

# Lecture 12

## The steady-state KF

### Performance of KF

Giancarlo Ferrari Trecate<sup>1</sup>

<sup>1</sup>Dependable Control and Decision Group  
École Polytechnique Fédérale de Lausanne (EPFL), Switzerland  
[giancarlo.ferraritrecate@epfl.ch](mailto:giancarlo.ferraritrecate@epfl.ch)

# Summary - previous lecture

## Linear Gaussian setting

$$\begin{aligned}x_{k+1} &= Ax_k + Bu_k + w_k & w_k &\sim N(0, W) & W &\geq 0 \\y_k &= Cx_k + v_k & v_k &\sim N(0, V) & V &> 0 \\x_0 &\sim N(\bar{x}_0, \Sigma_0)\end{aligned}$$

Standing statistical assumptions:  $x_0, w_i, v_j$  independent  $\forall i, j$

## KF equations

$$\begin{aligned}\hat{x}_{k+1|k} &= A\hat{x}_{k|k-1} + Bu_k + \underbrace{A\Sigma_{k|k-1}C^T [C\Sigma_{k|k-1}C^T + V]^{-1}}_{L_k} (y_k - C\hat{x}_{k|k-1}) \\ \Sigma_{k+1|k} &= W + A\Sigma_{k|k-1}A^T - A\Sigma_{k|k-1}C^T [C\Sigma_{k|k-1}C^T + V]^{-1} C\Sigma_{k|k-1}A^T \\ \Sigma_{0|-1} &= \Sigma_0\end{aligned}$$

# Steady-state Kalman predictor

## Problems

- Is  $\Sigma_{k+1|k}$  converging to a matrix  $\bar{\Sigma}$  as  $k \rightarrow +\infty$ ?  
 $\Leftrightarrow$  If yes,  $L_k$ , converges to a matrix  $\bar{L}$  as well
- Is  $\bar{\Sigma}$  positive definite?  
 $\Leftrightarrow$  If yes the error becomes, asymptotically, a stationary process
- Is  $A - \bar{L}C$  Schur?

All answers provided by **LQ control through duality**

# Steady-state Kalman predictor

## Theorem

Let  $B_q$  such that  $W = B_q B_q^T$ . If

- 1)  $(A, B_q)$  is reachable
- 2)  $(A, C)$  is observable

then

- A) the optimal steady-state predictor is

$$\hat{x}_{k+1|k} = A\hat{x}_{k|k-1} + Bu_k + \bar{L}[y_k - C\hat{x}_{k|k-1}] = \\ (A - \bar{L}C)\hat{x}_{k|k-1} + Bu_k + \bar{L}y_k$$

where

$$\bar{L} = A\bar{\Sigma}C^T [C\bar{\Sigma}C^T + V]^{-1}$$

and  $\bar{\Sigma}$  is the unique positive definite solution of the ARE

$$\bar{\Sigma} = A\bar{\Sigma}A^T + W - A\bar{\Sigma}C^T [C\bar{\Sigma}C^T + V]^{-1} C\bar{\Sigma}A^T$$

- B) The predictor is AS, that is  $\rho(A - \bar{L}C) < 1$

## Remarks

- From LQ control
  - ▶ Observability of  $(A, C)$  guarantees uniqueness of the solution  $\bar{\Sigma} \geq 0$  to the ARE
  - ▶ Reachability of  $(A, B_q)$  guarantees that  $\bar{\Sigma} > 0$
- The steady-state predictor (often termed the KF) is optimal (minimizes the error variance) only in the asymptotic regime

