# Stochastic Adaptive Control (02421)

Lecture 3

Tobias K. S. Ritschel

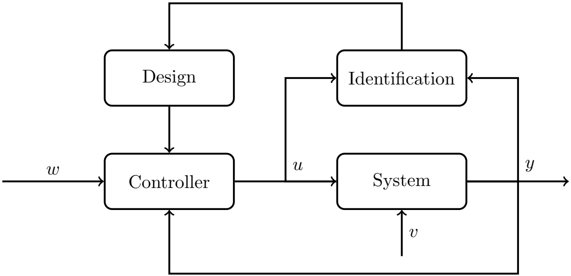
Section for Dynamical Systems

Department of Applied Mathematics and Computer Science

Technical University of Denmark

# Lecture Plan

1. System theory
2. Stochastics
3. **State estimation 1**
4. State estimation 2
5. Optimal control 1
6. System identification 1 + adaptive control 1
7. External models + prediction
8. Optimal control 2
9. Optimal control 3
10. System identification 2
11. System identification 3 + model validation
12. System identification 4 + adaptive control 2
13. Adaptive control 3



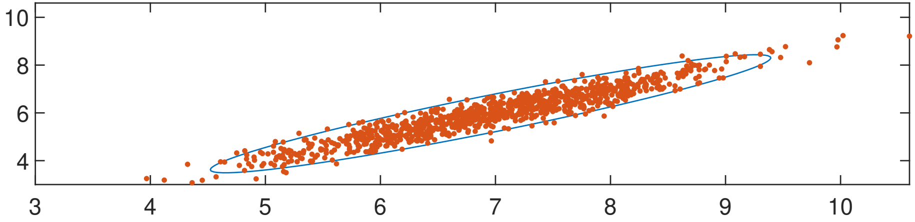
# Today's Topics

- Follow-up from last lecture
- The projection theorem
- Derivation of the Kalman filter
- Stationary Kalman filters
- State estimation errors

# Follow-Up from Last Lecture

Question 2.6: Sketch the 95% confidence interval for the two last signals

Answer: Use the Matlab function Niveau.m uploaded to DTU Learn




Normally distributed random variable

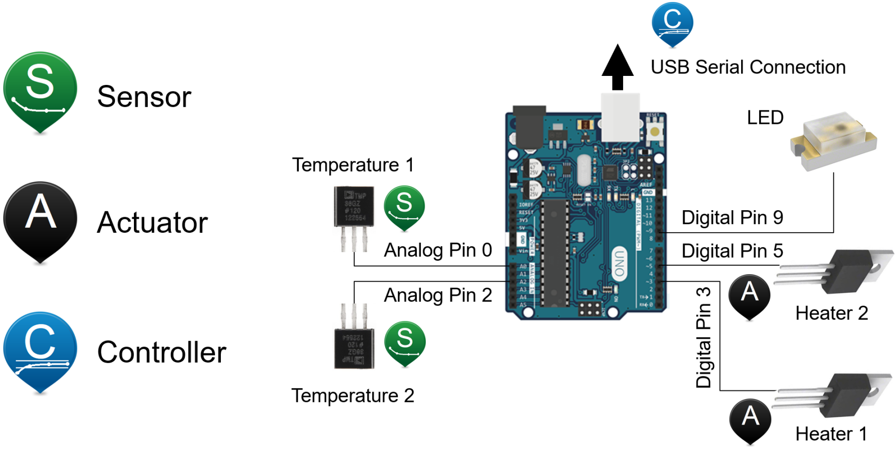
$$X \sim N(m, P) \tag{1}$$

Plot confidence interval using Niveau($m$, $P^{-1}$, $f$) where $f$ is the square root of the $\chi^2$ quantile level with 2 degrees of freedom

