CS281A/Stat241A: Statistical Learning Theory

Lecture 22: Kalman Filtering and Extensions

Lecturer: Michael I. Jordan *Scribes: Heer and Pederson*

1 Kalman Filtering

As in the HMM model, Y_t is the observed state of the system at time t and is a Gaussian random variable. The latent random variable, X_t is also a Gaussian random variable. Also like the HMM, there is a dependency between X_t and X_{t+1}. The final state of the system is indexed by T.

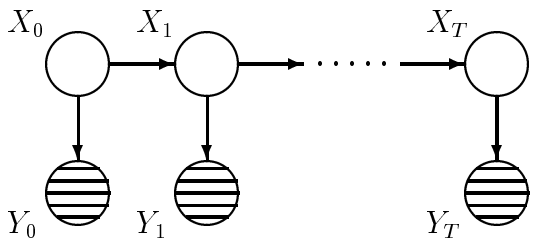


Figure 1: Graphical Model for Kalman Filtering

1.1 Goals

1. Filtering
   In the Filtering problem, we know the observed states up to the current time, t, and would like to characterize the latent state of the system at time t. For example, we would like to compute

   $$P(x_t \mid y_0, y_1, \ldots, y_t)$$

   This allows us to sequentially update the estimates for parameters as more data is obtained.

2. Smoothing
   In the Smoothing problem, we have observations for $1, 2, \ldots T$ and would like to characterize X_t for any $t \leq T$. Using probability notation, we would like to compute

   $$P(x_t \mid y_0, y_1, \ldots, y_T)$$

3. Likelihood

   $$P(y_0, y_1, \ldots, y_T)$$

1.2 Model

All the random variables in this model are Gaussian. First we introduce an error term, $W_t \sim N(0, Q)$, that is independent of X and Y. The latent random variable, X_t is influenced by this error term and the previous random variable as follows:

$$X_{t+1} = AX_t + GW_t$$

After defining the covariance of X_t to be P_t, we know the distribution of X_{t+1}. (Recall that the sum of Gaussian random variables is Gaussian)

$$X_{t+1} \sim N\left(AE[X_t], AP_tA^T + GQG^T\right)$$

Notice that if $X_0 \sim N(0, P_0)$, then the mean of all the X_t's will be zero.

The expression for the covariance of X_t is also called the *Lyapunov Equation*.

Now introduce another error term, $\eta \sim N(0, R)$ that is independent of X_t. The observed random variable, Y_t depends on this error term and X_t as follows:

$$Y_t = CX_t + \eta_t$$

1.3 Notation

$$\begin{aligned} \hat{X}_{t|t} &\triangleq E\left[X_t \,|\, Y_0, Y_1, \ldots, Y_t\right] \\ P_{t|t} &\triangleq E\left[(X_t - \hat{X}_{t|t})(X_t - \hat{X}_{t|t})^T \,|\, Y_0, Y_1, \ldots, Y_t\right] \end{aligned}$$

In these, the first subscript t refers to the index of X, and the second to the amount of observations that are being conditioned on.

Together these specify $P(x_t \,|\, y_0, y_1, \ldots, y_t)$, because we know that X_t is a Gaussian and the mean and covariance are sufficient to describe this probability distribution.

1.4 Implementation

The implementation of the Kalman Filter has two sections. First, given observations up to time t, we would like to find the probability distribution for the next state of the latent variable, X_{t+1}. This is called the *Time Update*. Then we would like to update our measure of the probability of X_{t+1} after getting the next observation, y_{t+1}.

It is clear that this process can be repeated for all t, and we will see that at every t, we can solve the Filtering problem. Also, because both X and Y are Gaussian, in order to characterize the probability distribution, it is sufficient to find the mean and covariance.

