for each of the filters are found from the first row of the F and G matrices. For $r = 20$, the impulse responses are shown below.



For $r = 200$, the impulse response looks like this.



In comparing the two cases, we see that the Wiener filter is more aggressive with measurements when it only has a few to work with. On the other hand, we observe that both filters are nearly identical when $r = 200$, as we expect from the setup of the problem.

2. ***Two Guards***

We have two guards in a corridor, and a sensor which blinks whenever either of the guards passes it. We model the corridor as a sequence of 10 slots numbered from 1 to 10, and the motion of the guards as a random walk, as follows.

- If a guard is at position $1 < i < 10$ at time t, then at time $t+1$ its position is $i+1$ with probability r, $i-1$ with probability r, and it stays at position i with probability $1-2r$.
- If the guard is at position 1 at time t then at time $t+1$ it is at position 2 with probability r, and stays at position 1 with probability $1-r$.
- If the guard is at position 10 at time t then at time $t+1$ it is at position 9 with probability r, and stays at position 10 with probability $1-r$.

We'll use $r = 0.07$. The sensor is at position 7. We measure a $y = 1$ if either or both guards is in position 7, otherwise we measure $y = 2$.

(a) Suppose guard i is at position $a_i(t)$ at time t. The sequence $a_1(0), a_1(1), a_1(2), \ldots$ is a Markov process, and similarly so is a_2. Both have the same transition matrix. What is it?

(b) Let P_1 be the transition matrix from part(a). Let

$$x(t) = 10(a_1(t) - 1) + a_2(t)$$

The sequence of random variables $x(0), x(1), \ldots$ is a Markov process. Let P be its transition matrix. What is P? (Hint: you can express it in terms of P_1.)

(c) Let G be the sensor transition matrix

$$G_{ij} = \text{Prob}(y(t) = j \mid x(t) = i)$$

What is G?

(d) The system starts with $a_1(0) = 7$ and $a_2(0) = 4$. Simulate the system on time interval $t = 0, 1, \ldots, 20$, and plot a_i, y as a function of time.

(e) Construct the Kalman filter. Assume that the initial state is known exactly. The posterior pdf $q_{t|t-1}$ is a vector in $\mathbb{R}^{100}$, which is a pmf on the 10×10 space of the pairs of possible positions of the guards. (Use Matlab's `reshape` function.) For the same sequence as in the previous part, plot the minimum probability of error estimate of $x(t)$ given $y(0), \ldots, y(t)$ of the two guards positions as a function of time.

(f) Starting in the same initial state as above, we measure the sequence of y's to be

$$y = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 2 & 2 & 2 & 2 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

Plot as a density the posterior pdf $q_{t|t-1}$ at times $t = 9$ and $t = 17$. Interpret these plots, i.e., what do they say about the locations of the guards?

(g) For the same data as in the previous part, give the sequence of minimum-probability-of-error estimates of $x(t)$ given $y(0), \ldots, y(t)$ of the positions of the two guards as a function of time.

Solution.

(a) The transition matrix is

$$P_1 = \begin{bmatrix} 1-r & r & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ r & 1-2r & r & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & r & 1-2r & r & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & r & 1-2r & r & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & r & 1-2r & r & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & r & 1-2r & r & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & r & 1-2r & r & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & r & 1-2r & r & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & r & 1-2r & r \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & r & 1-r \end{bmatrix}$$

(b) Note that for each $x(t)$, there is a unique $a_1(t)$ and $a_2(t)$. Thus, if $x(t) = i$, let $a_1(t) = i_1$ and $a_2(t) = i_2$ be the corresponding positions of the guards. As a result, we have

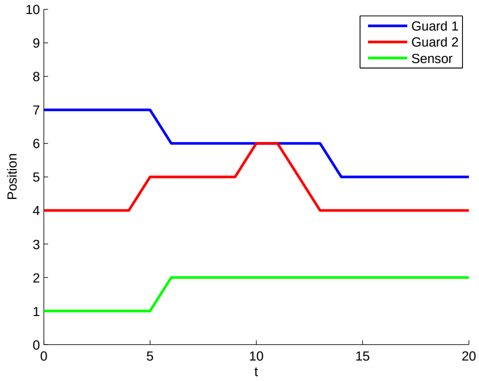
$$\begin{aligned} P(i,j) &= \text{Prob}(x(t+1) = j | x(t) = i) \\ &= \text{Prob}(a_1(t+1) = j_1, a_2(t+2) = j_2 | a_1(t) = i_1, a_2(t) = i_2) \\ &= \text{Prob}(a_1(t+1) = j_1 | a_1(t) = i_1)\,\text{Prob}(a_2(t+1) = j_2 | a_2(t) = i_2) \\ &= P_1(i_1, j_1) P_1(i_2, j_2) \end{aligned}$$

where the factorization arises because the guards move independently.

