Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Institute for Dynamic Systems and Control
IDSC
Institut für Dynamische Systeme und Regelungstechnik

Recursive Estimation

Raffaello D'Andrea

Spring 2021

Problem Set 4: Kalman Filtering

Last updated: May 5, 2021

Notes:

- **Notation:** Unless otherwise noted, x, y, and z denote random variables, p_x denotes the probability density function of x, and $p_{x|y}$ denotes the conditional probability density function of x conditioned on y. Note that shorthand (as introduced in Lecture 1 and 2) and longhand notation is used. The expected value of x and its variance is denoted by $\mathrm{E}[x]$ and $\mathrm{Var}[x]$, and $\Pr(Z)$ denotes the probability that the event Z occurs. A normally distributed random variable x with mean μ and variance σ^2 is denoted by $x \sim \mathcal{N}(\mu, \sigma^2)$.
- Please report any errors found in this problem set to the teaching assistants (hofermat@ethz.ch or csferrazza@ethz.ch).

Problem Set

Problem 1

Let $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^n$ be CRVs with a joint Gaussian distribution:

$$\begin{bmatrix} x \\ y \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}, \begin{bmatrix} \Sigma_x & \Sigma_{xy} \\ \Sigma_{xy}^T & \Sigma_y \end{bmatrix}\right).$$

In class, it was shown that a linear combination of two jointly GRVs is a GRV (Property 2). Let $z = x + y$. Calculate the mean μ_z and variance Σ_z of z.

Problem 2

In class, the Kalman filter measurement update equations (Step 2) were derived:

$$\hat{x}_m(k) = \hat{x}_p(k) + P_m(k)H^T(k)R^{-1}(k)\big(\bar{z}(k) - H(k)\hat{x}_p(k)\big) \tag{1}$$

$$P_m(k) = \big(P_p^{-1}(k) + H^T(k)R^{-1}(k)H(k)\big)^{-1}. \tag{2}$$

Furthermore, the alternative measurement update equations were stated:

$$\begin{aligned} K(k) &= P_p(k)\, H^T(k) \left(H(k)P_p(k)H^T(k) + R(k)\right)^{-1} \\ \hat{x}_m(k) &= \hat{x}_p(k) + K(k)\left(\bar{z}(k) - H(k)\hat{x}_p(k)\right) \\ P_m(k) &= \left(I - K(k)H(k)\right) P_p(k) \left(I - K(k)H(k)\right)^T + K(k)\, R(k)\, K^T(k). \end{aligned}$$

Use the matrix inversion lemma (given below) to show that the alternative equations are equivalent to equations (1) and (2).
Matrix inversion lemma: If A, D, and $D^{-1} + CA^{-1}B$ are nonsingular, then $A + BDC$ is nonsingular and

$$(A + BDC)^{-1} = A^{-1} - A^{-1}B(D^{-1} + CA^{-1}B)^{-1}CA^{-1}.$$

Problem 3 (adapted from *D. Simon, Optimal State Estimation, 2006*)

A radioactive particle mass has a half-life of τ seconds. At each time step k, the number of emitted particles $x(k)$ is half of what it was one time step ago, but there is some error $v(k)$ (zero-mean with variance Q) in the number of emitted particles due to background radiation. At each time step, the number of emitted particles is counted. The instrument used to count the number of emitted particles has a random error of $w(k)$ (zero-mean with variance R). Assume that $\{v(k)\}$, $\{w(k)\}$, and $x(0)$ are mutually independent.

a) Write the linear system equations for this system.

b) Design a Kalman filter to estimate the number of emitted particles at each time step.

c) What is the steady-state Kalman gain K when $Q = R$? What is the steady-state Kalman gain when $Q = 2R$? Give an intuitive explanation for why the steady-state gain changes the way it does when the ratio of Q to R changes.

Problem 4 (adapted from *D. Simon, Optimal State Estimation, 2006*)

Suppose that you have a fish tank with x_p piranhas and x_g guppies. Once per week, you put guppy food into the tank (which the piranhas do not eat). During each week, the piranhas eat some of the guppies. The birth rate of the piranhas is proportional to the guppy population, and the death rate of the piranhas is proportional to their own population (due to overcrowding). Therefore, at week k, $x_p(k) = x_p(k-1) + c_1 x_g(k-1) - c_2 x_p(k-1) + v_p(k-1)$, where c_1 and c_2 are proportionality constants and $v_p(k-1)$ is Gaussian distributed noise (zero-mean with variance one) that accounts for mismodeling. The birth rate of the guppies is proportional to the food supply u, and the death rate of the guppies is proportional to the piranha population. Therefore, $x_g(k) = x_g(k-1) + u(k-1) - c_3 x_p(k-1) + v_g(k-1)$, where c_3 is a proportionality constant and $v_g(k-1)$ is Gaussian distributed noise (zero-mean with variance one) that accounts for mismodeling. Every week, you count the piranhas and guppies. You can count the piranhas accurately, because they are large, but your guppy count is corrupted by Gaussian noise (zero-mean with variance one). Assume that $c_1 = 1$ and $c_2 = c_3 = 1/2$, and that the noise $\{v_p(k)\}, \{v_g(k)\}$, and the initial guppy/piranha count are mutually independent.

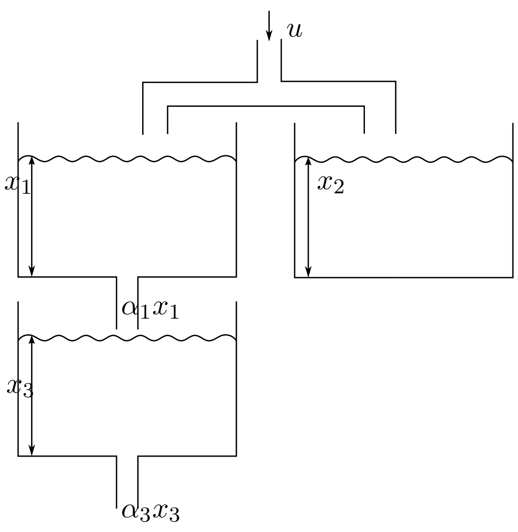
a) Write down a linear state-space model for this system.

b) Suppose that at the initial time you have a perfect count for x_p and x_g. Using a Kalman filter to estimate the guppy population, what is the variance of your guppy population estimate after one week? What is the variance after two weeks?

c) What is the ratio of the expected piranha population to the expected guppy population when they reach steady-state?

Problem 5

Consider the following water tank system:



A simplified model of the discrete-time system dynamics can be written as follows:

$$x(k) = \underbrace{\begin{bmatrix} 1-\alpha_1 & 0 & 0 \\ 0 & 1 & 0 \\ \alpha_1 & 0 & 1-\alpha_3 \end{bmatrix}}_{=:\,A} x(k-1) + \underbrace{\begin{bmatrix} 0.5 \\ 0.5 \\ 0 \end{bmatrix}}_{=:\,B} u(k-1) + v(k-1)$$

where the system state is $x(k) := (x_1(k), x_2(k), x_3(k))$. The control input is $u(k-1) \in \mathbb{R}$. The process noise $v(k-1) \in \mathbb{R}^3$ has the Gaussian distribution: $v(k-1) \sim \mathcal{N}(0, Q)$. The initial state has the Gaussian distribution $x(0) \sim \mathcal{N}(x_0, P_0)$. There are water level sensors installed in two of the three tanks. The measurement equation is

$$z(k) = \underbrace{\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{=:\, H} x(k) + w(k)\,.$$

with measurement noise $w(k) \in \mathbb{R}^2$ with the Gaussian distribution $w(k) \sim \mathcal{N}(0, R)$. The random variables $x(0), \{v(\cdot)\}$, and $\{w(\cdot)\}$ are mutually independent. The numerical values of the parameters above are

$$x_0 = \begin{bmatrix} 5 \\ 5 \\ 5 \end{bmatrix}, \quad P_0 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad Q = \begin{bmatrix} 1/40 & 0 & 0 \\ 0 & 1/10 & 0 \\ 0 & 0 & 1/5 \end{bmatrix}, \quad R = \begin{bmatrix} 1/2 & 0 \\ 0 & 1/2 \end{bmatrix}, \quad \begin{array}{l} \alpha_1 = 0.1, \\ \alpha_3 = 0.2. \end{array}$$

All the parameters and the system dynamics are constant in time. Therefore, this system is a linear time-invariant (LTI) system.

a) Determine if the variance of a time-varying Kalman Filter (KF) for this system converges: Check the conditions for the existence of a unique positive-semidefinite (PSD) steady-state prior variance P_∞.

b) Compute P_∞ by solving the Discrete Algebraic Riccati Equation (DARE) (you can use the MATLAB command `dare`).

c) Implement a time-varying KF for tracking the tank levels $x(k)$, for example in MATLAB. Further implement a simulation of the system dynamics that you can use to generate measurements that serve as input to your KF. Set the control input to $u(k) = 5$ for all k. Simulate for $k = 1, 2, \ldots, 1000$, i.e. for a long time, and answer the following questions:

1. What value does the prior variance $P_p(k)$ of the KF converge to? Compare this value to P_∞ obtained by solving the DARE. Plot the posterior mean $x_m(k)$ of the KF for large k. Discuss whether $x_m(k)$ converges to a steady-state value, analogous to the KF variance.
2. Inspect the structure of P_∞ (which entries are different from zero?). Explain the structure of P_∞ based on the physics of the problem.

   Based on the insight about the structure of the KF variance, can you think of an implementation of the KF that is computationally less expensive?
3. Initialize the Kalman Filter with different $P_m(0) \neq P_0$ and run the simulation. What values do the variances attain for large k?

d) Now let the control input be $u(k) = 5\,|\sin(k)|$. What changes in the estimator mean $\hat{x}_m(k)$? What changes in the prior variance $P_p(k)$?