1.4.1 Time Update

In this step, we would like to update $P(x_t \,|\, y_0, y_1, \ldots, y_t)$ and find $P(x_{t+1} \,|\, y_0, y_1, \ldots, y_t)$. Recall

$$X_{t+1} = AX_t + GW_t$$

Take the conditional expectation of both sides, conditioning on $y_0, y_1, \ldots, y_t$

$$\begin{aligned} E\left[X_{t+1} \mid y_0, \ldots, y_t\right] &= \hat{X}_{t+1|t} \\ &= A\hat{X}_{t|t} + GE\left[W_t \mid y_0, \ldots, y_t\right] \\ &= A\hat{X}_{t|t} + G\,0 \\ &= A\hat{X}_{t|t} \end{aligned}$$

Now for the covariance term:

$$\begin{aligned} P_{t+1|t} &= E\left[(X_{t+1} - \hat{X}_{t+1|t})(X_{t+1} - \hat{X}_{t+1|t})^T \mid y_0, y_1, \ldots, y_t\right] \\ &= E\left[(AX_t - GW_t - A\hat{X}_{t|t})(AX_t - GW_t - A\hat{X}_{t|t})^T \mid y_0, y_1, \ldots, y_t\right] \\ &= E\left[(A(X_t - \hat{X}_{t|t}) - GW_t)(A(X_t - \hat{X}_{t|t}) - GW_t)^T \mid y_0, y_1, \ldots, y_t\right] \\ &= E\Big[(A(X_t - \hat{X}_{t|t})(X_t - A\hat{X}_{t|t}))^T A^T - GW_t(X_t - \hat{X}_{t|t})^T A^T \\ &\qquad -A(X_t - \hat{X}_{t|t})W_t^T G^T + GW_tW_t^T G^T \mid y_0, y_1, \ldots, y_t\Big] \\ &= E\left[A(X_t - \hat{X}_{t|t})(X_t - \hat{X}_{t|t})^T A^T + GW_tW_t^T G^T \mid y_0, y_1, \ldots, y_t\right] \\ &= AP_{t|t}A^T + GQG^T \end{aligned}$$

The update equations for the mean and the covariance in this step are known as the Riccati equations and need no additional data.

1.4.2 Measurement Update

Now we would like to update $P(x_{t+1} \mid y_0, \ldots, y_t)$ and find $P(x_{t+1} \mid y_0, \ldots, y_{t+1})$. First, we will find mean and covariance of Y_{t+1} conditional on $Y_0, \ldots, Y_t$, and then use this to find the distribution of X_{t+1} conditional on $Y_0, \ldots, Y_{t+1}$.

Using the same notation as before, we introduce $\hat{Y}_{t|t}$.

$$\begin{aligned} \hat{Y}_{t+1|t} &\triangleq E\left[Y_{t+1} \mid y_0, y_1, \ldots, y_t\right] \\ &= E\left[X_{t+1} + \eta \mid y_0, y_1, \ldots, y_t\right] \\ &= C\hat{X}_{t+1|t} \end{aligned}$$

$$\begin{aligned} E\left[(Y_{t+1} - \hat{Y}_{t+1|t})(Y_{t+1} - \hat{Y}_{t+1|t})^T \mid y_0, \ldots, y_t\right] &= E\left[(CX_{t+1} + \eta - C\hat{X}_{t+1|t})(CX_{t+1} + \eta - C\hat{X}_{t+1|t})^T \mid y_0, \ldots, y_t\right] \\ &= CP_{t+1|t}C^T + R \end{aligned}$$

$$\begin{aligned} E\left[(Y_{t+1} - \hat{Y}_{t+1|t})(X_{t+1} - \hat{X}_{t+1|t})^T \mid y_0, \ldots, y_t\right] &= E\left[(C(X_{t+1} - \hat{X}_{t+1|t}) + \eta)(X_{t+1} - \hat{X}_{t+1|t})^T \mid y_0, \ldots, y_t\right] \\ &= CP_{t+1|t} \end{aligned}$$

Now we have the mean and covariance of X_{t+1} and Y_{t+1} conditional on $Y_0, \ldots, Y_t$ and would like to find the mean and covariance of X_{t+1} conditional on $Y_0, \ldots, Y_{t+1}$

