Adaptive Systems - Lecture 8

Prelude to the Kalman Filter

Joachim Dahl
joachim@kom.aau.dk

April 3, 2006

The Kalman filtering framework

Typical **state-space** model:

$$
\begin{aligned}
x_{k+1} &= A_k x_k + B_k u_k + w_k \\
y_k &= C_k x_k + v_k
\end{aligned}
$$

where x_k is the **state-variable**, u_k is the input signal, y_k is the output, and w_k and v_k are noise-sources all at time k. The recursive model is initialized at time $k = 0$.

The Kalman filter computes the **MMSE estimate of the state-vector** x_{k+1} based on the past input $\{u_0, u_1, \ldots, u_k\}$ and output $\{y_0, y_1, \ldots, y_k\}$ for Gaussian noise-sources.

For non Gaussian noise-sources, the Kalman filter computes an LMMSE estimate.

State-space model for auto-regressive model AR(2)

$$y_k = \phi_1 y_{k-1} + \phi_2 y_{k-2} + w_k.$$

Formulated using a state-space model:

$$x_{k+1} = \begin{bmatrix} \phi_1 & 1 \\ \phi_2 & 0 \end{bmatrix} x_k + \begin{bmatrix} 1 \\ 0 \end{bmatrix} w_k$$
$$y_k = \begin{bmatrix} 1 & 0 \end{bmatrix} x_k.$$

Let's verify that the state-space model implements an AR(2) model,

$$x^1_{k+1} = \phi_1 x^1_k + x^2_k + w_k$$
$$x^2_{k+1} = \phi_2 x^1_k.$$

Changing time-index in the last equation $x^2_k = \phi_2 x^1_{k-1}$ and inserting into the first equations gives the wanted result for $y_k = x^1_k$.

State-space model for moving average model MA(2)

$$y_k = w_k + \theta_1 w_{k-1} + \theta_2 w_{k-2}.$$

Formulated using a state-space model:

$$x_{k+1} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} x_k + \begin{bmatrix} 1 \\ \theta_1 \\ \theta_2 \end{bmatrix} w_k$$

$$y_k = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} x_k.$$

Changing time-indices and combining the equations

$$x^1_{k+1} = x^2_k + w_k$$

$$x^2_{k+1} = x^3_k + \theta_1 w_k$$

$$x^3_{k+1} = \theta_2 w_k$$

gives the result.

State-space model for ARMA(2)

$$y_k = \phi_1 y_{k-1} + \phi_2 y_{k-2} w_k + \theta_1 w_{k-1} + \theta_2 w_{k-2}.$$

Formulated using a state-space model:

$$x_{k+1} = \begin{bmatrix} \phi_1 & 1 & 0 \\ \phi_2 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} x_k + \begin{bmatrix} 1 \\ \theta_1 \\ \theta_2 \end{bmatrix} w_k$$

$$y_k = [\ 1 \quad 0 \quad 0\ ] x_k.$$

Changing time-indices and combining the equations

$$x^1_{k+1} = \phi_1 x^1_k + x^2_k + w_k$$

$$x^2_{k+1} = \phi_2 x^1_k + x^3_k + \theta_1 w_k$$

$$x^3_{k+1} = \theta_2 w_k$$

gives the result.

Rational transfer functions

Consider

$$H(z) = \frac{Y(z)}{W(z)} = \frac{b_0 + b_1 z^{-1} + b_2 z^{-2}}{1 + a_1 z^{-1} + a_2 z^{-2} + a_3 z^{-3}}.$$

One realization of a state-space model:

$$x_{k+1} = \begin{bmatrix} -a_1 & 1 & 0 \\ -a_2 & 0 & 1 \\ -a_3 & 0 & 0 \end{bmatrix} x_k + \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix} w_k$$

$$y_k = [\begin{array}{ccc} 1 & 0 & 0 \end{array}] x_k.$$

Another state-space model:

$$x_{k+1} = \begin{bmatrix} -a_1 & -a_2 & -a_3 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} x_k + \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} w_k$$

$$y_k = [\begin{array}{ccc} b_0 & b_1 & b_2 \end{array}] x_k.$$

Proper transfer functions

If the numerator and denominator have same degree we use polynomial division.

$$H(z) = \frac{1/8z^3 + 1/2z^2 + 1/2z + 1/8}{z^3 + 1/3z} = \frac{1/2z^2 + 11/24z + 1/8}{z^3 + 1/3z} + 1/8$$

We realize it as

$$\begin{aligned} x_{k+1} &= Ax_k + Bw_k \\ y_k &= Cx_k + Dw_k \end{aligned}$$

with $D = 1/8$, $i.e.$,

$$\begin{aligned} x_{k+1} &= \begin{bmatrix} 0 & -1/3 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} x_k + \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} w_k \\ y_k &= \begin{bmatrix} 1/2 & 11/24 & 1/8 \end{bmatrix} x_k + 1/8w_k. \end{aligned}$$

