Introduction to Nonlinear Control

Stability, control design, and estimation

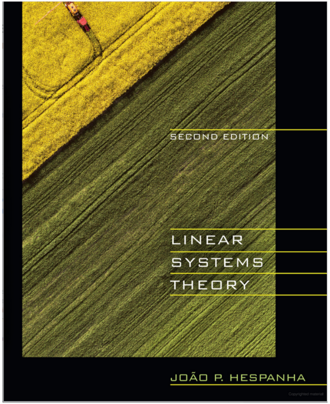
Christopher M. Kellett & Philipp Braun
School of Engineering,
Australian National University, Canberra, Australia

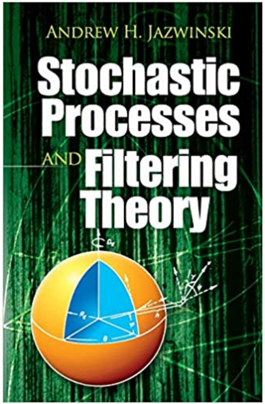
Part III:

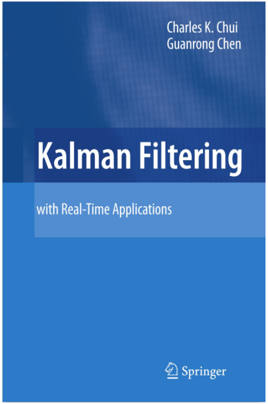
Chapter 16: Classical Observer Design

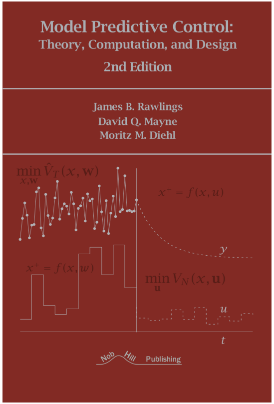
Australian National University

Classical Observer Design









Classical Observer Design

1. Luenberger Observers

2. Minimum Energy Estimator (Continuous time setting)

3. The discrete time Kalman filter
   - Least squares & minimum variance solution
   - A prediction-correction formulation
   - The steady-state Kalman filter

Classical Observer Design

So far:

- The concepts so far rely on the knowledge of the state $x \in \mathbb{R}^n$.
- The full state x is in general not known and only the output $y \in \mathbb{R}^p$ is available.

$\rightsquigarrow$ A controller design can not, in general, rely on the full state x.

$\rightsquigarrow$ An estimate $\hat{x}$ of the state needs to be derived (observability, detectibility)

- If $\hat{x}(t) \to x(t)$ for $t \to \infty$, can $\hat{x}$ be used for the definition of a feedback controller $u(\hat{x})$?

Classical Observer Design

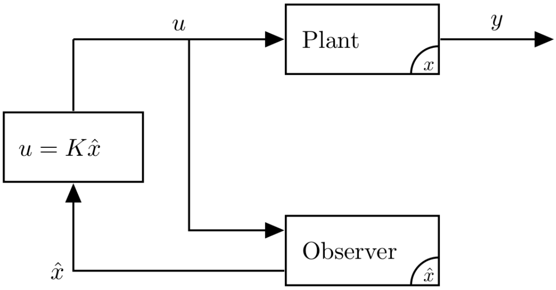
So far:

- The concepts so far rely on the knowledge of the state $x \in \mathbb{R}^n$.
- The full state x is in general not known and only the output $y \in \mathbb{R}^p$ is available.

$\rightsquigarrow$ A controller design can not, in general, rely on the full state x.

$\rightsquigarrow$ An estimate $\hat{x}$ of the state needs to be derived (observability, detectibility)

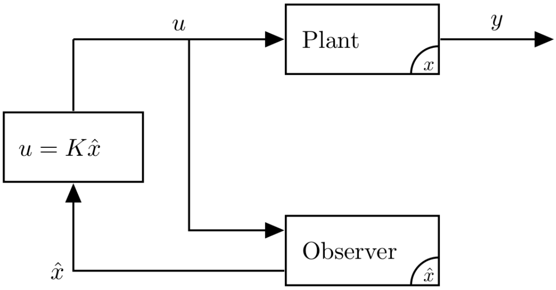
- If $\hat{x}(t) \to x(t)$ for $t \to \infty$, can $\hat{x}$ be used for the definition of a feedback controller $u(\hat{x})$?



Classical Observer Design

So far:

- The concepts so far rely on the knowledge of the state $x \in \mathbb{R}^n$.
- The full state x is in general not known and only the output $y \in \mathbb{R}^p$ is available.
- ⇝ A controller design can not, in general, rely on the full state x.
- ⇝ An estimate $\hat{x}$ of the state needs to be derived (observability, detectibility)
- If $\hat{x}(t) \to x(t)$ for $t \to \infty$, can $\hat{x}$ be used for the definition of a feedback controller $u(\hat{x})$?



Consider Linear systems:

$$\begin{aligned} \dot{x} &= Ax + Bu, \qquad x(0) \in \mathbb{R}^n, \\ y &= Cx + Du \end{aligned}$$

- We assume that $y \in \mathbb{R}^p$ and $u \in \mathbb{R}^m$ are known, while the internal state $x \in \mathbb{R}^n$ and the initial condition $x(0)$ are unknown.
- Assume that the matrix A is Hurwitz.

Classical Observer Design

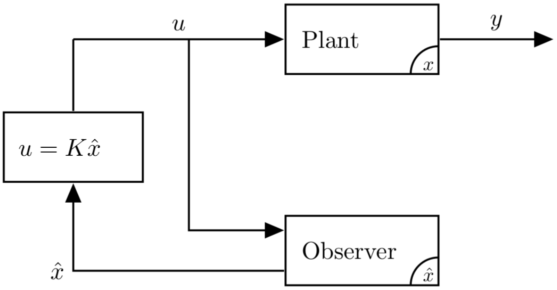
So far:

- The concepts so far rely on the knowledge of the state $x \in \mathbb{R}^n$.
- The full state x is in general not known and only the output $y \in \mathbb{R}^p$ is available.

⇝ A controller design can not, in general, rely on the full state x.

⇝ An estimate $\hat{x}$ of the state needs to be derived (observability, detectibility)

- If $\hat{x}(t) \to x(t)$ for $t \to \infty$, can $\hat{x}$ be used for the definition of a feedback controller $u(\hat{x})$?



Consider Linear systems:

$$\dot{x} = Ax + Bu, \qquad x(0) \in \mathbb{R}^n,$$
$$y = Cx + Du$$

- We assume that $y \in \mathbb{R}^p$ and $u \in \mathbb{R}^m$ are known, while the internal state $x \in \mathbb{R}^n$ and the initial condition $x(0)$ are unknown.
- Assume that the matrix A is Hurwitz.

- Introduce observer dynamics as a copy of the system

$$\dot{\hat{x}} = A\hat{x} + Bu, \qquad \hat{x}(0) \in \mathbb{R}^n$$

  - $\hat{x} \in \mathbb{R}^n$ estimate of the state $x \in \mathbb{R}^n$
  - Estimation error $e = x - \hat{x}$

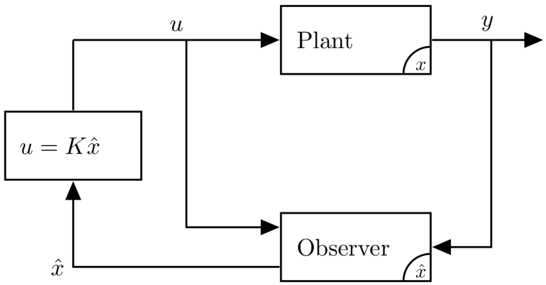
- Error dynamics:

$$\dot{e} = \dot{x} - \dot{\hat{x}} = Ax + Bu - A\hat{x} - Bu = A(x - \hat{x}) = Ae$$
$$\hat{x}(t) \to x(t) \quad \Leftrightarrow \quad e(t) \to 0 \quad \Leftrightarrow A \text{ Hurwitz}$$

Section 1

Luenberger Observers

Luenberger Observers



Consider

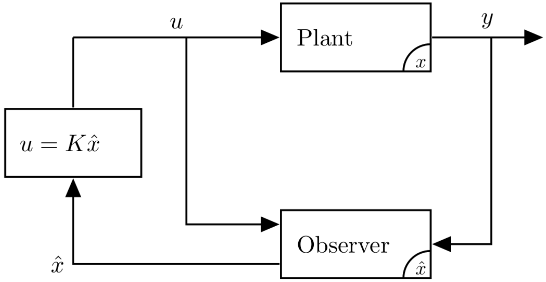
$$\dot{x} = Ax + Bu, \qquad x(0) \in \mathbb{R}^n, \qquad (\text{or } x^+ = Ax + Bu)$$
$$y = Cx + Du$$

Define observer dynamics:

$$\dot{\hat{x}} = A\hat{x} + Bu - L(y - \hat{y}),$$
$$\hat{y} = C\hat{x} + Du.$$

- *output injection term* $L \in \mathbb{R}^{n \times p}$

Luenberger Observers



Consider

$$\dot{x} = Ax + Bu, \qquad x(0) \in \mathbb{R}^n, \qquad (\text{or } x^+ = Ax + Bu)$$
$$y = Cx + Du$$

Define observer dynamics:

$$\dot{\hat{x}} = A\hat{x} + Bu - L(y - \hat{y}),$$
$$\hat{y} = C\hat{x} + Du.$$

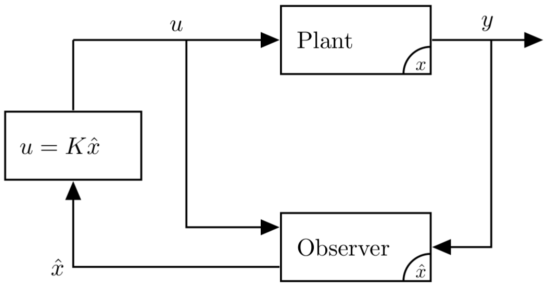
- *output injection term $L \in \mathbb{R}^{n \times p}$*

- Estimation error: $e = x - \hat{x}$
- Error dynamics:

$$\begin{aligned} \dot{e} &= \dot{x} - \dot{\hat{x}} \\ &= Ax + Bu - A\hat{x} - Bu + L(Cx + Du - C\hat{x} - Du) \\ &= Ae + LCe = (A + LC)e \end{aligned}$$

⇝ The error dynamics are independent of u

Luenberger Observers



Consider

$$\dot{x} = Ax + Bu, \qquad x(0) \in \mathbb{R}^n, \qquad (\text{or } x^+ = Ax + Bu)$$
$$y = Cx + Du$$

Define observer dynamics:

$$\dot{\hat{x}} = A\hat{x} + Bu - L(y - \hat{y}),$$
$$\hat{y} = C\hat{x} + Du.$$

- *output injection term* $L \in \mathbb{R}^{n \times p}$

- Estimation error: $e = x - \hat{x}$
- Error dynamics:

$$\begin{aligned} \dot{e} &= \dot{x} - \dot{\hat{x}} \\ &= Ax + Bu - A\hat{x} - Bu + L(Cx + Du - C\hat{x} - Du) \\ &= Ae + LCe = (A + LC)e \end{aligned}$$

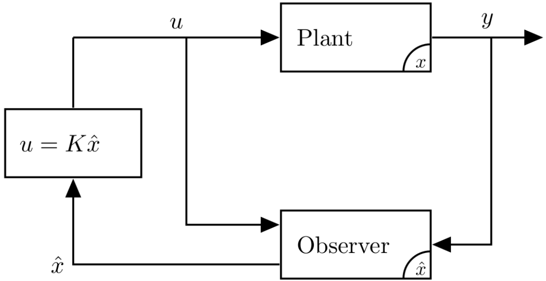
⇝ The error dynamics are independent of u

- $A + LC$ has the same eigenvalues as $(A + LC)^T = A^T + C^T L^T$

⇝ If (A, C) is observable, the poles of $A + LC$ can be placed arbitrarily, i.e., L can be defined such that $A + LC$ is Hurwitz.

⇝ If (A, C) is detectable, then there exists L such that $A + LC$ is Hurwitz.