Recall the following relationship between jointly distributed Gaussians and the conditional distribution:

If $\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}\right)$, then $X_{1|2} \sim N\left(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right)$

$$\text{mean} = \begin{pmatrix} \hat{X}_{t+1|t} \\ \hat{Y}_{t+1|t} \end{pmatrix} \quad \text{Covariance} = \begin{pmatrix} P_{t+1|t} & P_{t+1|t}\,C^T \\ C\,P_{t+1|t} & C\,P_{t+1|t}\,C^T + R \end{pmatrix}$$

$$\hat{X}_{t+1|t+1} = \hat{X}_{t+1|t} + P_{t+1|t}\,C^T(C\,P_{t+1|t}\,C^T + R)^{-1}\,(y_{t+1} - \hat{y}_{t+1})$$

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

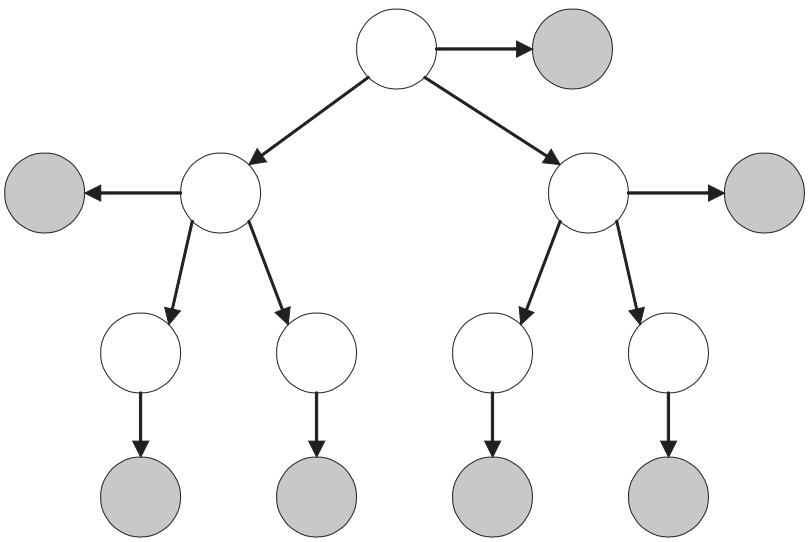
The correction term of the $\hat{X}_{t+1|t}$ equation above, $P_{t+1|t}\,C^T(C\,P_{t+1|t}\,C^T + R)^{-1}$, is called the *Kalman Gain Matrix*. If this is replaced by a scalar ρ, this is remniscent of LMS.

To summarize, the Kalman Filter has two steps, the Time Update and the Measurement Update and the equations to perform these are:

$$\begin{aligned}
\hat{X}_{t+1|t} &= A\hat{X}_{t|t} \\
P_{t+1|t} &= A\,P_{t|t}\,A^T + GQG^T \\
\hat{X}_{t+1|t+1} &= \hat{X}_{t+1|t} + P_{t+1|t}\,C^T(C\,P_{t+1|t}\,C^T + R)^{-1}\,(y_{t+1} - \hat{y}_{t+1}) \\
P_{t+1|t+1} &= P_{t+1|t} - P_{t+1|t}\,C^T(C\,P_{t+1|t}\,C^T + R)^{-1}C\,P_{t+1|t}
\end{aligned}$$

2 Other Graphical Models

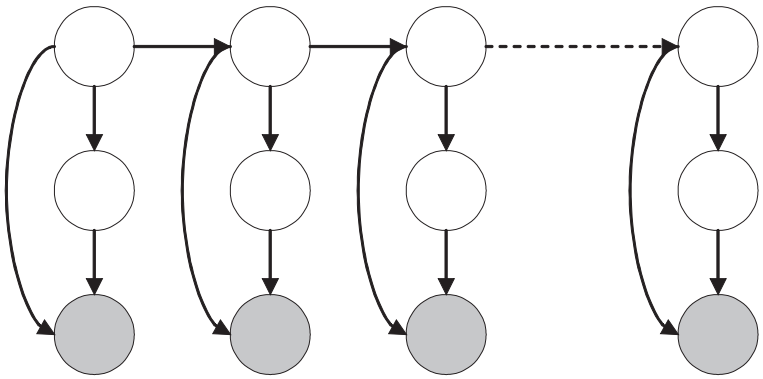
2.1 Multi-Resolution Models



(Alan Willsky) compute $E[x_t|y]$, EM. Solve via Sum-Product algorithm.