See, e.g., Example 3.16 in the Danish book

# Demonstration

# Temperature control laboratory (TCLab)




Link: `https://apmonitor.com/pdc/index.php/Main/ArduinoTemperatureControl`

# TCLab model

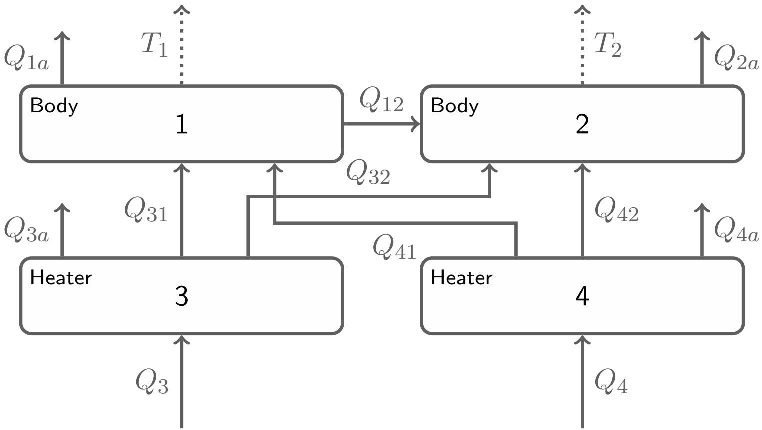



Figure: Four-compartment model of TCLab device.

# State Estimation

# Filters and Estimation

Objective: Obtain estimate, $\hat{x}_t$, of the signal $x_t$ based on measurements

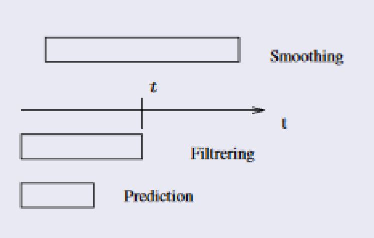
$$Y_{0:N} = \begin{bmatrix} y_0 & y_1 & \cdots & y_N \end{bmatrix} \quad (2)$$

and a state-output relation, e.g.,

$$y_t = Cx_t + e_t \quad (3)$$

Different types of estimation

1. Smoothing ($t < t_N$): Use both past and future data to estimate the states
2. Filtering: ($t = t_N$): Estimate the current states based on current and past data
3. Prediction: ($t > t_N$): Predict future states based on past data



# Filter Theory

Stochastic discrete-time system

$$x_{t+1} = Ax_t + Bu_t + v_t, \qquad x_0 \sim N(m_0, P_0), \qquad v_t \sim N(0, R_1), \tag{4}$$

$$y_t = Cx_t + e_t, \qquad e_t \sim N(0, R_2) \tag{5}$$

We will only consider filtering and prediction in this lecture

# Filter Theory - The Good Estimate

**Core concepts** of filter design

1. Characteristics of the signal and noise
2. Observation model (relation between $y$, $x$, $e$)
3. Criterion (what is a good estimate)
4. Restrictions (what information is available)

**Characteristics:** Nature of the states, dynamics, and noises

**Observation:** Relation between the output, $y$, the state $x$, and the noise

**The criterion:** A good estimate minimizes the expected squared deviation

$$\mathbb{E}[||x - \hat{x}||^2] \tag{6}$$

**Restrictions:** What data, $Y$, is available (filter, predict, or smoothe?)

# The filter problem: A good estimator

The law of total expectation

$$\mathbb{E}[g(x)] = \mathbb{E}_Y[\mathbb{E}[g(x)|Y]] \tag{7}$$

Introduce "inner" objective function

$$\begin{aligned} J &= \mathbb{E}[\|x-\hat{x}\|^2] && (8)\\ &= \mathbb{E}[(x-\hat{x})^T(x-\hat{x})] && (9)\\ &= \mathbb{E}_Y[\mathbb{E}[(x-\hat{x})^T(x-\hat{x})|Y]] = \mathbb{E}_Y[J_{in}] && (10) \end{aligned}$$

Inner objective function

$$\begin{aligned} J_{in} &= \mathbb{E}[x^Tx - \hat{x}^Tx - x^T\hat{x} + \hat{x}^T\hat{x}|Y], && (11)\\ &= \mathbb{E}[x^Tx|Y] - \hat{x}^T\mathbb{E}[x|Y] - \mathbb{E}[x|Y]^T\hat{x} + \hat{x}^T\hat{x} && (12) \end{aligned}$$

