Kalman filter - for linear systems

CA9/CA3 course Nonlinear Control Systems

Torben Knudsen

Automation and Control
Department of Electronic Systems
Aalborg University
Denmark

AALBORG UNIVERSITY
DENMARK

Outline

State estimation part of the course

State estimation problem for linear systems

Estimators based on a probabilistic models

Kalman filter for the linear DD problem

Possibilities with the Kalman filter

Challenges using the Kalman filter

State estimation part of the course
Introduction

- *State estimation* is part of the CA9/CA3 course: Nonlinear Control Systems.
- State estimation is used for estimating something which is not directly measured, or which is measured with lots of noise, from something else which is measured.
- State estimates can be used for diagnostics, fault detection or control as e.g. state space control.
- State estimators can be based on *deterministic models* and are then called observers.
- Here the state estimators are based on *stochastic models*.

Literature and lectures

- Good literature for this course is hard to find.
- We will use the same literature as last year: Mohinder S. Grewal and Angus P. Andrews, Kalman Filtering, Wiley, 2008.
- The lectures will not follow the book closely but will try to give alternative expositions as well and sometimes use other notation.
- The focus is on discrete time models. Continuous time models will only be covered where appropriate.

Given a state space model and *input/output* measurement up to and including time t estimate the *state* at time s. The estimation are given specific names depending on the relation between t and s.

- $s < t$ State *smoothing*.
- $s = t$ State *filtering*.
- $s > t$ State *prediction*.

The discrete-discrete (DD) estimation problem

Assume the system:

$$x_{k+1} = \Phi_k x_k + \Gamma_k u_k + w_k \ , \ w_k \in \mathrm{NID}(\underline{0}, Q_k) \ ,$$
$$y_k = H_k x_k + D_k u_k + v_k \ , \ v_k \in \mathrm{NID}(\underline{0}, R_k) \ ,$$
$$\mathrm{E}(w(k)v(l)^{\mathtt{T}}) = \underline{\underline{0}}$$

Given measurements and initial values:

$$\mathcal{Y}_k \triangleq y_0, y_1, y_2 \ldots y_k \ , \ u_0, u_1, u_2 \ldots u_k$$
$$x_0 \in \mathrm{N}(\hat{x}_0, P_0)$$

Find: $\hat{x}_k$ which minimized the mean square error

$$\mathrm{E}((x_k - \hat{x}_k)^{\mathtt{T}} M (x_k - \hat{x}_k)) \ , \ M > 0$$

The continuous-discrete (CD) estimation problem

The state model is in continuous time i.e. a stochastic differential equation (SDE):

$$
\begin{aligned}
dx(t) &= (F(t)x(t) + B(t)u(t))dt + d\omega(t) \ , \\
&\qquad \omega(t) \in \mathrm{W}(Q(t)) \ , \\
y(t_k) &= H_k x(t_k) + D_k u(t_k) + v_k \ , \ v_k \in \mathrm{NID}(\underline{0}, R_k)
\end{aligned}
$$

Given measurements and initial values:

$$
\begin{aligned}
&\mathcal{Y}_k \triangleq y(t_0), y(t_1), y(t_2) \ldots y(t_k) \ , \ u(s), t_0 \leq s \leq t_k \\
&x(t_0) \in \mathrm{N}(\hat{x}(t_0), P(t_0))
\end{aligned}
$$

Find: $\hat{x}_k$ which minimized the mean square error

$$
\mathrm{E}((x_k - \hat{x}_k)^{\mathtt{T}} M (x_k - \hat{x}_k)) \ , \ M > 0
$$

Using standard calculus for ODE the SDE can be rewritten as follows:

$$
\begin{aligned}
&dx(t) = (F(t)x(t) + B(t)u(t))dt + d\omega(t) \Rightarrow \\
&\dot{x}(t) = F(t)x(t) + B(t)u(t) + w(t)\ , \\
&w(t) = \frac{d\omega(t)}{dt}
\end{aligned}
$$

- This simplified formulation is wrong from a strict mathematical point of view.
- To make any sense $w(t)$ must be *continuous white noise* which is a unphysical signal with constant spectrum for all frequencies and consequently an infinite variance.

The continuous-continuous (CC) estimation problem

- The problem where continuous measurements is assumed is called the continuous-continuous estimation problem and the solution is called the Kalman-Busy filter.
- In practice the measurements are always in discrete time.
- Continuous measurements with continuous time white noise is also unphysical in some sense.