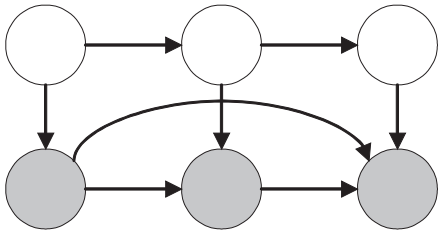
(Robert Nowak) relates Sum-Product to wavelets

2.2 HMM with Gaussian Mixture observables

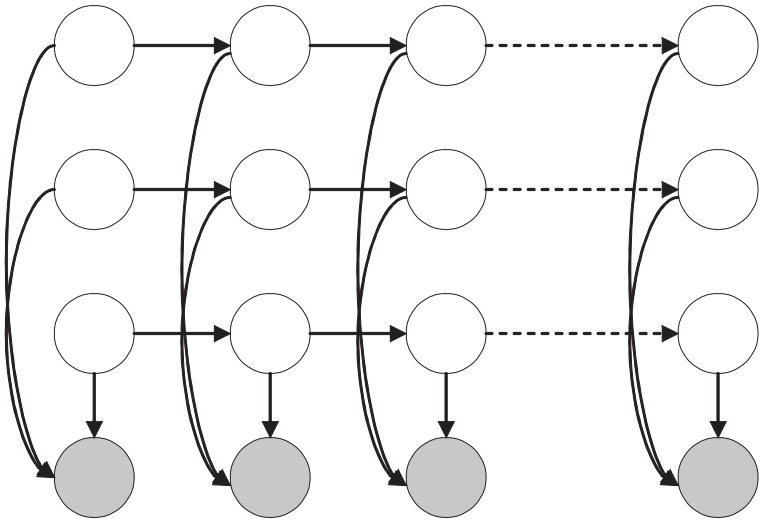


Widely used in speech recognition.