Optimal estimate

$$\begin{aligned} \nabla_{\hat{x}} J_{in} &= 2\hat{x} - 2\mathbb{E}[x|Y] = 0, && (13)\\ \hat{x} &= \mathbb{E}[x|Y] && (14) \end{aligned}$$

# The projection theorem

# Filter Theory - Projection Theorem

Normally distributed vector

$$Z = \begin{bmatrix} X \\ Y \end{bmatrix} \sim N\left( \begin{bmatrix} m_x \\ m_y \end{bmatrix}, \begin{bmatrix} P_x & P_{xy} \\ P_{xy}^T & P_y \end{bmatrix} \right) \tag{15}$$

**Projection theorem**: The conditional distribution $X|Y \sim N(m_{x|y}, P_{x|y})$ is

$$m_{x|y} = m_x + P_{xy} P_y^{-1} (y - m_y), \tag{16}$$

$$P_{x|y} = P_x - P_{xy} P_y^{-1} P_{xy}^T, \tag{17}$$

$$X - \hat{x} \perp Y \tag{18}$$

# Filter Theory - Proof of Projection Theorem

Probability density functions

$$f_{X,Y}(x,y) = \frac{1}{(2\pi)^{n_x+n_y}\sqrt{\det(P_z)}} e^{-\frac{1}{2}(z-m_z)^T P_z^{-1}(z-m_z)}, \tag{19}$$

$$f_Y(y) = \frac{1}{(2\pi)^{n_y}\sqrt{\det(P_y)}} e^{-\frac{1}{2}(y-m_y)^T P_y^{-1}(y-m_y)} \tag{20}$$

Probability density function of conditional normal distribution

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} \tag{21}$$

$$= \sqrt{\frac{\det(P_y)}{(2\pi)^{n_x}\det(P_z)}} e^{-\frac{1}{2}(z-m_z)^T P_z^{-1}(z-m_z)+\frac{1}{2}(y-m_y)^T P_y^{-1}(y-m_y)}$$

$$= \kappa e^{-\frac{1}{2}\alpha} \tag{22}$$

# Filter Theory - Proof of Projection Theorem

Schur complement

$$D = P_x - P_{xy} P_y^{-1} P_{xy}^T \tag{23}$$

Use Woodbury matrix identity on $P_z^{-1}$

$$P_z^{-1} = \begin{bmatrix} D^{-1} & -D^{-1} P_{xy} P_y^{-1} \\ -P_y^{-1} P_{xy}^T D^{-1} & P_y^{-1} + P_y^{-1} P_{xy}^T D^{-1} P_{xy} P_y^{-1} \end{bmatrix} \tag{24}$$

Determinant

$$\det(P_z) = \det(P_y) \det(D) \qquad \Leftrightarrow \qquad \frac{\det(P_y)}{\det(P_z)} = \frac{1}{\det(D)} \tag{25}$$

Factor

$$\kappa = \sqrt{\frac{\det(P_y)}{(2\pi)^{n_x} \det(P_z)}} = \frac{1}{\sqrt{(2\pi)^{n_x} \det(D)}} \tag{26}$$

Exponent

$$\alpha = (z - m_z)^T P_z^{-1} (z - m_z) - (y - m_y)^T P_y^{-1} (y - m_y) \tag{27}$$

$$= [x - (m_x + P_{xy} P_y^{-1} (y - m_y))]^T D^{-1} [x - (m_x + P_{xy} P_y^{-1} (y - m_y))] \tag{28}$$

# Filter Theory - Proof of Projection Theorem

Mean and covariance of conditional distribution

$$\mathbb{E}[X|Y] = m_{x|y} = m_x + P_{xy}P_y^{-1}(y - m_y) \quad (29)$$

$$\mathrm{Cov}(X|Y) = P_{x|y} = D = P_x - P_{xy}P_y^{-1}P_{xy}^T \quad (30)$$

Covariance (are the variables independent?)