Notice: You do not need to worry about continuous time white noise and details on SDE's in this course.

Estimation from a one dimensional distribution

Given the density function $f(x)$ for a scalar stochastic variable x find the estimate $\hat{x}$ that minimize the mean square error (MSE) $\mathrm{E}((x - x_e)^2)$ where x_e is any estimate.

$$
\begin{aligned}
L(x_e) &= \mathrm{E}\left((x - x_e)^2\right) \\
&= \mathrm{E}\left(((x - \mu) - (x_e - \mu))^2\right) \ , \ \mu \triangleq \mathrm{E}(x) \\
&= \mathrm{E}\left((x - \mu)^2\right) + \mathrm{E}\left((x_e - \mu)^2\right) \qquad \Rightarrow \\
&\quad - 2\,\mathrm{E}\left((x - \mu)(x_e - \mu)\right) \\
&= \mathrm{V}(x) + (x_e - \mu)^2
\end{aligned}
$$

$$
\hat{x} \triangleq \arg\min_{x_e} L(x_e) = \mu
$$

Notice: MSE is not the only possible criteria.

Estimation from a multi dimensional distribution

Given the density function $f(x)$ for a stochastic vector x find the estimate $\hat{x}$ that minimize the mean square error (MSE) $\mathrm{E}((x - x_e)^\mathsf{T} M (x - x_e))$ where x_e is any estimate and $M > 0$.

$$
\begin{aligned}
L(x_e) &= \mathrm{E}\left((x - x_e)^\mathsf{T} M (x - x_e)\right) \\
&= \mathrm{E}\left(((x - \mu) - (x_e - \mu))^\mathsf{T} M ((x - \mu) - (x_e - \mu))\right) \\
&= \mathrm{E}\left((x - \mu)^\mathsf{T} M (x - \mu)\right) + (x_e - \mu)^\mathsf{T} M (x_e - \mu) \Rightarrow
\end{aligned}
$$

$$
\hat{x} \triangleq \arg\min_{x_e} L(x_e) = \mu
$$

Estimation from a conditional distribution

Given the *conditional* density function $f(x|y)$ for a stochastic vector x given y find the estimate $\hat{x}$ that minimize the *conditional* mean square error (MSE) $\mathrm{E}((x-x_e)^\mathsf{T} M(x-x_e)|y)$ where x_e is any estimate and $M > 0$.

- The *conditional* distribution is still just a distribution.
- The solution is therefore the same as for other distribution i.e. the mean value which now is a conditional mean

$$L(x_e) = \mathrm{E}\left((x-x_e)^\mathsf{T} M(x-x_e)|y\right) \quad \Rightarrow$$
$$\hat{x} \triangleq \arg\min_{x_e} L(x_e) = \mathrm{E}(x|y)$$

Using unconditional loss

The conditional mean estimator will also minimize the *unconditional* loss

$$
\begin{aligned}
L_u(x_e) &\triangleq \mathrm{E}\left((x - x_e)^\mathsf{T} M (x - x_e)\right) \\
&= \mathrm{E}\left(\mathrm{E}\left((x - x_e)^\mathsf{T} M (x - x_e)|y\right)\right)
\end{aligned}
$$

The estimator $\hat{x} = \mathrm{E}(x|y)$ that minimizes the conditional MSE for any given y will also minimize the mean (with respect to y) MSE.
Notice: Above the general rule
$\mathrm{E}(g(x, y)) = \mathrm{E}(\mathrm{E}(g(x, y)|y))$ was used.

The error is unbiased

$$\begin{aligned}
\mathrm{E}\left((x-\hat{x})\right) &= \mathrm{E}\left(\mathrm{E}\left((x-\hat{x})|y\right)\right) \\
&= \mathrm{E}\left(\mathrm{E}\left((x-\mathrm{E}(x|y))|y\right)\right) \\
&= \mathrm{E}\left(\underline{0}\right) \\
&= \underline{0}
\end{aligned}$$

Notice:

- An estimator with bias would be a bad estimator.