2.3 Auto-Regressive HMMs

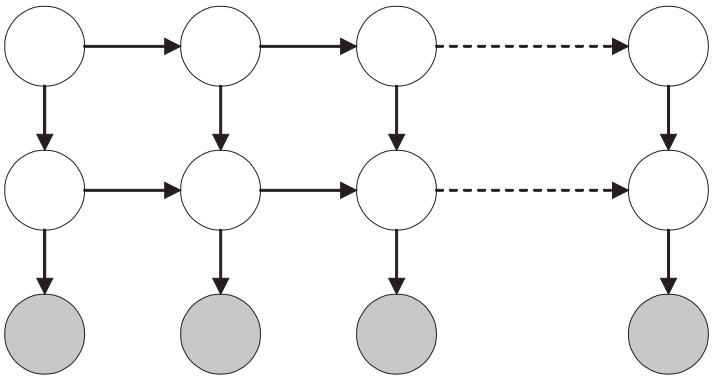


2.4 Factorial HMM



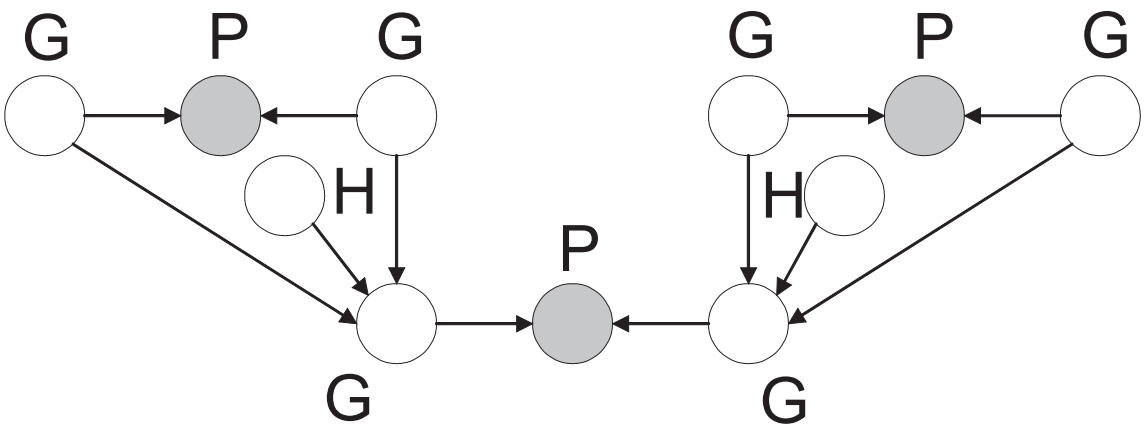
Inference: Is it feasible to collapse the hidden Markov chains into supernodes? NO. If we have 5 chains with 10 states/node, we have 10^5 states in a supernode $\implies A = 10^5$ x 10^5 matrix. By using the factorial approach we have 10 x 10 x 5 = 500 parameters.

2.5 Switching State-Space Model



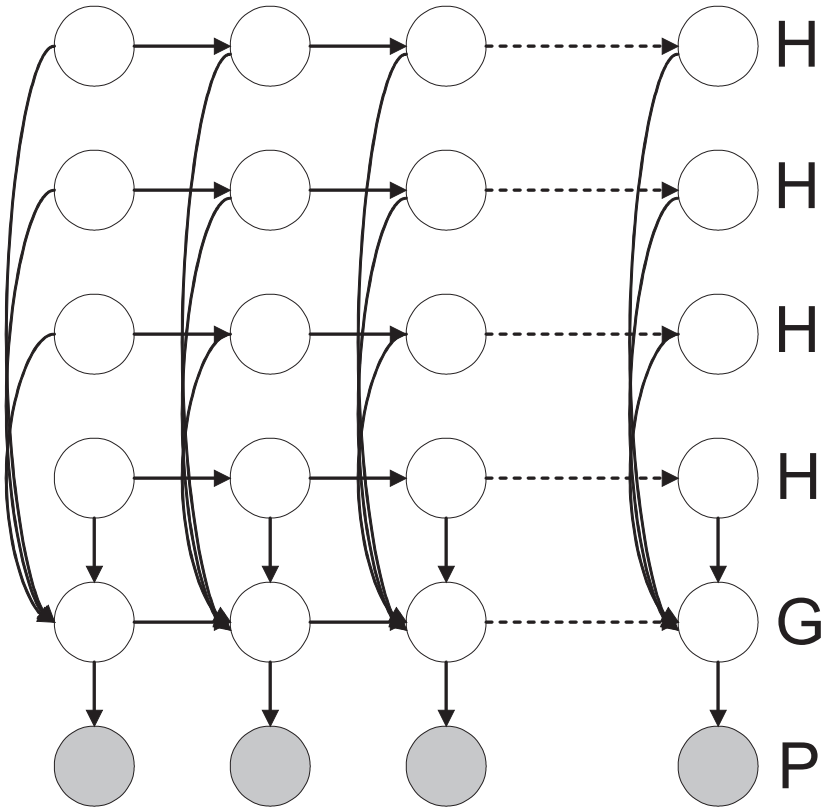
Top chain is made of discrete nodes, middle chain of continuous nodes.

2.6 Pedigrees



G - genotype, P - phenotype, H - haploid

Running elimination on this graph is also known as peeling. To deal with the multi-locus case, we unwind the model into a factorial HMM.


H
H
H
H
G
P