Time Series Analysis

Lecture 7: State Space Model - Estimation

Tohid Ardeshiri

Linköping University
Division of Statistics and Machine Learning

October 2, 2019

Kalman filter

Kalman filter is an algorithm that uses time series data, containing statistical noise and unknown innovations, and produces estimates of latent (hidden) process that tend to be more accurate than those based on a single observations using a probabilistic framework.

$$\mathbf{z}_t = A\mathbf{z}_{t-1} + e_t,$$
$$\mathbf{x}_t = C\mathbf{z}_t + \nu_t,$$

Kalman filtering output is

$$f(\mathbf{z}_t|\mathbf{x}_{1:t}).$$

That is, it computes the the posterior density of $\mathbf{z}_t$ using the observations up to time t.

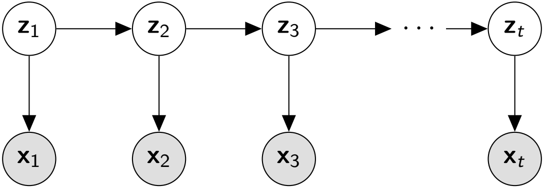
Kalman filtering recursion

1. initial estimate at $t = 1 \rightarrow N(\mathbf{z}_1; m_0, P_0)$
2. observation update using $\mathbf{x}_t$ and $\mathbf{x}_t = C\mathbf{z}_t + \nu_t \rightarrow N(\mathbf{z}_t; m_{t|t}, P_{t|t})$
3. prediction using $\mathbf{z}_{t+1} = A\mathbf{z}_t + e_{t+1} \rightarrow N(\mathbf{z}_t; m_{t+1|t}, P_{t+1|t})$
4. $t \leftarrow t + 1$
5. go to 2

State Space models - Time varying

State space models can be time-varying

$$\mathbf{z}_t = A_t \mathbf{z}_{t-1} + e_t, \qquad e_t \sim N(0, Q_t)$$
$$\mathbf{x}_t = C_t \mathbf{z}_t + \nu_t, \qquad \nu_t \sim N(0, R_t)$$



State space models with known deterministic input

State space model with input $\mathbf{u}$.

$$\mathbf{z}_t = A\mathbf{z}_{t-1} + B\mathbf{u}_{t-1} + e_t,$$
$$\mathbf{x}_t = C\mathbf{z}_t + \nu_t,$$

Initialization:

$$f(\mathbf{z}_1) = N(\mathbf{z}_1; m_{1|0}, P_{1|0})$$

Inputs: A, B, C, Q, R, $\mathbf{u}_{1:T}$, $\mathbf{x}_{1:T}$, $m_{1|0}$, $P_{1|0}$
for $t = 1$ to T **do**
Kalman filter observation update step
$\quad K_t \leftarrow P_{t|t-1}C^{\mathrm{T}}(CP_{t|t-1}C^{\mathrm{T}} + R)^{-1}$
$\quad m_{t|t} \leftarrow m_{t|t-1} + K_t(\mathbf{x}_t - Cm_{t|t-1})$
$\quad P_{t|t} \leftarrow P_{t|t-1} - K_tCP_{t|t-1}$
Kalman filter prediction step
$\quad m_{t+1|t} \leftarrow Am_{t|t} + B\mathbf{u}_t$
$\quad P_{t+1|t} \leftarrow AP_{t|t}A^{\mathrm{T}} + Q$
end for
Outputs: $m_{t|t}$ **and** $P_{t|t}$ **for** $t = 1 : T$

Kalman Smoothing

The purpose of Kalman smoothing is to compute the marginal posterior distribution of $\mathbf{z}_t$ at time t after receiving observations up to time T where $T > t$:

$$f(\mathbf{z}_t|\mathbf{x}_{1:T}) = N(\mathbf{z}_t; m_{t|T}, P_{t|T})$$

The RTS smoother uses a Kalman filter in its forward path. In its backwards path it updates the densities using the relation

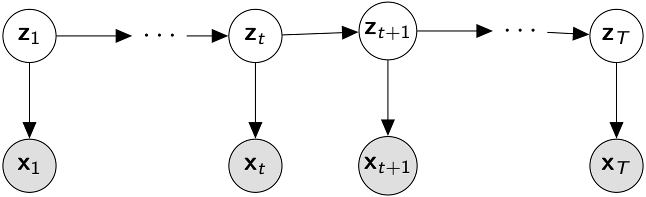
$$\mathbf{z}_t = A_{t-1}\mathbf{z}_{t-1} + e_t$$