Transfer function of a state-space model

Let us derive the transfer function of the state-space model

$$\begin{aligned} x_{k+1} &= Ax_k + Bw_k \\ y_k &= Cx_k + Dw_k. \end{aligned}$$

Using the z-transform we get

$$\begin{aligned} zX(z) &= AX(z) + BW(z) \\ Y(z) &= CX(z) + DW(z) \end{aligned}$$

which gives the result

$$Y(z) = \left[C(zI - A)^{-1}B + D\right] W(z).$$

The stochastic setup

We consider a jointly Gaussian variable (X, Y) where X and Y are vector valued random variables with mean and covariance

$$E\begin{pmatrix} X \\ Y \end{pmatrix} = \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}, \quad E\left[\begin{pmatrix} X-\mu_x \\ Y-\mu_y \end{pmatrix}\begin{pmatrix} X-\mu_x \\ Y-\mu_y \end{pmatrix}^T\right] = \begin{bmatrix} R_{xx} & R_{xy} \\ R_{yx} & R_{yy} \end{bmatrix}.$$

Let us estimate X based on knowledge of Y as,

$$\hat{x} = E[X|Y=y] = \int_{-\infty}^{\infty} x p_{X|Y}(x|y) dx$$

$i.e.$, we define the estimate as a conditional mean.

Conditional mean is conditional MMSEE

The conditional mean minimizes the conditional mean-squared error. To see this, let z be any estimate of X.

$$\begin{aligned}
\epsilon &= E[(X-z)^T(X-z)|Y=y] \\
&= E[X^TX|Y=y] - 2z^TE[X|Y=y] + z^Tz \\
&= (z - E[X|Y=y])^T(z - E[X|Y=y]) + E[X^TX|Y=y] \\
&\qquad - E[X|Y=y]^TE[X|Y=y].
\end{aligned}$$

Only the first term depends on z, so

$$z = \hat{x} = E[X|Y=y]$$

minimizes the conditional mean-squared error.

Unconditional MMSEE

In terms of minimum error variance, the conditional mean

$$\hat{X}(y) = E[X \mid Y = y]$$

is optimal, $i.e.$,

$$E_{X|Y}[\|X - \hat{X}(y)\|^2 \mid Y = y] \leq E_{X|Y}[\|X - Z(y)\|^2 \mid Y = y]$$

for any function Z that may depend on y. Taking expection over Y on both side gives us

$$E_{X,Y}[\|X - \hat{X}(Y)\|^2] \leq E_{X,Y}[\|X - Z(Y)\|^2]$$

or loosely

$$E[\|X - \hat{x}\|^2] \leq E[\|X - z\|^2]$$

$i.e.$, the conditional mean is also optimal in the unconditional minimum variance sense.

What does the Kalman filter do?

For a state-space model

$$x_{k+1} = A_k x_k + B_k u_k + w_k$$
$$y_k = C_k x_k + v_k$$

the Kalman filter minimizes the conditional error variance

$$E[\|x_{k+1} - \hat{x}_{k+1}\|^2 \mid y_0, \ldots, y_k, u_0, \ldots, u_k]$$

with

$$\hat{x}_{k+1} = E[x_{k+1} \mid y_0, \ldots, y_k, u_0, \ldots, u_k].$$

The Kalman filter does this in a recursive way, $i.e.$, $\hat{x}_{k+1}$ can be computed using only $\hat{x}_k$, y_k and u_k.

The jointly Gaussian conditional distribution

The conditional distribution can be written using Bayes' rule as

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

which for a jointly Gaussian distribution is equal to

$$f_{X|Y}(x|y) = \frac{1}{(2\pi)^{m/2}} \frac{\left| \begin{array}{cc} R_{XX} & R_{XY} \\ R_{YX} & R_{YY} \end{array} \right|^{-1/2}}{|R_{YY}|^{-1/2}}$$

$$\frac{\exp\left\{ -1/2 \left[\begin{array}{c} x-\mu_x \\ y-\mu_y \end{array} \right]^T \left[\begin{array}{cc} R_{XX} & R_{XY} \\ R_{YX} & R_{YY} \end{array} \right]^{-1} \left[\begin{array}{c} x-\mu_x \\ y-\mu_y \end{array} \right] \right\}}{\exp\left\{ -1/2(y-\mu_y)^T R_{YY}(y-\mu_y) \right\}}$$

Cholesky factorizations - a useful sidestep

Consider a positive definite matrix A partitioned as

$$A = \begin{bmatrix} A_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}.$$

We wish to factor the matrix as $A = UDU^T$, with U unit-diagonal upper-triangular and D a positive diagonal matrix.

$$\begin{aligned} A &= \begin{bmatrix} U_{11} & u_{12} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} D_{11} & 0 \\ 0 & d_{22} \end{bmatrix} \begin{bmatrix} U_{11}^T & 0 \\ u_{12}^T & 1 \end{bmatrix} \\ &= \begin{bmatrix} d_{22}u_{12}u_{12}^T + U_{11}D_{11}U_{11}^T & d_{22}u_{12}^T \\ d_{22}u_{12}^T & d_{22} \end{bmatrix} \end{aligned}$$