For the remaining parts of the problem, use $u(k) = 5$ for all k again.

e) Consider the modified measurement equations that result from moving the water level sensor from the second to the first tank:

$$z(k) = \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{=:\, H} x(k) + w(k)\,.$$

1. For the modified system, check the conditions for the existence of a unique PSD steady-state prior variance P_∞. Explain the result based on the modified measurement equation and the system dynamics.
2. Modify the simulation and the time-varying KF accordingly, and simulate until $k = 1000$ again. What is the prior variance $P_p(k)$ for large k now? Explain your findings.

f) Let the measurement equation be identical to part e), but now consider the modified system dynamics where an additional pipe connects the second to the third tank. The modified system dynamics are

$$x(k) = \underbrace{\begin{bmatrix} 1-\alpha_1 & 0 & 0 \\ 0 & 1-\alpha_2 & 0 \\ \alpha_1 & \alpha_2 & 1-\alpha_3 \end{bmatrix}}_{=:A} x(k-1) + \underbrace{\begin{bmatrix} 0.5 \\ 0.5 \\ 0 \end{bmatrix}}_{=:B} u(k-1) + v(k-1)$$

with $\alpha_2 = 0.5$.

Modify the simulation and the time-varying KF accordingly, and simulate until $k = 1000$ again. Explain how the modified dynamics affect the prior variance $P_p(k)$ for large k. Reason about the difference to the result from part e).

g) For this last part, the flow through the additional pipe between the second and the third tank is controlled by a valve which opens and closes periodically. This is modeled by replacing α_2 in the dynamics defined in part f) by the time-varying

$$\alpha_2(k) = \begin{cases} 0.5 & \text{for } k = 0, 3, 6, 9, \dots \\ 0 & \text{else.} \end{cases}$$

With this modification, the system is not time-invariant anymore. Modify the simulation and the time-varying KF accordingly and simulate until $k = 1000$ again. Observe the impact of the time-varying $\alpha_2(k)$ on the variance. Does the variance converge?

Problem 6

The Balancing Cube (see Figure 1) is a dynamic sculpture that can balance autonomously on any of its corners through the action of six rotating arms on its inner faces. The control system that stabilizes the cube about an equilibrium configuration consists of a Kalman Filter (KF) that is combined with a state-feedback controller. In this problem, you shall design and simulate a KF for the Balancing Cube. In the last problem set of this class, you will combine this KF with a suitable state-feedback controller and simulate the full control system.

The system dynamics about the equilibrium configuration shown in Figure 1 (the cube stands upright on one of its corners and all arms point downward) can be approximated by the linear model

$$x(k) = Ax(k-1) + Bu(k-1) + v(k-1), \tag{3}$$

where $x(k) \in \mathbb{R}^{16}$ is the state vector, $u(k-1) \in \mathbb{R}^6$ are the system inputs (torques applied to the rotating arms), and $v(k-1) \in \mathbb{R}^{16}$ is process noise (capturing model uncertainty such as linearization errors) with zero mean and variance Q. The system state is given by the angles of the six arms $(x_1, x_3, \dots, x_{11})$, the angular velocities of the arms $(x_2, x_4, \dots, x_{12})$, the cube body roll and pitch angles (x_{13}, x_{15}), and the roll and pitch angular rates (x_{14}, x_{16}).[1] The initial state

[1] The multi-body system's additional degree of freedom corresponding to rotation about the gravity axis (yaw) is neglected in this model since it is irrelevant for balancing.

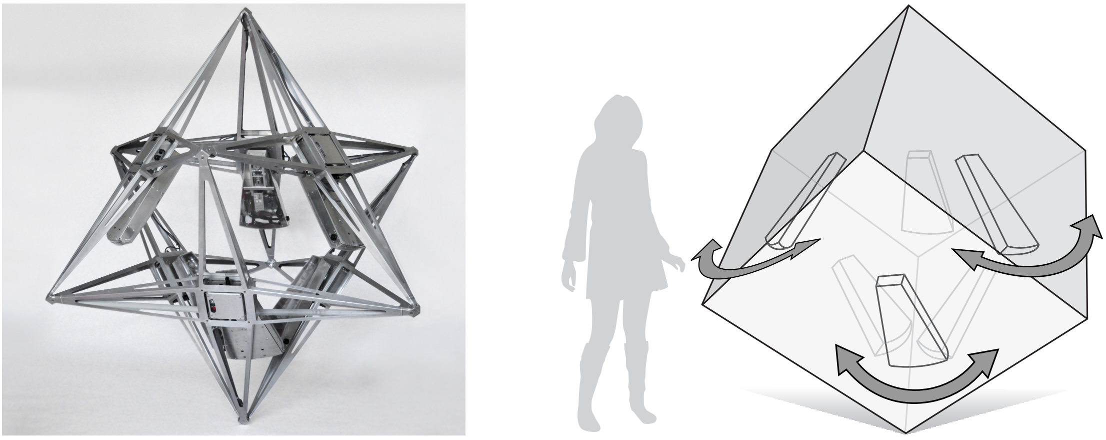


Figure 1: LEFT: The photo shows the Balancing Cube as it is balancing on one of its corners. RIGHT: The abstraction visualizes the cube body and the six rotating arms on its inner faces. The cube was built at IDSC, where it is used as a test bed for research in state estimation and control. (See `www.cube.ethz.ch` *for a video and more information on how the system works.)*

of the system is not known exactly. We model this imperfect knowledge as the initial state $x(0)$ having zero mean and variance $\text{Var}[x(0)] = 3 \cdot 10^{-4} I$ (I is the identity matrix). For the purpose of this problem, we consider the uncontrolled system with $u(k) = 0$ for all k.

Each arm is equipped with an angle encoder, which measures the arm's angle, and with a rate gyro sensor, from which the pitch and roll angular rates of the cube body can be inferred.[2] That is, the measurements by the sensors on arm i are ($i = 1, 2, \ldots, 6$):

$$z_i(k) = \begin{bmatrix} x_{2i-1}(k) \\ x_{14}(k) \\ x_{16}(k) \end{bmatrix} + \begin{bmatrix} w_{\text{encoder}}(k) \\ w_{\text{gyro}}(k) \\ w_{\text{gyro}}(k) \end{bmatrix},$$

where $w_{\text{encoder}}(k)$ and $w_{\text{gyro}}(k)$ is zero-mean sensor noise. The sensor noise variance[3] on the pitch and roll angular rate measurements is $\text{Var}[w_{\text{gyro}}(k)] = 2 \cdot 10^{-5}$. The arm angle sensors involve a small quantization error that stems from a finite encoder resolution; we approximately account for this error by assuming encoder noise with variance $\text{Var}[w_{\text{encoder}}(k)] = 1 \cdot 10^{-6}$. All measurements are combined in the vector $\bar{z}(k) \in \mathbb{R}^{18}$:

$$z(k) = \begin{bmatrix} z_1^T(k) & z_2^T(k) & \ldots & z_6^T(k) \end{bmatrix}^T = Hx(k) + w(k). \tag{4}$$

Numerical values for the matrices A, B, and H are available on the class webpage (MATLAB file `CubeModel.mat`).

a) Is the system stable? Is the system observable?

b) Assume $Q = 10^{-6} I$. Implement a time-varying Kalman Filter that estimates the state $x(k)$ of the Balancing Cube based on the model (3), (4) and measurements $\bar{z}(k)$. In addition, design and implement the corresponding steady-state Kalman Filter to estimate the state. Simulate the system as well as both estimators (in parallel, i.e. using the same initial mean and variance, and using the same measurements as input) and plot the estimation error for both estimators.

[2] In reality, the rate gyro sensors measure the angular rotation rate of the cube body in the body frame of reference. The angular rates in the body frame relate to the pitch and roll rate through a nonlinear transformation, which is applied to a rate gyro measurement before using it in the Kalman Filter. For the sake of simplicity within this problem, we neglect this transformation.

[3] Obtained from specifications of the rate gyro sensors.

c) Try different values for Q. How do the state estimates change qualitatively? In particular, what happens to the convergence rate of the steady-state Kalman Filter and the squared estimation error?

Problem 7 (adapted from Problem 4 of the final exam 2012)

Consider the deterministic, nonlinear, discrete-time process model

$$x(k) = x^2(k-1) \tag{5}$$

where $x(k)$ is the scalar system state. The initial state $x(0)$ is uniformly distributed on the interval $[0, 1]$.

a) Initialize an Extended Kalman Filter (EKF) for tracking the state $x(k)$ with the appropriate values for the mean $\hat{x}_m(0)$ and variance $P_m(0)$. Compute the prior mean $\hat{x}_p(1)$ with the EKF equations.

b) Calculate $\mathrm{E}[x(1)]$ from the given distribution of $x(0)$ and the process model (5). What underlying assumption in the EKF prior mean update results in $\mathrm{E}[x(1)] \neq \hat{x}_p(1)$?

For the remainder of the problem, let the process model be

$$x(k) = x^2(k-1) + v^2(k-1) \tag{6}$$

where $v(k-1)$ is process noise that is uniformly distributed on the interval $[-1, 1]$. Like in a), the initial state $x(0)$ is uniformly distributed on $[0, 1]$. The noise $\{v(\cdot)\}$ and the initial state $x(0)$ are mutually independent.

c) Calculate the prior mean $\hat{x}_p(1)$ and variance $P_p(1)$ with the EKF equations.

d) Calculate the expected value $\mathrm{E}[x(1)]$ and variance $\mathrm{Var}[x(1)]$ from the model (6) and the given distributions of $x(0)$ and $v(0)$.