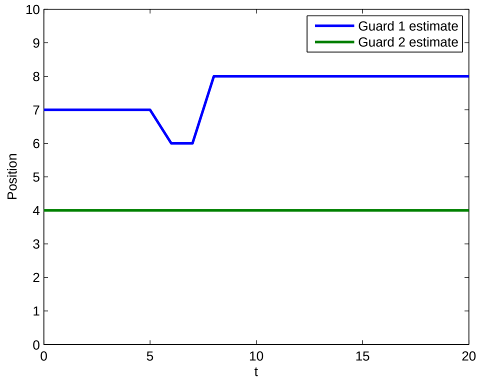
(c) Since the sensor is deterministic, its transition matrix is

$$\begin{aligned} G(i,j) &= \text{Prob}(y(t) = j | x(t) = i) \\ &= \text{Prob}(y(t) = j | a_1(t) = i_1, a_2(t) = i_2) \\ &= \begin{cases} 1 & j = 1, \text{ and } (i_1 = 7 \text{ or } i_2 = 7) \\ 1 & j = 2, \text{ and } (i_1 \neq 7 \text{ and } i_2 \neq 7) \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

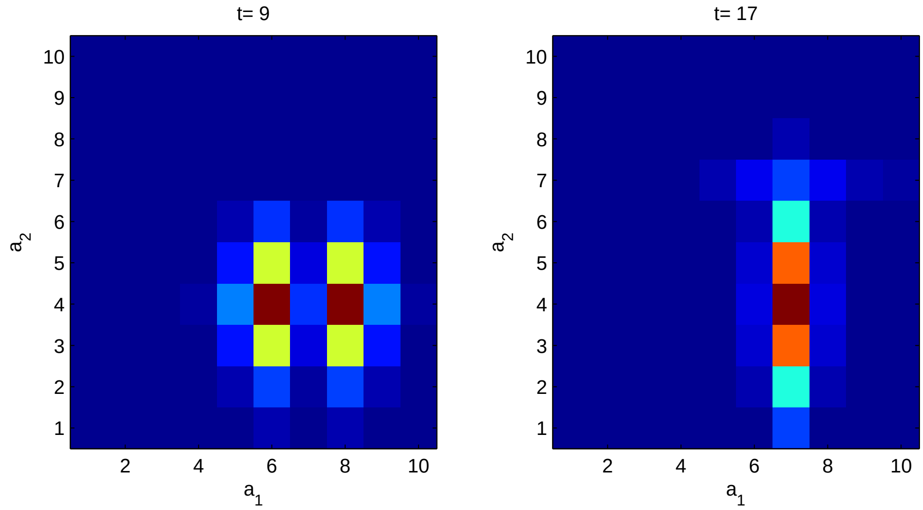
(d) The following plot shows the position of the guards and the sensor output for a particular simulation of the system.



(e) Using the Kalman filter, the minimum probability of error estimates for the positions of the guards for the above simulation are plotted below.



(f) The posterior distributions $q_{t|t-1}$ when $t = 9$ and $t = 17$ are plotted below. Note that since measurement $y(8) = 2$, there is very little probability that $a_1(9) = 7$ or $a_2(9) = 7$. Contrast this with $t = 17$, where $y(16) = 1$, indicating that someone is very likely at position 7 at time $t = 17$. These facts are displayed in the following plots.



(g) The following plot shows the minimum probability of error estimates for the positions of the guards for the same set of measurements. Note that at times $t = 6, 7$, it is equally likely for guard 1 to be at positions 6 or 8. In the plot below, we've estimated guard 1 to be at position 6 at those times.