$$\mathrm{Cov}(X - m_{x|y}, Y) = \mathrm{Cov}(X, Y) - P_{xy}P_y^{-1}\,\mathrm{Cov}(Y, Y) \quad (31)$$

$$= P_{xy} - P_{xy}P_y^{-1}P_y = 0 \quad (32)$$

As $X$ and $Y$ are Gaussian, they are independent

# Derivation of the Kalman filter

# State estimation

Stochastic discrete-time system

$$x_{t+1} = Ax_t + Bu_t + v_t, \qquad x_0 \sim N(\hat{x}_0, P_0), \qquad v_t \sim N(0, R_1), \tag{33}$$

$$y_t = Cx_t + e_t, \qquad e_t \sim N(0, R_2) \tag{34}$$

$v_t \perp x_s$ for all $s \leq t$ and $e_t \perp x_s$ for all $s$

Mean and covariance of joint distribution

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} |Y_{t-1} \sim N\left( \begin{bmatrix} \times \\ \times \end{bmatrix}, \begin{bmatrix} \times & \times \\ \times & \times \end{bmatrix} \right), \quad Y_t = \begin{bmatrix} Y_{t-1} \\ y_t \end{bmatrix} \tag{35}$$

Conditional state distributions

$$x_t|Y_{t-1} \sim N(\hat{x}_{t|t-1}, P_{t|t-1}) \tag{36}$$

$$x_t|Y_t = x_t|y_t, Y_{t-1} \sim N(\hat{x}_{t|t}, P_{t|t}) \tag{37}$$

# State estimation: A recursion

Measurement equation

$$y_t = Cx_t + e_t, \qquad e_t \sim N(0, R_2), \qquad e_t \perp x_s \tag{38}$$

Mean

$$\mathbb{E}[y_t|Y_{t-1}] = C\mathbb{E}[x_t|Y_{t-1}] + \mathbb{E}[e_t|Y_{t-1}] = C\hat{x}_{t|t-1} \tag{39}$$

Covariance

$$\begin{aligned}\mathrm{Cov}(y_t|Y_{t-1}) &= C\,\mathrm{Cov}(x_t|Y_{t-1})C^T + C\,\mathrm{Cov}(x_t, e_t|Y_{t-1}) \\ &\quad + \mathrm{Cov}(e_t, x_t|Y_{t-1})C^T + \mathrm{Cov}(e_t|Y_{t-1}) = CP_{t|t-1}C^T + R_2\end{aligned} \tag{40}$$

Cross-covariance

$$\mathrm{Cov}(y_t, x_t|Y_{t-1}) = C\,\mathrm{Cov}(x_t|Y_{t-1}) + C\,\mathrm{Cov}(x_t, e_t|Y_{t-1}) \tag{41}$$

$$= CP_{t|t-1} \tag{42}$$

Conditional distribution

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} |Y_{t-1} \sim N\left( \begin{bmatrix} \hat{x}_{t|t-1} \\ C\hat{x}_{t|t-1} \end{bmatrix}, \begin{bmatrix} P_{t|t-1} & P_{t|t-1}C^T \\ CP_{t|t-1} & CP_{t|t-1}C^T + R_2 \end{bmatrix} \right) \tag{43}$$

# State Estimation 4: The Current Estimate

Conditional distribution

$$x_t|y_t, Y_{t-1} = x_t|Y_t \sim N(\hat{x}_{t|t}, P_{t|t}) \tag{44}$$

Use projection theorem

$$\hat{x}_{t|t} = \hat{x}_{t|t-1} + P_{t|t-1}C^T(CP_{t|t-1}C^T + R_2)^{-1}(y_t - C\hat{x}_{t|t-1}), \tag{45}$$

$$P_{t|t} = P_{t|t-1} - P_{t|t-1}C^T(CP_{t|t-1}C^T + R_2)^{-1}CP_{t|t-1} \tag{46}$$

## State Estimation 5: Prediction Estimate

State equation

$$x_{t+1} = Ax_t + Bu_t + v_t, \quad v_t \sim N(0, R_1), \quad v_t \perp x_s \text{ for all } s \leq t \tag{47}$$

