EM formula for extended Kalman Filter

Ruiqi Zhong

June 21, 2017

Paramters:
$\hat{z_0}$ initial state mean
P_0 initial state covariance
A transition matrix
B control matrix
Q transition covariance matrix
R observation noise matrix

Variables:
S There are in total S sequence generated.
T_s there are in total $(T_s + 1)$ steps/measurements of the s^{th} sequence.
$z_{s,t}$ the hidden state of s^{th} sequence at time t.
$x_{s,t}$ the observation of s^{th} sequence at time t.
$\epsilon_{s,t}$ the transition noise of s^{th} sequence at time t.
$w_{s,t}$ the observation noise of s^{th} sequence at time t.
$u_{s,t}$ the control of s^{th} sequence at time t.

Model Generation:
for s in range(S):
    $z_{s,0} \sim \mathcal{N}(\hat{z_0}, P_0)$
    for t in range(T_s):
        $\epsilon_{s,t} \sim \mathcal{N}(0, Q)$
        $w_{s,t} \sim \mathcal{N}(0, P)$
        $z_{s,t+1} = Az_{s,t} + Bu_{s,t} + \epsilon_{s,t}$
        $x_{s,t} = z_{s,t} + w_{s,t}$

Expected Log Likelihood:
$\mathbb{E}_Z[L(X|A, B, Q, R, \hat{z_0}, P_0)] = -\frac{1}{2}[\sum_{s=0}^{S-1} \log|P_0| + Tr(P_0^{-1} \sum_{s=0}^{S-1} (z_{s,0} - \hat{z_0})(z_{s,0} - \hat{z_0})^T)$
$+ \sum_{s=0}^{S-1} T_s \log|Q| + Tr(Q^{-1}(\sum_{s=0}^{S-1} \sum_{t=0}^{T_s-1} (z_{s,t+1} - Az_{s,t} - Bu_{s,t})(z_{s,t+1} - Az_{s,t} - Bu_{s,t})^T))$
$+ \sum_{s=0}^{S-1} (T_s + 1) \log|R| + Tr(R^{-1}(\sum_{s=0}^{S-1} \sum_{t=0}^{T_s} (x_{s,t} - z_{s,t})(x_{s,t} - z_{s,t})^T)] + \text{const}$

Estimation/Inference:
Define $o_{s,0} = 0$
$o_{s,t+1} = Ao_{s,t} + Bu_{s,t}$

By mathematical induction, we can easily get $z_{s,t+1} - o_{s,t+1} = A(z_{s,t} - o_{s,t}) + \epsilon_{s,t}$
Also, we have $x_{s,t} - o_{s,t} = z_{s,t} - o_{s,t} + w_{s,t}$
Let $x'_{s,t} = x_{s,t} - o_{s,t}$, $z'_{s,t} = z_{s,t} - o_{s,t}$, and since each sequence is independent, we can estimate each sequence seperately using the standard Kalmn Filter techniques, since $z'_{s,t+1} \sim Az'_{s,t} + \epsilon_{s,t}$, $x'_{s,t} = z'_{s,t} + w_{s,t}$

Maximization:
Here we ignore the multiplicative constant
$\frac{\partial L}{\partial \hat{z_0}} = \sum_{s=0}^{S-1}(\hat{z_0} - z_{s,0})(\hat{z_0} - z_{s,0})^T(P_0^{-1} + P_0^{-1}) = 0$
$\Rightarrow \hat{z_0} = \frac{1}{S}\sum_{s=0}^{S-1} z_{s,0}$
$\frac{\partial L}{\partial P_0} = \sum_{s=0}^{S-1}(P_0^{-1} - P_0^{-1}(z_{s,0} - \hat{z_0})(z_{s,0} - \hat{z_0})^T P_0^{-1}) = 0$
$\Rightarrow P_0 = \frac{1}{S}\sum_{s=0}^{S-1}(z_{s,0} - \hat{z_0})(z_{s,0} - \hat{z_0})^T$
$\frac{\partial L}{\partial [A \quad B]} = \sum_{s=0}^{S-1}\sum_{t=0}^{T_s-1} \begin{bmatrix} z_{s,t} \\ u_{s,t} \end{bmatrix} \begin{bmatrix} z_{s,t} \\ u_{s,t} \end{bmatrix}^T [A \quad B]^T Q^{-1} - \begin{bmatrix} z_{s,t} \\ u_{s,t} \end{bmatrix} z_{s,t+1}^T Q^{-1}$
$\Rightarrow [A \quad B] = (\sum_{s=0}^{S-1}\sum_{t=0}^{T_s-1} \begin{bmatrix} z_{s,t} \\ u_{s,t} \end{bmatrix} z_{s,t+1}^T)^T (\sum_{s=0}^{S-1}\sum_{t=0}^{T_s-1} \begin{bmatrix} z_{s,t} \\ u_{s,t} \end{bmatrix} \begin{bmatrix} z_{s,t} \\ u_{s,t} \end{bmatrix}^T)^{-1}$
$\frac{\partial L}{\partial Q} = \sum_{s=0}^{S-1} T_s Q^{-1} - Q^{-1}(\sum_{s=0}^{S-1}\sum_{t=0}^{T_s-1}(z_{s,t+1} - Az_{s,t} - Bu_{s,t})(z_{s,t+1} - Az_{s,t} - Bu_{s,t})^T)Q^{-1}$
$\Rightarrow Q = \frac{\sum_{s=0}^{S-1}\sum_{t=0}^{T_s}(z_{s,t+1} - Az_{s,t} - Bu_{s,t})(z_{s,t+1} - Az_{s,t} - Bu_{s,t})^T)}{\sum_{s=0}^{S-1} T_s}$
$\frac{\partial L}{\partial R} = \sum_{s=0}^{S-1}(T_s + 1)R^{-1} - R^{-1}[\sum_{s=0}^{S-1}\sum_{t=0}^{T_s}(x_{s,t} - z_{s,t})(x_{s,t} - z_{s,t})^T]R^{-1}$
$\Rightarrow R = \frac{\sum_{s=0}^{S-1}\sum_{t=0}^{T_s}(x_{s,t} - z_{s,t})(x_{s,t} - z_{s,t})^T}{\sum_{s=0}^{S-1}(T_s + 1)}$

We observe that to calculate these values, we only need $\mathbb{E}[z_{s,t}|X], \mathbb{E}[z_{s,t}z_{s,t+1}^T|X], \mathbb{E}[z_{s,t}z_{s,t}^T|X]$, which can be easily computed during the estimation step.