Statistics for Data Science

MSc Data Science WiSe 2019/20

Prof. Dr. Dirk Ostwald

Bayesian inference

(13) Bayesian filtering

Bibliographic remarks

The material presented in this section follows Särkkä (2013) and Ostwald et al. (2014).

Bayesian filtering

- Foundations
- Bayesian filtering and the Kalman filter
- Bayesian smoothing and the Rauch-Tung-Striebel smoother
- Example

Bayesian filtering

- **Foundations**
- Bayesian filtering and the Kalman filter
- Bayesian smoothing and the Rauch-Tung-Striebel smoother
- Example

Bayesian filtering methods

- generalize Bayesian inference in static parameter scenarios to dynamic parameter (states) scenarios,
- aim to produce accurate estimates of latent states of time-varying systems based on multiple observational input,
- are a primary means to link dynamic models based on stochastic differential equations with statistical inference techniques (cf. Ostwald et al. (2014)),
- form the core methodology in the field of *data assimilation* as originating from the atmospheric and oceanographic sciences,
- form core assumptions in modern brain theories such as the *free energy principle* and *active inference*,
- provide the foundation for state-of-the-art reinforcement learning-based artificial intelligence in POMDP scenarios.

Application areas of Bayesian filtering methods include

- the global positioning system (GPS),
- target tracking,
- integrated inertial navigation (Apollo 11 lunar module),
- weather prediction,
- general sensor fusion problems,
- computational cognitive neuroscience and brain imaging,
- partially observable Markov decision processes (POMDP).

Definition (Probabilistic state space model)

A probabilistic state space model is a sequence of conditional distributions

$$S_0 \sim p(s_0), S_t \sim p(s_t|s_{t-1}) \text{ and } X_t \sim p(x_t|s_t) \text{ for } t = 1, 2, ..., T, \tag{1}$$

where

- t denotes a discrete time-point,
- S_t an n-dimensional random vector modeling the *system state* at t,
- X_t an m-dimensional random vector modeling a *system observation* at t,
- $p(s_t|s_{t-1})$ is the *dynamic model*, describing the system's evolution laws, and
- $p(x_t|s_t)$ is the *observation model*.

Remarks

- $p(s_t|s_{t-1})$ and $p(x_t|s_t)$ can be PMFs, PDFs, or mixtures of both.
- We here often refer to $p(s_t|s_{t-1})$ and $p(x_t|s_t)$ as distributions.

Remarks

- A probabilistic state space model has the joint distribution form

$$p(s_{0:T}, x_{1:T}) = p(s_0) \prod_{t=1}^{T} p(s_t|s_{t-1}) p(x_t|s_t). \tag{2}$$

- In a Bayesian inference sense, the prior distribution is

$$p(s_{0:T}) = p(s_0) \prod_{t=1}^{T} p(s_t|s_{t-1}). \tag{3}$$

- In a Bayesian inference sense, the likelihood is

$$p(x_{1:T}|s_{0:T}) = \prod_{t=1}^{T} p(x_t|s_t). \tag{4}$$

- In principle, inference could be performed using Bayes' theorem.

$$p(s_{0:T}|x_{1:T}) = \frac{p(x_{1:T}|s_{0:T}) p(s_{0:T})}{p(x_{1:T})}. \tag{5}$$

Remarks

- Computing $p(s_{0:T}|x_{1:T})$ directly is inefficient and often unnecessary.
- Instead the following conditional marginal distributions are considered:

  Predicted distributions are the conditional marginal distributions of s_t for $t = 2, 3, ..., T$ given observations x_t up to $t-1$:

  $$p(s_t|x_{1:t-1}) \text{ for } t = 2, 3, ..., T. \tag{6}$$

  Filtered distributions are the conditional marginal distributions of s_t for $t = 1, 2, ..., T$ given observations x_t up to t:

  $$p(s_t|x_{1:t}) \text{ for } t = 1, 2, ..., T. \tag{7}$$

  Smoothed distributions are the conditional marginal distributions of s_t for $t = 1, 2, ..., T$ given observations x_t for all $t = 1, 2, ..., T$:

  $$p(s_t|x_{1:T}) \text{ for } t = 1, 2, ..., T. \tag{8}$$

- Predicted and filtered distributions are evaluated by *Bayesian filtering.*
- Smoothed distributions are evaluated by *Bayesian smoothing.*