For tracking the process (6), your colleague suggests implementing the modified model

$$y(k) = y^2(k-1) + \alpha + \eta(k-1) \tag{7}$$

in your EKF, where $y(k)$ is the scalar state, α is a constant bias, and $\eta(k)$ is process noise which is uniformly distributed on $[-\beta, \beta]$. The probability density function of $y(0)$ is identical to $p(x(0))$, i.e. uniform on $[0, 1]$. Furthermore, the noise $\{\eta(\cdot)\}$ and the initial state $y(0)$ are mutually independent.

e) Calculate α, β such that $\mathrm{E}[\alpha + \eta(k)] = \mathrm{E}\left[v^2(k)\right]$ and $\mathrm{Var}[\alpha + \eta(k)] = \mathrm{Var}\left[v^2(k)\right]$.

f) Initialize an EKF for tracking $y(k)$ with the appropriate values for the mean $\hat{y}_m(0)$ and the variance $\widetilde{P}_m(0)$. Then, calculate the prior mean $\hat{y}_p(1)$ and the prior variance $\widetilde{P}_p(1)$ with the EKF equations.

g) Show that $\mathrm{E}[y(1)] = \mathrm{E}[x(1)]$ and $\mathrm{Var}[y(1)] = \mathrm{Var}[x(1)]$. Does this imply that $p_{y(1)}(\xi) = p_{x(1)}(\xi)$ for all ξ?

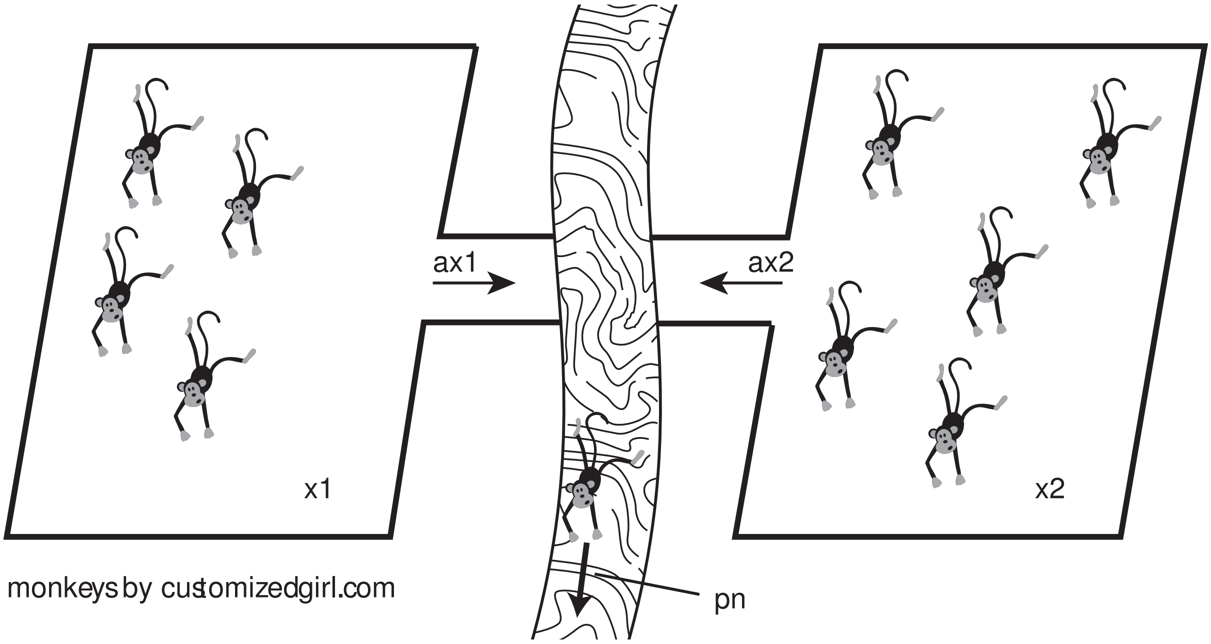
h) Show that the expected squared prediction error for $x(1)$ is strictly smaller for the EKF that uses the adapted model (7): $\mathrm{E}\left[(x(1) - \hat{y}_p(1))^2\right] < \mathrm{E}\left[(x(1) - \hat{x}_p(1))^2\right]$. Can you intuitively say why this inequality holds?

Problem 8

The logistic map is a nonlinear model for population dynamics in biology. The discrete-time dynamics are

$$x(k) = l(x(k-1)) := \mu\, x(k-1)(1 - x(k-1)) \tag{8}$$

where the state $x(k) \in [0, 1]$ captures the population at time k (think, for example, year k). The parameter $\mu \in [0, 4]$ captures the combined growth and shrinkage rate of the population. Small populations quickly grow (lots of food available per individual), while larger populations quickly shrink (food is scarce). Despite its simple appearance, the logistic map can exhibit complex behavior, for example chaos, depending on the choice of μ (read more about this on Wikipedia). In this problem, you implement an Extended Kalman Filter (EKF) for a dynamic system that consists of two coupled logistic maps (inspired by the paper: *"The Coupled Logistic Map: A Simple Model for the Effects of Spatial Heterogeneity on Population Dynamics"* by Lloyd, 1995). The idea of the dynamical system is illustrated below:



Two populations of monkeys live in fenced-off areas connected by a passage where the monkeys can reach the other area by swimming across a river. The monkey populations at the beginning of year k are $x(k)$, $y(k) \in [0, 1]$, and the growth/shrinkage during a year is governed by the logistic map (8). Just before the end of each year, when the populations grew or shrank to $l(x(k))$ and $l(y(k))$, a fraction of monkeys from each area attempts to cross the river. These fractions are $\alpha l(x(k))$ and $\alpha l(y(k))$, where $\alpha \in [0, 1]$ is a constant. Not all monkeys are successful in their endeavor: Random parts of the swimming monkeys are swept away by the river. The monkeys carried away are $v_x \alpha l(x(k))$ and $v_y \alpha l(y(k))$ where the noise signals $v_x(k)$ and $v_y(k)$ are drawn from uniform distributions on the interval $[0, a]$, where $a \in [0, 1]$ is a constant. The overall population dynamics are then given by

$$x(k) = (1 - \alpha)\, l(x(k-1)) + (1 - v_y(k-1))\, \alpha\, l(y(k-1))$$
$$y(k) = (1 - \alpha)\, l(y(k-1)) + (1 - v_x(k-1))\, \alpha\, l(x(k-1))$$

The logistic map parameter μ is constant. At time $k = 0$, two random populations $x(0)$ and $y(0)$ are set free in the two areas. Both $x(0)$ and $y(0)$ are drawn from a uniform distribution on $[0, 1]$. At the beginning of every year $k > 0$, undergraduate biology students are sent to the two areas to assess the monkey populations. Since monkeys are quite good at hiding, the students randomly underestimate the true populations. The population reports are therefore modeled as

$$z_x(k) = x(k)(1 - w_x(k))$$
$$z_y(k) = y(k)(1 - w_y(k))$$

where the noise signals $w_x(k), w_y(k)$ are drawn from uniform distributions on the interval $[0, b]$, where $b \in [0, 1]$ is a constant. The random variables $x(0), y(0), \{v_x(\cdot)\}, \{v_y(\cdot)\}, \{w_x(\cdot)\}, \{w_y(\cdot)\}$ are mutually independent.

a) Implement an EKF for tracking the monkey populations $x(k), y(k)$, for example in MATLAB. Further implement a simulation of the system dynamics that you can use to generate measurements that serve as input to your EKF. The numerical values of the constant parameters are $\mu = 3.9, \alpha = 0.07, a = 0.6$, and $b = 0.15$.

b) For this part, the constant parameter α is unknown. However, previous monkey studies found that the parameter is uniformly distributed on the interval $[0, 0.1]$. Adapt the simulation accordingly and extend the EKF from part a) to estimate the parameter α. The random variables $\alpha, x(0), y(0), \{v_x(\cdot)\}, \{v_y(\cdot)\}, \{w_x(\cdot)\}, \{w_y(\cdot)\}$ are mutually independent.

Additional Problems

The following problems are optional and we do not assume that you know these problem's specific results by heart. However, the problems let you further practice the concepts discussed in class.

Problem 9

Prove the *matrix inversion lemma* used in problem 2: If A, D, and $D^{-1}+CA^{-1}B$ are nonsingular, then $A + BDC$ is nonsingular and

$$(A + BDC)^{-1} = A^{-1} - A^{-1}B(D^{-1} + CA^{-1}B)^{-1}CA^{-1}.$$

Sample solutions

Problem 1

We define the GRV $c \in \mathbb{R}^{2n}$ as the concatenation

$$c := \begin{bmatrix} x \\ y \end{bmatrix}, \quad c \sim \mathcal{N}\left(\mu_c, \Sigma_c\right), \text{ with } \mu_c = \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}, \text{ and } \Sigma_c = \begin{bmatrix} \Sigma_x & \Sigma_{xy} \\ \Sigma_{xy}^T & \Sigma_y \end{bmatrix}.$$

Then, in matrix form

$$z = x + y = M \begin{bmatrix} x \\ y \end{bmatrix} = Mc$$

with $M = \begin{bmatrix} I & I \end{bmatrix}$ where $I \in \mathbb{R}^n$ is the identity matrix. The mean μ_z can be calculated by

$$\mu_z = \mathrm{E}\left[Mc\right] = M\mathrm{E}\left[c\right] = M\mu_c = M \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix} = \begin{bmatrix} I & I \end{bmatrix} \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix} = \mu_x + \mu_y.$$