Mean

$$\mathbb{E}[x_{t+1}|Y_t] = A\mathbb{E}[x_t|Y_t] + B\mathbb{E}[u_t|Y_t] + \mathbb{E}[v_t|Y_t] \tag{48}$$

$$= A\hat{x}_{t|t} + Bu_t \tag{49}$$

Covariance

$$\mathrm{Cov}(x_{t+1}|Y_t) = A\,\mathrm{Cov}(x_t|Y_t)A^T + \mathrm{Cov}(v_t|Y_t) + A\,\mathrm{Cov}(x_t, v_t|Y_t) \tag{50}$$

$$+ \mathrm{Cov}(v_t, x_t|Y_t)A^T = AP_{t|t}A^T + R_1 \tag{51}$$

Prediction estimate

$$\hat{x}_{t+1|t} = A\hat{x}_{t|t} + Bu_t, \tag{52}$$

$$P_{t+1|t} = AP_{t|t}A^T + R_1 \tag{53}$$

# Filter Theory - Kalman Filter

Data/measurement-update (inference)

$$\hat{x}_{t|t} = \hat{x}_{t|t-1} + \kappa_t(y_t - C\hat{x}_{t|t-1}), \tag{54}$$

$$\kappa_t = P_{t|t-1}C^T(CP_{t|t-1}C^T + R_2)^{-1}, \tag{55}$$

$$P_{t|t} = P_{t|t-1} - \kappa_t C P_{t|t-1} \tag{56}$$

Time-update (prediction)

$$\hat{x}_{t+1|t} = A\hat{x}_{t|t} + Bu_t, \qquad \hat{x}_{0|0} = \hat{x}_0, \tag{57}$$

$$P_{t+1|t} = AP_{t|t}A^T + R_1, \qquad P_{0|0} = P_0 \tag{58}$$

# Example: Pseudocode - Kalman Filter/Simulation Implementation

Initial values: $x_{0|-1}$, $P_{0|-1}$, $x_0$

---

for $t = 0, \ldots, N$

Measurement from true system:

$y_t = \texttt{Measurement}(x_t, e_t)$

---

Data update:

$[\hat{x}_{t|t}, P_{t|t}, \kappa_t] = \texttt{DataUpdate}(y_t, \hat{x}_{t|t-1}, P_{t|t-1}; C, R_2)$

---

Compute control:

$u_t = \texttt{Actuator}(\hat{x}_{t|t})$

Apply control:

$x_{t+1} = \texttt{Simulator}(x_t, u_t, v_t)$

---

Time update:

$[\hat{x}_{t+1|t}, P_{t+1|t}] = \texttt{TimeUpdate}(\hat{x}_{t|t}, P_{t|t}, u_t; A, B, R_1)$

end

## Example: Estimation of constant

Estimate scalar constant

$$x_{t+1} = x_t, \tag{59}$$

$$y_t = x_t + e_t, \quad e_t \in N(0, r_2) \tag{60}$$

Define $q_t = p_t^{-1}$

$$\kappa_t = \frac{p_t}{p_t + r_2} = \frac{1}{1 + r_2 q_t}, \tag{61}$$

$$p_{t+1} = (1 - \kappa_t) p_t = \left(1 - \frac{1}{1 + r_2 q_t}\right) p_t = \frac{r_2 q_t}{1 + r_2 q_t} p_t = \frac{r_2}{1 + r_2 q_t}, \tag{62}$$

$$q_{t+1} = \frac{1}{p_{t+1}} = \frac{1 + r_2 q_t}{r_2} = q_t + \frac{1}{r_2} = q_0 + \frac{t+1}{r_2}, \tag{63}$$

$$\hat{x}_{t+1} = \hat{x}_t + \kappa_t (y_t - \hat{x}_t) \tag{64}$$

If $q_0 = 0$ $(p_0 = \infty)$,

$$\hat{x}_{t+1} = \hat{x}_t + \frac{1}{1+t}(y_t - \hat{x}_t) \qquad \text{or} \qquad \hat{x}_t = \frac{1}{t} \sum_{i=0}^{t-1} y_i \tag{65}$$

# Two different forms of the Kalman filter

**Ordinary** Kalman filter

$$\underbrace{\begin{bmatrix} \hat{x}_{t|t} \\ P_{t|t} \end{bmatrix} \rightarrow \begin{bmatrix} \hat{x}_{t+1|t} \\ P_{t+1|t} \end{bmatrix} \rightarrow \begin{bmatrix} \hat{x}_{t+1|t+1} \\ P_{t+1|t+1} \end{bmatrix}}_{\text{Ordinary Kalman Filter}} \tag{66}$$