Remarks

- Probabilistic state space models are an example for the specification of high-dimensional probability distributions by means of conditional independence properties.
- For example, the definition of conditional probability implies

$$
\begin{aligned}
p(s_0, s_1, s_2, s_3) &= p(s_3|s_0, s_1, s_2)p(s_0, s_1, s_2) \\
&= p(s_3|s_0, s_1, s_2)p(s_2|s_0, s_1)p(s_0, s_1) \\
&= p(s_3|s_0, s_1, s_2)p(s_2|s_0, s_1)p(s_1|s_0)p(s_0).
\end{aligned} \tag{9}
$$

- The definition of conditional independence of s_t of all random variables except for s_{t-1} then render this

$$
p(s_0, s_1, s_2, s_3) := p(s_3|s_2)p(s_2|s_1)p(s_1|s_0)p(s_0). \tag{10}
$$

Remarks

Probabilistic state space models satisfy diverse conditional independence properties:

- The states $S_t, t = 0, 1, ..., T$ form a *Markov sequence*, i.e., S_t is conditionally independent of $S_{1:t-1}$ given S_{t-1},

$$p(s_t|s_{1:t-1}) = p(s_t|s_{t-1}). \tag{11}$$

- The states $S_t, t = 0, 1, ..., T$ are conditionally independent of $X_{1:t-1}$ given S_{t-1},

$$p(s_t|s_{1:t-1}, x_{1:t-1}) = p(s_t|s_{t-1}). \tag{12}$$

- The states $S_{t-1}, t = 1, 2, ..., T$ are conditionally independent of $S_{t:T}$ and $X_{t:T}$ given S_t

$$p(s_{t-1}|s_{t:T}, x_{t:T}) = p(s_{t-1}|s_t). \tag{13}$$

- The observations $X_t, t = 1, 2, ..., T$ are conditionally independent of the observations and state histories given S_t,

$$p(x_t|s_{1:t}, x_{1:t-1}) = p(x_t|s_t). \tag{14}$$

Bayesian filtering

- Foundations
- **Bayesian filtering and the Kalman filter**
- Bayesian smoothing and the Rauch-Tung-Striebel smoother
- Example

Definition (Bayesian filtering equations)

The *Bayesian filtering equations* are recursive equations for computing the predicted distributions $p(s_t|x_{1:t-1})$ and the filtered distributions $p(s_t|x_{1:t})$ for $t = 1, 2, ..., T$. They take the following form:

(1) *Prediction equation.* Given the filtered distribution at $t-1$, the predicted distribution at t can be computed using the the *Chapman-Kolmogorov* equation

$$p(s_t|x_{1:t-1}) = \int p(s_t|s_{t-1})p(s_{t-1}|x_{1:t-1})\, ds_{t-1}. \tag{15}$$

(2) *Filtering equation.* Given the predicted distribution at t, the filtered distribution at t can be computed using Bayes' theorem:

$$p(s_t|x_{1:t}) = \frac{p(x_t|s_t)p(s_t|x_{1:t-1})}{\int p(x_t|s_t)p(s_t|x_{1:t-1})\, ds_t}. \tag{16}$$

If some variables are discrete, integration is replaced by summation.

Definition (Bayesian filtering algorithm)

A *Bayesian filtering algorithm* is an iterative algorithm that computes the filtered distributions of a probabilistic state space model using the Bayesian filtering equations. It takes the following general form:

(0) *Initialization.* Set $p(s_0|x_{1:0}) := p(s_0)$.

For $t = 1, 2, ..., T$

(1) *Prediction.* Evaluate the predicted distribution

$$p(s_t|x_{1:t-1}) = \int p(s_t|s_{t-1})p(s_{t-1}|x_{1:t-1})\, ds_{t-1}. \tag{17}$$

(2) *Prediction.* Evaluate the filtered distribution

$$p(s_t|x_{1:t}) = \frac{p(x_t|s_t)p(s_t|x_{1:t-1})}{\int p(x_t|s_t)p(s_t|x_{1:t-1})\, ds_t}. \tag{18}$$

Definition (Linear Gaussian state space model, Kalman filter)

A *linear Gaussian state space models* is probabilistic state space model of the form

$$S_0 \sim N(m_0, \Sigma_0), S_t \sim N(As_{t-1}, Q), X_t \sim N(Bs_t, R) \text{ for } t = 1, 2, ..., T, \tag{19}$$

where

- S_0 is the n-dimensional Gaussian-distributed *initial state random vector* with expectation parameter $m_0 \in \mathbb{R}^n$ and covariance matrix parameter $\Sigma_0 \in \mathbb{R}^{n \times n}$p.d.,
- S_t is the n-dimensional Gaussian-distributed state random vector,
- $A \in \mathbb{R}^{n \times n}$ is the *state transition model matrix*,
- $Q \in \mathbb{R}^{n \times n}$ p.d. is the *state covariance matrix*,
- $B \in \mathbb{R}^{m \times n}$ is the *observation model matrix*, and
- $R \in \mathbb{R}^{m \times m}$ p.d. is the *observation covariance matrix.*

The *Kalman filter* is the Bayesian filtering algorithm for linear Gaussian state space models.