For the variance, we find

$$\begin{aligned}
\mathrm{Var}\left[z\right] &= \mathrm{E}\left[(z-\mu_z)(z-\mu_z)^T\right] = \mathrm{E}\left[(Mc - M\mu_c)(Mc - M\mu_c)^T\right] \\
&= \mathrm{E}\left[M(c-\mu_c)(c-\mu_c)^T M^T\right] = M\mathrm{E}\left[(c-\mu_c)(c-\mu_c)^T\right] M^T = M\mathrm{Var}\left[c\right] M^T \\
&= M\Sigma_c M^T = M \begin{bmatrix} \Sigma_x & \Sigma_{xy} \\ \Sigma_{xy}^T & \Sigma_y \end{bmatrix} M^T = \Sigma_x + \Sigma_y + \Sigma_{xy} + \Sigma_{xy}^T.
\end{aligned}$$

Problem 2

To simplify notation we drop all "k" arguments, and define $\bar{P} := P_m(k)$ and $P := P_p(k)$.
To show that the update equations for $\hat{x}_m(k)$ are the same, it suffices to show that $K = \bar{P}H^T R^{-1}$.
Using (2) and the matrix inversion lemma

$$\begin{aligned}
\bar{P}H^T R^{-1} &= \left(P^{-1} + H^T R^{-1} H\right)^{-1} H^T R^{-1} \\
&= \left(P - PH^T \left(R + HPH^T\right)^{-1} HP\right) H^T R^{-1} \quad \text{(by matrix inversion lemma)} \\
&= PH^T \left(I - \left(R + HPH^T\right)^{-1} HPH^T\right) R^{-1} \\
&= PH^T \left(R + HPH^T\right)^{-1} \left(R + HPH^T - HPH^T\right) R^{-1} \\
&= PH^T \left(R + HPH^T\right)^{-1} \\
&= K \quad \text{(by definition of } K\text{)}
\end{aligned}$$

as required.
The update equation for $\bar{P}$ is then

$$\begin{aligned}
\bar{P} &= \left(P^{-1} + H^T R^{-1} H\right)^{-1} \\
&= P - PH^T \left(R + HPH^T\right)^{-1} HP \quad \text{(by matrix inversion lemma)} \\
&= P - KHP \quad \text{(by definition of } K\text{)} \\
&= \left(I - KH\right) P \\
&= \left(I - KH\right) P - \left(I - KH\right) PH^T K^T + \left(I - KH\right) PH^T K^T \\
&= \left(I - KH\right) P \left(I - KH\right)^T + \left(PH^T - KHPH^T\right) K^T \\
&= \left(I - KH\right) P \left(I - KH\right)^T + \left(PH^T - K\left(HPH^T + R\right) + KR\right) K^T \\
&= \left(I - KH\right) P \left(I - KH\right)^T + \left(PH^T - PH^T + KR\right) K^T \quad \text{(by definition of } K\text{)} \\
&= \left(I - KH\right) P \left(I - KH\right)^T + KRK^T.
\end{aligned}$$

Problem 3

a) System equations:

$$x(k) = \frac{1}{2}x(k\text{–}1) + v(k\text{–}1) \qquad \mathrm{E}\left[v(k)\right] = 0,\ \mathrm{Var}\left[v(k)\right] = Q$$
$$z(k) = x(k) + w(k) \qquad \mathrm{E}\left[w(k)\right] = 0,\ \mathrm{Var}\left[w(k)\right] = R$$

b) Kalman filter:

Step 1:

$$\begin{aligned}\hat{x}_p(k) &= \frac{1}{2}\hat{x}_m(k\text{–}1)\\ P_p(k) &= \frac{1}{4}P_m(k\text{–}1) + Q\end{aligned}$$

Step 2:

$$\begin{aligned}K(k) &= P_p(k)\left(P_p(k) + R\right)^{-1} = \frac{P_p(k)}{P_p(k) + R}\\ \hat{x}_m(k) &= \hat{x}_p(k) + K(k)\left(\bar{z}(k) - \hat{x}_p(k)\right)\\ P_m(k) &= \left(1 - K(k)\right)P_p(k)\left(1 - K(k)\right) + K(k)RK(k)\\ &= \left(1 - K(k)\right)^2 P_p(k) + K^2(k)R\end{aligned}$$

Initialization:

$$\hat{x}_m(0) = x_0, \qquad P_m(0) = P_0 \quad \text{(initial data not given in this problem)}$$

c) We obtain the steady-state estimation variance P_∞ from the Discrete Algebraic Riccati Equation:

$$\begin{aligned}P_\infty &= AP_\infty A^T - AP_\infty H^T\left(HP_\infty H^T + R\right)^{-1}HP_\infty A^T + Q\\ &= \frac{1}{4}P_\infty - \frac{1}{4}P_\infty^2\left(P_\infty + R\right)^{-1} + Q\\ \Leftrightarrow \quad & \frac{3}{4}P_\infty^2 + \frac{3}{4}P_\infty R = QP_\infty + RQ - \frac{1}{4}P_\infty^2\\ \Leftrightarrow \quad & P_\infty^2 + \left(\frac{3}{4}R - Q\right)P_\infty - RQ = 0\\ \Leftrightarrow \quad & P_\infty = -\frac{1}{2}\left(\frac{3}{4}R - Q\right) \pm \frac{1}{2}\sqrt{\left(\frac{3}{4}R - Q\right)^2 + 4RQ}\end{aligned}$$

We analyze the solution for the given cases.

- $Q = R$:

$$\begin{aligned}P_\infty &= \left(\frac{1}{8} \pm \frac{1}{2}\sqrt{\left(-\frac{1}{4}\right)^2 + 4}\right)R\\ &\approx 1.133\,R \quad \text{or} \quad -0.883\,R\end{aligned}$$

  Since only the positive solution makes sense (P_∞ and R are variances), we have

$$K = \frac{P_\infty}{P_\infty + R} \approx \frac{1.133}{2.133} \approx 0.53.$$

- $Q = 2R$:

$$
\begin{aligned}
P_\infty &= \left(\frac{5}{8} \pm \frac{1}{2} \sqrt{\left(-\frac{5}{4}\right)^2 + 8} \right) R \\
&\approx 2.171\, R \quad \text{(only positive solution considered)} \\
K &\approx 0.68.
\end{aligned}
$$

- In general, as the process noise increases relative to the measurement noise, the gain K increases, i.e. the filter puts more emphasis on the measurements rather than the model-based predictions.

Problem 4

a) A linear time-invariant model for this problem is

$$
\begin{aligned}
x(k) := \begin{bmatrix} x_p(k) \\ x_g(k) \end{bmatrix} &= \underbrace{\begin{bmatrix} 1 - c_2 & c_1 \\ -c_3 & 1 \end{bmatrix}}_{=:A} \begin{bmatrix} x_p(k{-}1) \\ x_g(k{-}1) \end{bmatrix} + \underbrace{\begin{bmatrix} 0 \\ 1 \end{bmatrix}}_{=:B} u(k{-}1) + v(k{-}1) \\
z(k) &= x(k) + w(k)
\end{aligned}
$$

with

$$
\begin{aligned}
\mathrm{E}\left[v(k)\right] = 0,\ \mathrm{Var}\left[v(k)\right] = \mathrm{E}\left[v(k)v^T(k)\right] &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} =: Q \\
\mathrm{E}\left[w(k)\right] = 0,\ \mathrm{Var}\left[w(k)\right] = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} &=: R.
\end{aligned}
$$

With the given values $c_1 = 1,\ c_2 = c_3 = 1/2$, we get

$$
\begin{aligned}
x(k) &= \begin{bmatrix} 0.5 & 1 \\ -0.5 & 1 \end{bmatrix} x(k{-}1) + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u(k{-}1) + v(k{-}1) \\
z(k) &= x(k) + w(k).
\end{aligned}
$$

b) The Kalman filter is initialized with $P_m(0) = 0$. The variance after one week is $P_m(1)$, which we calculate with the Kalman filter equations:

$$
\begin{aligned}
&\text{Step 1:} && P_p(1) = A P_m(0) A^T + Q = Q = I \\
&\text{Step 2:} && K(1) = I\,(I + R)^{-1} \\
& && \quad = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{2} \end{bmatrix} \\
& && P_m(1) = (I - K(1)I)\, P_p(1)\, (I - K(1)I)^T + K(1) R K^T(1) \\
& && \quad = \begin{bmatrix} 0 & 0 \\ 0 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & \frac{1}{2} \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{2} \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & \frac{1}{2} \end{bmatrix}
\end{aligned}
$$

Therefore the variance of the guppy population is $\frac{1}{2}$ after one week.

The variance after two weeks is given by $P_m(2)$, which is computed as follows:

$$\text{Step 1:} \qquad P_p(2) = \begin{bmatrix} 0.5 & 1 \\ -0.5 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 0.5 & -0.5 \\ 1 & 1 \end{bmatrix} + I$$

$$= \begin{bmatrix} 0 & \frac{1}{2} \\ 0 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 0.5 & -0.5 \\ 1 & 1 \end{bmatrix} + I = \begin{bmatrix} 1.5 & 0.5 \\ 0.5 & 1.5 \end{bmatrix}$$

$$\text{Step 2:} \qquad K(2) = \frac{1}{2} \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} \left(\frac{1}{2} \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right)^{-1} = \begin{bmatrix} 1 & 0 \\ \frac{1}{7} & \frac{4}{7} \end{bmatrix}$$

$$P_m(2) = \begin{bmatrix} 0 & 0 \\ -\frac{1}{7} & \frac{3}{7} \end{bmatrix} \cdot \frac{1}{2} \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} 0 & -\frac{1}{7} \\ 0 & \frac{3}{7} \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ \frac{1}{7} & \frac{4}{7} \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & \frac{1}{7} \\ 0 & \frac{4}{7} \end{bmatrix}$$

$$= \begin{bmatrix} 0 & 0 \\ 0 & \frac{4}{7} \end{bmatrix}.$$

The variance of the guppy population after two weeks is $\frac{4}{7}$.

c) At steady-state, we have $\mathrm{E}[x(k)] = \mathrm{E}[x(k-1)] =: \bar{x}$, so that

$$\bar{x} = \mathrm{E}[x(k)] = A\,\mathrm{E}[x(k-1)] + Bu(k-1) = A\bar{x} + Bu$$

$$\bar{x} = (I - A)^{-1} Bu = \left(\begin{bmatrix} 0.5 & -1 \\ 0.5 & 0 \end{bmatrix} \right)^{-1} \begin{bmatrix} 0 \\ 1 \end{bmatrix} u = \begin{bmatrix} 2 \\ 1 \end{bmatrix} u.$$

The ratio between piranhas and guppies is

$$\frac{\bar{x}_p}{\bar{x}_g} = \frac{2}{1}.$$

Problem 5

A MATLAB script implementing all simulation parts of the problem is available on the class website.

a) The conditions for the existence of a unique positive-semidefinite (PSD) P_∞ were discussed in Lecture 8, *The Kalman Filter as State Observer*, and are

1. The pair (A, H) is detectable and
2. The pair (A, G) is stabilizable, where G is any G such that $Q = GG^T$.

The given Q is positive definite, $Q > 0$, from which directly follows that condition 2 is fulfilled (you may also check this formally by showing that the controllability matrix from A, G has full column rank, from which follows that (A, G) is reachable, and therefore also stabilizable).