Luenberger Observers



Consider

$$\dot{x} = Ax + Bu, \qquad x(0) \in \mathbb{R}^n, \qquad (\text{or } x^+ = Ax + Bu)$$
$$y = Cx + Du$$

Define observer dynamics:

$$\dot{\hat{x}} = A\hat{x} + Bu - L(y - \hat{y}),$$
$$\hat{y} = C\hat{x} + Du.$$

- *output injection term $L \in \mathbb{R}^{n \times p}$*

- **Estimation error:** $e = x - \hat{x}$
- **Error dynamics:**

$$\begin{aligned} \dot{e} &= \dot{x} - \dot{\hat{x}} \\ &= Ax + Bu - A\hat{x} - Bu + L(Cx + Du - C\hat{x} - Du) \\ &= Ae + LCe = (A + LC)e \end{aligned}$$

⇝ The error dynamics are independent of u

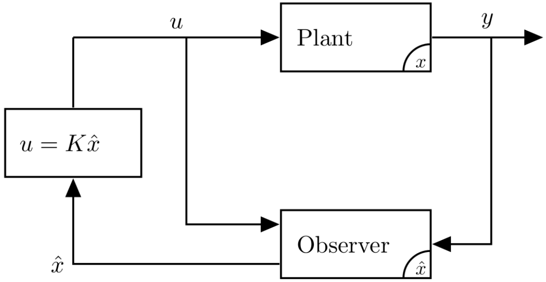
- $A + LC$ has the same eigenvalues as $(A + LC)^T = A^T + C^T L^T$

⇝ If (A, C) is observable, the poles of $A + LC$ can be placed arbitrarily, i.e., L can be defined such that $A + LC$ is Hurwitz.

⇝ If (A, C) is detectable, then there exists L such that $A + LC$ is Hurwitz.

- See pole placement
- x can be approximated through $\hat{x}$

Luenberger Observers



Consider

$$\dot{x} = Ax + Bu, \qquad x(0) \in \mathbb{R}^n, \qquad (\text{or } x^+ = Ax + Bu)$$
$$y = Cx + Du$$

Define observer dynamics:

$$\dot{\hat{x}} = A\hat{x} + Bu - L(y - \hat{y}),$$
$$\hat{y} = C\hat{x} + Du.$$

- *output injection term $L \in \mathbb{R}^{n \times p}$*

- Estimation error: $e = x - \hat{x}$
- Error dynamics:

$$\begin{aligned}\dot{e} &= \dot{x} - \dot{\hat{x}} \\ &= Ax + Bu - A\hat{x} - Bu + L(Cx + Du - C\hat{x} - Du) \\ &= Ae + LCe = (A + LC)e\end{aligned}$$

⇝ The error dynamics are independent of u

- $A + LC$ has the same eigenvalues as $(A + LC)^T = A^T + C^T L^T$

⇝ If (A, C) is observable, the poles of $A + LC$ can be placed arbitrarily, i.e., L can be defined such that $A + LC$ is Hurwitz.

⇝ If (A, C) is detectable, then there exists L such that $A + LC$ is Hurwitz.

- See pole placement
- x can be approximated through $\hat{x}$
- Controller design $u = K\hat{x}$?

Luenberger Observers & Controller design

Consider

$$\dot{x} = Ax + Bu, \qquad x(0) \in \mathbb{R}^n$$
$$y = Cx + Du$$

Define observer dynamics:

$$\dot{\hat{x}} = A\hat{x} + Bu - L(y - \hat{y}),$$
$$\hat{y} = C\hat{x} + Du.$$

Error dynamics:

$$\dot{e} = (A + LC)e$$

Controller design:

$$u(\hat{x}) = K\hat{x}, \qquad K \in \mathbb{R}^{m \times n}$$

Luenberger Observers & Controller design

Consider

$$\begin{aligned} \dot{x} &= Ax + Bu, \qquad x(0) \in \mathbb{R}^n \\ y &= Cx + Du \end{aligned}$$

Define observer dynamics:

$$\begin{aligned} \dot{\hat{x}} &= A\hat{x} + Bu - L(y - \hat{y}), \\ \hat{y} &= C\hat{x} + Du. \end{aligned}$$

Error dynamics:

$$\dot{e} = (A + LC)e$$

Controller design:

$$u(\hat{x}) = K\hat{x}, \qquad K \in \mathbb{R}^{m \times n}$$

We can rewrite the plant dynamics:

$$\begin{aligned} \dot{x} &= Ax + Bu(\hat{x}) = Ax + BK\hat{x} \\ &= Ax + BK(x + e) \\ &= (A + BK)x + BKe. \end{aligned}$$

Luenberger Observers & Controller design

Consider

$$\dot{x} = Ax + Bu, \qquad x(0) \in \mathbb{R}^n$$
$$y = Cx + Du$$

Define observer dynamics:

$$\dot{\hat{x}} = A\hat{x} + Bu - L(y - \hat{y}),$$
$$\hat{y} = C\hat{x} + Du.$$

Error dynamics:

$$\dot{e} = (A + LC)e$$

Controller design:

$$u(\hat{x}) = K\hat{x}, \qquad K \in \mathbb{R}^{m \times n}$$

We can rewrite the plant dynamics:

$$\begin{aligned} \dot{x} &= Ax + Bu(\hat{x}) = Ax + BK\hat{x} \\ &= Ax + BK(x + e) \\ &= (A + BK)x + BKe. \end{aligned}$$

Overall closed loop system

$$\begin{bmatrix} \dot{x} \\ \dot{e} \end{bmatrix} = \begin{bmatrix} A + BK & BK \\ 0 & A + LC \end{bmatrix} \begin{bmatrix} x \\ e \end{bmatrix}$$

- If (A, B) is controllable and (A, C) is observable, we can place the poles of the closed-loop system arbitrarily by choosing K and L.
- The convergence $|x(t)| \to 0$ and $|e(t)| \to 0$ for $t \to \infty$ can be guaranteed by designing L and K individually. $\rightsquigarrow$ *separation principle*
- (The separation principle is only true for the asymptotic behavior)

Luenberger Observers & Controller design

Consider

$$\dot{x} = Ax + Bu, \qquad x(0) \in \mathbb{R}^n$$
$$y = Cx + Du$$

Define observer dynamics:

$$\dot{\hat{x}} = A\hat{x} + Bu - L(y - \hat{y}),$$
$$\hat{y} = C\hat{x} + Du.$$

Error dynamics:

$$\dot{e} = (A + LC)e$$

Controller design:

$$u(\hat{x}) = K\hat{x}, \qquad K \in \mathbb{R}^{m \times n}$$

We can rewrite the plant dynamics:

$$\begin{aligned}\dot{x} &= Ax + Bu(\hat{x}) = Ax + BK\hat{x} \\ &= Ax + BK(x + e) \\ &= (A + BK)x + BKe.\end{aligned}$$

Overall closed loop system

$$\begin{bmatrix} \dot{x} \\ \dot{e} \end{bmatrix} = \begin{bmatrix} A + BK & BK \\ 0 & A + LC \end{bmatrix} \begin{bmatrix} x \\ e \end{bmatrix}$$

- If (A, B) is controllable and (A, C) is observable, we can place the poles of the closed-loop system arbitrarily by choosing K and L.
- The convergence $|x(t)| \to 0$ and $|e(t)| \to 0$ for $t \to \infty$ can be guaranteed by designing L and K individually. $\rightsquigarrow$ *separation principle*
- (The separation principle is only true for the asymptotic behavior)

Alternative representation in terms of x and $\hat{x}$:

$$\begin{bmatrix} \dot{x} \\ \dot{\hat{x}} \end{bmatrix} = \begin{bmatrix} A & BK \\ -LC & A + BK + LC \end{bmatrix} \begin{bmatrix} x \\ \hat{x} \end{bmatrix}$$

$\rightsquigarrow$ While the separation principle is not visible the dynamics capture the same information.

Luenberger Observers & Controller design (Linearization pendulum; upright position)

Example

- Linearization of the pendulum on a cart in the upright position

$$A = \begin{bmatrix} 0 & 0 & 1.00 & 0 \\ 0 & 0 & 0 & 1.00 \\ 0 & 3.27 & -0.07 & -0.03 \\ 0 & 6.54 & -0.03 & -0.07 \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ 0 \\ 0.67 \\ 0.33 \end{bmatrix}$$

with output

$$C = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix};$$

i.e., only the position of the cart and the angle of the pendulum are available as measurements.

- Feedback gain

$$K = \begin{bmatrix} 7.34 & -140.84 & 15.47 & -60.54 \end{bmatrix}$$

ensures that the closed loop matrix $A + BK$ is Hurwitz and has the eigenvalues $\{-4, -3, -2, -1\}$.

- Initial conditions:

$$x_0 = [1, 1, 1, 1]^T, \qquad \hat{x} = [1, 1, 0, 0]^T$$

The observer gain

$$L = \begin{bmatrix} -2.90 & -1.07 \\ -3.75 & -6.49 \\ -2.58 & -6.96 \\ -8.53 & -16.64 \end{bmatrix}$$

ensures that $A+LC$ has the same eigenvalues as $A+BK$.

Luenberger Observers & Controller design (Linearization pendulum; upright position)

Example

- Linearization of the pendulum on a cart in the upright position

$$A = \begin{bmatrix} 0 & 0 & 1.00 & 0 \\ 0 & 0 & 0 & 1.00 \\ 0 & 3.27 & -0.07 & -0.03 \\ 0 & 6.54 & -0.03 & -0.07 \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ 0 \\ 0.67 \\ 0.33 \end{bmatrix}$$

with output

$$C = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix};$$

i.e., only the position of the cart and the angle of the pendulum are available as measurements.

- Feedback gain

$$K = \begin{bmatrix} 7.34 & -140.84 & 15.47 & -60.54 \end{bmatrix}$$

ensures that the closed loop matrix $A + BK$ is Hurwitz and has the eigenvalues $\{-4, -3, -2, -1\}$.

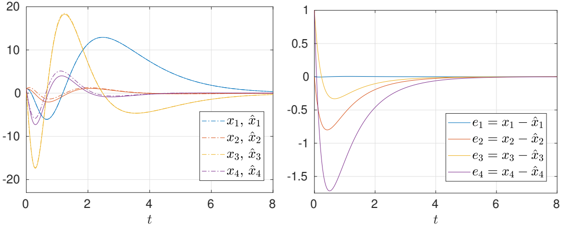
- Initial conditions:

$$x_0 = [1, 1, 1, 1]^T, \qquad \hat{x} = [1, 1, 0, 0]^T$$

The observer gain

$$L = \begin{bmatrix} -2.90 & -1.07 \\ -3.75 & -6.49 \\ -2.58 & -6.96 \\ -8.53 & -16.64 \end{bmatrix}$$

ensures that $A + LC$ has the same eigenvalues as $A + BK$.



Luenberger Observers & Controller design (Linearization pendulum; upright position)

Example

- Linearization of the pendulum on a cart in the upright position

$$A = \begin{bmatrix} 0 & 0 & 1.00 & 0 \\ 0 & 0 & 0 & 1.00 \\ 0 & 3.27 & -0.07 & -0.03 \\ 0 & 6.54 & -0.03 & -0.07 \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ 0 \\ 0.67 \\ 0.33 \end{bmatrix}$$

with output

$$C = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix};$$

i.e., only the position of the cart and the angle of the pendulum are available as measurements.

- Feedback gain

$$K = \begin{bmatrix} 7.34 & -140.84 & 15.47 & -60.54 \end{bmatrix}$$

ensures that the closed loop matrix $A + BK$ is Hurwitz and has the eigenvalues $\{-4, -3, -2, -1\}$.

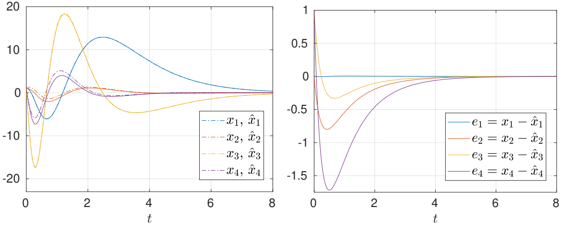
- Initial conditions:

$$x_0 = [1, 1, 1, 1]^T, \qquad \hat{x} = [1, 1, 0, 0]^T$$

The observer gain

$$L = \begin{bmatrix} -2.90 & -1.07 \\ -3.75 & -6.49 \\ -2.58 & -6.96 \\ -8.53 & -16.64 \end{bmatrix}$$

ensures that $A + LC$ has the same eigenvalues as $A + BK$.



Note that:

- The convergence is independent of the initial condition $x_0, \hat{x}_0$ since for linear systems local results are also global and the stability properties of the linear system solely depend on the properties of the closed loop matrix.

Section 2

Minimum Energy Estimator (Continuous time setting)

Minimum Energy Estimator (Continuous time setting)

Recall

- We have discussed optimal control for linear systems
- Can we derive an optimal estimator (in terms of minimal energy)?

Minimum Energy Estimator (Continuous time setting)

Recall

- We have discussed optimal control for linear systems
- Can we derive an optimal estimator (in terms of minimal energy)?

Consider a perturbed linear system:

$$\dot{x} = Ax + Bu + \bar{B}v$$
$$y = Cx + Du + w$$

- *disturbance*: $v(\cdot) : \mathbb{R} \to \mathbb{R}^q$ (model uncertainty)
- *measurement noise*: $w(\cdot) : \mathbb{R} \to \mathbb{R}^p$ (depending on the sensors)

Minimum Energy Estimator (Continuous time setting)

Recall

- We have discussed optimal control for linear systems
- Can we derive an optimal estimator (in terms of minimal energy)?