The error is uncorrelated with the known information

This is also called the orthogonality principle.

$$\begin{aligned} \mathrm{E}\left((x-\hat{x})y^{\mathsf{T}}\right) &= \mathrm{E}\left(\mathrm{E}\left((x-\hat{x})y^{\mathsf{T}}|y\right)\right) \\ &= \mathrm{E}\left(\mathrm{E}\left((x-\mathrm{E}(x|y))|y\right)y^{\mathsf{T}}\right) \\ &= \mathrm{E}\left(\underline{0}y^{\mathsf{T}}\right) \\ &= \underline{\underline{0}} \end{aligned}$$

Notice:

- correlation between $(x-\hat{x})$ and y would mean that there were more information to extract.
- correlation between $(x-\hat{x})$ and x is not zero.
- but correlation between $(x-\hat{x})$ and $\hat{x}$ is zero because $\hat{x}$ is a function of y.

Conditional Gaussian distributions

Assume x, y is jointly Gaussian then $x|y$ is Gaussian with mean and covariance

$$\mathrm{E}(x|y) = \mathrm{E}(x) + \mathrm{Cov}(x,y)\,\mathrm{Cov}(y)^{-1}(y - \mathrm{E}(y)) \ ,$$

$$\mathrm{Cov}(x|y) = \mathrm{Cov}(x) - \mathrm{Cov}(x,y)\,\mathrm{Cov}(y)^{-1}\,\mathrm{Cov}(y,x) \ ,$$

$$\mathrm{Cov}(x|y) \triangleq \mathrm{E}(\tilde{x}\tilde{x}^{\mathsf{T}}|y) = \mathrm{E}(\tilde{x}\tilde{x}^{\mathsf{T}}) \ , \ \tilde{x} \triangleq x - \mathrm{E}(x|y)$$

Independent measurements

Assume x, u, v is jointly Gaussian and that u, v is independent then

$$\mathrm{E}(x|u,v) = \mathrm{E}(x|u) + \mathrm{E}(x|v) - \mathrm{E}(x)$$

Notation

The notation used is not always the same as in Grewal.

$$
\begin{aligned}
\hat{x}_{k|k} &\triangleq \hat{x}_k^{(+)} \triangleq \mathrm{E}(x_k|\mathcal{Y}_k) \ , \\
\hat{x}_{k|k-1} &\triangleq \hat{x}_k^{(-)} \triangleq \mathrm{E}(x_k|\mathcal{Y}_{k-1}) \ , \\
\tilde{x}_{k|k} &\triangleq \tilde{x}_k^{(+)} \triangleq x_k - \hat{x}_{k|k} \ , \\
\tilde{x}_{k|k-1} &\triangleq \tilde{x}_k^{(-)} \triangleq x_k - \hat{x}_{k|k-1} \ , \\
P_{k|k} &\triangleq P_k^{(+)} \triangleq \mathrm{E}(\tilde{x}_{k|k}\tilde{x}_{k|k}^{\mathtt{T}}) \ , \\
P_{k|k-1} &\triangleq P_k^{(-)} \triangleq \mathrm{E}(\tilde{x}_{k|k-1}\tilde{x}_{k|k-1}^{\mathtt{T}})
\end{aligned}
$$

Main idea

The derivation builds on the following (which would be hard to figure out on your own):

- The LMS estimator is $\hat{x}_k = \mathrm{E}(x_k|\mathcal{Y}_k)$
- The state space model is recursive in time then the KF should also be recursive in time.
- Given the current estimate $\hat{x}_{k|k}$ it is easy to find predictions $\hat{x}_{k+j|k}$, $j \geq 1$ including the error covariance.
- The KF is split into a *measurement update* using y_k to update from $\hat{x}_{k|k-1}$ to $\hat{x}_{k|k}$ and a *time update* to update from $\hat{x}_{k|k}$ to $\hat{x}_{k+1|k}$ using no new measurements.

Time update

Assume $\hat{x}_{k|k}$ and $P_{k|k}$ known then

$$
\begin{aligned}
x_{k+1} &= \Phi_k x_k + \Gamma_k u_k + w_k \Rightarrow \\
\hat{x}_{k+1|k} &= \Phi_k \hat{x}_{k|k} + \Gamma_k u_k \Rightarrow \\
\tilde{x}_{k+1|k} &= \Phi_k \tilde{x}_{k|k} + w_k \Rightarrow \\
P_{k+1|k} &= \Phi_k P_{k|k} \Phi_k^{\mathsf{T}} + Q_k \ .
\end{aligned}
$$

