CS281A/Stat241A: Statistical Learning Theory

State Space Models (11/04/04)

Lecturer: Michael I. Jordan *Scribes: Ankit Jain and Hua Chen*

1 State Space Models

Continuing from the last lecture, given $\hat{x}_{t|t}$ and $P_{t+1|t}$, we can compute $\hat{x}_{t+1|t}$ and $P_{t+1|t}$ as:

$$\begin{aligned} \hat{x}_{t+1|t+1} &= \mathbb{E}[x_t + 1 | y_0 . y_1, \ldots y_t] \\ &= \mathbb{E}[Ax_t + Gw_t | y_0, y_1, \ldots, y_t] \\ &= A\mathbb{E}[x_t | y_0, y_1, \ldots y_t] = A\hat{x}_{t|t} \end{aligned} \tag{1}$$

$$\begin{aligned} P_{t+1|t} &= \mathbb{E}[(x_{t+1} - \hat{x}_{t+1|t})(x_{t+1} - \hat{x}_{t+1|t})^T | y_0, y_1, \ldots, y_t] \\ &= \mathbb{E}[(A(x_t - \hat{x}_{t|t}) + Gw_t)(A(x_t - \hat{x}_{t|t}) + Gw_t)^T | y_0, \ldots, y_t] \\ &= AP_{t|t}A^T + GQG^T \end{aligned} \tag{2}$$

The equations (1) and (2) are called the time update equations. Our goal is to find the joint distribution of (x_{t+1}, y_{t+1}) given $(y_1, \ldots y_t)$. Now,

$$\begin{aligned} \mathbb{E}[(y_{t+1} | y_0, \ldots, y_t)] &= \mathbb{E}[Cx_{t+1} + v_t | y_0, \ldots, y_t] = C\hat{x}_{t+1|t} \\ Var(y_{t+1} | y_0, \ldots, y_t) &= CP_{t+1|t}C^T + R \end{aligned}$$

Also the covariance is given by:

$$\begin{aligned} \mathbb{E}[(y_{t+1} - \hat{y}_{t+1|t})(x_{t+1} - \hat{x}_{t+1})^T | y_0, \ldots, y_t] &= \mathbb{E}[(Cx_{t+1|t} + v_{t+1} - C\hat{x}_{t+1|t})(x_{t+1} - \hat{x}_{t+1})^T | y_0, \ldots, y_t] \\ &= CP_{t+1|t} \end{aligned}$$

<u>Summary</u>: The joint distribution of (x_{t+1}, y_{t+1}) given $(y_1, \ldots y_t)$ is Gaussian with mean $\begin{pmatrix} \hat{X}_{t+1|t} \\ C\hat{X}_{t+1|t} \end{pmatrix}$ and covariance $\begin{pmatrix} P_{t+1|t} & P_{t+1|t}C^T \\ CP_{t+1|t} & CP_{t+1|t}C^T + R \end{pmatrix}$. Also the Kalman Filter (Time updates (3) & (4) and measurement updates (5) & (6)) are given by:

$$\hat{x}_{t+1|t+1} = \hat{x}_{t+1|t} + \underbrace{P_{t+1|t}C^T(CP_{t+1|t}C^T + R)^{-1}}_{\text{Kalman gain Matrix } K_{t+1}}(y_{t+1} - C\hat{x}_{t+1|t}) \tag{3}$$

$$P_{t+1|t+1} = P_{t+1|t} - P_{t+1|t}C^T(CP_{t+1|t}C^T + R)^{-1}CP_{t+1|t} \tag{4}$$