# Example 1

$$\hat{x}_{k+1|k} = A\hat{x}_{k|k-1} + Bu_k + \underbrace{A\Sigma_{k|k-1}C^T [C\Sigma_{k|k-1}C^T + V]^{-1}}_{L_k} (y_k - C\hat{x}_{k|k-1})$$
$$\Sigma_{k+1|k} = W + A\Sigma_{k|k-1}A^T - A\Sigma_{k|k-1}C^T [C\Sigma_{k|k-1}C^T + V]^{-1} C\Sigma_{k|k-1}A^T$$
$$\Sigma_{0|-1} = \Sigma_0 \quad \hat{x}_{0|-1} \sim \mathcal{N}(\bar{x}_0, \Sigma_0)$$

$$x^+ = Ax + w$$

$$y = v \quad y = cx + v \rightarrow c = 0$$

$$C = 0 \Rightarrow L_k = 0 \Rightarrow \begin{cases} \hat{x}_{k+1|k} = A\hat{x}_{k|k-1} \\ \hat{x}_{0|-1} = \bar{x}_0 \end{cases}$$

- $y$  is just noise
- No information in  $y$  about  $x \Rightarrow$  the best option is to follow the free evolution of the system from  $\bar{x}_0$

1h21/ ↓

## Example 2

**Problem:** estimate  $\bar{x} \in \mathbb{R}$  from noisy measurements collected online

**KF setting:** define a fake dynamics  $\overset{W}{\rightarrow}$   
 $+ w_k \quad w_k \sim \mathcal{WGN}(0, 0)$

$$x_{k+1} = x_k \quad V \rightarrow A = 1 \quad B = 0 \quad C = 1$$

$$y_k = x_k + v_k$$

$$x_1 \sim N(\bar{x}_1, 1) \quad (\text{initial time } k=1 \text{ instead of } k=0)$$

**Problem data:**  $\Sigma_1 = 1, W = 0, V = 1, \bar{x}_1 = 1$

Use a time-varying Kalman predictor

$$\Sigma_{k+1|k} = A \Sigma_{k|k-1} A^T + W - A \Sigma_{k|k-1} C^T \left[ C \Sigma_{k|k-1} C^T + V \right]^{-1} C \Sigma_{k|k-1} A^T$$

$$\Rightarrow \Sigma_{k+1|k} = \Sigma_{k|k-1} - \frac{\Sigma_{k|k-1}^2}{\Sigma_{k|k-1} + 1} = \frac{\Sigma_{k|k-1}}{\Sigma_{k|k-1} + 1} \quad (*)$$

$$\text{For } \Sigma_{1,0} = 1, \quad \Sigma_{2,1} = \frac{1}{1+1} = \frac{1}{2}, \quad \Sigma_{3,2} = \frac{\frac{1}{2}}{\frac{1}{2}+1} = \frac{\frac{1}{2}}{\frac{3}{2}} = \frac{1}{3}$$

## Example 2 - ctd.

$$L_k = A \Sigma_{k|k-1} C^T \left[ C \Sigma_{k|k-1} C^T + V \right]^{-1} \quad (**)$$
$$\Rightarrow L_k = \frac{\Sigma_{k|k-1}}{\Sigma_{k|k-1} + 1}$$

- Setting  $\Sigma_{1|0} = 1$ , from (\*), one has  $\Sigma_{k|k-1} = \frac{1}{k}$ . Then

$$L_k = \frac{1}{k+1}$$

## Example 2 - ctd.

### Kalman Predictor

$$\hat{x}_{k+1|k} = \hat{x}_{k|k-1} + L_k e_k = \hat{x}_{k|k-1} + \frac{e_k}{k+1}$$

$$e_k = y_k - C\hat{x}_{k|k-1}$$

Setting  $\hat{x}_{1|0} = \bar{x}_1$  one has

$$\hat{x}_{k|k-1} = \frac{\bar{x}_1 + y_1 + y_2 + \dots + y_{k-1}}{k}$$

- average of all data (including  $\bar{x}_1$ )  
 $\hookrightarrow$  Best option for estimating a constant from measurements with the same noise variance.
- When new measurements are available, the estimate improves  
 $\hookrightarrow$  it makes sense that  $L_k \rightarrow 0$
- Since  $W = B_q B_q^T \Rightarrow B_q = 0$ , the pair  $(A, B_q)$  is not reachable  
 $\hookrightarrow$  the steady-state predictor is not guaranteed to be AS.  
 $\hookrightarrow$  indeed it is not because  $\bar{L} = 0$  and, hence