For the first condition, we calculate the observability matrix

$$\begin{bmatrix} H \\ HA \\ HA^2 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \\ \alpha_1 & 0 & 1-\alpha_3 \\ 0 & 1 & 0 \\ \alpha_1(2-\alpha_1-\alpha_3) & 0 & (1-\alpha_3)^2 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0.1 & 0 & 0.8 \\ 0 & 1 & 0 \\ 0.17 & 0 & 0.64 \end{bmatrix}$$

which has full column rank. The system is therefore observable, which implies detectability. The conditions for the existence of a unique PSD steady-state prior variance are satisfied.

b) The MATLAB command `dare(A',H',Q,R)` yields

$$P_\infty = \begin{bmatrix} 0.1294 & 0.0000 & 0.0205 \\ 0.0000 & 0.2791 & -0.0000 \\ 0.0205 & -0.0000 & 0.3306 \end{bmatrix}.$$

c)

1. The prior variance converges to P_∞ that was obtained by solving the the DARE in part b). The posterior mean $x_m(k)$ does not converge: The means of $x_1(k)$ and $x_3(k)$ converge, but the mean of $x_2(k)$ does not since its dynamics are unstable. Note that even though the mean of $x_2(k)$ does not converge, its variance does converge.

2. The steady-state variance matrix has the following structure:

$$P_\infty = \begin{bmatrix} * & 0 & * \\ 0 & * & 0 \\ * & 0 & * \end{bmatrix}$$

   where $*$ denotes a nonzero entry. The covariance of x_2 and x_1 as well as of x_2 and x_3 is zero, i.e. the random variable x_2 is independent of x_1, x_3. This is because: 1) the dynamics of the second tank and the other two tanks are decoupled, except for the known control input; 2) the measurements are decouples as well; and 3) the process noise and the measurement noise variance matrices are diagonal. Specifically, the system can be rewritten as two separate systems:

$$\begin{aligned} \begin{bmatrix} x_1(k) \\ x_3(k) \end{bmatrix} &= \begin{bmatrix} 1-\alpha_1 & 0 \\ \alpha_1 & 1-\alpha_3 \end{bmatrix} \begin{bmatrix} x_1(k-1) \\ x_3(k-1) \end{bmatrix} + \begin{bmatrix} 0.5 \\ 0 \end{bmatrix} u(k-1) + v_{13}(k-1) \\ z_3(k) &= x_3(k) + w_3(k) \end{aligned}$$

   with

$$v_{13}(k) \sim \mathcal{N}\left(0, \begin{bmatrix} 1/40 & 0 \\ 0 & 1/5 \end{bmatrix}\right), \quad w_3(k) \sim \mathcal{N}(0, 1/2)$$

   and

$$\begin{aligned} x_2(k) &= x_2(k-1) + 0.5u(k-1) + v_2(k-1) \\ z_2(k) &= x_2(k) + w_2(k) \end{aligned}$$

   with

$$v_2(k) \sim \mathcal{N}(0, 1/10), \quad w_2(k) \sim \mathcal{N}(0, 1/2).$$

   Therefore, the two systems can also be tracked with two separate KFs.

3. For any PSD $P_m(0) \neq P_0$ that we choose, we find that the KF prior variance still converges to P_∞ for large k. This is implied by (A, H) being detectable and (A, G) being stabilizable as shown in part a). That is, for any PSD $P_m(0)$, $P_p(k) \to P_\infty$ for $k \to \infty$. However, since $P_m(0) \neq P_0$, the distributions captured by the mean and variance of the KF do not represent the actual state distributions anymore, and optimality in the MMSE-sense is lost as well.

d) The mean $\hat{x}_m(k)$ oscillates. The variance is unaffected because it does not depend on the control input: The control input is known to the KF (i.e. deterministic) and, therefore, it does not affect the variance of the estimate.

e) 1. The process noise condition, condition 2 in part a) is still fulfilled. We recompute the observability matrix:

$$\begin{bmatrix} H \\ HA \\ HA^2 \end{bmatrix} = \begin{bmatrix} 1.0 & 0 & 0 \\ 0 & 0 & 1.0 \\ 0.9 & 0 & 0 \\ 0.1 & 0 & 0.8 \\ 0.81 & 0 & 0 \\ 0.17 & 0 & 0.64 \end{bmatrix}$$

Note that the second column contains only zeros, and, therefore, the observability matrix does not have full column rank and the system is not observable. Specifically, the state x_2 is unobservable and not detectable because: 1) There is no measurement of x_2; 2) the dynamics of x_2 are not stable (if they were, the system would be unobservable but detectable); and 3) the dynamics of x_2 are decoupled from x_1, x_3. To formally verify this, we apply the Popov-Belevitch-Hautus (PBH) test. The eigenvalues of the matrix A are $\lambda_1 = 1 - \alpha_1, \lambda_2 = 1, \lambda_3 = 1 - \alpha_3$. The matrix

$$\begin{bmatrix} A - \lambda I \\ H \end{bmatrix} = \begin{bmatrix} 1 - \alpha_1 - \lambda & 0 & 0 \\ 0 & 1 - \lambda & 0 \\ \alpha_1 & 0 & 1 - \alpha_3 - \lambda \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

is not full column rank for $\lambda_2 = 1$, and the system is therefore not detectable.

2. The effect of the system not being detectable can also be observed in the simulation, where the entry P_{22} of the variance matrix tends to infinity.

f) The added pipe causes the variance of x_2 to converge.The conditions for the existence of a unique PSD steady-state prior variance are now satisfied. There are two reasons for this, each one would suffice on its own: Firstly, the system dynamics with the added pipe are now stable (all eigenvalues of A are of magnitude less than one). Secondly, the observability matrix now has full rank:

$$\begin{bmatrix} H \\ HA \\ HA^2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0.90 & 0 & 0 \\ 0.10 & 0.5 & 0.8 \\ 0.81 & 0 & 0 \\ 0.17 & 0.65 & 0.64 \end{bmatrix}.$$

g) Due to the time-varying system dynamics, the variance does not converge to a constant value, but to a periodic solution. The variance of x_2 increases whenever the valve is closed, as the fill level of tank 2 cannot be determined during these phases. It is then reduced in the time step where $\alpha_2(k) \neq 0$.

Problem 6

a) The system is unstable. In particular, it has two poles with magnitude greater than one (both at 1.0290). The two unstable poles correspond to the two unstable degrees of freedom of the cube body when standing on a corner (pitch and roll).

The observability matrix has full rank. Hence, the system is observable.

The corresponding MATLAB computations can be found in the MATLAB file accompanying this problem.

b) See the accompanying MATLAB file for the implementation of both KFs.

Notice that the steady-state KF estimates converge to the estimates of the time-varying KF; the filters basically differ only in their transient behavior. Hence, one does not lose much performance with the steady-state implementation if one mostly cares about the long-term performance of the estimator.

c) Generally speaking, by increasing the process noise, the KF trusts the process model less. This tends to increase the convergence rate of the steady-state KF, but also tends to amplify the noise in the estimates.

In the accompanying MATLAB file, we consider $Q = 10^{-3}I$, $Q = 10^{-6}I$, and $Q = 10^{-9}I$ for the process noise variance to illustrate this behavior. In particular, we compute the poles of the error dynamics (with greater process noise variance, (some) poles become faster) and the squared estimation error (tends to increase with more process noise).