Consider a perturbed linear system:

$$\dot{x} = Ax + Bu + \bar{B}v$$
$$y = Cx + Du + w$$

- *disturbance*: $v(\cdot) : \mathbb{R} \to \mathbb{R}^q$ (model uncertainty)
- *measurement noise*: $w(\cdot) : \mathbb{R} \to \mathbb{R}^p$ (depending on the sensors)

The *minimum energy estimation* problem:

- For given $u(\cdot)$, $y(\cdot)$, find $\bar{x} : \mathbb{R}_{\leq t_0} \to \mathbb{R}^n$ for $t_0 \geq 0$, which satisfies the dynamics

$$\dot{\bar{x}} = A\bar{x} + Bu + \bar{B}v$$
$$y = C\bar{x} + Du + w$$

an which minimizes the cost function

$$J_{\text{MEE}}(\bar{x}(t_0), v(\cdot)) = \int_{-\infty}^{t_0} w(\tau)^T Q w(\tau) + v(\tau)^T R v(\tau) d\tau$$

Minimum Energy Estimator (Continuous time setting)

Recall

- We have discussed optimal control for linear systems
- Can we derive an optimal estimator (in terms of minimal energy)?

Consider a perturbed linear system:

$$\dot{x} = Ax + Bu + \bar{B}v$$
$$y = Cx + Du + w$$

- *disturbance*: $v(\cdot) : \mathbb{R} \to \mathbb{R}^q$ (model uncertainty)
- *measurement noise*: $w(\cdot) : \mathbb{R} \to \mathbb{R}^p$ (depending on the sensors)

The *minimum energy estimation* problem:

- For given $u(\cdot)$, $y(\cdot)$, find $\bar{x} : \mathbb{R}_{\leq t_0} \to \mathbb{R}^n$ for $t_0 \geq 0$, which satisfies the dynamics

$$\dot{\bar{x}} = A\bar{x} + Bu + \bar{B}v$$
$$y = C\bar{x} + Du + w$$

an which minimizes the cost function

$$J_{\text{MEE}}(\bar{x}(t_0), v(\cdot)) = \int_{-\infty}^{t_0} w(\tau)^T Q w(\tau) + v(\tau)^T R v(\tau) d\tau$$

Note that:

- Design parameters: $Q \in \mathcal{S}^p_{>0}$, $R \in \mathcal{S}^q_{>0}$
- $J_{\text{MEE}}(\bar{x}(t_0), v(\cdot))$ is a function of $v(\cdot)$ but not $w(\cdot)$:

$$J_{\text{MEE}}(\bar{x}(t_0), v(\cdot)) = \int_{-\infty}^{t_0} v^T R v + (C\bar{x} + Du - y)^T Q (C\bar{x} + Du - y) d\tau$$

$\rightsquigarrow$ Given past $u(\cdot)$ and $y(\cdot)$, the optimization problem aims to find the disturbance $v(\cdot)$ with minimum energy and an estimated state $\bar{x}(t_0)$ that explains the observed inputs and outputs.

Minimum Energy Estimator (Continuous time setting)

Recall

- We have discussed optimal control for linear systems
- Can we derive an optimal estimator (in terms of minimal energy)?

Consider a perturbed linear system:

$$\dot{x} = Ax + Bu + \bar{B}v$$
$$y = Cx + Du + w$$

- *disturbance*: $v(\cdot) : \mathbb{R} \to \mathbb{R}^q$ (model uncertainty)
- *measurement noise*: $w(\cdot) : \mathbb{R} \to \mathbb{R}^p$ (depending on the sensors)

The *minimum energy estimation* problem:

- For given $u(\cdot)$, $y(\cdot)$, find $\bar{x} : \mathbb{R}_{\leq t_0} \to \mathbb{R}^n$ for $t_0 \geq 0$, which satisfies the dynamics

$$\dot{\bar{x}} = A\bar{x} + Bu + \bar{B}v$$
$$y = C\bar{x} + Du + w$$

an which minimizes the cost function

$$J_{\mathrm{MEE}}(\bar{x}(t_0), v(\cdot)) = \int_{-\infty}^{t_0} w(\tau)^T Q w(\tau) + v(\tau)^T R v(\tau) d\tau$$

Note that:

- Design parameters: $Q \in \mathcal{S}^p_{>0}$, $R \in \mathcal{S}^q_{>0}$
- $J_{\mathrm{MEE}}(\bar{x}(t_0), v(\cdot))$ is a function of $v(\cdot)$ but not $w(\cdot)$:

$$J_{\mathrm{MEE}}(\bar{x}(t_0), v(\cdot)) = \int_{-\infty}^{t_0} v^T R v + (C\bar{x} + Du - y)^T Q (C\bar{x} + Du - y) d\tau$$

⇝ Given past $u(\cdot)$ and $y(\cdot)$, the optimization problem aims to find the disturbance $v(\cdot)$ with minimum energy and an estimated state $\bar{x}(t_0)$ that explains the observed inputs and outputs.

- To ensure that the problem is well-defined we assume

$$\bar{x}(t) \to 0,\ u(t) \to 0,\ w(t) \to 0,\ y(t) \to 0 \text{ for } t \to -\infty$$

- Q large: penalize noise $w(\cdot)$; neglect disturbance
- R large: penalize disturbance $v(\cdot)$; neglect noise

Minimum Energy Estimator (Continuous time setting)

Optimization problem

$$V_{\mathrm{MEE}}(\bar{x}_0) = \min_{v(\cdot):\mathbb{R}\to\mathbb{R}^q} J_{\mathrm{MEE}}(\bar{x}_0, v(\cdot))$$

$$\text{subject to } \dot{x} = Ax + Bu + \bar{B}v$$

Additionally define

$\bar{\mathcal{X}} = \{x : \mathbb{R}_{\leq t_0} \to \mathbb{R}^n\}$ and $\mathcal{V} = \{v : \mathbb{R}_{\leq t_0} \to \mathbb{R}^q\}$.

Minimum Energy Estimator (Continuous time setting)

Optimization problem

$$V_{\mathrm{MEE}}(\bar{x}_0) = \min_{v(\cdot):\mathbb{R}\to\mathbb{R}^q} J_{\mathrm{MEE}}(\bar{x}_0, v(\cdot))$$

$$\text{subject to } \dot{x} = Ax + Bu + \bar{B}v$$

Additionally define

$\bar{\mathcal{X}} = \{x : \mathbb{R}_{\leq t_0} \to \mathbb{R}^n\}$ and $\mathcal{V} = \{v : \mathbb{R}_{\leq t_0} \to \mathbb{R}^q\}$.

Definition (Feedback invariant)

Consider the system

$$\dot{\bar{x}} = A\bar{x} + Bu + \bar{B}v$$
$$y = C\bar{x} + Du + w,$$

$t_0 \in \mathbb{R}_{\geq 0}$ and fixed $u(\cdot) : \mathbb{R}_{<t_0} \to \mathbb{R}^m$, $y(\cdot) : \mathbb{R}_{<t_0} \to \mathbb{R}^p$.
A functional $H : \bar{\mathcal{X}} \times \mathcal{V} \to \mathbb{R}$ is called feedback invariant if for all solution pairs $(\bar{x}_1(\cdot), v_1(\cdot)), (\bar{x}_2(\cdot), v_2(\cdot)) \in \bar{\mathcal{X}} \times \mathcal{V}$ with $\bar{x}_1(t_0) = \bar{x}_2(t_0)$ the equation

$$H(\bar{x}_1(\cdot), v_1(\cdot)) = H(\bar{x}_2(\cdot), v_2(\cdot))$$

holds.

Minimum Energy Estimator (Continuous time setting)

Optimization problem

$$V_{\text{MEE}}(\bar{x}_0) = \min_{v(\cdot):\mathbb{R}\to\mathbb{R}^q} J_{\text{MEE}}(\bar{x}_0, v(\cdot))$$

$$\text{subject to } \dot{x} = Ax + Bu + \bar{B}v$$

Additionally define

$\bar{\mathcal{X}} = \{x : \mathbb{R}_{\leq t_0} \to \mathbb{R}^n\}$ and $\mathcal{V} = \{v : \mathbb{R}_{\leq t_0} \to \mathbb{R}^q\}$.

Definition (Feedback invariant)

Consider the system

$$\dot{\bar{x}} = A\bar{x} + Bu + \bar{B}v$$
$$y = C\bar{x} + Du + w,$$

$t_0 \in \mathbb{R}_{\geq 0}$ and fixed $u(\cdot) : \mathbb{R}_{<t_0} \to \mathbb{R}^m$, $y(\cdot) : \mathbb{R}_{<t_0} \to \mathbb{R}^p$. A functional $H : \bar{\mathcal{X}} \times \mathcal{V} \to \mathbb{R}$ is called feedback invariant if for all solution pairs $(\bar{x}_1(\cdot), v_1(\cdot)), (\bar{x}_2(\cdot), v_2(\cdot)) \in \bar{\mathcal{X}} \times \mathcal{V}$ with $\bar{x}_1(t_0) = \bar{x}_2(t_0)$ the equation

$$H(\bar{x}_1(\cdot), v_1(\cdot)) = H(\bar{x}_2(\cdot), v_2(\cdot))$$

holds.

Theorem (Feedback invariant)

Consider the linear system for a $y(\cdot) : \mathbb{R}_{\leq t_0} \to \mathbb{R}^p$ and $u(\cdot) : \mathbb{R}_{\leq t_0} \to \mathbb{R}^m$ for $t_0 > 0$. Then, for every symmetric matrix $\bar{P} \in \mathcal{S}^n$, differentiable signal $\beta(\cdot) : \mathbb{R}_{\leq t_0} \to \mathbb{R}^n$, and a scalar $H_0 \in \mathbb{R}^n$ (which does not depend on $\bar{x}(\cdot)$ and $v(\cdot)$), the functional

$$\begin{aligned} H(\bar{x}(\cdot), v(\cdot)) &= H_0 \\ &+ \int_{-\infty}^{t_0} \left(A\bar{x} + Bu + \bar{B}v - \dot{\beta}\right)^T P\,(\bar{x} - \beta) \\ &+ (\bar{x} - \beta)^T P \left(A\bar{x} + Bu + \bar{B}v - \dot{\beta}\right) d\tau \\ &- (\bar{x}(t_0) - \beta(t_0))^T P\,(\bar{x}(t_0) - \beta(t_0)) \end{aligned}$$

is a feedback invariant as long as $\lim_{t\to-\infty}(\bar{x}(t) - \beta(t)) = 0$.

Minimum Energy Estimator (Continuous time setting)

- Perturbed linear system:

$$\dot{x} = Ax + Bu + \bar{B}v$$
$$y = Cx + Du + w$$

- Optimization problem

$$V_{\text{MEE}}(\bar{x}_0) = \min_{v(\cdot):\mathbb{R}\to\mathbb{R}^q} J_{\text{MEE}}(\bar{x}_0, v(\cdot))$$
$$\text{subject to } \dot{x} = Ax + Bu + \bar{B}v$$

⇝ 'Optimal disturbance' $v(\cdot) : \mathbb{R} \to \mathbb{R}^q$:

- 'Optimal disturbance' defines 'optimal estimated dynamics':

$$\dot{\bar{x}} = A\bar{x} + Bu + \bar{B}v$$
$$y = C\bar{x} + Du + w$$

Minimum Energy Estimator (Continuous time setting)

- Perturbed linear system:

$$\dot{x} = Ax + Bu + \bar{B}v$$
$$y = Cx + Du + w$$

- Optimization problem

$$V_{\mathrm{MEE}}(\bar{x}_0) = \min_{v(\cdot):\mathbb{R}\to\mathbb{R}^q} J_{\mathrm{MEE}}(\bar{x}_0, v(\cdot))$$
$$\text{subject to } \dot{x} = Ax + Bu + \bar{B}v$$

⇝ 'Optimal disturbance' $v(\cdot) : \mathbb{R} \to \mathbb{R}^q$:

- 'Optimal disturbance' defines 'optimal estimated dynamics':

$$\dot{\bar{x}} = A\bar{x} + Bu + \bar{B}v$$
$$y = C\bar{x} + Du + w$$

Theorem (The minimum energy estimator)

- *Consider the perturbed linear system and assume that* $(A, \bar{B})$ *is controllable and* (A, C) *is detectable.*
- *Consider the optimization problem where the cost function is defined through positive definite matrices* $Q \in \mathcal{S}^p_{>0}$ *and* $R \in \mathcal{S}^q_{>0}$.
- *Then there exists* $S \in \mathcal{S}^n_{>0}$ *to the dual algebraic Riccati equation*

$$AS + SA^T + \bar{B}R^{-1}\bar{B}^T - SC^TQCS = 0$$

such that $A - LC$ *is Hurwitz, where* $L = SC^TQ$.