$$\hat{x}_{k+1|k} = \hat{x}_{k|k-1}$$

which implies  $e_{k+1} = e_k$

$$\begin{aligned} \hat{x}_{2|1} &= \bar{x}_1 + \frac{y_1 - \bar{x}_1}{2} = \frac{2\bar{x}_1 + y_1 - \bar{x}_1}{2} = \frac{\bar{x}_1 + y_1}{2} \\ \hat{x}_{3|2} &= \hat{x}_{2|1} + \frac{y_2 - \hat{x}_{2|1}}{3} = \frac{\bar{x}_1 + y_1}{2} + \frac{y_2 - \frac{\bar{x}_1 + y_1}{2}}{3} = \frac{3\bar{x}_1 + 3y_1 + 2y_2 - \bar{x}_1 - y_1}{6} \\ &= \frac{2\bar{x}_1 + 2y_1 + 2y_2}{6} = \frac{\bar{x}_1 + y_1 + y_2}{3} \end{aligned}$$

## Example 3: comparison of Luenberger and Kalman estimators

Noiseless system

$$x_{k+1} = ax_k, \quad x_k \in \mathbb{R} \quad (1)$$

$$y_k = x_k \rightarrow C = 1 \quad (2)$$

$$x_0 = \bar{x}$$

Luenberger observer

$$\hat{x}_{k+1} = a\hat{x}_k + L(y_k - \hat{x}_k)$$

$$\downarrow \text{error } e_k = x_k - \hat{x}_k$$

$$e_{k+1} = (a - L)e_k$$

- Dead beat estimator for  $a = L$  (best choice in the deterministic setting)

Consider now noisy measurements, replacing (2) with

$$y_k = x_k + v_k \quad v_k \sim \text{WGN}(0, 1)$$

## Example 3 - ctd.

Error

$$e_{k+1} = (a - L)e_k - Lv_k$$

Error variance (assuming  $E[\hat{x}_0] = \bar{x}$ )

$$\text{Var}[e_{k+1}] = (a - L)^2 \text{Var}[e_k] + L^2$$

*Prove @ home*

Assume  $|a - L| < 1$ . Which  $L$  minimises the steady-state variance  $E$ ?

$E$  verifies

$$E = (a - L)^2 E + L^2 \rightarrow E = \frac{L^2}{1 - (a - L)^2}$$

The optimal gain is  $\bar{L} = \frac{a^2 - 1}{a}$  (check at home)

The dead beat estimator is no longer optimal!

## Example 3 - ctd.

We show that  $\bar{L}$  is the gain of the steady-state KF

For  $A = a$ ,  $B = 0$ ,  $C = 1$ ,  $\text{Var}[w_k] = 0$ ,  $\text{Var}[v_k] = 1$ , under the usual statistical assumptions we have the KF formulae

$$L_k = (a \Sigma_{k|k-1}) (\Sigma_{k|k-1} + 1)^{-1} \xrightarrow{\Sigma_{k|k-1} \rightarrow a^{2k-1}} \frac{a(a^{2k-1})}{(a^{2k-1}) + 1} \approx$$

$$\Sigma_{k+1|k} = a^2 \Sigma_{k|k-1} - L_k^2 (\Sigma_{k|k-1} + 1) = \frac{a(a^{2k-1})}{a^{2k}} =$$

$$= \frac{a^2 - 1}{a}$$

Then  $\Sigma_{k+1|k} = \frac{a^2 \Sigma_{k|k-1}}{1 + \Sigma_{k|k-1}}$  and for  $k \rightarrow +\infty$  one has

$$\Sigma_{k|k-1} \rightarrow \bar{\Sigma} = a^2 - 1 \text{ (shown in the exercise session)}$$

$$\text{Therefore } L_k \rightarrow \frac{a^2 - 1}{a}$$

## Generalization: correlated noise

$$x^+ = Ax + Bu + w \quad (1)$$

$$y = Cx + v \quad (2)$$

Usual assumptions, except that, for  $\xi = \begin{bmatrix} w \\ v \end{bmatrix}$

$$E \begin{bmatrix} \xi & \xi^T \end{bmatrix} = \begin{bmatrix} W & Z \\ Z^T & V \end{bmatrix} \quad Z \neq 0$$