RTS Smoother's derivation

Assume $f(\mathbf{z}_{t+1}|\mathbf{x}_{1:T})$ is available as in

$$f(\mathbf{z}_{t+1}|\mathbf{x}_{1:T}) = N(\mathbf{z}_{t+1}; m_{t+1|T}, P_{t+1|T})$$

For example $f(\mathbf{z}_T|\mathbf{x}_{1:T})$ which is the filtering density of $\mathbf{z}_T$ is available after filtering.



The objective is to compute $f(\mathbf{z}_t, \mathbf{z}_{t+1}|\mathbf{x}_{1:T})$.

RTS Smoother's derivation

The joint posterior $f(\mathbf{z}_t, \mathbf{z}_{t+1}|\mathbf{x}_{1:t})$ can be written as

$$\begin{aligned} f(\mathbf{z}_t, \mathbf{z}_{t+1}|\mathbf{x}_{1:t}) =& N(\mathbf{z}_t; m_{t|t}, P_{t|t}) N(\mathbf{z}_{t+1}; A\mathbf{z}_t, Q) \\ =& N\left(\begin{bmatrix} \mathbf{z}_t \\ \mathbf{z}_{t+1} \end{bmatrix}, \begin{bmatrix} m_{t|t} \\ Am_{t|t} \end{bmatrix}, \begin{bmatrix} P_{t|t} & P_{t|t}A^{\mathrm{T}} \\ AP_{t|t} & AP_{t|t}A^{\mathrm{T}} + Q \end{bmatrix} \right) \end{aligned}$$

Using the conditioning property of the multivariate normal distribution $f(\mathbf{z}_t|\mathbf{z}_{t+1}, \mathbf{x}_{1:t})$ can be computed as a normal density as given in the following:

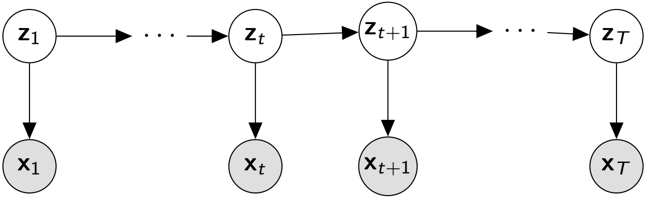
$$f(\mathbf{z}_t|\mathbf{z}_{t+1}, \mathbf{x}_{1:t}) = N(\mathbf{z}_t; \tilde{m}_t, \tilde{P}_t)$$

where $\tilde{m}_t$ is a function of $\mathbf{z}_{t+1}$.

RTS Smoother's derivation

Note the Markov property

$$f(\mathbf{z}_t|\mathbf{z}_{t+1},\mathbf{x}_{1:T}) = f(\mathbf{z}_t|\mathbf{z}_{t+1},\mathbf{x}_{1:t})$$



Assume $f(\mathbf{z}_{t+1}|\mathbf{x}_{1:T})$ is available as in

$$f(\mathbf{z}_{t+1}|\mathbf{x}_{1:T}) = N(\mathbf{z}_{t+1}; m_{t+1|T}, P_{t+1|T})$$

Recall that

$$\begin{aligned} f(\mathbf{z}_{t+1},\mathbf{z}_t|\mathbf{x}_{1:T}) &= f(\mathbf{z}_{t+1}|\mathbf{x}_{1:T}) f(\mathbf{z}_t|\mathbf{z}_{t+1},\mathbf{x}_{1:T}) \\ &= f(\mathbf{z}_{t+1}|\mathbf{x}_{1:T}) f(\mathbf{z}_t|\mathbf{z}_{t+1},\mathbf{x}_{1:t}) \\ &= N(\mathbf{z}_{t+1}; m_{t+1|T}, P_{t+1|T}) N(\mathbf{z}_t; \tilde{m}_t, \tilde{P}_t) \end{aligned}$$