- *The minimum energy estimator is given by*

$$\dot{\hat{x}} = A\hat{x} + Bu + L(y - C\hat{x} - Du)$$

and the initial condition $\hat{x}(t_0) = \bar{x}_0$, $t_0 \in \mathbb{R}_{\geq 0}$.

Minimum Energy Estimator (Continuous time setting)

Example (Pendulum)

The linearization at the stable equilibrium $[x_1^e, x_2^e]^T = [\theta^e, \dot{\theta}^e]^T = [\pi, 0]^T$;

$$\dot{x} = Ax + Bu + \bar{B}v$$
$$y = Cx + w$$

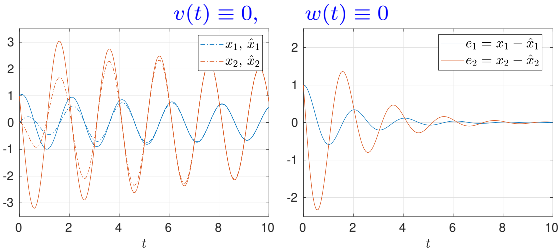
$$A = \begin{bmatrix} 0 & 1 \\ -\frac{mg\ell}{J+m\ell^2} & -\frac{\gamma}{J+m\ell^2} \end{bmatrix}, \; B = \begin{bmatrix} 0 \\ \frac{\ell}{J+m\ell^2} \end{bmatrix}, \; C = \begin{bmatrix} 1 & 0 \end{bmatrix}$$

$$\bar{B} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \; Q = 1, \; R = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Constants: $m = \ell = 1$, $J = 0$, $g = 9.81$, and $\gamma = 0.1$.
Observer gain:

$$L = \begin{bmatrix} 0.9548 \\ -0.0441 \end{bmatrix}$$

Eigenvalues of $A - LC$: $\lambda_{1,2} = -0.5274 \pm 3.0957j$
Initialization: $x_0 = [1, 1]^T$ and $\hat{x}_0 = [0, 0]^T$



Minimum Energy Estimator (Continuous time setting)

Example (Pendulum)

The linearization at the stable equilibrium $[x_1^e, x_2^e]^T = [\theta^e, \dot{\theta}^e]^T = [\pi, 0]^T$;

$$\dot{x} = Ax + Bu + \bar{B}v$$
$$y = Cx + w$$

$$A = \begin{bmatrix} 0 & 1 \\ -\frac{mg\ell}{J+m\ell^2} & -\frac{\gamma}{J+m\ell^2} \end{bmatrix}, \; B = \begin{bmatrix} 0 \\ \frac{\ell}{J+m\ell^2} \end{bmatrix}, \; C = \begin{bmatrix} 1 & 0 \end{bmatrix}$$

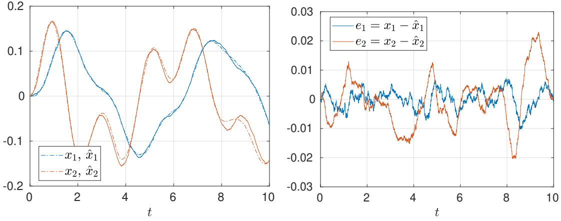
$$\bar{B} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \; Q = 1, \; R = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

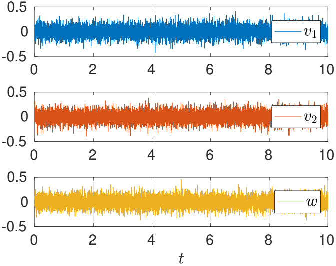
Constants: $m = \ell = 1$, $J = 0$, $g = 9.81$, and $\gamma = 0.1$.
Observer gain:

$$L = \begin{bmatrix} 0.9548 \\ -0.0441 \end{bmatrix}$$

Eigenvalues of $A - LC$: $\lambda_{1,2} = -0.5274 \pm 3.0957j$
Initialization: $x_0 = [1, 1]^T$ and $\hat{x}_0 = [0, 0]^T$

v, w Gaussian white noise, zero mean, standard dev. 0.1





Minimum Energy Estimator (Continuous time setting)

Concluding remarks:

- Here, we have derived the *minimum energy estimator* using a *deterministic setting*
- In the *stochastic setting* the minimum energy estimator is known as (cont. time) Kalman filter.
- Under certain assumptions on disturbances $v(t)$, $w(t)$ in the system dynamics

$$\begin{aligned} \dot{x} &= Ax + Bu + \bar{B}v \\ y &= Cx + Du + w \end{aligned}$$

  equivalences between the minimum energy estimator and the Kalman filter can be derived.

Minimum Energy Estimator (Continuous time setting)

Concluding remarks:

- Here, we have derived the *minimum energy estimator* using a *deterministic setting*
- In the *stochastic setting* the minimum energy estimator is known as (cont. time) Kalman filter.
- Under certain assumptions on disturbances $v(t)$, $w(t)$ in the system dynamics

$$\begin{aligned} \dot{x} &= Ax + Bu + \bar{B}v \\ y &= Cx + Du + w \end{aligned}$$

  equivalences between the minimum energy estimator and the Kalman filter can be derived.
- In particular, assume $v(\cdot)$ and $w(\cdot)$ represent functions of zero-mean Gaussian white noise with covariance matrices satisfying

$$\begin{aligned} \mathrm{E}[v(t)v(\tau)^T] &= \delta(t-\tau)R^{-1}, \\ \mathrm{E}[w(t)w(\tau)^T] &= \delta(t-\tau)Q^{-1}, \end{aligned}$$

  for all $t, \tau \in \mathbb{R}$ and $Q \in \mathcal{S}^p_{>0}$, $R \in \mathcal{S}^q_{>0}$.
- Additionally, $\mathrm{E}[v(t)w(\tau)^T] = 0 \qquad \forall\, t, \tau \in \mathbb{R}$.

Here:

- *expected value*: $\mathrm{E}[\cdot]$:
- *Dirac delta function*: $\delta : \mathbb{R} \to \mathbb{R} \cup \{\infty\}$

$$\delta(t) = \begin{cases} \infty, & t = 0 \\ 0, & t \neq 0 \end{cases} \qquad \text{and} \qquad \int_{-\infty}^{\infty} \delta(t)dt = 1.$$

Minimum Energy Estimator (Continuous time setting)

Concluding remarks:

- Here, we have derived the *minimum energy estimator* using a *deterministic setting*
- In the *stochastic setting* the minimum energy estimator is known as (cont. time) Kalman filter.
- Under certain assumptions on disturbances $v(t)$, $w(t)$ in the system dynamics

$$\dot{x} = Ax + Bu + \bar{B}v$$
$$y = Cx + Du + w$$

equivalences between the minimum energy estimator and the Kalman filter can be derived.

- In particular, assume $v(\cdot)$ and $w(\cdot)$ represent functions of zero-mean Gaussian white noise with covariance matrices satisfying

$$\mathrm{E}[v(t)v(\tau)^T] = \delta(t-\tau)R^{-1},$$
$$\mathrm{E}[w(t)w(\tau)^T] = \delta(t-\tau)Q^{-1},$$

for all $t, \tau \in \mathbb{R}$ and $Q \in \mathcal{S}^p_{>0}$, $R \in \mathcal{S}^q_{>0}$.

- Additionally, $\mathrm{E}[v(t)w(\tau)^T] = 0 \qquad \forall\, t, \tau \in \mathbb{R}$.

Here:

- *expected value*: $\mathrm{E}[\cdot]$:
- *Dirac delta function*: $\delta : \mathbb{R} \to \mathbb{R} \cup \{\infty\}$

$$\delta(t) = \begin{cases} \infty, & t = 0 \\ 0, & t \neq 0 \end{cases} \quad \text{and} \quad \int_{-\infty}^{\infty} \delta(t)dt = 1.$$

Under these conditions

- $\hat{x}$ obtained through the minimum energy estimator minimizes the expected value

$$\lim_{t\to\infty} \mathrm{E}\left[|x(t) - \hat{x}(t)|^2\right] \tag{1}$$

⇝ The Kalman filter is derived based on (1)

Section 3

The discrete time Kalman filter

The discrete time Kalman filter

Consider

$$x(k+1) = Ax(k) + Bu(k) + \bar{B}v(k),$$
$$y(k) = Cx(k) + w(k).$$

- $(v(k))_{k\in\mathbb{N}} \subset \mathbb{R}^q$, $(w(k))_{k\in\mathbb{N}} \subset \mathbb{R}^p$: unknown disturbances and measurement noise.

Goal: For a finite set of measurements $y(0), \ldots, y(k)$, define a state observer

$$\hat{x}(k+1) = A\hat{x}(k) + Bu(k) + \bar{B}\hat{v}(k), \quad \hat{x}(0) = \hat{x}_0$$
$$y(k) = C\hat{x}(k) + \hat{w}(k)$$

and sequences $\hat{v}(\cdot)$, $\hat{w}(\cdot)$, to be determined.

- $\hat{v}(k)$, $\hat{w}(k)$ will be defined such that $\hat{x}(k)$ is optimal w.r.t. assumptions on $v(\cdot)$ and $w(\cdot)$, and w.r.t. the measured output $y(0), \ldots, y(k)$.

The discrete time Kalman filter

Consider

$$x(k+1) = Ax(k) + Bu(k) + \bar{B}v(k),$$
$$y(k) = Cx(k) + w(k).$$

- $(v(k))_{k\in\mathbb{N}} \subset \mathbb{R}^q$, $(w(k))_{k\in\mathbb{N}} \subset \mathbb{R}^p$: unknown disturbances and measurement noise.

Goal: For a finite set of measurements $y(0), \ldots, y(k)$, define a state observer

$$\hat{x}(k+1) = A\hat{x}(k) + Bu(k) + \bar{B}\hat{v}(k), \quad \hat{x}(0) = \hat{x}_0$$
$$y(k) = C\hat{x}(k) + \hat{w}(k)$$

and sequences $\hat{v}(\cdot)$, $\hat{w}(\cdot)$, to be determined.

- $\hat{v}(k)$, $\hat{w}(k)$ will be defined such that $\hat{x}(k)$ is optimal w.r.t. assumptions on $v(\cdot)$ and $w(\cdot)$, and w.r.t. the measured output $y(0), \ldots, y(k)$.
- *Variance*: $\mathrm{Var}(\cdot)$
- *Expected value*: $\mathrm{E}[\cdot]$

Assumption

$v : \mathbb{N} \to \mathbb{R}^q$, $w : \mathbb{N} \to \mathbb{R}^p$ sequences of zero-mean Gaussian white noise such that $\mathrm{Var}(v(k)) = Q^{-1} \in \mathcal{S}^q_{>0}$ and $\mathrm{Var}(w(k)) = R^{-1} \in \mathcal{S}^p_{>0}$ and $\mathrm{E}\left[v(k)w(j)^T\right] = 0$ for all $k, j \in \mathbb{N}_0$.
Additionally, the initial state is assumed to be independent of $v(k)$ and $w(k)$ in the sense that $\mathrm{E}\left[x_0 v(k)^T\right] = 0$ and $\mathrm{E}\left[x_0 w(k)^T\right] = 0$ for all $k \in \mathbb{N}_0$.

The discrete time Kalman filter

Consider

$$x(k+1) = Ax(k) + Bu(k) + \bar{B}v(k),$$
$$y(k) = Cx(k) + w(k).$$

- $(v(k))_{k\in\mathbb{N}} \subset \mathbb{R}^q$, $(w(k))_{k\in\mathbb{N}} \subset \mathbb{R}^p$: unknown disturbances and measurement noise.

Goal: For a finite set of measurements $y(0), \ldots, y(k)$, define a state observer

$$\hat{x}(k+1) = A\hat{x}(k) + Bu(k) + \bar{B}\hat{v}(k), \quad \hat{x}(0) = \hat{x}_0$$
$$y(k) = C\hat{x}(k) + \hat{w}(k)$$

and sequences $\hat{v}(\cdot)$, $\hat{w}(\cdot)$, to be determined.