From $\hat{x}_{k|k-1}$ to $\hat{y}_{k|k-1}$

$$
\begin{aligned}
y_k &= H_k x_k + D_k u_k + v_k \Rightarrow \\
\hat{y}_{k|k-1} &= H_k \hat{x}_{k|k-1} + D_k u_k \Rightarrow \\
\tilde{y}_{k|k-1} &= H_k \tilde{x}_{k|k-1} + v_k \Rightarrow \\
\mathrm{P}^y_{k|k-1} &= H_k P_{k|k-1} H_k^{\mathsf{T}} + R_k \ .
\end{aligned}
$$

Measurements update - estimate

Assume $\hat{x}_{k|k-1}$ and $P_{k|k-1}$ and $y(k)$ known then

$$\begin{aligned}\hat{x}_{k|k} &= \mathrm{E}(x_k|\mathcal{Y}_k)\\ &= \mathrm{E}(x_k|y_k,\mathcal{Y}_{k-1})\\ &= \mathrm{E}(x_k|\tilde{y}_{k|k-1},\mathcal{Y}_{k-1})\\ &= \hat{x}_{k|k-1} + \mathrm{E}(x_k|\tilde{y}_{k|k-1}) - \mathrm{E}(x_k)\ .\end{aligned}$$

$$\mathrm{E}(x_k|\tilde{y}_{k|k-1}) = \mathrm{E}(x_k) + \\ \mathrm{Cov}(x_k,\tilde{y}_{k|k-1})\,\mathrm{Cov}(\tilde{y}_{k|k-1})^{-1}\tilde{y}_{k|k-1}$$

$$
\begin{aligned}
&\mathrm{Cov}(x_k, \tilde{y}_{k|k-1}) \\
&\quad = \mathrm{Cov}(\hat{x}_{k|k-1} + \tilde{x}_{k|k-1}, \tilde{y}_{k|k-1}) \\
&\quad = \mathrm{Cov}(\hat{x}_{k|k-1}, \tilde{y}_{k|k-1}) + \mathrm{Cov}(\tilde{x}_{k|k-1}, \tilde{y}_{k|k-1}) \\
&\quad = \mathrm{Cov}(\tilde{x}_{k|k-1}, \tilde{y}_{k|k-1}) \\
&\quad = \mathrm{E}(\tilde{x}_{k|k-1}\tilde{y}_{k|k-1}) \\
&\quad = \mathrm{E}(\tilde{x}_{k|k-1}(H_k\tilde{x}_{k|k-1} + v_k)^{\mathtt{T}}) \\
&\quad = \mathrm{E}(\tilde{x}_{k|k-1}\tilde{x}_{k|k-1}^{\mathtt{T}}H_k^{\mathtt{T}}) \\
&\quad = P_{k|k-1}H_k^{\mathtt{T}}
\end{aligned}
$$

Collecting the above

$$
\hat{x}_{k|k} = \hat{x}_{k|k-1} + P_{k|k-1}H_k^{\mathtt{T}}(H_kP_{k|k-1}H_k^{\mathtt{T}} + R_k)^{-1}\tilde{y}_{k|k-1}
$$

Measurements update - error variance

$$
\begin{aligned}
K_k &\triangleq P_{k|k-1}H_k^\mathsf{T}(H_kP_{k|k-1}H_k^\mathsf{T}+R_k)^{-1} \Rightarrow \\
\hat{x}_{k|k} &= \hat{x}_{k|k-1}+K_k\tilde{y}_{k|k-1} \Rightarrow \\
\tilde{x}_{k|k} &= \tilde{x}_{k|k-1}-K_k\tilde{y}_{k|k-1} \\
&= \tilde{x}_{k|k-1}-K_k(H_k\tilde{x}_{k|k-1}+v_k) \\
&= (I-K_kH_k)\tilde{x}_{k|k-1}-K_kv_k \quad \Rightarrow \\
P_{k|k} &= (I-K_kH_k)P_{k|k-1}(I-K_kH_k)^\mathsf{T}+K_kR_kK_k^\mathsf{T} \\
&= (I-K_kH_k)P_{k|k-1}
\end{aligned}
$$

Kalman filter algorithm

Initial conditions:

$$x_0 \in \mathrm{N}(\hat{x}_{0|-1}, P_{0|-1})$$

Measurements update after receiving y_k and u_k:

$$\hat{y}_{k|k-1} = H_k \hat{x}_{k|k-1} + D_k u_k \ ,$$

$$\tilde{y}_{k|k-1} = y_k - \hat{y}_{k|k-1} \ ,$$

$$K_k \triangleq P_{k|k-1} H_k^{\mathsf{T}} (H_k P_{k|k-1} H_k^{\mathsf{T}} + R_k)^{-1} \ ,$$