**Predictive** Kalman filter

$$\underbrace{\begin{bmatrix} \hat{x}_{t|t-1} \\ P_{t|t-1} \end{bmatrix} \rightarrow \begin{bmatrix} \hat{x}_{t|t} \\ P_{t|t} \end{bmatrix} \rightarrow \begin{bmatrix} \hat{x}_{t+1|t} \\ P_{t+1|t} \end{bmatrix}}_{\text{Predictive Kalman Filter}} \tag{67}$$

# Ordinary Kalman filter

Time-update (prediction)

$$\hat{x}_{t+1|t} = A\hat{x}_{t|t} + Bu_t, \qquad \hat{x}_{0|0} = \hat{x}_0, \tag{68}$$

$$P_{t+1|t} = AP_{t|t}A^T + R_1, \qquad P_{0|0} = P_0 \tag{69}$$

Data-update (inference)

$$\hat{x}_{t|t} = \hat{x}_{t|t-1} + \kappa_t(y_t - C\hat{x}_{t|t-1}), \tag{70}$$

$$\kappa_t = P_{t|t-1}C^T(CP_{t|t-1}C^T + R_2)^{-1}, \tag{71}$$

$$P_{t|t} = P_{t|t-1} - \kappa_t CP_{t|t-1} \tag{72}$$

**Ordinary** Kalman filter

$$\hat{x}_{t|t} = (I - \kappa_t C)(A\hat{x}_{t-1|t-1} + Bu_{t-1}) + \kappa_t y_t, \tag{73}$$

$$P_{t|t} = AP_{t-1|t-1}A^T + R_1 - \kappa_t C(AP_{t-1|t-1}A^T + R_1), \tag{74}$$

$$\kappa_t = (AP_{t-1|t-1}A^T + R_1)C^T(C(AP_{t-1|t-1}A^T + R_1)C^T + R_2)^{-1} \tag{75}$$

# Predictive Kalman filter

Time-update (prediction)

$$\hat{x}_{t+1|t} = A\hat{x}_{t|t} + Bu_t, \qquad \hat{x}_{0|0} = \hat{x}_0, \tag{76}$$

$$P_{t+1|t} = AP_{t|t}A^T + R_1, \qquad P_{0|0} = P_0 \tag{77}$$

Data-update (inference)

$$\hat{x}_{t|t} = \hat{x}_{t|t-1} + \kappa_t(y_t - C\hat{x}_{t|t-1}), \tag{78}$$

$$\kappa_t = P_{t|t-1}C^T(CP_{t|t-1}C^T + R_2)^{-1}, \tag{79}$$

$$P_{t|t} = P_{t|t-1} - \kappa_t CP_{t|t-1} \tag{80}$$

**Predictive** Kalman filter

$$\hat{x}_{t+1|t} = (A - K_tC)\hat{x}_{t|t-1} + Bu_t + K_ty_t, \tag{81}$$

$$P_{t+1|t} = AP_{t|t-1}A^T + R_1 - K_tCP_{t|t-1}A^T, \tag{82}$$

$$K_t = A\kappa_t = AP_{t|t-1}C^T(CP_{t|t-1}C^T + R_2)^{-1} \tag{83}$$

# Stationary Kalman filters

# Stationary Kalman Filters

Stationary covariance of the predictive Kalman filter

$$\boxed{P_\infty^p = AP_\infty^p A^T + R_1 - AP_\infty^p C^T (CP_\infty^p C^T + R_2)^{-1} CP_\infty^p A^T} \tag{84}$$

Stationary covariance of the ordinary Kalman filter

$$\boxed{\begin{aligned} P_\infty^o = {} & AP_\infty^o A^T + R_1 - (AP_\infty^o A^T + R_1)C^T \\ & (C(AP_\infty^o A^T + R_1)C^T + R_2)^{-1} C(AP_\infty^o A^T + R_1) \end{aligned}} \tag{85}$$