Problem 7

a) We initialize the mean and variance according to the specified PDF of $x(0)$:

$$\begin{aligned}
\hat{x}_m(0) &= \mathrm{E}\left[x(0)\right] = \frac{1}{2} \\
P_m(0) &= \mathrm{Var}\left[x(0)\right] = \mathrm{E}\left[x^2(0)\right] - \mathrm{E}\left[x(0)\right]^2 \\
&= \int_{-\infty}^{\infty} x^2(0) f(x(0)) dx(0) - \frac{1}{4} = \int_0^1 x^2(0) dx(0) - \frac{1}{4} = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}.
\end{aligned}$$

The prior mean is

$$\hat{x}_p(k) = \hat{x}_m^2(k\text{-}1) \quad \Rightarrow \quad \hat{x}_p(1) = \hat{x}_m^2(0) = \frac{1}{4}.$$

b) The expected value is

$$\mathrm{E}\left[x(1)\right] = \mathrm{E}\left[x^2(0)\right] = \int_0^1 x^2(0) dx(0) = \frac{1}{3} \neq \hat{x}_p(1).$$

The underlying assumption that we make when calculating the prior mean update in the EKF is that the nonlinear function that describes the process model commutes with the expected value. Let, for example, $x(k) = q(x(k\text{-}1), v(k\text{-}1))$ be this nonlinear function. In the EKF prior mean update, we assume that $\mathrm{E}\left[x(k)\right] = \mathrm{E}\left[q(x(k\text{-}1), v(k\text{-}1))\right]$ is approximately $q(\mathrm{E}\left[x(k\text{-}1)\right], \mathrm{E}\left[v(k\text{-}1)\right])$. The above calculations highlight this approximation: Indeed, $\mathrm{E}\left[x(1)\right] \neq \hat{x}_p(1)$. In practice, computing the actual expected value and variance from the distributions is usually not possible due to computational intractability of the general nonlinear case. The EKF is a trade-off between accuracy and computational tractability.

c) Since the distribution of $x(0)$ is identical to part a), the initial mean and variance are also the same:

$$\hat{x}_m(0) = \mathrm{E}\left[x(0)\right] = \frac{1}{2}, \qquad P_m(0) = \mathrm{Var}\left[x(0)\right] = \frac{1}{12}.$$

We define

$$x(k) = q(x(k\text{-}1), v(k\text{-}1)) := x^2(k\text{-}1) + v^2(k\text{-}1).$$

The prior mean is

$$\hat{x}_p(1) = q(\hat{x}_m(0), \mathrm{E}\left[v(0)\right]) = \hat{x}_m^2(0) + \mathrm{E}\left[v(0)\right]^2 = \frac{1}{4}.$$

Linearizing the above process model about the state $x(k{-}1)$ and process noise $v(k{-}1)$, we obtain

$$\begin{aligned}
A(k{-}1) &= \frac{\partial q}{\partial x}\left(\hat{x}_m(k{-}1), \mathrm{E}\left[v(k{-}1)\right]\right) = 2\hat{x}_m(k{-}1) \\
L(k{-}1) &= \frac{\partial q}{\partial v}\left(\hat{x}_m(k{-}1), \mathrm{E}\left[v(k{-}1)\right]\right) = 2\mathrm{E}\left[v(k{-}1)\right] = 0 \\
P_p(k) &= A(k{-}1)P_m(k{-}1)A^{\mathrm{T}}(k{-}1) + L(k{-}1)Q(k{-}1)L^{\mathrm{T}}(k{-}1) \\
&= 4\hat{x}_m^2(k{-}1)P_m(k{-}1) + 0 \\
P_p(1) &= 4\hat{x}_m^2(0)P_m(0) = \frac{1}{12}.
\end{aligned}$$

d) The expected value is

$$\begin{aligned}
\mathrm{E}\left[x(1)\right] &= \mathrm{E}\left[x^2(0) + v^2(0)\right] = \mathrm{E}\left[x^2(0)\right] + \mathrm{E}\left[v^2(0)\right] \\
&= \frac{1}{3} + \int_{-\infty}^{\infty} \lambda^2 p_{v(0)}(\lambda)\, d\lambda = \frac{1}{3} + \frac{1}{2}\int_{-1}^{1} \lambda^2 d\lambda = \frac{1}{3} + \frac{1}{3} = \frac{2}{3}.
\end{aligned}$$

The variance is

$$\begin{aligned}
\mathrm{Var}\left[x(1)\right] &= \mathrm{E}\left[x^2(1)\right] - \mathrm{E}\left[x(1)\right]^2 \\
&= \mathrm{E}\left[\left(x^2(0) + v^2(0)\right)^2\right] - \frac{4}{9} \\
&= \mathrm{E}\left[x^4(0)\right] + 2\mathrm{E}\left[x^2(0)v^2(0)\right] + \mathrm{E}\left[v^4(1)\right] - \frac{4}{9} \\
&= \mathrm{E}\left[x^4(0)\right] + 2\mathrm{E}\left[x^2(0)\right]\mathrm{E}\left[v^2(0)\right] + \mathrm{E}\left[v^4(0)\right] - \frac{4}{9} \quad \text{(by independence)} \\
&= \frac{1}{5} + \frac{2}{9} + \frac{1}{5} - \frac{4}{9} \\
&= \frac{8}{45}
\end{aligned}$$

where we used

$$\mathrm{E}\left[x^4(0)\right] = \int_0^1 \lambda^4 d\lambda = \frac{1}{5} \quad \text{and} \quad \mathrm{E}\left[v^4(0)\right] = \frac{1}{2}\int_{-1}^{1} \lambda^4 d\lambda = \frac{1}{5}.$$

e) We first match the expected value:

$$\begin{aligned}
\mathrm{E}\left[\alpha + \eta(k)\right] &= \mathrm{E}\left[v^2(k)\right] \\
\alpha &= \mathrm{E}\left[v^2(k)\right] \\
\alpha &= \frac{1}{3}.
\end{aligned}$$

We then match the variance:

$$\begin{aligned}
\operatorname{Var}[\alpha+\eta(k)] &= \operatorname{Var}\left[v^2(k)\right]\\
\operatorname{Var}[\eta(k)] &= \mathrm{E}\left[v^4(k)\right]-\mathrm{E}\left[v^2(k)\right]^2\\
\mathrm{E}\left[\eta^2(k)\right] &= \mathrm{E}\left[v^4(k)\right]-\alpha^2 \quad (\eta \text{ zero-mean})\\
\frac{1}{3}\beta^2 &= \frac{1}{5}-\alpha^2\\
\beta &= \frac{2}{\sqrt{15}}.
\end{aligned}$$

f) Since the distributions of $y(0)$ and $x(0)$ are identical, we find that

$$\hat{y}_m(0)=\hat{x}_m(0)=\frac{1}{2}, \quad \widetilde{P}_m(0)=P_m(0)=\frac{1}{12}.$$

The prior mean is

$$\hat{y}_p(1)=\hat{y}_m^2(0)+\alpha=\frac{1}{4}+\frac{1}{3}=\frac{7}{12}$$

which is quite close to the actual expected value $\mathrm{E}[x(1)]=2/3=8/12$. Let

$$y(k)=\tilde{q}(y(k-1),\eta(k-1)):=y^2(k-1)+\alpha+\eta(k-1).$$

The linearization yields

$$\begin{aligned}
\widetilde{A}(0) &= \frac{\partial\tilde{q}}{\partial y}\left(\hat{y}_m(0),\mathrm{E}[\eta(0)]\right)=2\hat{y}_m(0)=1\\
\widetilde{L}(0) &= \frac{\partial\tilde{q}}{\partial \eta}\left(\hat{y}_m(0),\mathrm{E}[\eta(0)]\right)=1.
\end{aligned}$$

The prior variance is therefore

$$\begin{aligned}
\widetilde{P}_p(1) &= \widetilde{A}(0)\widetilde{P}_m(0)\widetilde{A}^{\mathrm{T}}(0)+\widetilde{L}(0)Q(0)\widetilde{L}^{\mathrm{T}}(0)\\
&= \widetilde{P}_m(0)+\operatorname{Var}[\eta]=\frac{1}{12}+\mathrm{E}\left[v^4(k)\right]-\alpha^2=\frac{1}{12}+\frac{1}{5}-\frac{1}{9}=\frac{31}{180}
\end{aligned}$$

which is quite close to the actual variance $\operatorname{Var}[x(1)]=8/45=32/180$.

g) The expected value is

$$\begin{aligned}
\mathrm{E}[y(1)] &= \mathrm{E}\left[y^2(0)+\alpha+\eta(0)\right]\\
&= \mathrm{E}\left[y^2(0)\right]+\alpha\\
&= \mathrm{E}\left[x^2(0)\right]+\mathrm{E}\left[v^2(0)\right]\\
&= \mathrm{E}\left[x^2(0)+v^2(0)\right]\\
&= \mathrm{E}[x(1)].
\end{aligned}$$

The variance of $x(1)$ is $\operatorname{Var}[x(1)]=\mathrm{E}\left[x^2(1)\right]-\mathrm{E}[x(1)]^2$ and analogously for $y(1)$. Since we just showed that the expected values are the same, we only have to show that $\mathrm{E}\left[y^2(1)\right]=$

$\mathrm{E}\left[x^2(1)\right]$, in order to show that the variances are indeed the same:

$$\begin{aligned}
\mathrm{E}\left[y^2(1)\right] &= \mathrm{E}\left[\left(y^2(0)+\alpha+\eta(0)\right)^2\right] \\
&= \mathrm{E}\left[y^4(0)\right] + 2\mathrm{E}\left[y^2(0)\left(\alpha+\eta(0)\right)\right] + \mathrm{E}\left[\left(\alpha+\eta(0)\right)^2\right] \\
&= \mathrm{E}\left[x^4(0)\right] + 2\mathrm{E}\left[x^2(0)\right]\left(\alpha+\mathrm{E}\left[\eta(0)\right]\right) + \mathrm{E}\left[\left(\alpha+\eta(0)\right)^2\right] \quad \text{(by independence)} \\
&= \mathrm{E}\left[x^4(0)\right] + 2\mathrm{E}\left[x^2(0)\right]\mathrm{E}\left[v^2(0)\right] + \mathrm{Var}\left[\alpha+\eta(0)\right] + \mathrm{E}\left[\alpha+\eta(0)\right]^2 \\
&= \mathrm{E}\left[x^4(0)\right] + 2\mathrm{E}\left[x^2(0)\right]\mathrm{E}\left[v^2(0)\right] + \mathrm{Var}\left[v^2(0)\right] + \mathrm{E}\left[v^2(0)\right]^2 \\
&= \mathrm{E}\left[x^4(0)\right] + 2\mathrm{E}\left[x^2(0)\right]\mathrm{E}\left[v^2(0)\right] + \mathrm{E}\left[v^4(0)\right] \\
&= \mathrm{E}\left[\left(x^2(0)+v^2(0)\right)^2\right] \\
&= \mathrm{E}\left[x^2(1)\right].
\end{aligned}$$