RTS Smoother's derivation **Whiteboard**

where

$$
\begin{aligned}
G_t &= P_{t|t} A_t^{\mathrm{T}} (A P_{t|t} A^{\mathrm{T}} + Q)^{-1} = P_{t|t} A_t^{\mathrm{T}} P_{t+1|t}^{-1} \\
\tilde{m}_t &= m_{t|t} + G_t(\mathbf{z}_{t+1} - A m_{t|t}) \\
\tilde{P}_t &= P_{t|t} - G_t(A P_{t|t} A^{\mathrm{T}} + Q) G_t^{\mathrm{T}} = P_{t|t} - G_t P_{t+1|t} G_t^{\mathrm{T}}
\end{aligned}
$$

Hence,

$$
\begin{aligned}
& f(\mathbf{z}_{t+1}, \mathbf{z}_t | \mathbf{x}_{1:T}) = N(\mathbf{z}_{t+1}; m_{t+1|T}, P_{t+1|T}) N(\mathbf{z}_t; \tilde{m}_t, \tilde{P}_t) \\
& = N\left(\begin{bmatrix} \mathbf{z}_t \\ \mathbf{z}_{t+1} \end{bmatrix}, \begin{bmatrix} m_{t|t} + G_t(m_{t+1|T} - A m_{t|t}) \\ m_{t+1|T} \end{bmatrix}, \begin{bmatrix} G_t P_{t+1|T} G_t^{\mathrm{T}} + \tilde{P}_t & G_t P_{t+1|T} \\ P_{t+1|T} G_t^{\mathrm{T}} & P_{t+1|T} \end{bmatrix} \right)
\end{aligned}
$$

RTS Smoother's derivation **Whiteboard**

The smoothing density's parameters is given by

$$
\begin{aligned}
G_t &= P_{t|t}A_t^{\mathrm{T}}(AP_{t|t}A^{\mathrm{T}} + Q)^{-1} = P_{t|t}A_t^{\mathrm{T}}P_{t+1|t}^{-1} \\
m_{t|T} &= m_{t|t} + G_t(m_{t+1|T} - Am_{t|t}) \\
P_{t|T} &= \tilde{P}_t + G_tP_{t+1|T}G_t^{\mathrm{T}} = P_{t|t} - G_tP_{t+1|t}G_t^{\mathrm{T}} + G_tP_{t+1|T}G_t^{\mathrm{T}} \\
&= P_{t|t} + G_t(P_{t+1|T} - P_{t+1|t})G_t^{\mathrm{T}}
\end{aligned}
$$

RTS smoother's backwards recursion

Prove the backwards recursion of the RTS smoother for following state space model with initial prior on the state $f(\mathbf{z}_1) = N(\mathbf{z}_1; m_0, P_0)$

$$\begin{aligned} \mathbf{z}_t &= A_{t-1}\mathbf{z}_{t-1} + e_t, & e_t &\sim N(0, Q_t) \\ \mathbf{x}_t &= C_t\mathbf{z}_t + \nu_t, & \nu_t &\sim N(0, R_t) \end{aligned}$$

1: **Inputs:** A_t, Q_t, $m_{t|t}$, $P_{t|t}$, $m_{t+1|t}$, $P_{t+1|t}$ for $1 \leq t \leq T$
initialization
2: **for** $t =$ T-1 down to 1 **do**
3: $\quad G_t \leftarrow P_{t|t} A_t^{\mathrm{T}} P_{t+1|t}^{-1}$
4: $\quad m_{t|T} \leftarrow m_{t|t} + G_t(m_{t+1|T} - A_t m_{t|t})$
5: $\quad P_{t|T} \leftarrow P_{t|t} + G_t(P_{t+1|T} - P_{t+1|t}) G_t^{\mathrm{T}}$
6: **end for**
7: **Outputs:** $m_{t|T}$, $P_{t|T}$

State Space models - Estimation

We consider three approaches.