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k \tilde{y}_{k|k-1} \ ,$$

$$P_{k|k} = (I - K_k H_k) P_{k|k-1} (I - K_k H_k)^{\mathsf{T}} + K_k R_k K_k^{\mathsf{T}} \ .$$

Time update from k to $k+1$:

$$\hat{x}_{k+1|k} = \Phi_k \hat{x}_{k|k} + \Gamma_k u_k \ ,$$

$$P_{k+1|k} = \Phi_k P_{k|k} \Phi_k^{\mathsf{T}} + Q_k \ .$$

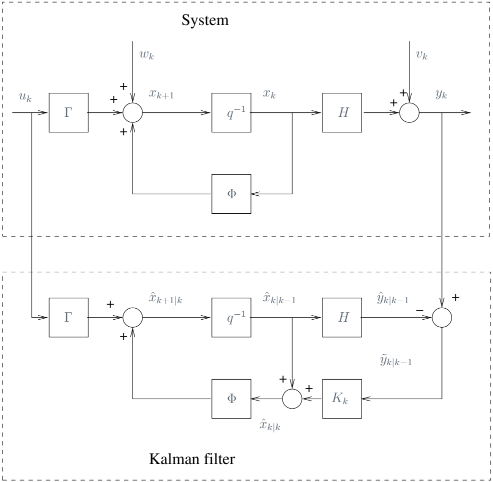
Initial conditions:
Assume $\hat{x}_{k|k}, P_{k|k}$ know,

Multi step prediction is then obtained for $j \geq 0$ by:

$$\hat{x}_{k+j+1|k} = \Phi_{k+j}\hat{x}_{k+j|k} + \Gamma_{k+j}u_{k+j} \ ,$$
$$P_{k+j+1|k} = \Phi_{k+j}P_{k+j|k}\Phi_{k+j}^{\mathsf{T}} + Q_{k+j} \ ,$$
$$\hat{y}_{k+j+1|k} = H_{k+j+1}\hat{x}_{k+j+1|k} + D_{k+j+1}u_{k+j+1} \ ,$$
$$P_{k+j+1|k}^{y} = H_{k+j+1}P_{k+j+1|k}H_{k+j+1}^{\mathsf{T}} + R_{k+j+1} \ .$$

Multi step prediction can be interpreted as: simulation starting with initial state $\hat{x}_{k|k}$

Block diagram interpretation



Other important features of Kalman filter

- The filter gain K_k and error covariances $P_{k|k}, P_{k|k-1}, P^y_{k|k-1}$ does not depend on measurements.
- If the model is time invariant the Kalman filter also converges to a time invariant filter given by the solution to the *discrete time Riccati equation*

$$P \triangleq \lim_{k\to\infty} P_{k|k-1} \ ,$$
$$P = \Phi P \Phi^{\mathsf{T}} - \Phi P H^{\mathsf{T}}(HPH^{\mathsf{T}} + R)^{-1} HP\Phi^{\mathsf{T}} + Q$$

- The measurement errors $\tilde{y}_{k|k-1}, \tilde{y}_{j|j-1}, k \neq j$ are independent and uncorrelated because for e.g. $j < k$, $\tilde{y}_{k|k-1}$ is independent of $\mathcal{Y}_{k-1}$ and $\tilde{y}_{j|j-1}$ is a linear function of $\mathcal{Y}_j \subseteq \mathcal{Y}_{k-1}$ for $j < k$.

The Kalman filter is a fantastic flexible tool

- The kalman filter can serve many purposes.
- Creative use of the parameters makes it possible to tailor the Kalman filter to the problem.
- State estimation, smoothing and prediction is the basic application.
- Parameters estimation for ARX models is a less standard problem that can be solved.

Challenges using the Kalman filter

- If the process and measurement noise is not white the filter is not optimal.
  - Colored noise is included by augmenting the original state space model.
  - Correlation between process and measurement noise can also be included in the Kalman filter.
- All parameters must be known.
- Especially noise covariances are difficult to get.
  - Some times they can be found from *first principles*.
  - If input, output and states can be measured they can be estimated.
  - If only input and output, but not states, can be measured only a *innovation* model can be estimated.
  - Noise covariances can also be interpreted as tuning parameters.
- The system must of course be *observable*.