Theorem (Kalman filter)

(0) *Initialization.* Set $p(s_0|x_{1:0}) := p(s_0) := N(s_0; m_0, \Sigma_0)$.

For $t = 1, 2, ..., T$

(1) *Prediction.* Evaluate the predicted distribution as

$$p(s_t|x_{1:t-1}) = N(s_t; \tilde{m}_t, \tilde{\Sigma}_t) \text{ with } \tilde{m}_t := Am_{t-1} \text{ and } \tilde{\Sigma}_t := A\Sigma_{t-1}A^T + Q \quad (20)$$

(2) *Filtering.* Evaluate the filtered distribution as

$$p(s_t|x_{1:t}) = N(s_t; m_t, \Sigma_t) \text{ with } m_t := \tilde{m}_t + K_t v_t \text{ and } \Sigma_t := \tilde{\Sigma}_t - K_t U_t K_t^T, \quad (21)$$

where

$v_t := x_t - B\tilde{m}_t$ is the *observation prediction error*,
$U_t := B\tilde{\Sigma}_t B_t^T + R$ is the *predicted observation covariance*, and
$K_t := \tilde{\Sigma}_t B^T U_t^{-1}$ is the *Kalman gain matrix.*

Remarks

- The update formulas of the Kalman filter can be derived using the joint and conditional Gaussian distributions theorems (Lecture (3))

Theorem (Joint Gaussian distributions)

Given an m-dimensional random vector X distributed according to a Gaussian distribution with PDF

$$p_X : \mathbb{R}^m \to \mathbb{R}_{>0},\ x \mapsto p_X(x) := N(x; \mu_x, \Sigma_{xx}) \text{ for } \mu_x \in \mathbb{R}^m, \Sigma_{xx} \in \mathbb{R}^{m\times m}, \tag{22}$$

a matrix $A \in \mathbb{R}^{n\times m}$, a vector $b \in \mathbb{R}^n$, and a n-dimensional random vector Y conditionally distributed according to a Gaussian distribution with conditional PDF

$$p_{y|X}(\cdot|x) : \mathbb{R}^n \to \mathbb{R}_{>0},\ y \mapsto p_{Y|X}(y|x) := N(y; AX + b, \Sigma_{yy}) \text{ for } \Sigma_{yy} \in \mathbb{R}^{n\times n} \tag{23}$$

the $m + n$-dimensional random vector (X, Y) is distributed according to a Gaussian distribution with joint PDF

$$p_{X,Y} : \mathbb{R}^{m+n} \to \mathbb{R}_{>0},\ (x, y) \mapsto p_{X,Y}(x, y) = N\left((x, y); \mu_{x,y}, \Sigma_{x,y}\right), \tag{24}$$

where $\mu_{x,y} \in \mathbb{R}^{n+m}$ and $\Sigma_{x,y} \in \mathbb{R}^{m+n\times m+n}$, and in particular

$$\mu_{x,y} = \begin{pmatrix} \mu_x \\ A\mu_x + b \end{pmatrix} \text{ and } \Sigma_{x,y} = \begin{pmatrix} \Sigma_{xx} & \Sigma_{xx}A^T \\ A\Sigma_{xx} & \Sigma_{yy} + A\Sigma_{xx}A^T \end{pmatrix}. \tag{25}$$

Note that the parameters of the Gaussian joint distribution can be computed from the parameters of the Gaussian marginal and Gaussian conditional distributions.