1. (Variational Bayes)
   T. Ardeshiri, E. Özkan, U. Orguner and F. Gustafsson, "**Approximate Bayesian Smoothing with Unknown Process and Measurement Noise Covariances**," in IEEE Signal Processing Letters, vol. 22, no. 12, pp. 2450-2454, Dec. 2015.
2. Direct maximum likelihood estimate
3. Expectation maximization (EM)

Variational Bayes smoothing with unknwn time varying R_t and Q_t

Consider a Linear and Gaussian state space model with parameters

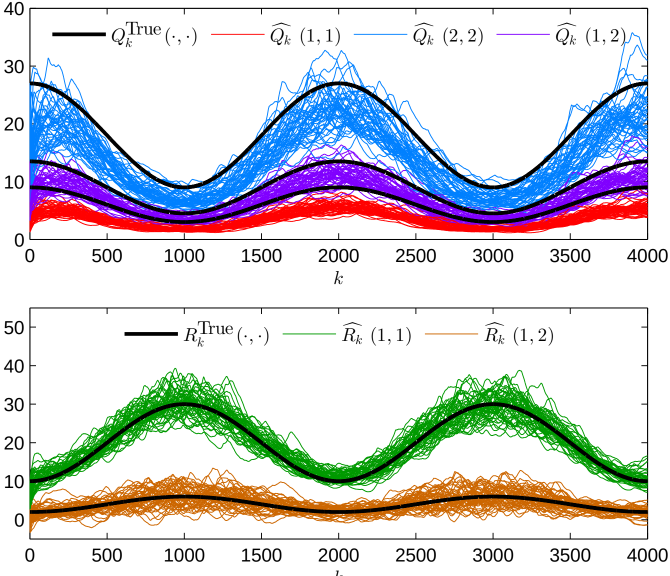
$$A_k = \mathrm{Diag}(a, a), \qquad Q_0 = \mathrm{Diag}(q, q),$$

$$a = \begin{bmatrix} 1 & \tau \\ 0 & 1 \end{bmatrix}, \qquad q = \sigma_\nu^2 \begin{bmatrix} \tau^3/3 & \tau^2/2 \\ \tau^2/2 & \tau \end{bmatrix},$$

$$R_k^{\mathrm{True}} = \left(2 - \cos\left(\frac{4\pi k}{K}\right)\right) R_0, \qquad R_0 = \sigma_e^2 \begin{bmatrix} 5 & 1 \\ 1 & 5 \end{bmatrix},$$

$$Q_k^{\mathrm{True}} = \left(\frac{2}{3} + \frac{1}{3}\cos\left(\frac{4\pi k}{K}\right)\right) Q_0, \qquad C_k = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

Variational Bayes smoothing with unknwn time varying R_t and Q_t



Maximum likelihood methods **Whiteboard**

Let $\theta = \{A, C, R, Q, m_0, P_0\}$ denote the unknown state space parameters

$$f(\mathbf{x}_{1:T}|\theta) = f(\mathbf{x}_1|\theta) f(\mathbf{x}_2|\mathbf{x}_1, \theta) f(\mathbf{x}_3|\mathbf{x}_{1:2}, \theta) \cdots f(\mathbf{x}_T|\mathbf{x}_{1:T-1}, \theta)$$

where

$$f(\mathbf{x}_{t+1}|\mathbf{x}_{1:t}, \theta) = \int f(\mathbf{x}_{t+1}|\mathbf{z}_{t+1}, \mathbf{x}_{1:t}, \theta) f(\mathbf{z}_{t+1}|\mathbf{x}_{1:t}, \theta)\, \mathrm{d}\mathbf{z}_{t+1}$$

This can be computed using the Kalman filter

$$\begin{aligned} f(\mathbf{x}_{t+1}|\mathbf{x}_{1:t}, \theta) &= \int f(\mathbf{x}_{t+1}|\mathbf{z}_{t+1}, \mathbf{x}_{1:t}, \theta) f(\mathbf{z}_{t+1}|\mathbf{x}_{1:t}, \theta)\, \mathrm{d}\mathbf{z}_{t+1} \\ &= \int N(\mathbf{x}_{t+1}; C\mathbf{z}_{t+1}, R) N(\mathbf{z}_{t+1}; m_{t+1|t}, P_{t+1|t})\, \mathrm{d}\mathbf{z}_{t+1} \\ &= N(\mathbf{x}_{t+1}; Cm_{t+1|t}, CP_{t+1|t}C^{\mathrm{T}} + R) \end{aligned}$$