- $\hat{v}(k)$, $\hat{w}(k)$ will be defined such that $\hat{x}(k)$ is optimal w.r.t. assumptions on $v(\cdot)$ and $w(\cdot)$, and w.r.t. the measured output $y(0), \ldots, y(k)$.
- *Variance*: $\mathrm{Var}(\cdot)$
- *Expected value*: $\mathrm{E}[\cdot]$

Assumption

$v : \mathbb{N} \to \mathbb{R}^q$, $w : \mathbb{N} \to \mathbb{R}^p$ sequences of zero-mean Gaussian white noise such that $\mathrm{Var}(v(k)) = Q^{-1} \in \mathcal{S}^q_{>0}$ and $\mathrm{Var}(w(k)) = R^{-1} \in \mathcal{S}^p_{>0}$ and $\mathrm{E}\left[v(k)w(j)^T\right] = 0$ for all $k, j \in \mathbb{N}_0$.
Additionally, the initial state is assumed to be independent of $v(k)$ and $w(k)$ in the sense that $\mathrm{E}\left[x_0 v(k)^T\right] = 0$ and $\mathrm{E}\left[x_0 w(k)^T\right] = 0$ for all $k \in \mathbb{N}_0$.

Additionally, assume that

- (A, B, C) is controllable and observable
- A is nonsingular (if not, define $u = Kx + \tilde{u}$ with $A + BK$ nonsingular)

The discrete time Kalman filter (2)

Consider

$$x(k+1) = Ax(k) + Bu(k) + \bar{B}v(k),$$
$$y(k) = Cx(k) + w(k).$$

- $(v(k))_{k\in\mathbb{N}} \subset \mathbb{R}^q$, $(w(k))_{k\in\mathbb{N}} \subset \mathbb{R}^p$: unknown disturbances and measurement noise.

Goal: For a finite set of measurements $y(0), \ldots, y(k)$, define a state observer

$$\hat{x}(k+1) = A\hat{x}(k) + Bu(k) + \bar{B}\hat{v}(k), \quad \hat{x}(0) = \hat{x}_0$$
$$y(k) = C\hat{x}(k) + \hat{w}(k)$$

and sequences $\hat{v}(\cdot)$, $\hat{w}(\cdot)$, to be determined.

- $\hat{v}(k)$, $\hat{w}(k)$ will be defined such that $\hat{x}(k)$ is optimal w.r.t. assumptions on $v(\cdot)$ and $w(\cdot)$, and w.r.t. the measured output $y(0), \ldots, y(k)$.
- *Variance*: $\mathrm{Var}(\cdot)$
- *expectation*: $\mathrm{E}[\cdot]$

- Split the estimated state $\hat{x} = \hat{x}_d + \hat{x}_s$ (deterministic and stochastic component)

$$\hat{x}_d(k+1) = A\hat{x}_d(k) + Bu(k), \qquad \hat{x}_{d,0} = \hat{x}_0$$
$$\hat{y}_d(k) = C\hat{x}_d(k)$$

and

$$\hat{x}_s(k+1) = A\hat{x}_s(k) + \bar{B}\hat{v}(k) \qquad \hat{x}_{s,0} = 0$$
$$\hat{y}_s(k) = C\hat{x}_s(k) + \hat{w}(k),$$

The discrete time Kalman filter (2)

Consider

$$x(k+1) = Ax(k) + Bu(k) + \bar{B}v(k),$$
$$y(k) = Cx(k) + w(k).$$

- $(v(k))_{k\in\mathbb{N}} \subset \mathbb{R}^q$, $(w(k))_{k\in\mathbb{N}} \subset \mathbb{R}^p$: unknown disturbances and measurement noise.

Goal: For a finite set of measurements $y(0), \ldots, y(k)$, define a state observer

$$\hat{x}(k+1) = A\hat{x}(k) + Bu(k) + \bar{B}\hat{v}(k), \quad \hat{x}(0) = \hat{x}_0$$
$$y(k) = C\hat{x}(k) + \hat{w}(k)$$

and sequences $\hat{v}(\cdot)$, $\hat{w}(\cdot)$, to be determined.

- $\hat{v}(k)$, $\hat{w}(k)$ will be defined such that $\hat{x}(k)$ is optimal w.r.t. assumptions on $v(\cdot)$ and $w(\cdot)$, and w.r.t. the measured output $y(0), \ldots, y(k)$.
- *Variance*: $\mathrm{Var}(\cdot)$
- *expectation*: $\mathrm{E}[\cdot]$

- Split the estimated state $\hat{x} = \hat{x}_d + \hat{x}_s$ (deterministic and stochastic component)

$$\hat{x}_d(k+1) = A\hat{x}_d(k) + Bu(k), \qquad \hat{x}_{d,0} = \hat{x}_0$$
$$\hat{y}_d(k) = C\hat{x}_d(k)$$

and

$$\hat{x}_s(k+1) = A\hat{x}_s(k) + \bar{B}\hat{v}(k) \qquad \hat{x}_{s,0} = 0$$
$$\hat{y}_s(k) = C\hat{x}_s(k) + \hat{w}(k),$$

The discrete time Kalman filter (2)

Consider

$$x(k+1) = Ax(k) + Bu(k) + \bar{B}v(k),$$
$$y(k) = Cx(k) + w(k).$$

- $(v(k))_{k\in\mathbb{N}} \subset \mathbb{R}^q$, $(w(k))_{k\in\mathbb{N}} \subset \mathbb{R}^p$: unknown disturbances and measurement noise.

Goal: For a finite set of measurements $y(0), \dots, y(k)$, define a state observer

$$\hat{x}(k+1) = A\hat{x}(k) + Bu(k) + \bar{B}\hat{v}(k), \quad \hat{x}(0) = \hat{x}_0$$
$$y(k) = C\hat{x}(k) + \hat{w}(k)$$

and sequences $\hat{v}(\cdot)$, $\hat{w}(\cdot)$, to be determined.

- $\hat{v}(k)$, $\hat{w}(k)$ will be defined such that $\hat{x}(k)$ is optimal w.r.t. assumptions on $v(\cdot)$ and $w(\cdot)$, and w.r.t. the measured output $y(0), \dots, y(k)$.
- *Variance*: $\mathrm{Var}(\cdot)$
- *expectation*: $\mathrm{E}[\cdot]$

- Split the estimated state $\hat{x} = \hat{x}_d + \hat{x}_s$ (deterministic and stochastic component)

$$\hat{x}_d(k+1) = A\hat{x}_d(k) + Bu(k), \qquad \hat{x}_{d,0} = \hat{x}_0$$
$$\hat{y}_d(k) = C\hat{x}_d(k)$$

and

$$\hat{x}_s(k+1) = A\hat{x}_s(k) + \bar{B}\hat{v}(k) \qquad \hat{x}_{s,0} = 0$$
$$\hat{y}_s(k) = C\hat{x}_s(k) + \hat{w}(k),$$

- For given $\hat{x}_d(0) = \hat{x}_{d,0}$ and $(u(k))_{k\in\mathbb{N}}$ it holds that

$$\hat{x}_d(k) = A^k\hat{x}_{d,0} + \sum_{i=1}^{k} A^{k-i}Bu(i-1) \quad k \in \mathbb{N}$$

- $\hat{x}_s$ cannot be computed because it depends on $\hat{v}(\cdot)$ and $\hat{w}(\cdot)$.

The discrete time Kalman filter (2)

Consider

$$x(k+1) = Ax(k) + Bu(k) + \bar{B}v(k),$$
$$y(k) = Cx(k) + w(k).$$

- $(v(k))_{k\in\mathbb{N}} \subset \mathbb{R}^q$, $(w(k))_{k\in\mathbb{N}} \subset \mathbb{R}^p$: unknown disturbances and measurement noise.

Goal: For a finite set of measurements $y(0), \dots, y(k)$, define a state observer

$$\hat{x}(k+1) = A\hat{x}(k) + Bu(k) + \bar{B}\hat{v}(k), \quad \hat{x}(0) = \hat{x}_0$$
$$y(k) = C\hat{x}(k) + \hat{w}(k)$$

and sequences $\hat{v}(\cdot)$, $\hat{w}(\cdot)$, to be determined.

- $\hat{v}(k)$, $\hat{w}(k)$ will be defined such that $\hat{x}(k)$ is optimal w.r.t. assumptions on $v(\cdot)$ and $w(\cdot)$, and w.r.t. the measured output $y(0), \dots, y(k)$.
- *Variance*: $\mathrm{Var}(\cdot)$
- *expectation*: $\mathrm{E}[\cdot]$

- Split the estimated state $\hat{x} = \hat{x}_d + \hat{x}_s$ (deterministic and stochastic component)

$$\hat{x}_d(k+1) = A\hat{x}_d(k) + Bu(k), \qquad \hat{x}_{d,0} = \hat{x}_0$$
$$\hat{y}_d(k) = C\hat{x}_d(k)$$

and

$$\hat{x}_s(k+1) = A\hat{x}_s(k) + \bar{B}\hat{v}(k) \qquad \hat{x}_{s,0} = 0$$
$$\hat{y}_s(k) = C\hat{x}_s(k) + \hat{w}(k),$$

- For given $\hat{x}_d(0) = \hat{x}_{d,0}$ and $(u(k))_{k\in\mathbb{N}}$ it holds that

$$\hat{x}_d(k) = A^k\hat{x}_{d,0} + \sum_{i=1}^{k} A^{k-i}Bu(i-1) \quad k \in \mathbb{N}$$

- $\hat{x}_s$ cannot be computed because it depends on $\hat{v}(\cdot)$ and $\hat{w}(\cdot)$.

$\leadsto$ We look for $\hat{v}(\cdot)$, $\hat{w}(\cdot)$ which describe the mismatch

$$\hat{y}_s(k) = y(k) - \hat{y}_d(k)$$

between the measured output $y(k)$ and the deterministic output $\hat{y}_d(k)$ in an optimal way.

The discrete time Kalman filter (Least squares & minimum variance solution)

It holds that (for all $0 \leq j \leq k \in \mathbb{N}$):

$$\begin{aligned}\hat{x}_s(k) &= A\hat{x}_s(k-1) + \bar{B}\hat{v}(k-1)\\ &= A^{k-j}\hat{x}_s(j) + \sum_{i=j+1}^{k} A^{k-i}\bar{B}\hat{v}(i-1)\end{aligned}$$

or equivalently

$$\hat{x}_s(j) = A^{j-k}\hat{x}_s(k) - \sum_{i=j+1}^{k} A^{j-i}\bar{B}\hat{v}(i-1).$$

Moreover: $(j \in \{0, \ldots, k\})$

$$\begin{aligned}\hat{y}_s(j) &= C\hat{x}_s(j) + \hat{w}(j)\\ &= CA^{j-k}\hat{x}_s(k) + \hat{w}(j) - \sum_{i=j+1}^{k} CA^{j-i}\bar{B}\hat{v}(i-1)\end{aligned}$$

The discrete time Kalman filter (Least squares & minimum variance solution)

It holds that (for all $0 \leq j \leq k \in \mathbb{N}$):

$$\hat{x}_s(k) = A\hat{x}_s(k-1) + \bar{B}\hat{v}(k-1)$$

$$= A^{k-j}\hat{x}_s(j) + \sum_{i=j+1}^{k} A^{k-i}\bar{B}\hat{v}(i-1)$$

or equivalently

$$\hat{x}_s(j) = A^{j-k}\hat{x}_s(k) - \sum_{i=j+1}^{k} A^{j-i}\bar{B}\hat{v}(i-1).$$

Moreover: $(j \in \{0,\ldots,k\})$

$$\hat{y}_s(j) = C\hat{x}_s(j) + \hat{w}(j)$$

$$= CA^{j-k}\hat{x}_s(k) + \hat{w}(j) - \sum_{i=j+1}^{k} CA^{j-i}\bar{B}\hat{v}(i-1)$$

In vector form: $(j \in \{0,\ldots,k\})$

$$\Lambda_k^j = \Phi_k^j \hat{x}_s^j(k) + \Gamma_k^j$$

Where

$$\Lambda_k^j = \begin{bmatrix} \hat{y}_s(0) \\ \hat{y}_s(1) \\ \vdots \\ \hat{y}_s(j) \end{bmatrix}, \qquad \Phi_k^j = \begin{bmatrix} CA^{-k} \\ CA^{1-k} \\ \vdots \\ CA^{j-k} \end{bmatrix},$$

$$\Gamma_k^j = \begin{bmatrix} \hat{w}(0) - \sum_{i=1}^{k} CA^{1-i}\bar{B}\hat{v}(i-1) \\ \hat{w}(1) - \sum_{i=2}^{k} CA^{2-i}\bar{B}\hat{v}(i-1) \\ \vdots \\ \hat{w}(j) - \sum_{i=j+1}^{k} CA^{k-i}\bar{B}\hat{v}(i-1) \end{bmatrix}$$

The discrete time Kalman filter (Least squares & minimum variance solution)

It holds that (for all $0 \leq j \leq k \in \mathbb{N}$):

$$\hat{x}_s(k) = A\hat{x}_s(k-1) + \bar{B}\hat{v}(k-1)$$

$$= A^{k-j}\hat{x}_s(j) + \sum_{i=j+1}^{k} A^{k-i}\bar{B}\hat{v}(i-1)$$

or equivalently

$$\hat{x}_s(j) = A^{j-k}\hat{x}_s(k) - \sum_{i=j+1}^{k} A^{j-i}\bar{B}\hat{v}(i-1).$$