No, this does not imply that $p_{y(1)}(\xi) = p_{x(1)}(\xi)$ for all ξ. For example, let $x \sim \mathcal{N}(0,1)$ and y be uniformly distributed on $[-\sqrt{3}, \sqrt{3}]$. The mean and variance of x and y are the same, but their probability density functions are not.

h) For the following derivation, note that $\hat{y}_m(0) = \hat{x}_m(0)$ and $\mathrm{E}\left[x^2(0)\right] = \mathrm{Var}\left[x(0)\right] + \mathrm{E}\left[x(0)\right]^2$.

$$\begin{aligned}
\mathrm{E}\left[\left(x(1)-\hat{x}_p(1)\right)^2\right] &> \mathrm{E}\left[\left(x(1)-\hat{y}_p(1)\right)^2\right] \\
\mathrm{E}\left[x^2(1) - 2x(1)\hat{x}_p(1) + \hat{x}_p^2(1)\right] &> \mathrm{E}\left[x^2(1) - 2x(1)\hat{y}_p(1) + \hat{y}_p^2(1)\right] \\
\mathrm{E}\left[x^2(1)\right] - 2\mathrm{E}\left[x(1)\right]\hat{x}_p(1) + \hat{x}_p^2(1) &> \mathrm{E}\left[x^2(1)\right] - 2\mathrm{E}\left[x(1)\right]\hat{y}_p(1) + \hat{y}_p^2(1) \\
-2\mathrm{E}\left[x(1)\right]\hat{x}_p(1) + \hat{x}_p^2(1) &> -2\mathrm{E}\left[x(1)\right]\hat{y}_p(1) + \hat{y}_p^2(1) \\
-2\mathrm{E}\left[x(1)\right]\hat{x}_m^2(0) + \hat{x}_m^4(0) &> -2\mathrm{E}\left[x(1)\right]\left(\hat{y}_m^2(0)+\alpha\right) + \left(\hat{y}_m^2(0)+\alpha\right)^2 \\
-2\mathrm{E}\left[x(1)\right]\hat{x}_m^2(0) + \hat{x}_m^4(0) &> -2\mathrm{E}\left[x(1)\right]\left(\hat{x}_m^2(0)+\alpha\right) + \left(\hat{x}_m^2(0)+\alpha\right)^2 \\
0 &> -2\alpha\mathrm{E}\left[x(1)\right] + 2\alpha\hat{x}_m^2(0) + \alpha^2 \\
0 &> -2\alpha\left(\mathrm{E}\left[x^2(0)\right] + \mathrm{E}\left[v^2(0)\right]\right) + 2\alpha\hat{x}_m^2(0) + \alpha^2 \\
0 &> -2\alpha\left(\mathrm{E}\left[x^2(0)\right] + \alpha\right) + 2\alpha\hat{x}_m^2(0) + \alpha^2 \\
0 &> -2\alpha\left(\mathrm{Var}\left[x(0)\right] + \mathrm{E}\left[x(0)\right]^2\right) + 2\alpha\hat{x}_m^2(0) - \alpha^2 \\
0 &> -2\alpha\left(\mathrm{Var}\left[x(0)\right] + \hat{x}_m^2(0)\right) + 2\alpha\hat{x}_m^2(0) - \alpha^2 \\
0 &> -2\alpha\mathrm{Var}\left[x(0)\right] - \alpha^2.
\end{aligned}$$

Both summands are strictly smaller than zero because $\alpha > 0$ and $\mathrm{Var}\left[x(0)\right] > 0$. The inequality therefore holds, which shows that the expected squared estimation error is smaller for the estimator using the adapted model (7) compared to the estimator using the actual system dynamics.

Intuitively, this is because the EKF using the actual model (6) underestimates (effectively ignores) the contribution of the process noise $v(0)$ on the prior mean and variance. This example illustrates that it may be beneficial to implement a modified process model in an EKF in order to achieve better state prediction performance.

Problem 8

In the lecture, the EKF equations were derived for zero-mean noise, i.e. $\mathrm{E}\left[v(k)\right] = \mathrm{E}\left[w(k)\right] = 0$. In this problem, both the process and the measurement noise are not zero-mean. In the following, you find the slightly modified EKF equations for this case. The modifications stem from linearizing the process and measurement equations in the EKF derivation about the mean of the noise (now different from zero), which results in using the expected value of the noise in the respective equations (instead of zero).

a) We define the system state as $s(k) := (x(k), y(k))$. First, we initialize the EKF with the initial posterior mean $\hat{s}_m(0)$ and variance $P_m(0)$ according to the specified probability density functions (PDFs) of the initial populations $x(0)$ and $y(0)$. For $x(0)$ we obtain

$$\begin{aligned} \mathrm{E}[x(0)] &= \frac{1}{2} \\ \mathrm{Var}[x(0)] &= \mathrm{E}\left[x^2(0)\right] - \mathrm{E}[x(0)]^2 \\ &= \frac{1}{3} - \frac{1}{4} = \frac{1}{12}. \end{aligned}$$

Since the PDFs of $x(0)$ and $y(0)$ are identical, we find

$$\hat{s}_m(0) := \begin{bmatrix} \hat{x}_m(0) \\ \hat{y}_m(0) \end{bmatrix} = \mathrm{E}[s(0)] = \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \end{bmatrix}, \qquad P_m(0) = \mathrm{Var}[s(0)] = \begin{bmatrix} \frac{1}{12} & 0 \\ 0 & \frac{1}{12} \end{bmatrix}.$$

Next, we calculate the specific prior update equations. We define the process noise vector as $v(k-1) := (v_x(k-1), v_y(k-1))$ and the nonlinear system dynamics as

$$\begin{aligned} s(k) = q(s(k-1), v(k-1)) &:= \begin{bmatrix} (1-\alpha)\, l(x(k-1)) + (1 - v_y(k-1))\, \alpha\, l(y(k-1)) \\ (1-\alpha)\, l(y(k-1)) + (1 - v_x(k-1))\, \alpha\, l(x(k-1)) \end{bmatrix} \\ l(x(k-1)) &:= \mu\, x(k-1)(1 - x(k-1)). \end{aligned}$$

The prior mean $\hat{s}_p(k) := (\hat{x}_p(k), \hat{y}_p(k))$ is therefore

$$\hat{s}_p(k) = q(\hat{s}_m(k-1), \mathrm{E}[v(k-1)]) = \begin{bmatrix} (1-\alpha)\, l(\hat{x}_m(k-1)) + \left(1 - \frac{a}{2}\right) \alpha\, l(\hat{y}_m(k-1)) \\ (1-\alpha)\, l(\hat{y}_m(k-1)) + \left(1 - \frac{a}{2}\right) \alpha\, l(\hat{x}_m(k-1)) \end{bmatrix}.$$

Note the expected value of the process noise used in the prior mean update. For the prior variance update, we linearize the nonlinear dynamics about the posterior mean and expected value of the process noise signals and obtain

$$A(k-1) := \frac{\partial q}{\partial s}(\hat{s}_m(k-1), \mathrm{E}[v(k-1)]) = \begin{bmatrix} (1-\alpha)\, l'(\hat{x}_m(k-1)) & \left(1 - \frac{a}{2}\right) \alpha\, l'(\hat{y}_m(k-1)) \\ \left(1 - \frac{a}{2}\right) \alpha\, l'(\hat{x}_m(k-1)) & (1-\alpha)\, l'(\hat{y}_m(k-1)) \end{bmatrix} \quad (9)$$

with $l'(\xi) := \frac{d}{d\xi} l(\xi) = \mu(1 - 2\xi)$. For the partial with respect to the process noise, we obtain

$$L(k-1) := \frac{\partial q}{\partial v}(\hat{s}_m(k-1), \mathrm{E}[v(k-1)]) = \begin{bmatrix} 0 & -\alpha l(\hat{y}_m(k-1)) \\ -\alpha l(\hat{x}_m(k-1)) & 0 \end{bmatrix}$$

The prior variance can then be calculated with

$$P_p(k) = A(k-1) P_m(k-1) A^{\mathrm{T}}(k-1) + L(k-1) Q L^{\mathrm{T}}(k-1)$$

where the constant process noise variance

$$Q = \frac{a^2}{12} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

is given by the PDFs of $v_x(k-1)$ and $v_y(k-1)$. Next, we calculate the measurement update equations. We define the measurement vector $z(k) := (z_x(k), z_y(k))$, the measurement noise vector $w(k) := (w_x(k), w_y(k))$, and the measurement equation

$$z(k) = h(s(k), w(k)) := \begin{bmatrix} x(k)(1 - w_x(k)) \\ y(k)(1 - w_y(k)) \end{bmatrix}.$$

For the calculation of the Kalman gain $K(k)$, we linearize the measurement equation about the prior mean and expected value of the measurement noise:

$$H(k) := \frac{\partial h}{\partial s}\left(\hat{s}_p(k), \mathrm{E}\left[w(k)\right]\right) = \begin{bmatrix} 1-\frac{b}{2} & 0 \\ 0 & 1-\frac{b}{2} \end{bmatrix}$$

$$M(k) := \frac{\partial h}{\partial w}\left(\hat{s}_p(k), \mathrm{E}\left[w(k)\right]\right) = \begin{bmatrix} -\hat{x}_p(k) & 0 \\ 0 & -\hat{y}_p(k) \end{bmatrix}.$$

