Learning an Outlier-Robust Kalman Filter: A Summary

Jo-Anne Ting

joanneti@usc.edu

January 8, 2008

1 The Kalman filter

Given observed data $\mathbf{z}_{1:N}$ and hidden states $\boldsymbol{\theta}_{1:N}$ over N time steps, the Kalman filter is a linear-Guassian state-space model and can be written as follows:

$$\begin{aligned} \mathbf{z}_k &= \mathbf{C}\theta_k + \mathbf{v}_k \\ \theta_k &= \mathbf{A}\theta_{k-1} + \mathbf{r}_k \end{aligned} \tag{1}$$

where $\mathbf{A} \in \Re^{d_2 \times d_2}$ is the state transition matrix, $\mathbf{C} \in \Re^{d_1 \times d_2}$ is the output matrix, and $\mathbf{v}_k \in \Re^{d_1 \times 1}$ and $\mathbf{r}_k \in \Re^{d_2 \times 1}$ are the observation and state noise vectors, respectively, at timestep k. Note that:

$$\begin{aligned} \mathbf{v}_k &\sim N(0, \mathbf{R}) \\ \mathbf{r}_k &\sim N(0, \mathbf{Q}) \end{aligned}$$

where $\mathbf{R}$ and $\mathbf{Q}$ are diagonal covariances for the observation noise and state noise. The Kalman filter propagation and update equations are, for $k = 1, ..., N$:

Propagation:

$$\boldsymbol{\theta}'_k = \mathbf{A}\langle\boldsymbol{\theta}_{k-1}\rangle \tag{2}$$

$$\boldsymbol{\Sigma}'_k = \mathbf{A}\boldsymbol{\Sigma}_{k-1}\mathbf{A}^T + \mathbf{Q} \tag{3}$$

Update:

$$\mathbf{S}'_k = \left(\mathbf{C}\boldsymbol{\Sigma}'_k\mathbf{C}^T + \mathbf{R}\right)^{-1} \tag{4}$$

$$K'_k = \boldsymbol{\Sigma}'_k\mathbf{C}^T\mathbf{S}'_k \tag{5}$$

$$\langle\boldsymbol{\theta}_k\rangle = \boldsymbol{\theta}'_k + K'_k\left(\mathbf{z}_k - \mathbf{C}\boldsymbol{\theta}'_k\right) \tag{6}$$

$$\boldsymbol{\Sigma}_k = \left(\mathbf{I} - K'_k\mathbf{C}\right)\boldsymbol{\Sigma}'_k \tag{7}$$

where $\langle\boldsymbol{\theta}_k\rangle$ is the posterior mean vector of the state $\boldsymbol{\theta}_k$, $\boldsymbol{\Sigma}_k$ is the posterior covriance of $\boldsymbol{\theta}_k$, and $\mathbf{S}'_k$ is the covariance of the residual prediction error—all at time step k.

We can re-write the equations above by substituting the update steps, Eqs. (4)-(7), into the propagation step, Eqs. (2)-(3), to get an expression for the posterior covariance and posterior mean of the state, $\boldsymbol{\Sigma}_k$ and $\boldsymbol{\theta}_k$:

$$\boxed{\begin{aligned}\boldsymbol{\Sigma}_k &= \left(\left(\mathbf{A}\boldsymbol{\Sigma}_{k-1}\mathbf{A}^T + \mathbf{Q}\right)^{-1} + \mathbf{C}^T\mathbf{R}^{-1}\mathbf{C}\right)^{-1} \\ \langle\boldsymbol{\theta}_k\rangle &= \boldsymbol{\Sigma}_k\left(\mathbf{A}\boldsymbol{\Sigma}_{k-1}\mathbf{A}^T + \mathbf{Q}\right)^{-1}\mathbf{A}\langle\boldsymbol{\theta}_{k-1}\rangle + \boldsymbol{\Sigma}_k\mathbf{C}^T\mathbf{R}^{-1}\mathbf{z}_k\end{aligned}} \tag{8}$$

2 A Weighted Outlier-Robust Kalman filter

We can take the model in Eq. (1) and introduce a scalar weight w_k for each observed data sample $\mathbf{z}_k$, such that the variance of $\mathbf{z}_k$ is weighted with w_k. The conditional prior distributions for the system are then:

$$\begin{aligned}\mathbf{z}_k|\theta_k, w_k &\sim N\left(\mathbf{C}\theta_k, \frac{1}{w_k}\mathbf{R}\right) \\ \theta_k|\theta_{k-1} &\sim N\left(\mathbf{A}\theta_{k-1}, \mathbf{Q}\right) \\ w_k &\sim G\left(a_{w_k}, b_{w_k}\right)\end{aligned} \tag{9}$$

where the weights are modeled to be Gamma distributed random variables to ensure positive values. The joint probability for the hidden and observed variables is:

$$p(\theta_{1:N}, \mathbf{z}_{1:N}) = \prod_{k=1}^{N}\left[p(\theta_k|\theta_{k-1})p(\mathbf{z}_k|\theta_k, w_k)p(w_k)\right]p(\theta_0) \tag{10}$$