**Trick:** add and subtract  $ZV^{-1}y_k$  in (1)

$$\begin{aligned} x^+ &= Ax + Bu + w + ZV^{-1}y - ZV^{-1}Cx - ZV^{-1}v \\ &= \underbrace{(A - ZV^{-1}C)}_{\bar{A}} x + \underbrace{Bu + ZV^{-1}y}_{\bar{u}} + \bar{w} \end{aligned}$$

where  $\bar{w} = w - ZV^{-1}v$

## Generalization: correlated noise

- Key property:  $\bar{w}_k$  and  $v_k$  are uncorrelated

$$\text{check } E \left[ \bar{w}_k \quad v_k^T \right] = \underbrace{E \left[ w_k \quad v_k^T \right]}_Z - ZV^{-1} \underbrace{E \left[ v_k \quad v_k^T \right]}_V = 0$$

- $\text{Var} [\bar{w}_k] = W - ZV^{-1}Z^T \geq 0$

The system

$$x^+ = \bar{A}x + \bar{u} + \bar{w} \quad (3)$$

$$y = Cx + v \quad (4)$$

verifies the assumption that  $\bar{w}$  and  $v$  are uncorrelated

↪ design the KF for (3) and (4)

### Remark

Note that  $\bar{u}_k$  is known at each  $k = 0, 1, \dots$

The idea is to **dump a known term** into the **input** to **remove correlations** *zh(1) ↓*

# Parameter tuning in KF

Input data (under the standard assumptions on noise)

$$\text{Var}[w_k] = W$$

$$\text{Var}[v_k] = V$$

$$\text{Var}[x_0] = \Sigma_0 (= \Sigma_{0|-1})$$

- $V$  models the **sensor accuracy**: usually **known**
- $W$  accounts for **unknown disturbances** and **model mismatch** } **Seldom known apriori**  
↳ **trial and error**
  - ▶  $W = 0$ : bad choice because  $L_k \rightarrow 0$   
Motivation: perfect plant knowledge ( $W = 0$ )  $\rightarrow$  the optimal solution is to do open-loop estimation after the initial condition errors have died out.  
↳ not realistic as **some process noise is always present in reality!**
- $\text{Var}[x_0]$  often unknown as well, but less critical than  $\sim W$  (reason: If  $\Sigma_{k+1|k}$  converges, this initial condition is forgotten)
- **Common choice**:  $V$  and  $W$  diagonal

## Remark

In several applications,  $w$ , is not white

- generalizations of KF to non-white noise exist
- practical trick: make  $w$  "bigger" for introducing robustness against non-whiteness of  $w$ 
  - ▶ does not always work but it is reasonable

- Steady-state  $KF$

- ▶ output noise vs. process noise  $\rightarrow$  set  $V$  "bigger" than  $W \rightarrow$  "slower" convergence (the filter does not trust the measured outputs)

# Performance of KF

Accuracy is indicated by the error  $x_k - \hat{x}_k$ , but  $x_k$  is unknown.

## Idea

Consider the innovation sequence  $\nu_k = y_k - C\hat{x}_{k|k-1}$

## Recall

If the system and noise models are perfect (i.e.  $(A, B, C)$ ,  $\Sigma_0$ ,  $\bar{x}_0$ ,  $V$  and  $W$  are perfectly known), then

- $\nu_k \sim \mathcal{N}(0, S_k)$ , where  $S_k = V + C\Sigma_{k|k-1}C^T$ .
- $\nu_{k_1}$  and  $\nu_{k_2}$  are independent if  $k_1 \neq k_2$ .

**There is a mismatch between the real and the assumed model if:**

- $\nu_k$  has not zero mean or
- $\nu_{k_1}$  and  $\nu_{k_2}$ ,  $k_1 \neq k_2$ , are correlated or
- $\text{Var}[\nu_k]$  is not  $S_k$ .