Theorem (Conditional Gaussian distributions)

Given an $m+n$-dimensional random vector (X, Y) distributed according to a Gaussian distribution with PDF

$$p_{X,Y} : \mathbb{R}^{m+n} \to \mathbb{R}_{>0}, (x, y) \mapsto p_{X,Y}(x, y) := N\left((x, y); \mu_{x,y}, \Sigma_{x,y}\right), \tag{26}$$

where

$$\mu_{x,y} = \begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}, \Sigma_{x,y} = \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{pmatrix}, \tag{27}$$

for $x, \mu_x \in \mathbb{R}^m, y, \mu_y \in \mathbb{R}^n$ and $\Sigma_{xx} \in \mathbb{R}^{m \times m}, \Sigma_{xy} \in \mathbb{R}^{m \times n}, \Sigma_{yy} \in \mathbb{R}^{n \times n}$, the distribution of X given Y has an m-dimensional conditional PDF

$$p_{X|Y}(\cdot|y) : \mathbb{R}^m \to \mathbb{R}_{>0}, x \mapsto p_{X|Y}(x|y) := N(x; \mu_{x|y}, \Sigma_{x|y}), \tag{28}$$

where

$$\mu_{x|y} = \mu_x + \Sigma_{xy}\Sigma_{yy}^{-1}(Y - \mu_y) \in \mathbb{R}^m \tag{29}$$

and

$$\Sigma_{x|y} = \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx} \in \mathbb{R}^{m \times m}. \tag{30}$$

Note that with the previous statements, the parameters of conditional and marginal Gaussian PDFs can be computed from the parameters of complementary marginal and conditional Gaussian PDFs.

Proof

We assume the existence of the parameters of the filtered state distribution

$$p(s_{t-1}|x_{1:t-1}) = N(s_{t-1}; m_{t-1}, \Sigma_{t-1}) \text{ for } t = 2, 3, ..., T, \tag{31}$$

and for $t = 1$ we set $p(s_0|x_{1:0}) := p(s_0) = N(s_0; m_0, \Sigma_0)$.

Prediction

For the conditional joint distribution of s_{t-1} and s_t given $x_{1:t-1}$ we have with the conditional independence properties of probabilistic state space models and the joint distribution theorem for multivariate Gaussians (cf. Lecture (3))

$$\begin{aligned} p(s_{t-1}, s_t|x_{1:t-1}) &= p(s_t|s_{t-1}, x_{1:t-1})p(s_{t-1}|x_{1:t-1}) \\ &= p(s_t|s_{t-1})p(s_{t-1}|x_{1:t-1}) \\ &= N(s_t; As_{t-1}, Q)N(s_{t-1}; m_{t-1}, \Sigma_{t-1}) \\ &= N\left(\begin{pmatrix} s_{t-1} \\ s_t \end{pmatrix}; \begin{pmatrix} m_{t-1} \\ Am_{t-1} \end{pmatrix}, \begin{pmatrix} \Sigma_{t-1} & \Sigma_{t-1}A^T \\ A\Sigma_{t-1} & A\Sigma_{t-1}A^T + Q \end{pmatrix}\right) \end{aligned} \tag{32}$$

With the properties of joint multivariate Gaussian distributions, the conditional marginal distribution of s_t, i.e., the Chapman-Kolmogorov equation thus evaluates to

$$\begin{aligned} p(s_t|x_{1:t-1}) &= \int p(s_t|s_{t-1})p(s_{t-1}|x_{1:t-1})\, ds_{t-1} \\ &= N\left(s_t; Am_{t-1}, A\Sigma_{t-1}A^T + Q\right) \end{aligned} \tag{33}$$

Proof (cont.)

We thus have

$$p(s_t|x_{1:t-1}) = N(s_t; \tilde{m}_t, \tilde{\Sigma_t}) \text{ with } \tilde{m}_t := Am_t \text{ and } \tilde{\Sigma}_t := A\Sigma_{t-1}A^T + Q. \tag{34}$$

Filtering

For the conditional joint distribution of s_t and x_t given $x_{1:t-1}$ we then have with the conditional independence properties of probabilistic state space models and the joint distribution theorem for multivariate Gaussian distributions (cf. Lecture (4))

$$\begin{aligned} p(s_t, x_t|x_{1:t-1}) &= p(x_t|s_t, x_{1:t-1})p(s_t|x_{1:t-1}) \\ &= p(x_t|s_t)p(s_t|x_{1:t-1}) \\ &= N(x_t; Bs_t, R)N(s_t; \tilde{m}_t, \tilde{\Sigma}_t) \\ &= N\left(\begin{pmatrix} s_t \\ x_t \end{pmatrix}; \begin{pmatrix} \tilde{m}_t \\ B\tilde{m}_t \end{pmatrix}, \begin{pmatrix} \tilde{\Sigma}_t & \tilde{\Sigma}_t B^T \\ B\tilde{\Sigma}_t & B\tilde{\Sigma}_t B^T + R \end{pmatrix}\right) \end{aligned} \tag{35}$$

With the conditional distribution theorem for multivariate Gaussian distributions (cf. Lecture (3)) it then follows that

$$p(s_t|x_{1:t}) = N\left(s_t; m_t, \Sigma_t\right) \tag{36}$$

with

$$\begin{aligned} m_t &= \tilde{m}_t + \tilde{\Sigma}_t B^T \left(B\tilde{\Sigma}_t B^T + R\right)^{-1} (x_t - B\tilde{m}_t) \\ \Sigma_t &= \tilde{\Sigma}_t + \tilde{\Sigma}_t B^T \left(B\tilde{\Sigma}_t B^T + R\right)^{-1} B\tilde{\Sigma}_t \end{aligned} \tag{37}$$

Proof (cont.)