$$\hat{x}_{t+1|t} = A\hat{x}_{t|t} \tag{5}$$

$$P_{t+1|t} = AP_{t|t}A^T + GQG^T \tag{6}$$

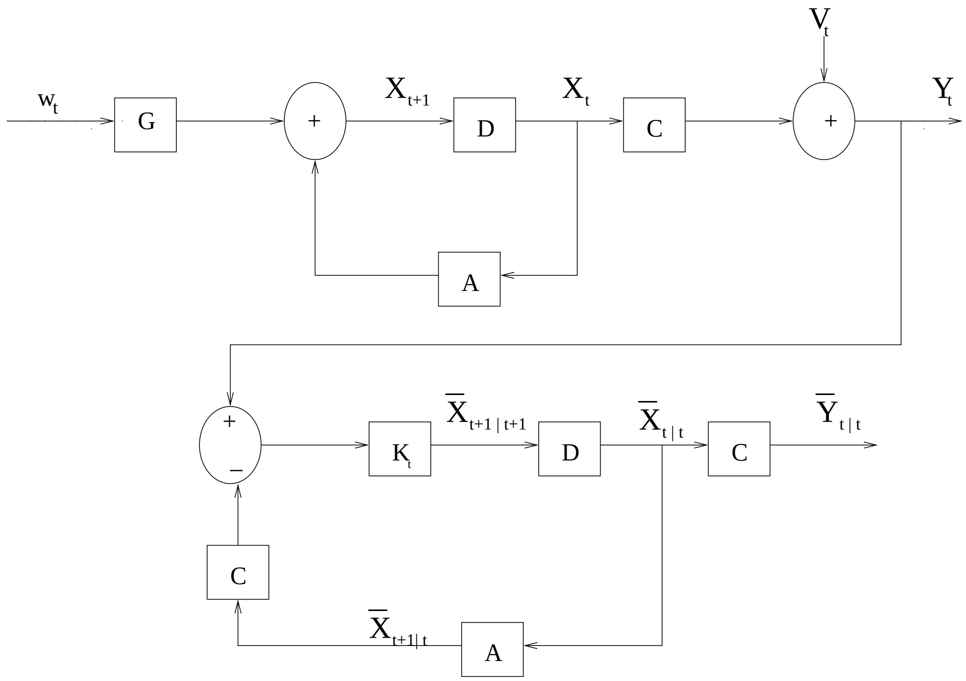

Figure 1: Control diagram for state space models

2 Relationship to LMS

In the above section we have defined the Kalman gain matrix K_{t+1}. Using the matrix inversion formulas we can write the matrix as:

$$\begin{aligned}
K_{t+1} &= P_{t+1|t}C^T(CP_{t+1|t}C^T + R)^{-1} \\
&= (P_{t+1|t}^{-1} + C^TRC)^{-1}C^TR^{-1} \\
&= (P_{t+1|t} + P_{t+1|t}C^T(CP_{t+1|t}C^T + R)^{-1}CP_{t+1|t})C^TR^{-1} \\
&= P_{t+1|t+1}C^TR^{-1}
\end{aligned}$$

Using this the time update equation can be written as:

$$\hat{x}_{t+1|t+1} = A\hat{x}_{t|t} + P_{t+1|t+1}C^TR^{-1}(y_{t+1} - CA\hat{x}_{t|t})$$

To make a connection with the LMS algorithm, if we let $A = I, G = 0$, then the dynamical equation $x_{t+1} = x_t + Gw_t$, reduces to the statement that the "state" is a constant. Let θ denotes this constant. Now if we replace the matrix C by the time varying vector x_t^T, then the state space model becomes

$$y_t = x_t^T\theta + v_t$$

In this case the Kalman filtering equation becomes:

$$\hat{\theta}_{t+1} = \hat{\theta}_t + P_{t+1}R^{-1}(y_{t+1} - x_t^T\hat{\theta}_t)x_t \tag{7}$$

Equation (7) combined with the update for P_{t+1} is referred to as the *recursive least squares (RLS) algorithm.* Comparing it with LMS:

$$\hat{\theta}_{t+1} = \hat{\theta}_t + \mu(y_{t+1} - x_t^T\hat{\theta}_t)x_t \tag{8}$$

Noting that $P_{t+1}R^{-1}$ is like μ in LMS. Thus LMS is like an approximation to Kalman Filter.

3 Raunch-Tung-Striebel (RTS) Smoothing

The goal in smoothing is to compute $p(x_t|y_0, \cdots, y_T)$. Define $\hat{x}_{t|T} = E(x_t|y_0, \cdots, y_T)$

There are a similar set of update equations as above, including all of the observations to time T.

Equations:

$$\begin{cases} \hat{x}_{t|T} = \hat{x}_{t|t} + L_t(\hat{x}_{t+1|T} - \hat{x}_{t+1|t}) \\ P_{t|T} = P_{t|t} + L_t(P_{t+1|T} - P_{t+1|t})L_t^T \\ L_t = P_{t|t}A^T P_{t+1|t}^{-1} \end{cases}$$