# Innovation-based tests

## Simplifying assumption

$\nu_k$  is scalar (hence  $\nu_k$  is scalar as well)

## Tests

- **Test 1:** Consider the confidence interval of a Gaussian distribution,  $\nu_k \in [-2\sqrt{S_k}, 2\sqrt{S_k}]$  with probability 95%.
- **Test 2:** Consider that  $\tilde{\nu}_k := S_k^{-\frac{1}{2}} \nu_k$ , called the normalized innovation sequence, is WGN(0,1).
  - ▶ The autocorrelation function is  $\tilde{r}(\tau) = E[\tilde{\nu}_k \tilde{\nu}_{k+\tau}]$ , for any  $k$ . *can't be computed  $\rightarrow$  estimate it using samples*
  - ▶ Consider the normalized approximation of  $\tilde{r}$  given by  $\gamma(\tau) = \frac{\hat{r}(\tau)}{\hat{r}(0)}$ , where  $\hat{r}(\tau) = \frac{1}{N} \sum_{k=1}^{N-\tau} \tilde{\nu}_k \tilde{\nu}_{k+\tau}$ . One has that, for  $\tau > 0$ ,

$\sqrt{N}\gamma(\tau)$  converges in distribution to  $\mathcal{N}(0, 1)$ , as  $N \rightarrow +\infty$ .

Thus,  $P(-\frac{2}{\sqrt{N}} \leq \gamma(\tau) \leq \frac{2}{\sqrt{N}}) \approx 0.95$ , for any  $\tau > 0$ .

*N: n° of samples*

# Example<sup>1</sup>

Applying KF to the following system

$$x_k = \begin{bmatrix} \text{position} \\ \text{velocity} \end{bmatrix}$$

→ discretization of  $\frac{d(\text{position})}{dt} = \text{velocity}$

$$x_{k+1} = \begin{bmatrix} 1 & \Delta T \\ 0 & 1 \end{bmatrix} x_k + w_k, \quad \Delta T = 1, \quad (1)$$

$$y_k = \begin{bmatrix} 1 & 0 \end{bmatrix} x_k + v_k, \quad (2)$$

↳ noisy measurement of the position

where

$$W_k = 0.01 \begin{bmatrix} \Delta T^3/3 & \Delta T^2/2 \\ \Delta T^2/2 & \Delta T \end{bmatrix}, \quad V_k = 0.1.$$

<sup>1</sup> Ian Reid. *Estimation II*

## Example (ctd.)

- Performance of the time-varying KF under perfect system model

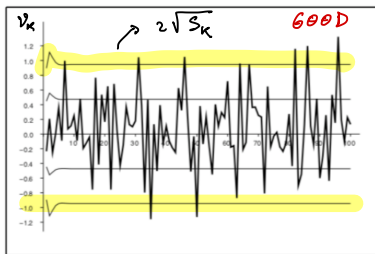


Figure 1 Innovation and innovation std bounds

Test 1

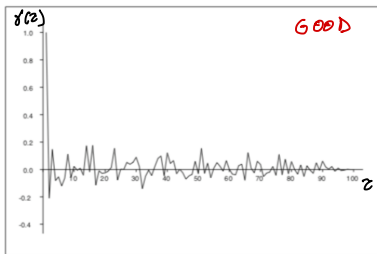


Figure 2 Plot of  $\gamma(\tau)$

Test 2

## Remark

- Figure 1 :  $\nu_k$  is consistent with  $S_k$ .
- Figure 2 :  $\tilde{\nu}_k$  is white. Notice that  $\gamma(\tau) = \hat{r}(\tau)/\hat{r}(0)$  is the approximate normalized autocorrelation.