Relation between stationary covariances

$$P_\infty^p = AP_\infty^o A^T + R_1 \tag{86}$$

$$(P_\infty^o)^{-1} = (P_\infty^p)^{-1} + C^T R_2^{-1} C \tag{87}$$

# Stationary Kalman Filter - The Riccati Equation

Discrete Riccati equation

$$X_{t+1} = AX_tA^T + R_1 - AX_tC^T(CX_tC^T + R_2)^{-1}CX_tA^T \tag{88}$$

Discrete algebraic Riccati equation (DARE)

$$X = AXA^T + R_1 - AXC^T(CXC^T + R_2)^{-1}CXA^T \tag{89}$$

- If $(A, C)$ is observable, a positive semi-definite solution $X$ exists for each $X_0$
- If $(A, C)$ is observable, $(A, R)$ is reachable $(RR^T = R_1)$, $R_1 \succeq 0$, and $R_2 \succ 0$, the solution is unique and independent of $X_0$ and $A - KC$ is asymptotically stable (its eigenvalues are strictly within the unit circle)

Hint: Use Matlab's `idare` function

# Estimation errors

# Kalman Errors

Estimation errors (if the model is correct)

$$\tilde{x}_{t|t} = x_t - \hat{x}_{t|t}, \qquad \tilde{x}_{t|t} \sim N(0, P_{t|t}), \tag{90}$$

$$\tilde{x}_{t|t-1} = x_t - \hat{x}_{t|t-1}, \qquad \tilde{x}_{t|t-1} \sim N(0, P_{t|t-1}), \tag{91}$$

$$\epsilon_t = y_t - C\hat{x}_{t|t-1}, \qquad \epsilon_t \sim N(0, CP_{t|t-1}C^T + R_2) \tag{92}$$

The innovation errors are white ($\epsilon_s \perp \epsilon_t$ for $s \neq t$) and can be used for

1. model validation (i.e., validating estimates of $A$, $B$, $\ldots$)
2. system representation
3. fault detection

Discrete-time systems for estimation errors

$$\tilde{x}_{t+1|t+1} = (I - \kappa_{t+1}C)(A\tilde{x}_{t|t} + v_t) - \kappa_{t+1}e_{t+1}, \tag{93}$$

$$\tilde{x}_{t+1|t} = (A - K_tC)\tilde{x}_{t|t-1} - K_te_t + v_t \tag{94}$$

The relation between the Kalman gains is $K_t = A\kappa_t$

# Example of Prediction Error

Discrete-time system

$$x_{t+1} = 0.5x_t + v_t, \qquad v_t \sim N(0, 0.1), \tag{95}$$

$$y_t = x_t + e_t, \qquad e_t \sim N(0, 0.5) \tag{96}$$

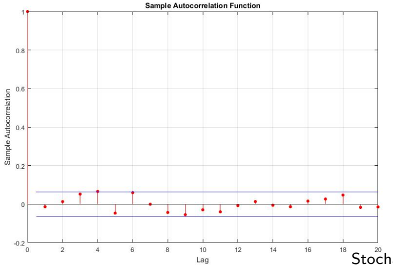
Prediction error

$$\epsilon_t \sim N(0, 0.625) \tag{97}$$

Empirical mean and variance

$$\mathbb{E}[\epsilon_t] = -0.0047, \qquad \mathrm{Var}(\epsilon_t) = 0.6218 \tag{98}$$

Empirical autocorrelation indicates that $\epsilon_t$ is white



# Theory - Standard Kalman Assumptions

The Kalman filter is designed for systems in the form

$$x_{t+1} = Ax_t + Bu_t + v_t, \quad (99)$$

$$y_t = Cx_t + e_t \quad (100)$$

It assumes the following noise distributions

1. $x_0 \sim N(\hat{x}_0, P_0)$
2. $v_t \sim N(0, P_v)$, white
3. $e_t \sim N(0, P_e)$, white
4. $\text{Cov}(v_t, e_t) = 0$
5. $v_t, e_t \perp x_s, \quad s \leq t$

In lecture 4, we will relax some of these assumptions

# Matlab example

# Questions

Questions?