$i.e.$, we have the outline of a recursive procedure for computing UDU^T,

$$d_{22} = a_{22}, \qquad u_{12} = \frac{a_{12}}{a_{22}}, \qquad U_{11}D_{11}U_{11}^T = A_{11} - d_{22}u_{12}u_{12}^T.$$

Block-diagonal factorization

Let's try the same idea for a block-diagonal factorization

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} = \begin{bmatrix} I & X \\ 0 & I \end{bmatrix} \begin{bmatrix} D_{11} & 0 \\ 0 & D_{22} \end{bmatrix} \begin{bmatrix} I & 0 \\ X^T & I \end{bmatrix}$$

$$= \begin{bmatrix} D_{11} + XD_{22}X^T & XD_{22} \\ D_{22}X^T & D_{22} \end{bmatrix}.$$

It follows that

$$D_{22} = A_{22}, \qquad X = A_{12}A_{22}^{-1}, \qquad D_{11} = A_{11} - A_{12}A_{22}^{-1}A_{21},$$

$i.e.,$

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} = \begin{bmatrix} I & A_{12}A_{22}^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} A_{11} - A_{12}A_{22}^{-1}A_{21} & 0 \\ 0 & A_{22} \end{bmatrix} \begin{bmatrix} I & 0 \\ A_{22}^{-1}A_{21} & I \end{bmatrix}$$

Inverse factorization

What we need is

$$A^{-1} = \begin{bmatrix} I & 0 \\ -A_{22}^{-1}A_{21} & I \end{bmatrix} \begin{bmatrix} (A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1} & 0 \\ 0 & A_{22}^{-1} \end{bmatrix} \begin{bmatrix} I & -A_{12}A_{22}^{-1} \\ 0 & I \end{bmatrix}$$

If we denote $S_{11} = (A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1}$ then we have

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}^{-1} = \begin{bmatrix} S_{11} & -S_{11}A_{12}A_{22}^{-1} \\ -A_{22}A_{21}S_{11} & A_{22}^{-1} + A_{22}^{-1}A_{21}S_{11}A_{12}A_{22}^{-1} \end{bmatrix}$$

and

$$\begin{bmatrix} x \\ y \end{bmatrix}^{T} \begin{bmatrix} S_{11} & -S_{11}A_{12}A_{22}^{-1} \\ -A_{22}A_{21}S_{11} & A_{22}^{-1} + A_{22}^{-1}A_{21}S_{11}A_{12}A_{22}^{-1} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} =$$

$$(x - A_{12}A_{22}^{-1}y)^{T}S_{11}(x - A_{12}A_{22}^{-1}y) + y^{T}A_{22}^{-1}y.$$

Back to the jointly Gaussian conditional distribution

Using the results for the inverse factorization we have that

$$\begin{bmatrix} x-\mu_x \\ y-\mu_y \end{bmatrix}^T \begin{bmatrix} R_{XX} & R_{XY} \\ R_{YX} & R_{YY} \end{bmatrix}^{-1} \begin{bmatrix} x-\mu_x \\ y-\mu_y \end{bmatrix} =$$

$$(x-\bar{x})^T(R_{XX}-R_{XY}R_{YY}^{-1}R_{YX})^{-1}(x-\bar{x}) + (y-\mu_y)^T R_{YY}^{-1}(y-\mu_y)$$

where $\bar{x} = \mu_x + R_{XY}R_{YY}^{-1}(y-\mu_y)$. In other words,

$$f_{X|Y}(x|y) = \frac{|R_{XX}-R_{XY}R_{YY}^{-1}R_{YX}|^{-1/2}}{(2\pi)^{m/2}} \times$$
$$\exp\{-1/2(x-\bar{x})^T(R_{XX}-R_{XY}R_{YY}^{-1}R_{YX})^{-1}(x-\bar{x})\}$$

describes a Gaussian distribution with

$$E[X \mid Y] = \mu_x + R_{XY}R_{YY}^{-1}(y-\mu_y)$$
$$\text{cov}(X \mid Y) = R_{XX} - R_{XY}R_{YY}^{-1}R_{YX}.$$

Conditioning for uncorrelated variables

Assume that $X, Y_1, Y_2, \ldots, Y_n$ are jointly Gaussian and that $Y_1, Y_2, \ldots, Y_n$ are mutually uncorrelated. Then

$$E[X \mid Y_1, Y_2, \ldots, Y_n] = E[X \mid Y_1] + E[X \mid Y_2] + \cdots + E[X \mid Y_n] + (n-1)\mu_x.$$

This follows from

$$E[X \mid Y_1, Y_2, \ldots, Y_n] = \mu_x + R_{XY} R_{YY}^{-1}(y - \mu_y)$$

by observing that R_{YY} is a diagonal matrix for uncorrelated variables $Y_1, \ldots, Y_n$.

In the next lecture we will combine these results and derive the solution the Kalman filter.