Lecture 12

The steady-state KF
Performance of KF

Giancarlo Ferrari Trecate[1]

[1]Dependable Control and Decision Group
École Polytechnique Fédérale de Lausanne (EPFL), Switzerland
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

$$\hat{x}_{k+1|k} = A\hat{x}_{k|k-1} + Bu_k + \underbrace{A\Sigma_{k|k-1}C^T\left[C\Sigma_{k|k-1}C^T + V\right]^{-1}}_{L_k}(y_k - C\hat{x}_{k|k-1})$$

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

$$\Sigma_{0|-1} = \Sigma_0$$

Steady-state Kalman predictor

Problems

- Is $\Sigma_{k+1|k}$ converging to a matrix $\bar{\Sigma}$ as $k \to +\infty$?
  $\hookrightarrow$ If yes, L_k, converges to a matrix $\bar{L}$ as well
- Is $\bar{\Sigma}$ positive definite?
  $\hookrightarrow$ 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

"Q" in LQR

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

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

where

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

for filtering

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

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

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

KF equations

$$\hat{x}_{k+1|k} = A\hat{x}_{k|k-1} + Bu_k + \underbrace{A\Sigma_{k|k-1}C^T\left[C\Sigma_{k|k-1}C^T + V\right]^{-1}}_{L_k}(y_k - C\hat{x}_{k|k-1})$$

$$\Sigma_{k+1|k} = W + A\Sigma_{k|k-1}A^T - A\Sigma_{k|k-1}C^T\left[C\Sigma_{k|k-1}C^T + V\right]^{-1}C\Sigma_{k|k-1}A^T$$
$$\Sigma_{0|-1} = \Sigma_0$$

$$x^+ = Ax + w$$
$$y = v + 0\,x$$

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

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

1h 411 ↓

Example 2

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

KF setting: define a fake dynamics

$$x_{k+1} = x_k \rightarrow A = 1 \quad B = 0 \quad C = 1$$
$$y_k = x_k + v_k$$
$$x_1 \sim N(\bar{x}_1, 1) \quad \text{(initial time } k = 1 \text{ instead of } k = 0\text{)}$$

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

Use a time-varying Kalman predictor

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

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

Example 2 - ctd.

KF equations

$$\hat{x}_{k+1|k} = A\hat{x}_{k|k-1} + Bu_k + \underbrace{A\Sigma_{k|k-1}C^T\left[C\Sigma_{k|k-1}C^T + V\right]^{-1}}_{L_k}(y_k - C\hat{x}_{k|k-1})$$

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

$$\Sigma_{0|-1} = \Sigma_0$$

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

$$\Rightarrow L_k = \frac{\Sigma_{k|k-1}}{\Sigma_{k|k-1}+1}$$

$$\Sigma_{2|1} = \frac{\Sigma_{1|0}}{\Sigma_{1|0}+1} = \frac{1}{1+1} = \frac{1}{2}$$

$$\Sigma_{3|2} = \frac{\Sigma_{2|1}}{\Sigma_{2|1}+1} = \frac{\frac{1}{2}}{\frac{1}{2}+1} = \frac{\frac{1}{2}}{\frac{3}{2}} = \frac{1}{3}$$

- 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.

$$\hat{x}_{2|1} = \bar{x}_1 + \frac{y_1 - C\bar{x}_1}{2} = \frac{2\bar{x}_1 + y_1 - \bar{x}_1}{2} = \frac{\bar{x}_1 + y_1}{2}$$

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

Kalman Predictor

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

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

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

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

- 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 $L_K \to 0$
- 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 $e_{k+1} = e_k$

Example 3: comparison of Luenberger and Kalman estimators

Noiseless system

$$x_{k+1} = ax_k, \quad x_k \in \mathbb{R} \tag{1}$$
$$y_k = x_k \tag{2}$$
$$x_0 = \bar{x}$$

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

$$y_k = x_k + v_k \qquad v_k \sim WGN(0, 1)$$

Example 3 - ctd.

Error

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

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

not an average!

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

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

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

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.

KF equations

$$\hat{x}_{k+1|k} = A\hat{x}_{k|k-1} + Bu_k + \underbrace{A\Sigma_{k|k-1}C^T\left[C\Sigma_{k|k-1}C^T + V\right]^{-1}}_{L_k}(y_k - C\hat{x}_{k|k-1})$$

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

$$\Sigma_{0|-1} = \Sigma_0$$

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

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

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

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

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

$$= a^2\Sigma_{k|k-1} - \Sigma^2_{k|k-1}\cdot\frac{a^2}{\Sigma_{k|k-1}+1} \rightarrow L_k^2\left(\Sigma_{k|k-1}+1\right)$$

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

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

Therefore $L_k \to \dfrac{a^2-1}{a}$

$$\frac{a\bar{\Sigma}}{\bar{\Sigma}+1} = \frac{a(a^2-1)}{a^2-1+1} = \frac{a^2-1}{a}$$

Generalization: correlated noise

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

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

Usual assumptions, except that, for $\xi = \begin{bmatrix} w \\ v \end{bmatrix}$ $\quad \mathbb{E}[\xi] = 0$