The Kalman gain is

$$K(k) = P_p(k)H^T(k)\left(H(k)P_p(k)H^T(k) + M(k)RM^T(k)\right)^{-1}$$

with the constant measurement variance

$$R = \frac{b^2}{12}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

given by the PDFs of $w_x(k)$ and $w_y(k)$. Given the measurement $\bar{z}(k)$ produced by the simulation, the posterior mean is

$$\hat{s}_m(k) = \hat{s}_p(k) + K(k)\left(\bar{z}(k) - h\left(\hat{s}_p(k), \mathrm{E}\left[w(k)\right]\right)\right)$$

with

$$h\left(\hat{s}_p(k), \mathrm{E}\left[w(k)\right]\right) = \begin{bmatrix} \hat{x}_p(k)(1-\frac{b}{2}) \\ \hat{y}_p(k)(1-\frac{b}{2}) \end{bmatrix}.$$

The posterior variance is

$$P_m(k) = (I - K(k)H(k))P_p(k)(I - K(k)H(k))^T + K(k)M(k)\,R\,M^T(k)K^T(k).$$

You may find a sample Matlab implementation of the EKF for this part on the class website.

b) In order to estimate the unknown constant parameter α, we extend and redefine the system state from a) to $s(k) = (x(k), y(k), \alpha(k))$ and then use an EKF to estimate this state. The initial mean and variance of the extended state follow from a) and the PDF of α given in the problem:

$$\hat{s}_m(0) := \begin{bmatrix} \hat{x}_m(k\text{-}1) \\ \hat{y}_m(k\text{-}1) \\ \hat{\alpha}_m(k\text{-}1) \end{bmatrix} = \mathrm{E}\left[s(0)\right] = \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \\ 0.05 \end{bmatrix}, \qquad P_m(0) = \mathrm{Var}\left[s(0)\right] = \begin{bmatrix} \frac{1}{12} & 0 & 0 \\ 0 & \frac{1}{12} & 0 \\ 0 & 0 & \frac{0.01}{12} \end{bmatrix}$$

We define the state dynamics as

$$s(k) = q(s(k\text{-}1), v(k\text{-}1)) := \begin{bmatrix} (1-\alpha(k\text{-}1))\,l(x(k\text{-}1)) + (1-v_y(k\text{-}1))\,\alpha(k\text{-}1)\,l(y(k\text{-}1)) \\ (1-\alpha(k\text{-}1))\,l(y(k\text{-}1)) + (1-v_x(k\text{-}1))\,\alpha(k\text{-}1)\,l(x(k\text{-}1)) \\ \alpha(k\text{-}1) \end{bmatrix}.$$

The parameter α is constant and its dynamics are not affected by uncertainty, i.e. process noise. Therefore, $v(k) := (v_x(k), v_y(k))$ and Q remain identical to part a). The prior mean is redefined to $\hat{s}_p(k) := (\hat{x}_p(k), \hat{y}_p(k), \hat{\alpha}_p(k))$ and is

$$\hat{s}_p(k) = q(\hat{s}_m(k\text{-}1), \mathrm{E}\left[v(k\text{-}1)\right]) = \begin{bmatrix} (1-\hat{\alpha}_m(k\text{-}1))\,l(\hat{x}_m(k\text{-}1)) + \left(1-\frac{a}{2}\right)\hat{\alpha}_m(k\text{-}1)\,l(\hat{y}_m(k\text{-}1)) \\ (1-\hat{\alpha}_m(k\text{-}1))\,l(\hat{y}_m(k\text{-}1)) + \left(1-\frac{a}{2}\right)\hat{\alpha}_m(k\text{-}1)\,l(\hat{x}_m(k\text{-}1)) \\ \hat{\alpha}_m(k\text{-}1) \end{bmatrix}.$$

Linearizing the state dynamics for the prior update, we obtain

$$A(k{-}1) := \frac{\partial q}{\partial s}\left(\hat{s}_m(k{-}1), \mathrm{E}\left[v(k{-}1)\right]\right) = \begin{bmatrix} A_{xy}(k{-}1) & A_\alpha(k{-}1) \\ 0 \quad 0 & 1 \end{bmatrix}$$

where $A_{xy}(k{-}1)$ is the linearization from part a) with the posterior estimate $\hat{\alpha}_m(k{-}1)$ substituted for α

$$A_{xy}(k{-}1) := \begin{bmatrix} \left(1 - \hat{\alpha}_m(k{-}1)\right) l'(\hat{x}_m(k{-}1)) & \left(1 - \frac{a}{2}\right)\hat{\alpha}_m(k{-}1)\, l'(\hat{y}_m(k{-}1)) \\ \left(1 - \frac{a}{2}\right)\hat{\alpha}_m(k{-}1)\, l'(\hat{x}_m(k{-}1)) & \left(1 - \hat{\alpha}_m(k{-}1)\right) l'(\hat{y}_m(k{-}1)) \end{bmatrix}$$

and $A_\alpha(k{-}1)$ is the partial of the population dynamics with respect to $\alpha(k{-}1)$:

$$A_\alpha(k{-}1) := \begin{bmatrix} \frac{\partial x(k)}{\partial \alpha(k-1)}\left(\hat{s}_m(k{-}1), \mathrm{E}\left[v(k{-}1)\right]\right) \\ \frac{\partial y(k)}{\partial \alpha(k-1)}\left(\hat{s}_m(k{-}1), \mathrm{E}\left[v(k{-}1)\right]\right) \end{bmatrix} = \begin{bmatrix} -l(\hat{x}_m(k{-}1)) + (1 - \frac{a}{2}) l(\hat{y}_m(k{-}1)) \\ -l(\hat{y}_m(k{-}1)) + (1 - \frac{a}{2}) l(\hat{x}_m(k{-}1)) \end{bmatrix}.$$

The partial with respect to the process noise is:

$$L(k{-}1) := \frac{\partial q}{\partial v}\left(\hat{s}_m(k{-}1), \mathrm{E}\left[v(k{-}1)\right]\right) = \begin{bmatrix} 0 & -\hat{\alpha}_m(k{-}1) l(\hat{y}_m(k{-}1)) \\ -\hat{\alpha}_m(k{-}1) l(\hat{x}_m(k{-}1) & 0 \\ 0 & 0 \end{bmatrix}.$$

The prior variance update is then

$$P_p(k) = A(k{-}1) P_m(k{-}1) A^{\mathrm{T}}(k{-}1) + L(k{-}1) Q L^{\mathrm{T}}(k{-}1).$$

For the measurement update, the only modification to the equations stated in part a) is that the partial of the measurement equation with respect to the state is now

$$H(k) := \frac{\partial h}{\partial s}\left(\hat{s}_p(k), \mathrm{E}\left[w(k)\right]\right) = \begin{bmatrix} 1 - \frac{b}{2} & 0 & 0 \\ 0 & 1 - \frac{b}{2} & 0 \end{bmatrix}.$$

The remaining steps in the measurement update are identical to part a). You may find a sample Matlab implementation of the EKF for this part on the class website.

Problem 9

The matrix inversion lemma can be proven by solving the following matrix equation:

$$\begin{bmatrix} A & B \\ C & -D^{-1} \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} I \\ 0 \end{bmatrix}.$$

Expanding this yields

$$AX + BY = I \tag{10}$$

$$CX - D^{-1} Y = 0 \tag{11}$$

We rewrite (11) (D^{-1} is invertible),

$$D^{-1} Y = CX \quad \Leftrightarrow \quad Y = DCX.$$

Substituting this into (10) we obtain ($(A + BDC)$ is invertible, see below)

$$(A + BDC)\, X = I \quad \Leftrightarrow \quad X = (A + BDC)^{-1}, \tag{12}$$

which is the left-hand side of the identity under consideration.

On the other hand, we may solve (10) for X, $X = A^{-1}(I - BY)$ (A is invertible), and substitute into (11)

$$
\begin{aligned}
& CA^{-1}(I - BY) = D^{-1}Y \\
\Leftrightarrow \quad & CA^{-1} - CA^{-1}BY = D^{-1}Y \\
\Leftrightarrow \quad & CA^{-1} = \left(D^{-1} + CA^{-1}B\right)Y \\
\Leftrightarrow \quad & \left(D^{-1} + CA^{-1}B\right)^{-1} CA^{-1} = Y \quad ((D^{-1} + CA^{-1}B) \text{ is invertible}).
\end{aligned}
$$

Substituting this into (10), we obtain another equation for X,

$$
\begin{aligned}
& AX + B\left(D^{-1} + CA^{-1}B\right)^{-1} CA^{-1} = I \\
\Leftrightarrow \quad & X = A^{-1} - A^{-1}B\left(D^{-1} + CA^{-1}B\right)^{-1} CA^{-1},
\end{aligned}
$$

which is the right-hand side of the identity under consideration.
Note that if A, D, $D^{-1} + CA^{-1}B$ are all nonsingular, then $A + BDC$ is nonsingular, since one can write

$$
\begin{bmatrix} A & B \\ C & -D^{-1} \end{bmatrix} = \begin{bmatrix} I & 0 \\ CA^{-1} & I \end{bmatrix} \begin{bmatrix} A & 0 \\ 0 & -D^{-1} - CA^{-1}B \end{bmatrix} \begin{bmatrix} I & A^{-1}B \\ 0 & I \end{bmatrix}
$$

and

$$
\begin{bmatrix} A & B \\ C & -D^{-1} \end{bmatrix} = \begin{bmatrix} I & -BD \\ 0 & I \end{bmatrix} \begin{bmatrix} A + BDC & 0 \\ 0 & -D^{-1} \end{bmatrix} \begin{bmatrix} I & 0 \\ -DC & I \end{bmatrix},
$$

and therefore

$$
\det(A)\det\left(-D^{-1} - CA^{-1}B\right) = \det(A + BDC)\det\left(-D^{-1}\right).
$$

This implies that $\det(A + BDC) \neq 0$ if $\det(A)$, $\det\left(-D^{-1}\right)$, and $\det\left(-D^{-1} - CA^{-1}B\right)$ are all nonzero.