Definition of

$$v_t := x_t - B\tilde{m}_t \text{ and } U_t := B\tilde{\Sigma}_t B^T + R \tag{38}$$

then yields

$$\begin{aligned} m_t &= \tilde{m}_t + \tilde{\Sigma}_t B^T U_t^{-1} v_t \\ \Sigma_t &= \tilde{\Sigma}_t + \tilde{\Sigma}_t B^T U_t^{-1} B \tilde{\Sigma}_t. \end{aligned} \tag{39}$$

Furthermore, definition of

$$K_t := \tilde{\Sigma}_t B^T U_t^{-1} \tag{40}$$

yields

$$\begin{aligned} m_t &= \tilde{m}_t + K_t v_t \\ \Sigma_t &= \tilde{\Sigma}_t + K_t B \tilde{\Sigma}_t. \end{aligned} \tag{41}$$

Finally, with

$$K_t U_t K_t^T = K_t U_t U_t^{-1} B \tilde{\Sigma}_t^T = K_t B \tilde{\Sigma}_t \tag{42}$$

we have

$$\Sigma_t = \tilde{\Sigma}_t + K_t U_t K_t^T. \tag{43}$$

□

Bayesian filtering

- Foundations
- Bayesian filtering and the Kalman filter
- **Bayesian smoothing and the Rauch-Tung-Striebel smoother**
- Example

Definition (Bayesian smoothing equation)

The *Bayesian smoothing equation* is a recursive equation for computing the smoothed distributions $p(s_t|x_{1:T})$ for $t = T-1, T-2, ..., 1$.

Smoothing equation. Given the smoothed and predicted distributions at $t+1$, the smoothed distribution at t can be computed as

$$p(s_t|x_{1:T}) = p(s_t|x_{1:t}) \int \frac{p(s_{t+1}|s_t)p(s_{t+1}|x_{1:T})}{p(s_{t+1}|x_{1:t})}\, ds_{t+1} \tag{44}$$

If some variables are discrete, integration is replaced by summation.

Proof

We show the validity of the smoothing equation. To this end, we show that the conditional joint distribution of s_t and s_{t+1} given $x_{1:T}$ is given by

$$p(s_t, s_{t+1}|x_{1:T}) = p(s_t|x_{1:t})\frac{p(s_{t+1}|s_t)p(s_{t+1}|x_{1:T})}{p(s_{t+1}|x_{1:t})}, \tag{45}$$

from which the smoothing equation than follows immediately.

$$\begin{aligned} p(s_t|s_{t+1}, x_{1:T}) &= p(s_t|s_{t+1}, x_{1:t}) \\ &= \frac{p(s_t, s_{t+1}|x_{1:t})}{p(s_{t+1}|x_{1:t})} \\ &= \frac{p(s_{t+1}|s_t, x_{1:t})p(s_t|x_{1:t})}{p(s_{t+1}|x_{1:t})} \\ &= \frac{p(s_{t+1}|s_t)p(s_t|x_{1:t})}{p(s_{t+1}|x_{1:t})} \end{aligned} \tag{46}$$

and hence

$$\begin{aligned} p(s_t, s_{t+1}|x_{1:T}) &= p(s_t|s_{t+1}, x_{1:T})p(s_{t+1}|x_{1:T}) \\ &= p(s_t|s_{t+1}, x_{1:t})p(s_{t+1}|x_{1:T}) \\ &= p(s_t|x_{1:t})\frac{p(s_{t+1}|s_t)p(s_{t+1}|x_{1:T})}{p(s_{t+1}|x_{1:t})} \end{aligned} \tag{47}$$

□

Definition (Bayesian smoothing algorithm, RTS smoother)

A *Bayesian smoothing algorithm* is an iterative algorithm that computes the smoothed distributions of a probabilistic state space model using the filtered distributions and the Bayesian smoothing equations. It takes the following general form