We then want to find the values of $\boldsymbol{\theta}_{1:N}$, $\mathbf{w}$, $\mathbf{A}$, $\mathbf{C}$, $\mathbf{Q}$ and $\mathbf{R}$ to maximize the log complete evidence:

$$\begin{aligned}\log p(\boldsymbol{\theta}_{1:N}, \mathbf{z}_{1:N}) &= \sum_{k=1}^{N}\log p(\boldsymbol{\theta}_k|\boldsymbol{\theta}_{k-1}) + \sum_{k=1}^{N}\log p(\mathbf{z}_k|\boldsymbol{\theta}_k, w_k) + \sum_{k=1}^{N}\log p(w_k) + \log p(\boldsymbol{\theta}_0) \\ &= \frac{1}{2}\sum_{k=1}^{N}\log w_k - \frac{N}{2}\log\mathbf{R} - \frac{1}{2}\sum_{k=1}^{N}w_k\left(\mathbf{z}_k - \mathbf{C}\boldsymbol{\theta}_k\right)^T\mathbf{R}^{-1}\left(\mathbf{z}_k - \mathbf{C}\boldsymbol{\theta}_k\right) \\ &\quad - \frac{N}{2}\log\mathbf{Q} - \frac{1}{2}\sum_{k=1}^{N}\left(\boldsymbol{\theta}_k - \mathbf{A}\boldsymbol{\theta}_{k-1}\right)^T\mathbf{Q}^{-1}\left(\boldsymbol{\theta}_k - \mathbf{A}\boldsymbol{\theta}_{k-1}\right) \\ &\quad - \frac{1}{2}\log\mathbf{Q}_0 - \frac{1}{2}\left(\boldsymbol{\theta}_0 - \mathbf{A}\hat{\boldsymbol{\theta}}_0\right)^T\mathbf{Q}_0^{-1}\left(\boldsymbol{\theta}_0 - \mathbf{A}\hat{\boldsymbol{\theta}}_0\right) \\ &\quad + \sum_{k=1}^{N}\left(a_{w_k,0} - 1\right)\log w_k - \sum_{k=1}^{N}b_{w_k,0}w_k + \text{const}\end{aligned} \tag{11}$$

However, since we are assuming this model to be an online one, where at time step k, we only have access to data points $\mathbf{z}_{1:k}$, then we only have access to the log evidence of all the data points observed to date. That is to say:

$$\begin{aligned}
\log p(\boldsymbol{\theta}_{1:K}, \mathbf{z}_{1:K}) &= \sum_{k=1}^{K} \log p(\boldsymbol{\theta}_k|\boldsymbol{\theta}_{k-1}) + \sum_{k=1}^{K} \log p(\mathbf{z}_k|\boldsymbol{\theta}_k, w_k) + \sum_{k=1}^{K} \log p(w_k) + \log p(\boldsymbol{\theta}_0) \\
&= \frac{1}{2}\sum_{k=1}^{K} \log w_k - \frac{N}{2}\log \mathbf{R} - \frac{1}{2}\sum_{k=1}^{K} w_k \left(\mathbf{z}_k - \mathbf{C}\boldsymbol{\theta}_k\right)^T \mathbf{R}^{-1} \left(\mathbf{z}_k - \mathbf{C}\boldsymbol{\theta}_k\right) \\
&\quad - \frac{N}{2}\log \mathbf{Q} - \frac{1}{2}\sum_{k=1}^{K} \left(\boldsymbol{\theta}_k - \mathbf{A}\boldsymbol{\theta}_{k-1}\right)^T \mathbf{Q}^{-1} \left(\boldsymbol{\theta}_k - \mathbf{A}\boldsymbol{\theta}_{k-1}\right) \\
&\quad - \frac{1}{2}\log \mathbf{Q}_0 - \frac{1}{2}\left(\boldsymbol{\theta}_0 - \hat{\boldsymbol{\theta}}_0\right)^T \mathbf{Q}_0^{-1} \left(\boldsymbol{\theta}_0 - \hat{\boldsymbol{\theta}}_0\right) \\
&\quad + \sum_{k=1}^{N} (a_{w_k,0} - 1)\log w_k - \sum_{k=1}^{N} b_{w_k,0} w_k + \text{const}
\end{aligned} \tag{12}$$

We will make the factorial variational approximation over the hidden and unknown variables so that:

$$Q(\mathbf{w}, \boldsymbol{\theta}) = \prod_{k=1}^{N} Q(w_k) \prod_{k=0}^{N} Q(\boldsymbol{\theta}_k). \tag{13}$$

Note that the above factorization of $\boldsymbol{\theta}$ only considers the influence on each w_k from within its Markov blanket, conserving the Markov property that Kalman filters, by definition, have.