Moreover: ($j \in \{0, \ldots, k\}$)

$$\hat{y}_s(j) = C\hat{x}_s(j) + \hat{w}(j)$$

$$= CA^{j-k}\hat{x}_s(k) + \hat{w}(j) - \sum_{i=j+1}^{k} CA^{j-i}\bar{B}\hat{v}(i-1)$$

In vector form: ($j \in \{0, \ldots, k\}$)

$$\Lambda_k^j = \Phi_k^j \hat{x}_s^j(k) + \Gamma_k^j$$

Where

$$\Lambda_k^j = \begin{bmatrix} \hat{y}_s(0) \\ \hat{y}_s(1) \\ \vdots \\ \hat{y}_s(j) \end{bmatrix}, \qquad \Phi_k^j = \begin{bmatrix} CA^{-k} \\ CA^{1-k} \\ \vdots \\ CA^{j-k} \end{bmatrix},$$

$$\Gamma_k^j = \begin{bmatrix} \hat{w}(0) - \sum_{i=1}^{k} CA^{1-i}\bar{B}\hat{v}(i-1) \\ \hat{w}(1) - \sum_{i=2}^{k} CA^{2-i}\bar{B}\hat{v}(i-1) \\ \vdots \\ \hat{w}(j) - \sum_{i=j+1}^{k} CA^{k-i}\bar{B}\hat{v}(i-1) \end{bmatrix}$$

Note that:

- $j \in \{0, \ldots, k\}$ indicates that $y(0)$ to $y(j)$ are taken into account to calculate the stochastic part $\hat{x}_s^j(k)$
- $\Lambda_k^k = \Phi_k^k \hat{x}_s^k(k) + \Gamma_k^k$
- Λ_k^j contains mismatch between $y(\cdot)$ and $\hat{y}_d(\cdot)$

$\rightsquigarrow$ Find $\hat{x}_s^j(k)$ which fits the data in an optimal way

$\rightsquigarrow$ estimate of $x(k)$ through $\hat{x}(k) = \hat{x}_d(k) + \hat{x}_s^j(k)$

- Λ_k^j and Φ_k^j are known; Γ_k^j is not known

The discrete time Kalman filter (Least squares & minimum variance solution)

Note that:

- Λ_k^j, Φ_k^j are known; Γ_k^j is not known; ($\Lambda_k^k = \Phi_k^k \hat{x}_s^k(k) + \Gamma_k^k$)
- $(v(k))_{k\in\mathbb{N}}$ and $(w(k))_{k\in\mathbb{N}}$ sequences of Gaussian white noise with zero mean
- Find $\hat{x}_s^j(k)$ that minimizes the expected value

$$F(\hat{x}_s^j(k), W_k^j) = \mathrm{E}\left[|\Lambda_k^j - \Phi_k^j \hat{x}_s^j(k)|_{W_k^j}\right]$$

for $W_k^j \in \mathcal{S}_{>0}^{p(j+1)}$.

The discrete time Kalman filter (Least squares & minimum variance solution)

Note that:

- Λ_k^j, Φ_k^j are known; Γ_k^j is not known; $(\Lambda_k^k = \Phi_k^k \hat{x}_s^k(k) + \Gamma_k^k)$
- $(v(k))_{k\in\mathbb{N}}$ and $(w(k))_{k\in\mathbb{N}}$ sequences of Gaussian white noise with zero mean
- Find $\hat{x}_s^j(k)$ that minimizes the expected value

$$F(\hat{x}_s^j(k), W_k^j) = \mathrm{E}\left[|\Lambda_k^j - \Phi_k^j \hat{x}_s^j(k)|_{W_k^j}\right]$$

for $W_k^j \in \mathcal{S}_{>0}^{p(j+1)}$.

For $(\Phi_k^j)^T W_k^j \Phi_k^j$ nonsingular, it holds that:

$$F(\hat{x}_s^j(k), W_k^j) = \mathrm{E}\left[(\Lambda_k^j - \Phi_k^j \hat{x}_s^j(k))^T W_k^j (\Lambda_k^j - \Phi_k^j \hat{x}_s^j(k))\right]$$

$$= \mathrm{E}\Big[[(\Phi_k^j)^T W_k \Phi_k^j)\hat{x}_s^j(k) - (\Phi_k^j)^T W_k^j \Lambda_k^j]^T ((\Phi_k^j)^T W_k^j \Phi_k^j)^{-1}$$

$$\cdot [(\Phi_k^j)^T W_k^j \Phi_k^j)\hat{x}_s^j(k) - (\Phi_k^j)^T W_k^j \Lambda_k^j]\Big]$$

$$+ \mathrm{E}\left[(\Lambda_k^j)^T (I - W_k^j \Phi_k^j ((\Phi_k^j)^T W_k^j \Phi_k^j)^{-1} (\Phi_k^j)^T) W_k^j \Lambda_k^j\right]$$

Since second term is independent of $\hat{x}_s^j(k)$

$\rightsquigarrow \hat{x}_s^j(k) = \hat{x}_s^j(k; W_k^j) = ((\Phi_k^j)^T W_k^j \Phi_k^j)^{-1} (\Phi_k^j)^T W_k^j \Lambda_k^j$

The discrete time Kalman filter (Least squares & minimum variance solution)

Note that:

- Λ_k^j, Φ_k^j are known; Γ_k^j is not known; $(\Lambda_k^k = \Phi_k^k \hat{x}_s^k(k) + \Gamma_k^k)$
- $(v(k))_{k\in\mathbb{N}}$ and $(w(k))_{k\in\mathbb{N}}$ sequences of Gaussian white noise with zero mean
- Find $\hat{x}_s^j(k)$ that minimizes the expected value

$$F(\hat{x}_s^j(k), W_k^j) = \mathrm{E}\left[|\Lambda_k^j - \Phi_k^j \hat{x}_s^j(k)|_{W_k^j} \right]$$

for $W_k^j \in \mathcal{S}_{>0}^{p(j+1)}$.

For $(\Phi_k^j)^T W_k^j \Phi_k^j$ nonsingular, it holds that:

$$F(\hat{x}_s^j(k), W_k^j) = \mathrm{E}\left[(\Lambda_k^j - \Phi_k^j \hat{x}_s^j(k))^T W_k^j (\Lambda_k^j - \Phi_k^j \hat{x}_s^j(k)) \right]$$

$$= \mathrm{E}\Big[[(\Phi_k^j)^T W_k \Phi_k^j)\hat{x}_s^j(k) - (\Phi_k^j)^T W_k^j \Lambda_k^j]^T ((\Phi_k^j)^T W_k^j \Phi_k^j)^{-1} \cdot [(\Phi_k^j)^T W_k^j \Phi_k^j)\hat{x}_s^j(k) - (\Phi_k^j)^T W_k^j \Lambda_k^j] \Big]$$

$$+ \mathrm{E}\left[(\Lambda_k^j)^T (I - W_k^j \Phi_k^j ((\Phi_k^j)^T W_k^j \Phi_k^j)^{-1} (\Phi_k^j)^T) W_k^j \Lambda_k^j \right]$$

Since second term is independent of $\hat{x}_s^j(k)$

$\rightsquigarrow \hat{x}_s^j(k) = \hat{x}_s^j(k; W_k^j) = ((\Phi_k^j)^T W_k^j \Phi_k^j)^{-1} (\Phi_k^j)^T W_k^j \Lambda_k^j$

Question:

- How to define W_k^j in an optimal way?

$\rightsquigarrow$ Minimize the variance

$$\mathrm{Var}(x(k) - \hat{x}^j(k)) = \mathrm{Var}(x(k) - \hat{x}_d(k) - \hat{x}_s^j(k))$$

The discrete time Kalman filter (Least squares & minimum variance solution)

Note that:

- Λ_k^j, Φ_k^j are known; Γ_k^j is not known; ($\Lambda_k^k = \Phi_k^k \hat{x}_s^k(k) + \Gamma_k^k$)
- $(v(k))_{k\in\mathbb{N}}$ and $(w(k))_{k\in\mathbb{N}}$ sequences of Gaussian white noise with zero mean
- Find $\hat{x}_s^j(k)$ that minimizes the expected value

$$F(\hat{x}_s^j(k), W_k^j) = \mathrm{E}\left[|\Lambda_k^j - \Phi_k^j \hat{x}_s^j(k)|_{W_k^j}\right]$$

for $W_k^j \in \mathcal{S}_{>0}^{p(j+1)}$.

For $(\Phi_k^j)^T W_k^j \Phi_k^j$ nonsingular, it holds that:

$$F(\hat{x}_s^j(k), W_k^j) = \mathrm{E}\left[(\Lambda_k^j - \Phi_k^j \hat{x}_s^j(k))^T W_k^j (\Lambda_k^j - \Phi_k^j \hat{x}_s^j(k))\right]$$

$$= \mathrm{E}\Big[[(\Phi_k^j)^T W_k \Phi_k^j)\hat{x}_s^j(k) - (\Phi_k^j)^T W_k^j \Lambda_k^j]^T ((\Phi_k^j)^T W_k^j \Phi_k^j)^{-1}$$

$$\cdot [(\Phi_k^j)^T W_k^j \Phi_k^j)\hat{x}_s^j(k) - (\Phi_k^j)^T W_k^j \Lambda_k^j]\Big]$$

$$+ \mathrm{E}\left[(\Lambda_k^j)^T (I - W_k^j \Phi_k^j ((\Phi_k^j)^T W_k^j \Phi_k^j)^{-1} (\Phi_k^j)^T) W_k^j \Lambda_k^j\right]$$

Since second term is independent of $\hat{x}_s^j(k)$

$\rightsquigarrow \hat{x}_s^j(k) = \hat{x}_s^j(k; W_k^j) = ((\Phi_k^j)^T W_k^j \Phi_k^j)^{-1} (\Phi_k^j)^T W_k^j \Lambda_k^j$

Question:

- How to define W_k^j in an optimal way?

$\rightsquigarrow$ Minimize the variance

$$\mathrm{Var}(x(k) - \hat{x}^j(k)) = \mathrm{Var}(x(k) - \hat{x}_d(k) - \hat{x}_s^j(k))$$

The variance satisfies

$$\mathrm{Var}(x(k) - \hat{x}_d(k) - \hat{x}_s^j(k)) = [(\Phi_k^j)^T W_k^j \Phi_k^j]^{-1}$$

$$\cdot (\Phi_k^j)^T W_k^j \, \mathrm{E}[\Gamma_k^j (\Gamma_k^j)^T] W_k^j \Phi_k^j [(\Phi_k^j)^T W_k^j \Phi_k^j]^{-1}$$

The discrete time Kalman filter (Least squares & minimum variance solution)

Note that:

- Λ_k^j, Φ_k^j are known; Γ_k^j is not known; $(\Lambda_k^k = \Phi_k^k \hat{x}_s^k(k) + \Gamma_k^k)$
- $(v(k))_{k\in\mathbb{N}}$ and $(w(k))_{k\in\mathbb{N}}$ sequences of Gaussian white noise with zero mean
- Find $\hat{x}_s^j(k)$ that minimizes the expected value

$$F(\hat{x}_s^j(k), W_k^j) = \mathrm{E}\left[|\Lambda_k^j - \Phi_k^j \hat{x}_s^j(k)|_{W_k^j} \right]$$

for $W_k^j \in \mathcal{S}_{>0}^{p(j+1)}$.

For $(\Phi_k^j)^T W_k^j \Phi_k^j$ nonsingular, it holds that:

$$F(\hat{x}_s^j(k), W_k^j) = \mathrm{E}\left[(\Lambda_k^j - \Phi_k^j \hat{x}_s^j(k))^T W_k^j (\Lambda_k^j - \Phi_k^j \hat{x}_s^j(k)) \right]$$

$$= \mathrm{E}\Big[[(\Phi_k^j)^T W_k \Phi_k^j)\hat{x}_s^j(k) - (\Phi_k^j)^T W_k^j \Lambda_k^j]^T ((\Phi_k^j)^T W_k^j \Phi_k^j)^{-1}$$
$$\cdot [(\Phi_k^j)^T W_k^j \Phi_k^j)\hat{x}_s^j(k) - (\Phi_k^j)^T W_k^j \Lambda_k^j] \Big]$$
$$+ \mathrm{E}\left[(\Lambda_k^j)^T (I - W_k^j \Phi_k^j ((\Phi_k^j)^T W_k^j \Phi_k^j)^{-1} (\Phi_k^j)^T) W_k^j \Lambda_k^j \right]$$

Since second term is independent of $\hat{x}_s^j(k)$

$\rightsquigarrow$ $\hat{x}_s^j(k) = \hat{x}_s^j(k; W_k^j) = ((\Phi_k^j)^T W_k^j \Phi_k^j)^{-1} (\Phi_k^j)^T W_k^j \Lambda_k^j$

Question:

- How to define W_k^j in an optimal way?