(0) *Initialization.* Obtain the predicted distributions $p(s_{t+1}|x_{1:t})$ and the filtered distributions $p(s_t|x_{1:t})$ for $t = 1, ..., T$ using the Kalman filter

For $t = T-1, T-2, ..., 1$

(1) *Smoothing.* Evaluate the smoothed distribution

$$p(s_t|x_{1:T}) = p(s_t|x_{1:t}) \int \frac{p(s_{t+1}|s_t)p(s_{t+1}|x_{1:T})}{p(s_{t+1}|x_{1:t})}\, ds_{t+1} \tag{48}$$

The *Rauch-Tung-Striebel (RTS) smoother* is the Bayesian smoothing algorithm for a linear Gaussian state space model.

Theorem (Rauch-Tung-Striebel smoother)

(0) *Initialization.* Evaluate $p(s_t|x_{1:t}) = N(s_t; m_t, \Sigma_t)$ for $t = 1, \dots, T$ using the Kalman filter.

For $t = T - 1, 2, \dots, T$

(1) *Prediction.* Evaluate the predicted distributions as

$$p(s_{t+1}|x_t) = N(s_{t+1}; \bar{m}_{t+1}, \bar{\Sigma}_{t+1}) \tag{49}$$

with

$$\tilde{m}_{t+1} := Am_t \text{ and } \tilde{\Sigma}_{t+1} := A\Sigma_t A^T + Q. \tag{50}$$

(2) *Smoothing.* Evaluate the smoothed distributions as

$$p(s_t|x_{1:T}) = N(s_t; \bar{m}_t, \bar{\Sigma}_t) \tag{51}$$

with

$$\bar{m}_t := m_t + G_t(\bar{m}_{t+1} - \tilde{m}_{t+1}) \text{ and } \bar{\Sigma}_t := \Sigma_t + G_t(\bar{\Sigma}_{t+1} - \tilde{\Sigma}_{t+1})G_t^T, \tag{52}$$

where

$$G_t := \Sigma_t A^T \left(A\Sigma_t A^T + Q\right)^{-1}. \tag{53}$$

Remark

- The parameter update equations of the RTS smoother can be derived using the joint and conditional Gaussian distribution theorems.

Proof

The parameter update equations for the predicted distribution follow as for the Kalman filter. To derive the parameter update equations for the smoothed distribution, we make repeated use of the multivariate joint and conditional Gaussian distribution theorems to establish the parameters of the joint distribution $p(s_{t+1}, s_t|x_{1:T})$ and from it, the parameters of the smoothed distribution $p(s_t|x_{1:T})$. To this end, we first note that with the conditional independence properties of probabilistic state space models, we have

$$p(s_{t+1}, s_t|x_{1:T}) = p(s_t|s_{t+1}, x_{1:T})p(s_{t+1}|x_{1:T}) = p(s_t|s_{t+1}, x_{1:t})p(s_{t+1}|x_{1:T}) \tag{54}$$

We thus proceed as follows: in *Step (1)*, we evaluate the parameters of $p(s_t|s_{t+1}, x_{1:t})$. In *Step (2)*, we then evaluate the parameters of $p(s_{t+1}, s_t|x_{1:T})$. Finally, in *Step (3)* we evaluate the parameters of $p(s_t|x_{1:T})$.

Step (1) Evaluation of $p(s_t|s_{t+1}, x_{1:t})$

The parameters of $p(s_t|s_{t+1}, x_{1:t})$ can be evaluated based on the parameters of

$$p(s_t, s_{t+1}|x_{1:t}) = p(s_{t+1}|s_t, x_{1:t})p(s_t|x_{1:t}) \tag{55}$$

$$= p(s_{t+1}|s_t)p(s_t|x_{1:t}) \tag{56}$$

$$= N(s_{t+1}; As_t, Q)N(s_t; m_t, \Sigma_t), \tag{57}$$

which according to the joint multivariate Gaussian theorem is given by

$$p(s_t, s_{t+1}|x_{1:t}) = N\left(\begin{pmatrix} s_t \\ s_{t+1} \end{pmatrix}; \begin{pmatrix} m_t \\ Am_t \end{pmatrix}, \begin{pmatrix} \Sigma_t & \Sigma_t A^T \\ A\Sigma_t & A\Sigma_t A^T + Q \end{pmatrix}\right) \tag{58}$$

Proof (cont.)