Maximum likelihood methods **Whiteboard**

The negative logarithm of the likelihood becomes

$$
\begin{aligned}
l(\theta) &= -\sum_{t=1}^{T} \log f(\mathbf{x}_t | \mathbf{x}_{1:t-1}, \theta) \\
&= -\sum_{t=1}^{T} \log N(\mathbf{z}_{t+1}; Cm_{t+1|t}, CP_{t+1|t}C^{\mathrm{T}} + R) \\
&= \frac{1}{2}\sum_{t=1}^{T} \log |CP_{t+1|t}C^{\mathrm{T}} + R| \\
&+ \frac{1}{2}\sum_{t=1}^{T} (\mathbf{x}_t - Cm_{t+1|t})(CP_{t+1|t}C^{\mathrm{T}} + R)^{-1}(\mathbf{x}_t - Cm_{t+1|t})^{\mathrm{T}}
\end{aligned}
$$

which can be solved using for example Newton-Raphson method.

Maximum likelihood methods

The first two derivatives of the negative log-likelihood is computed with respect to the θ.

Then in the iterations of the Newton-Raphson method

1. An initial value for for θ is selected, say $\theta^{(0)}$.
2. A Kalman filter is run to compute the quantities for the first two derivatives of $l(\theta)$.
3. A new set of parameters are obtained from a Newton-Raphson procedure.
4. Iterations are repeated until convergence.

Expectation Maximization **Whiteboard**

- Expectation-maximization (EM) method can be used to compute the maximum likelihood (ML) estimate of the state space parameters.
- In the E (Expectation) step of the EM algorithm the conditional expectation of the joint log-likelihood is computed using the last estimates of the unknown parameters as in

$$\mathcal{Q} = E\left[\log f(\mathbf{z}_{1:T}, \mathbf{x}_{1:T}) \,\middle|\, \mathbf{x}_{1:T}, \theta^{(i)}\right] \tag{1}$$

where

$$\begin{aligned}\log f(\mathbf{z}_{1:T}, \mathbf{x}_{1:T}) = &\log N(\mathbf{z}_1; m_0, P_0) - \frac{T+1}{2}\log|R| \\ &- \frac{1}{2}\sum_{t=1}^{T}\operatorname{Tr}\left(R^{-1}(\mathbf{x}_t - C\mathbf{z}_t)(\mathbf{x}_t - C\mathbf{z}_t)^{\mathrm{T}}\right) - \frac{T}{2}\log|Q| \\ &- \frac{1}{2}\sum_{t=1}^{T-1}\operatorname{Tr}\left(Q^{-1}(\mathbf{z}_{t+1} - A\mathbf{z}_t)(\mathbf{z}_{t+1} - A\mathbf{z}_t)^{\mathrm{T}}\right) + c. \end{aligned} \tag{2}$$

Expectation Maximization **Whiteboard**

Therefore,

$$
\begin{aligned}
\mathcal{Q} = &-\frac{1}{2}E[(\mathbf{z}_0 - m_0)P_0^{-1}(\mathbf{z}_0 - m_0)^{\mathrm{T}} + \log|P_0|] \\
&- \frac{T+1}{2}\log|R| - \frac{1}{2}\operatorname{Tr}\left(R^{-1}\sum_{t=0}^{T} E[(\mathbf{x}_t - C\mathbf{z}_t)(\mathbf{x}_t - C\mathbf{z}_t)^{\mathrm{T}}|\mathbf{x}_{1:T}]\right) \\
&- \frac{T}{2}\log|Q| - \frac{1}{2}\operatorname{Tr}\left(Q^{-1}\sum_{t=0}^{T-1} E[(\mathbf{z}_{t+1} - A\mathbf{z}_t)(\mathbf{z}_{t+1} - A\mathbf{z}_t)|\mathbf{x}_{1:T}]\right) + c,
\end{aligned} \tag{3}
$$

Computing the Expectations **Whiteboard**

In order to compute the expectations the RTS smoother's posterior $f(\mathbf{z}_t|\mathbf{z}_{1:T})$ is used.
Then in the iterations of the EM method

1. An initial value for for θ is selected, say $\theta^{(0)}$.
2. A Kalman smoother is run using $\theta^{(i)}$
3. In the expectation step $\mathcal{Q}$ function as a function of θ is derived.
4. A new set of parameters $\theta^{(i+1)}$ are obtained from maximization of the $\mathcal{Q}$ function.
5. Iterations are repeated until convergence.

Read home

- Shumway and Stoffer, Chapter 6.3