## Example (ctd.)

- If the process noise variance  $W_k$  is underestimated by a factor 10

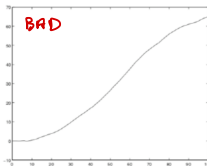


Figure 3 Estimated state

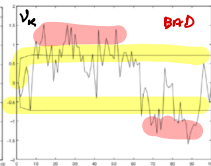


Figure 4 Innovation and innovation std bounds

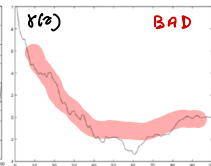


Figure 5 Plot of  $\gamma(\tau)$

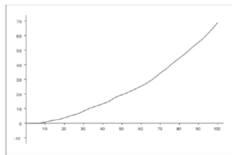


Figure 6 True state

### Remark

- The estimation of the true state (which is known only because we are simulating the system) is poor
- $\nu_k$  is not consistent with  $S_k$  and  $\tilde{\nu}_k$  is not white.

## Example (ctd.)

- If the measurement noise  $V_k$  is underestimated by a factor 10

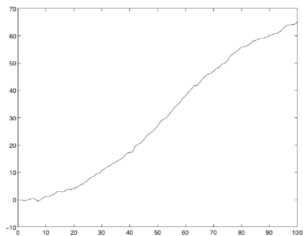


Figure 7 Estimated state

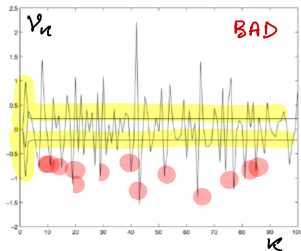


Figure 8 Innovation and innovation std bounds

TEST 1

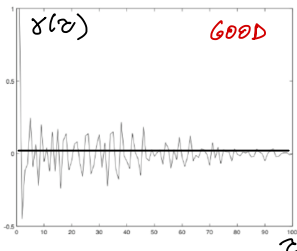


Figure 9 Plot of  $\gamma(r)$

TEST 2

### Remark

- $\nu_k$  exceeds the  $2\sigma$  bounds, but the autocorrelation sequence does not show obvious time correlation.

## Example (ctd.)

- Modelling error: suppose the motion follows a constant-acceleration model, however the KF utilizes the previous constant-velocity model. The true process is described by

$$\begin{array}{l} \left[ \begin{array}{l} \text{Position} \\ \text{Velocity} \\ \text{Acceleration} \end{array} \right] \end{array} \quad x_{k+1} = \begin{array}{c} \xrightarrow{\text{previous model}} \\ \left[ \begin{array}{cc|c} 1 & \Delta T & \Delta T^2/2 \\ 0 & 1 & \Delta T \\ \hline 0 & 0 & 1 \end{array} \right] \end{array} x_k + w_k, \quad (1)$$

$y_k = x_{1,k} + v_k = \text{noisy measurement of the position}$   
while the process noise is described by

$$W_k = 10^{-4} \begin{bmatrix} \Delta T^5/20 & \Delta T^4/8 & \Delta T^3/6 \\ \Delta T^4/8 & \Delta T^3/3 & \Delta T^2/2 \\ \Delta T^3/6 & \Delta T^2/2 & \Delta T \end{bmatrix}.$$

## Example (ctd.)

$$x^{\dagger} = x + T u + w \quad u = \text{velocity}$$
$$y = x + v$$

- The KF performance when the model is wrong

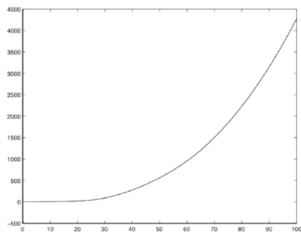


Figure 10 Estimated state

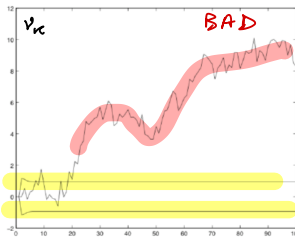


Figure 11 Innovation and innovation std bounds

Test 1

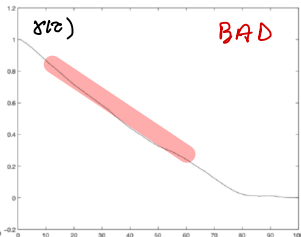


Figure 12 Plot of  $\gamma(\tau)$

Test 2

## Remark

- $E[v_k] \neq 0$
- The autocorrelation is non-negligible even for large  $\tau$