With the conditional multivariate Gaussian theorem, we then have

$$p(s_t|s_{t+1}, x_{1:t}) = N\left(s_t; \tilde{\bar{m}}_t, \tilde{\bar{\Sigma}}_t\right) \tag{59}$$

with

$$\tilde{\bar{m}}_t := m_t + \Sigma_t A^T \left(A\Sigma_t A^T + Q\right)^{-1} (s_{t+1} - Am_t) \tag{60}$$

and

$$\tilde{\bar{\Sigma}}_t := \Sigma_t - \Sigma_t A^T \left(A\Sigma_t A^T + Q\right)^{-1} A\Sigma_t. \tag{61}$$

Definition of

$$G_t := \Sigma_t A^T \left(A\Sigma_t A^T + Q\right)^{-1}, \tag{62}$$

allows for rewriting the above as

$$\tilde{\bar{m}}_t := m_t + G_t(s_{t+1} - Am_t) \text{ and } \tilde{\bar{\Sigma}}_t := \Sigma_t - G_t(A\Sigma_t A^T + Q)G_t^T \tag{63}$$

Step (2) Evaluation of $p(s_{t+1}, s_t|x_{1:T})$

Assuming the existence of

$$p(s_{t+1}|x_{1:T}) = N(s_{t+1}; \bar{m}_{t+1}, \bar{\Sigma}_{t+1}) \tag{64}$$

we thus obtain for $p(s_{t+1}, s_t|x_{1:T})$,

$$\begin{aligned} p(s_{t+1}, s_t|x_{1:T}) &= p(s_t|s_{t+1}, x_{1:t})p(s_{t+1}|x_{1:T}) \\ &= N\left(\begin{pmatrix} s_{t+1} \\ s_t \end{pmatrix}; \begin{pmatrix} \bar{m}_{t+1} \\ m_t + G_t(\bar{m}_{t+1} - Am_t) \end{pmatrix}, \begin{pmatrix} \bar{\Sigma}_{t+1} & \bar{\Sigma}_{t+1}G_t^T \\ G_t\bar{\Sigma}_{t+1} & G_t\bar{\Sigma}_{t+1}G_t^T + \tilde{\bar{\Sigma}}_t \end{pmatrix}\right) \end{aligned}$$

Proof (cont.)

Step (3) Evaluation of $p(s_t|x_{1:T})$

From the parameters of $p(s_{t+1}, s_t|x_{1:T})$, we can read of the parameters of the marginal distribution

$$p(s_t|x_{1:T}) = N(s_t; \bar{m}_t, \bar{\Sigma}_t) \tag{65}$$

as

$$\begin{aligned} \bar{m}_t &= m_t + G_t(\bar{m}_{t+1} - Am_t) \\ &= m_t + G_t(\bar{m}_{t+1} - \tilde{m}_{t+1}) \end{aligned} \tag{66}$$

and

$$\begin{aligned} \bar{\Sigma}_t &= \tilde{\tilde{\Sigma}}_t + G_t \bar{\Sigma}_{t+1} G_t^T \\ &= \Sigma_t - G_t(A\Sigma_t A^T + Q)G_t^T + G_t \bar{\Sigma}_{t+1} G_t^T \\ &= \Sigma_t + G_t(\bar{\Sigma}_{t+1} - A\Sigma_t A^T - Q)G_t^T \\ &= \Sigma_t + G_t(\bar{\Sigma}_{t+1} - \tilde{\Sigma}_{t+1})G_t^T \end{aligned} \tag{67}$$

□

Bayesian filtering

- Foundations
- Bayesian filtering and the Kalman filter
- Bayesian smoothing and the Rauch-Tung-Striebel smoother
- **Example**

Example (A latent Gaussian random walk)

A latent Gaussian random walk model is a linear Gaussian state space model

$$S_0 \sim N(m_0, \sigma_0^2), S_t \sim N(s_{t-1}, \sigma_q^2), X_t \sim N(s_t, \sigma_r^2) \text{ for } t = 1, 2, ..., T, \tag{68}$$