$\rightsquigarrow$ Minimize the variance

$$\mathrm{Var}(x(k) - \hat{x}^j(k)) = \mathrm{Var}(x(k) - \hat{x}_d(k) - \hat{x}_s^j(k))$$

The variance satisfies

$$\mathrm{Var}(x(k) - \hat{x}_d(k) - \hat{x}_s^j(k)) = [(\Phi_k^j)^T W_k^j \Phi_k^j]^{-1}$$
$$\cdot (\Phi_k^j)^T W_k^j \, \mathrm{E}[\Gamma_k^j (\Gamma_k^j)^T] W_k^j \Phi_k^j [(\Phi_k^j)^T W_k^j \Phi_k^j]^{-1}$$

for

$$W_k^j = (\Xi_k^j)^{-1} = \mathrm{E}\left[\Gamma_k^j (\Gamma_k^j)^T \right]^{-1} \in \mathcal{S}_{>0}^{p(j+1)}$$

the variance reduces to

$$\mathrm{Var}(x(k) - \hat{x}_d(k) - \hat{x}_s^j(k)) = [(\Phi_k^j)^T (\Xi_k^j)^{-1} \Phi_k^j]^{-1}$$

The discrete time Kalman filter (Least squares & minimum variance solution)

Note that:

- Λ_k^j, Φ_k^j are known; Γ_k^j is not known; $(\Lambda_k^k = \Phi_k^k \hat{x}_s^k(k) + \Gamma_k^k)$
- $(v(k))_{k\in\mathbb{N}}$ and $(w(k))_{k\in\mathbb{N}}$ sequences of Gaussian white noise with zero mean
- Find $\hat{x}_s^j(k)$ that minimizes the expected value

$$F(\hat{x}_s^j(k), W_k^j) = \mathrm{E}\left[|\Lambda_k^j - \Phi_k^j \hat{x}_s^j(k)|_{W_k^j}\right]$$

for $W_k^j \in \mathcal{S}_{>0}^{p(j+1)}$.

For $(\Phi_k^j)^T W_k^j \Phi_k^j$ nonsingular, it holds that:

$$F(\hat{x}_s^j(k), W_k^j) = \mathrm{E}\left[(\Lambda_k^j - \Phi_k^j \hat{x}_s^j(k))^T W_k^j (\Lambda_k^j - \Phi_k^j \hat{x}_s^j(k))\right]$$

$$= \mathrm{E}\left[[(\Phi_k^j)^T W_k \Phi_k^j)\hat{x}_s^j(k) - (\Phi_k^j)^T W_k^j \Lambda_k^j]^T ((\Phi_k^j)^T W_k^j \Phi_k^j)^{-1}\right.$$

$$\left.\cdot [(\Phi_k^j)^T W_k^j \Phi_k^j)\hat{x}_s^j(k) - (\Phi_k^j)^T W_k^j \Lambda_k^j]\right]$$

$$+ \mathrm{E}\left[(\Lambda_k^j)^T (I - W_k^j \Phi_k^j ((\Phi_k^j)^T W_k^j \Phi_k^j)^{-1} (\Phi_k^j)^T) W_k^j \Lambda_k^j\right]$$

Since second term is independent of $\hat{x}_s^j(k)$

$\rightsquigarrow \hat{x}_s^j(k) = \hat{x}_s^j(k; W_k^j) = ((\Phi_k^j)^T W_k^j \Phi_k^j)^{-1} (\Phi_k^j)^T W_k^j \Lambda_k^j$

Question:

- How to define W_k^j in an optimal way?

$\rightsquigarrow$ Minimize the variance

$$\mathrm{Var}(x(k) - \hat{x}^j(k)) = \mathrm{Var}(x(k) - \hat{x}_d(k) - \hat{x}_s^j(k))$$

The variance satisfies

$$\mathrm{Var}(x(k) - \hat{x}_d(k) - \hat{x}_s^j(k)) = [(\Phi_k^j)^T W_k^j \Phi_k^j]^{-1}$$

$$\cdot (\Phi_k^j)^T W_k^j \,\mathrm{E}[\Gamma_k^j (\Gamma_k^j)^T] W_k^j \Phi_k^j [(\Phi_k^j)^T W_k^j \Phi_k^j]^{-1}$$

for

$$W_k^j = (\Xi_k^j)^{-1} = \mathrm{E}\left[\Gamma_k^j (\Gamma_k^j)^T\right]^{-1} \in \mathcal{S}_{>0}^{p(j+1)}$$

the variance reduces to

$$\mathrm{Var}(x(k) - \hat{x}_d(k) - \hat{x}_s^j(k)) = [(\Phi_k^j)^T (\Xi_k^j)^{-1} \Phi_k^j]^{-1}$$

Optimal estimate of $x(k) - \hat{x}_d(k)$ based on $y(0)$ to $y(j)$:

$$\hat{x}_s^j(k) = [(\Phi_k^j)^T (\Xi_k^j)^{-1} \Phi_k^j]^{-1} (\Phi_k^j)^T (\Xi_k^j)^{-1} \Lambda_k^j,$$

For $j = k$, the index is omitted: $\hat{x}_s(k) = \hat{x}_s^k(k)$.

The discrete time Kalman filter (Least squares & minimum variance solution)

Note that:

- Λ_k^j, Φ_k^j are known; Γ_k^j is not known; $(\Lambda_k^k = \Phi_k^k \hat{x}_s^k(k) + \Gamma_k^k)$
- $(v(k))_{k\in\mathbb{N}}$ and $(w(k))_{k\in\mathbb{N}}$ sequences of Gaussian white noise with zero mean
- Find $\hat{x}_s^j(k)$ that minimizes the expected value

$$F(\hat{x}_s^j(k), W_k^j) = \mathrm{E}\left[|\Lambda_k^j - \Phi_k^j \hat{x}_s^j(k)|_{W_k^j}\right]$$

for $W_k^j \in \mathcal{S}_{>0}^{p(j+1)}$.

For $(\Phi_k^j)^T W_k^j \Phi_k^j$ nonsingular, it holds that:

$$F(\hat{x}_s^j(k), W_k^j) = \mathrm{E}\left[(\Lambda_k^j - \Phi_k^j \hat{x}_s^j(k))^T W_k^j (\Lambda_k^j - \Phi_k^j \hat{x}_s^j(k))\right]$$

$$= \mathrm{E}\Big[[(\Phi_k^j)^T W_k \Phi_k^j)\hat{x}_s^j(k) - (\Phi_k^j)^T W_k^j \Lambda_k^j]^T((\Phi_k^j)^T W_k^j \Phi_k^j)^{-1}$$

$$\cdot [(\Phi_k^j)^T W_k^j \Phi_k^j)\hat{x}_s^j(k) - (\Phi_k^j)^T W_k^j \Lambda_k^j]\Big]$$

$$+ \mathrm{E}\left[(\Lambda_k^j)^T (I - W_k^j \Phi_k^j ((\Phi_k^j)^T W_k^j \Phi_k^j)^{-1} (\Phi_k^j)^T) W_k^j \Lambda_k^j\right]$$

Since second term is independent of $\hat{x}_s^j(k)$

$\rightsquigarrow \hat{x}_s^j(k) = \hat{x}_s^j(k; W_k^j) = ((\Phi_k^j)^T W_k^j \Phi_k^j)^{-1} (\Phi_k^j)^T W_k^j \Lambda_k^j$

Question:

- How to define W_k^j in an optimal way?

$\rightsquigarrow$ Minimize the variance

$$\mathrm{Var}(x(k) - \hat{x}^j(k)) = \mathrm{Var}(x(k) - \hat{x}_d(k) - \hat{x}_s^j(k))$$

The variance satisfies

$$\mathrm{Var}(x(k) - \hat{x}_d(k) - \hat{x}_s^j(k)) = [(\Phi_k^j)^T W_k^j \Phi_k^j]^{-1}$$

$$\cdot (\Phi_k^j)^T W_k^j \, \mathrm{E}[\Gamma_k^j (\Gamma_k^j)^T] W_k^j \Phi_k^j [(\Phi_k^j)^T W_k^j \Phi_k^j]^{-1}$$

for

$$W_k^j = (\Xi_k^j)^{-1} = \mathrm{E}\left[\Gamma_k^j (\Gamma_k^j)^T\right]^{-1} \in \mathcal{S}_{>0}^{p(j+1)}$$

the variance reduces to

$$\mathrm{Var}(x(k) - \hat{x}_d(k) - \hat{x}_s^j(k)) = [(\Phi_k^j)^T (\Xi_k^j)^{-1} \Phi_k^j]^{-1}$$

Optimal estimate of $x(k) - \hat{x}_d(k)$ based on $y(0)$ to $y(j)$:

$$\hat{x}_s^j(k) = [(\Phi_k^j)^T (\Xi_k^j)^{-1} \Phi_k^j]^{-1} (\Phi_k^j)^T (\Xi_k^j)^{-1} \Lambda_k^j,$$

For $j = k$, the index is omitted: $\hat{x}_s(k) = \hat{x}_s^k(k)$.

$\rightsquigarrow$ Dim. of linear equation grows linearly with $k \in \mathbb{N}$

The discrete time Kalman filter (A prediction-correction formulation)

- Goal: Rewrite problem such that the complexity of the calculation of $\hat{x}(k)$ is independent of k.

The discrete time Kalman filter (A prediction-correction formulation)

- Goal: Rewrite problem such that the complexity of the calculation of $\hat{x}(k)$ is independent of k.

Derive a recursive formula to iteratively compute $\hat{x}(k)$:

$$\hat{\chi}(k) = A\hat{x}(k-1) + Bu(k-1) \qquad \text{(prediction step)}$$
$$\hat{x}(k) = \chi(k) + G_k(y(k) - C\hat{\chi}(k)) \qquad \text{(correction step)}$$

- How to define the Kalman gain matrices $G_k \in \mathbb{R}^{n\times p}$, $k \in \mathbb{N}$?

The discrete time Kalman filter (A prediction-correction formulation)

- Goal: Rewrite problem such that the complexity of the calculation of $\hat{x}(k)$ is independent of k.

Derive a recursive formula to iteratively compute $\hat{x}(k)$:

$$\hat{\chi}(k) = A\hat{x}(k-1) + Bu(k-1) \qquad \text{(prediction step)}$$
$$\hat{x}(k) = \chi(k) + G_k(y(k) - C\hat{\chi}(k)) \qquad \text{(correction step)}$$

- How to define the Kalman gain matrices $G_k \in \mathbb{R}^{n\times p}$, $k \in \mathbb{N}$?

It can be shown that:

$$G_k = P_k^{k-1} C^T [C P_k^{k-1} C^T + R^{-1}]^{-1}$$

where

$$P_k^{k-1} = A P_{k-1} A^T + \bar{B} Q^{-1} \bar{B}^T$$
$$P_k = [I - G_k C] P_k^{k-1}$$

and

$$P_0 = \mathrm{E}\left[(x_0 - \mathrm{E}[x_0])(x_0 - \mathrm{E}[x_0])^T\right] = \mathrm{Var}\,(x_0)\,.$$

The discrete time Kalman filter (A prediction-correction formulation)

- Goal: Rewrite problem such that the complexity of the calculation of $\hat{x}(k)$ is independent of k.

Derive a recursive formula to iteratively compute $\hat{x}(k)$:

$$\hat{\chi}(k) = A\hat{x}(k-1) + Bu(k-1) \qquad \text{(prediction step)}$$
$$\hat{x}(k) = \chi(k) + G_k(y(k) - C\hat{\chi}(k)) \qquad \text{(correction step)}$$

- How to define the Kalman gain matrices $G_k \in \mathbb{R}^{n\times p}$, $k \in \mathbb{N}$?

It can be shown that:

$$G_k = P_k^{k-1} C^T [C P_k^{k-1} C^T + R^{-1}]^{-1}$$

where

$$P_k^{k-1} = A P_{k-1} A^T + \bar{B} Q^{-1} \bar{B}^T$$
$$P_k = [I - G_k C] P_k^{k-1}$$

and

$$P_0 = \mathrm{E}\left[(x_0 - \mathrm{E}[x_0])(x_0 - \mathrm{E}[x_0])^T\right] = \mathrm{Var}\,(x_0)\,.$$

Input: $Q^{-1} = \mathrm{Var}(v(k))$, $R^{-1} = \mathrm{Var}(w(k))$, $\hat{x}(0) = \hat{x}_0$, $P_0 \in \mathcal{S}^n_{>0}$.

Output: Estimates $\hat{\chi}(k)$, $\hat{x}(k)$ of $x(k)$ for $k \in \mathbb{N}$.