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

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

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

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

Generalization: correlated noise

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

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

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

$E[\bar{w}] = 0$

$E\left[(w - ZV^{-1}v)(w - ZV^{-1}v)^T \right] =$

$= E[ww^T] + E[ZV^{-1}vv^TV^{-1}Z^T] +$

$-2E[wv^TV^{-1}Z^T] =$

$= W + ZV^{-1}\cancel{VV^{-1}}Z^T$

$-2\underbrace{E[wv^T]}_{Z}V^{-1}Z^T =$

$= W + ZV^{-1}Z^T - 2ZV^{-1}Z^T =$

$= W - ZV^{-1}Z^T$

The system

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

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

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 $k = 0, 1, \ldots$

The idea is to dump a known term into the input to remove correlations

Parameter tuning in KF

Input data (under the standard assumptions on noise)

$$
\begin{aligned}
Var\left[w_k\right] &= W \\
Var\left[v_k\right] &= V \\
Var\left[x_0\right] &= \Sigma_0 (= \Sigma_{0|-1})
\end{aligned}
$$

Linear Gaussian setting

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

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

- V models the sensor accuracy: usually known
- W accounts for $\left.\begin{array}{l}\text{unknown disturbances} \\ \text{model mismatch}\end{array}\right\}$ $\begin{array}{l}\text{Seldom known apriori} \\ \hookrightarrow \text{trial and error}\end{array}$
  - $W = 0$: 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.
    $\hookrightarrow$ not realistic as some process noise is always present in reality!
- $Var\left[x_0\right]$ 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), Σ_0, $\bar{x}_0$, V and W are perfectly known), then

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

reference properties for developing statistical tests.

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

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

Innovation-based tests

Simplifying assumption

y_k is scalar (hence ν_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.
    - 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 $\tau > 0$,

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

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

Example[1]

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

Applying KF to the following system

discretization of $\dfrac{d\,(\text{position})}{dt} = \text{velocity}$

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

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

where

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

[1] Ian Reid. *Estimation II*

Example (ctd.)

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

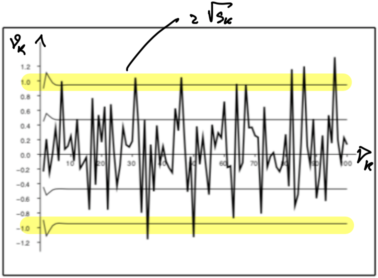

Figure 1 Innovation and innovation std bounds

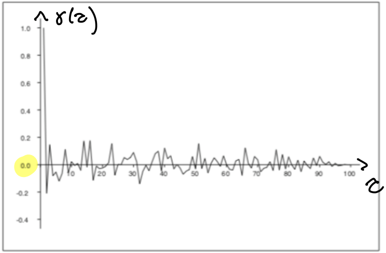

Figure 2 Plot of $\gamma(\tau)$

Remark

- Figure 1 : ν_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

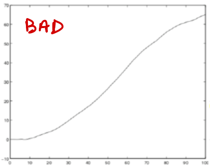

Figure 3 Estimated state

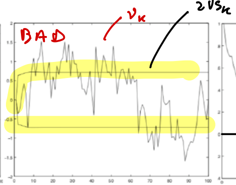

Figure 4 Innovation and innovation std bounds

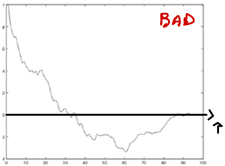

Figure 5 Plot of $\gamma(\tau)$

big difference

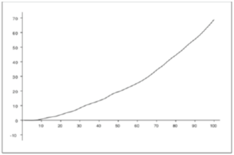

Figure 6 True state

Remark

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

Example (ctd.)

- If the measurment noise V_k is underestimated by a factor 10

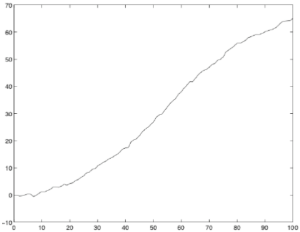

Figure 7 Estimated state

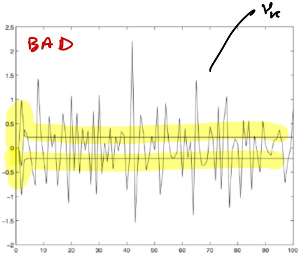

Figure 8 Innovation and innovation std bounds

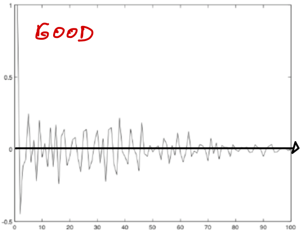

Figure 9 Plot of $\gamma(\tau)$

Remark

- ν_k exceeds the 2σ 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

position
velocity
acceleration

$$x_{k+1} = \begin{bmatrix} 1 & \Delta T & \Delta T^2/2 \\ 0 & 1 & \Delta T \\ 0 & 0 & 1 \end{bmatrix} x_k + w_k, \tag{1}$$

acceleration is integrated white noise

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.)

- The KF performance when the model is wrong

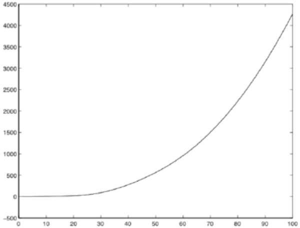

Figure 10 Estimated state

Figure 11 Innovation and innovation std bounds

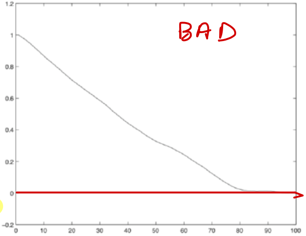

Figure 12 Plot of $\gamma(\tau)$

Remark

- $E[\nu_k] \neq 0$
- The autocorrelation is non-negligible even for "large" τ e.g in the interval $[1, 60]$

24 L12