We can then approximate the posteriors of $Q(\mathbf{w})$, $Q(\boldsymbol{\theta})$ and determine the point-estimated values of parameters $\mathbf{A}$, $\mathbf{C}$, $\mathbf{R}$ and $\mathbf{Q}$. The final EM update equations are listed below:

E-step:

$$\boldsymbol{\Sigma}_k = \left(\mathbf{Q}_k^{-1} + \langle w_k \rangle \mathbf{C}_k^T \mathbf{R}_k^{-1} \mathbf{C}_k\right)^{-1} \tag{14}$$

$$\langle \boldsymbol{\theta}_k \rangle = \boldsymbol{\Sigma}_k \left(\mathbf{Q}_k^{-1} \mathbf{A}_k \langle \boldsymbol{\theta}_{k-1} \rangle + \langle w_k \rangle \mathbf{C}_k^T \mathbf{R}_k^{-1} \mathbf{z}_k\right) \tag{15}$$

$$\langle w_k \rangle = \frac{a_{w_k,0} + \frac{1}{2}}{b_{w_k,0} + \left\langle \left(\mathbf{z}_k - \mathbf{C}_k\boldsymbol{\theta}_k\right)^T \mathbf{R}_k^{-1} \left(\mathbf{z}_k - \mathbf{C}_k\boldsymbol{\theta}_k\right) \right\rangle} \tag{16}$$

M-step:

$$\mathbf{C}_k = \left(\textstyle\sum_{i=1}^{k} \langle w_i \rangle \mathbf{z}_i \langle \boldsymbol{\theta}_i \rangle^T\right) \left(\textstyle\sum_{i=1}^{k} \langle w_i \rangle \left\langle \boldsymbol{\theta}_i \boldsymbol{\theta}_i^T \right\rangle\right)^{-1} \tag{17}$$

$$\mathbf{A}_k = \left(\textstyle\sum_{i=1}^{k} \langle \boldsymbol{\theta}_i \rangle \langle \boldsymbol{\theta}_{i-1} \rangle^T\right) \left(\textstyle\sum_{i=1}^{k} \left\langle \boldsymbol{\theta}_{i-1} \boldsymbol{\theta}_{i-1}^T \right\rangle\right)^{-1} \tag{18}$$

$$r_{km} = \tfrac{1}{k} \textstyle\sum_{i=1}^{k} \langle w_i \rangle \left\langle \left(\mathbf{z}_{im} - \mathbf{C}_k(m,:)\boldsymbol{\theta}_i\right)^2 \right\rangle \tag{19}$$

$$q_{kn} = \tfrac{1}{k} \textstyle\sum_{i=1}^{k} \left\langle \left(\boldsymbol{\theta}_{in} - \mathbf{A}_k(n,:)\boldsymbol{\theta}_{i-1}\right)^2 \right\rangle \tag{20}$$

where $m = 1, .., d_1$, $n = 1, .., d_2$; r_{km} is the mth coefficient of the vector $\mathbf{r}_k$; q_{kn} is the nth coefficient of the vector $\mathbf{q}_k$; $\mathbf{C}_k(m,:)$ is the mth row of the matrix $\mathbf{C}_k$; $\mathbf{A}_k(n,:)$ is the nth row of the matrix $\mathbf{A}_k$; and $a_{w_k,0}$ and $b_{w_k,0}$ are prior scale parameters for the weight w_k. (14) to (20) should be computed once for each time step k when the data sample $\mathbf{z}_k$ becomes available.

We can also re-write Eqs. (14)-(15) in the standard propagation and update form:

$$\begin{aligned}
&\textbf{Propagation:} \\
\boldsymbol{\theta}'_k &= \mathbf{A}\left\langle \boldsymbol{\theta}_{k-1} \right\rangle && (21) \\
\boldsymbol{\Sigma}'_k &= \mathbf{Q} && (22) \\
&\textbf{Update:} \\
\mathbf{S}'_k &= \left(\mathbf{C}\boldsymbol{\Sigma}'_k\mathbf{C}^T + \frac{1}{\langle w_k \rangle}\mathbf{R} \right)^{-1} && (23) \\
K'_k &= \boldsymbol{\Sigma}'_k\mathbf{C}^T\mathbf{S}'_k && (24) \\
\langle \boldsymbol{\theta}_k \rangle &= \boldsymbol{\theta}'_k + K'_k\left(\mathbf{z}_k - \mathbf{C}\boldsymbol{\theta}'_k \right) && (25) \\
\boldsymbol{\Sigma}_k &= \left(\mathbf{I} - K'_k\mathbf{C} \right)\mathbf{Q} && (26)
\end{aligned}$$

We can compare Eq. (8) to Eqs. (14)-(15) and note the differences. We can also re-write