Algorithm: For $k \in \mathbb{N}$:

1. Update the gain matrix G_k:
$$P_k^{k-1} = A P_{k-1} A^T + \bar{B} Q^{-1} \bar{B}^T,$$
$$G_k = P_k^{k-1} C^T [C P_k^{k-1} C^T + R^{-1}]^{-1},$$
$$P_k = [I - G_k C] P_k^{k-1}.$$
2. Update estimate (before $y(k)$ is available):
$$\hat{\chi}(k) = A\hat{x}(k-1) + Bu(k-1).$$
3. Measure the output: $y(k) = Cx(k) + w(k)$
4. Update estimate (after $y(k)$ is available):
$$\hat{x}(k) = \hat{\chi}(k) + G_k(y(k) - C\hat{\chi}(k)),$$
set $k = k + 1$ and go to step 1.

The discrete time Kalman filter (Additional comments)

Input: $Q^{-1} = \mathrm{Var}(v(k))$, $R^{-1} = \mathrm{Var}(w(k))$, $\hat{x}(0) = \hat{x}_0$, $P_0 \in \mathcal{S}^n_{>0}$.

Output: Estimates $\hat{\chi}(k)$, $\hat{x}(k)$ of $x(k)$ for $k \in \mathbb{N}$.

Algorithm: For $k \in \mathbb{N}$:

1. Update the gain matrix G_k:
$$P_k^{k-1} = AP_{k-1}A^T + \bar{B}Q^{-1}\bar{B}^T,$$
$$G_k = P_k^{k-1}C^T[CP_k^{k-1}C^T + R^{-1}]^{-1},$$
$$P_k = [I - G_k C]P_k^{k-1}.$$
2. Update estimate (before $y(k)$ is available):
$$\hat{\chi}(k) = A\hat{x}(k-1) + Bu(k-1).$$
3. Measure the output: $y(k) = Cx(k) + w(k)$
4. Update estimate (after $y(k)$ is available):
$$\hat{x}(k) = \hat{\chi}(k) + G_k(y(k) - C\hat{\chi}(k)),$$
set $k = k + 1$ and go to step 1.

The Kalman filter can be written as a discrete time system:

$$\begin{aligned}\hat{\chi}(k+1) &= A(\hat{\chi}(k) + Bu(k) + G_k(y(k) - C\hat{\chi}(k))) \\ &= (A - AG_kC)\hat{\chi}(k) + Bu(k) + AG_ky(k) \\ \hat{x}(k+1) &= A\hat{x}(k) + Bu(k) + G_{k+1}(y(k+1) - C(A\hat{x}(k) + Bu(k))) \\ &= (I - G_{k+1}C)(A\hat{x}(k) + Bu(k)) + G_{k+1}y(k+1)\end{aligned}$$

The Kalman filter can be applied to time varying systems (i.e., $A(k)$, $B(k)$, $\bar{B}(k)$, $C(k)$)

The discrete time Kalman filter (A prediction-correction formulation)

Example

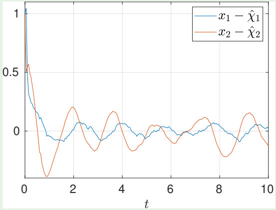
Consider $C = \begin{bmatrix} 1 & 0 \end{bmatrix}$ and

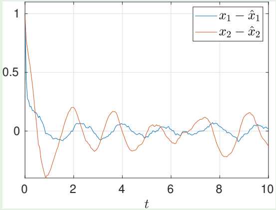
$$A = \begin{bmatrix} 1.000 & 0.050 \\ -0.491 & 0.995 \end{bmatrix}, \; B = \begin{bmatrix} 0 \\ 0.05 \end{bmatrix}, \; \bar{B} = \begin{bmatrix} 0.05 & 0 \\ 0 & 0.05 \end{bmatrix}$$

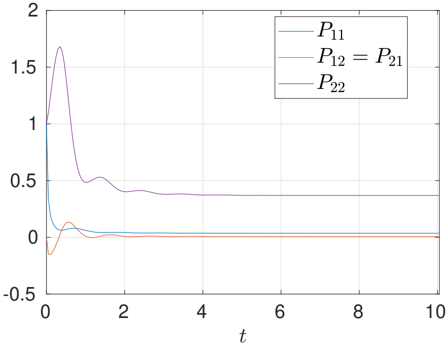
Additionally, let

$$R^{-1} = \frac{1}{2} \quad \text{and} \quad Q^{-1} = \frac{1}{2} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

(defined based on $v(k)$ and $w(k)$)







The discrete time Kalman filter (at steady-state)

Recall: Update the gain matrix G_k:

$$\begin{aligned} P_k^{k-1} &= AP_{k-1}A^T + \bar{B}Q^{-1}\bar{B}^T, \\ G_k &= P_k^{k-1}C^T[CP_k^{k-1}C^T + R^{-1}]^{-1}, \\ P_k &= [I - G_kC]P_k^{k-1}. \end{aligned}$$

Note that:

- Under certain conditions $G_k = G_\infty$, (i.e., $P_k = P_\infty$) converges to a steady-state

The discrete time Kalman filter (at steady-state)

Recall: Update the gain matrix G_k:

$$\begin{aligned} P_k^{k-1} &= AP_{k-1}A^T + \bar{B}Q^{-1}\bar{B}^T, \\ G_k &= P_k^{k-1}C^T[CP_k^{k-1}C^T + R^{-1}]^{-1}, \\ P_k &= [I - G_kC]P_k^{k-1}. \end{aligned}$$

Note that:

- Under certain conditions $G_k = G_\infty$, (i.e., $P_k = P_\infty$) converges to a steady-state

In particular, with $P_\infty = P_k = P_{k-1}$, $\Pi = P_k^{k-1}$:

$$\Pi = A\Pi A^T - A\Pi C^T(C\Pi C^T + R^{-1})^{-1}C\Pi A^T + \bar{B}Q^{-1}\bar{B}^T$$

$\rightsquigarrow$ discrete time algebraic Riccati equation

It holds that:

$$\begin{aligned} G_\infty &= \Pi C^T(C\Pi C^T + R^{-1})^{-1} \\ \tilde{G}_\infty &= A\Pi C^T(C\Pi C^T + R^{-1})^{-1} \\ P_\infty &= (I - G_\infty C)\Pi = (I - (\Pi C^T(C\Pi C^T + R^{-1})^{-1})C)\Pi \end{aligned}$$

The discrete time Kalman filter (at steady-state)

Recall: Update the gain matrix G_k:

$$P_k^{k-1} = AP_{k-1}A^T + \bar{B}Q^{-1}\bar{B}^T,$$
$$G_k = P_k^{k-1}C^T[CP_k^{k-1}C^T + R^{-1}]^{-1},$$
$$P_k = [I - G_kC]P_k^{k-1}.$$

Note that:

- Under certain conditions $G_k = G_\infty$, (i.e., $P_k = P_\infty$) converges to a steady-state

In particular, with $P_\infty = P_k = P_{k-1}, \Pi = P_k^{k-1}$:

$$\Pi = A\Pi A^T - A\Pi C^T(C\Pi C^T + R^{-1})^{-1}C\Pi A^T + \bar{B}Q^{-1}\bar{B}^T$$

$\rightsquigarrow$ discrete time algebraic Riccati equation

It holds that:

$$G_\infty = \Pi C^T(C\Pi C^T + R^{-1})^{-1}$$
$$\tilde{G}_\infty = A\Pi C^T(C\Pi C^T + R^{-1})^{-1}$$
$$P_\infty = (I - G_\infty C)\Pi = (I - (\Pi C^T(C\Pi C^T + R^{-1})^{-1})C)\Pi$$

Theorem

Consider the linear system

$$x(k+1) = Ax(k) + Bu(k) + \bar{B}v(k),$$
$$y(k) = Cx(k) + w(k).$$

and assume that $(A, \bar{B})$ is stabilizable and (A, C) is detectable. Additionally, let $R \in \mathcal{S}^p_{>0}$ and $Q \in \mathcal{S}^q_{>0}$. Then the Riccati equation has a unique positive definite solution $\Pi \in \mathcal{S}^n_{>0}$, and the matrix

$$A - \tilde{G}_\infty C = A - A\Pi C^T(C\Pi C^T + R^{-1})^{-1}C \quad (2)$$

is a Schur matrix.

The discrete time Kalman filter (at steady-state)

Recall: Update the gain matrix G_k:

$$P_k^{k-1} = AP_{k-1}A^T + \bar{B}Q^{-1}\bar{B}^T,$$
$$G_k = P_k^{k-1}C^T[CP_k^{k-1}C^T + R^{-1}]^{-1},$$
$$P_k = [I - G_kC]P_k^{k-1}.$$

Note that:

- Under certain conditions $G_k = G_\infty$, (i.e., $P_k = P_\infty$) converges to a steady-state

In particular, with $P_\infty = P_k = P_{k-1}$, $\Pi = P_k^{k-1}$:

$$\Pi = A\Pi A^T - A\Pi C^T(C\Pi C^T + R^{-1})^{-1}C\Pi A^T + \bar{B}Q^{-1}\bar{B}^T$$

$\rightsquigarrow$ discrete time algebraic Riccati equation

It holds that:

$$G_\infty = \Pi C^T(C\Pi C^T + R^{-1})^{-1}$$
$$\tilde{G}_\infty = A\Pi C^T(C\Pi C^T + R^{-1})^{-1}$$
$$P_\infty = (I - G_\infty C)\Pi = (I - (\Pi C^T(C\Pi C^T + R^{-1})^{-1})C)\Pi$$

Theorem

Consider the linear system

$$x(k+1) = Ax(k) + Bu(k) + \bar{B}v(k),$$
$$y(k) = Cx(k) + w(k).$$

and assume that $(A, \bar{B})$ is stabilizable and (A, C) is detectable. Additionally, let $R \in \mathcal{S}^p_{>0}$ and $Q \in \mathcal{S}^q_{>0}$. Then the Riccati equation has a unique positive definite solution $\Pi \in \mathcal{S}^n_{>0}$, and the matrix

$$A - \tilde{G}_\infty C = A - A\Pi C^T(C\Pi C^T + R^{-1})^{-1}C \qquad (2)$$

is a Schur matrix.

The steady-state Kalman filter reduces to

$$\hat{\chi}(k+1) = (A - \tilde{G}_\infty C)\hat{\chi}(k) + \tilde{G}_\infty y(k) + Bu(k)$$
$$\hat{x}(k+1) = (I - G_\infty C)(A\hat{x}(k) + Bu(k)) + G_\infty y(k+1)$$

$\rightsquigarrow$ The structure of the Luenberger observer or the minimum energy estimator is recovered

A hybrid Kalman filter

Input: Linear system

$$\dot{x}_c(t) = A_c x(t) + B_c u(t) + \bar{B}_c v_c(t), \qquad y_c(t) = C_c x(t) + w_c(t).$$

control input $u : \mathbb{R}_{\geq 0} \to \mathbb{R}^m$, positive definite matrices Q, R, initial estimates $\hat{x}(0) = \hat{x}_0$, $P_0 \in \mathcal{S}^n_{>0}$, and a sequence of discrete time steps $(\tau_k)_{k\in\mathbb{N}} \subset \mathbb{R}_{\geq 0}$, $\tau_k < \tau_{k+1}$, for all $k \in \mathbb{N}_0$.

Output: Continuous time and discrete time estimates $\hat{\chi}(t)$ and $\hat{x}(\tau_k)$ of the state $x(t)$.

Algorithm: For $k \in \mathbb{N}$:

1. Continuous time update: For $t \in [\tau_{k-1}, \tau_k]$ solve

$$\dot{P}(t) = A_c P(t) + P(t) A_c^T + \bar{B}_c Q^{-1} \bar{B}_c^T, \qquad P(\tau_{k-1}) = P_{k-1}$$
$$\dot{\hat{\chi}}(t) = A_c \hat{\chi}(t) + B_c u(t), \qquad \hat{\chi}(\tau_{k-1}) = \hat{x}(k-1).$$

2. Measure the output: $y_c(\tau_k) = C x_c(\tau_k) + w(\tau_k)$.
3. Discrete time update:

$$G_k = P(\tau_k) C_c^T (C_c P(\tau_k) C_c^T + R^{-1})^{-1},$$
$$P_k = (I - G_k C_c) P(\tau_k),$$
$$\hat{x}(k) = \hat{\chi}(\tau_k) + G_k (y_c(\tau_k) - C_c \hat{\chi}(\tau_k)).$$

Set $k = k + 1$ and go to step 1.

Introduction to Nonlinear Control

Stability, control design, and estimation

Christopher M. Kellett & Philipp Braun
School of Engineering,
Australian National University, Canberra, Australia

Part III:

Chapter 16: Classical Observer Design

Australian National University