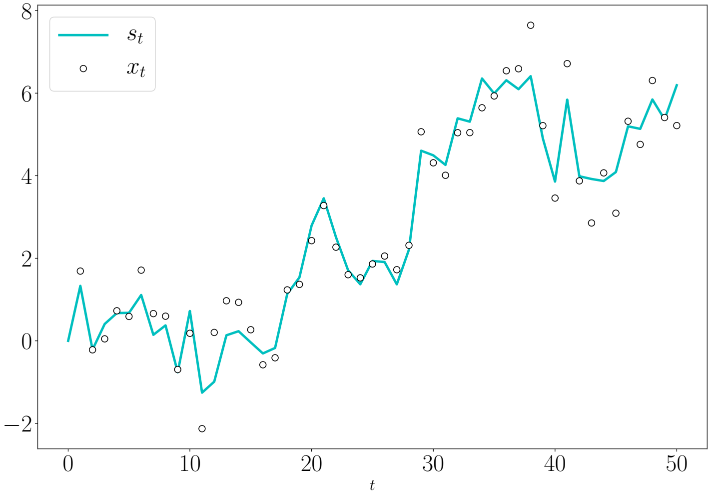
where S_0, S_t, X_t are real-valued random variables, $m_0 \in \mathbb{R}$ and $\sigma_q^2, \sigma_r^2 > 0$, and the state transition model and observation model matrices are given by $A := B := 1$.

A latent Gaussian random walk model can equivalently be written as

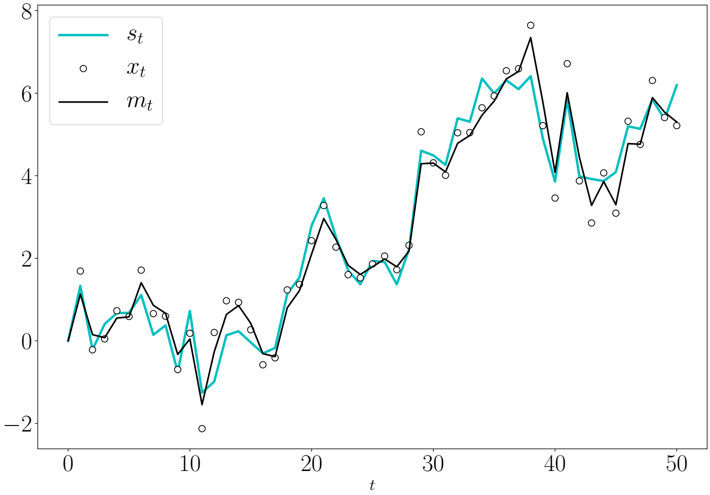
$$\begin{aligned} S_0 &\sim N(m_0, \sigma_0^2) \\ s_t &= s_{t-1} + \varepsilon_t, \varepsilon_t \sim N(0, \sigma_q^2) \\ x_t &= s_t + \eta_t, \quad \eta_t \sim N(0, \sigma_q^2), \end{aligned} \tag{69}$$

for $t = 1, 2, ..., T$ where the random variables ε_t and η_t are referred to as state and observation noise, respectively.

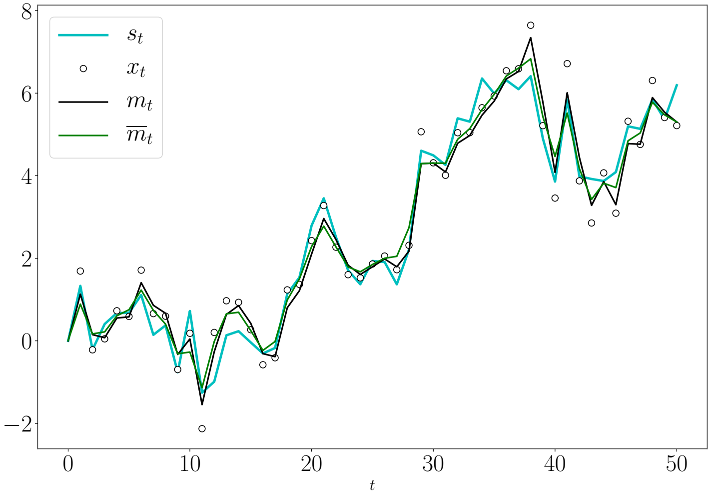
Example



Example



Example



Extensions for nonlinear dynamics and observation emissions (Särkkä, 2013)

- *Extended Kalman filter and RTS smoother*

  First-order Taylor approximation of nonlinear functions
- *Unscented Kalman filter and RTS smoother*

  Gaussian approximation of transformed distributions using mean and covariance.
- *Particle filtering and smoothing*

  Numerical sampling approaches

Parameter estimation

- Expectation-maximization for maximum likelihood estimation
- Bayesian estimation
- Variational estimation (Ostwald et al., 2014)

References

Ostwald, D., Kirilina, E., Starke, L., and Blankenburg, F. (2014). A tutorial on variational bayes for latent linear stochastic time-series models. *Journal of Mathematical Psychology*, 60:1–19.

Särkkä, S. (2013). *Bayesian filtering and smoothing*, volume 3. Cambridge University Press.