State Space Models

Shunxing Yan, Tiannuo Liang

School of Mathematical Sciences
University of Science and Technology of China

Dec 17, 2020

Table of Contents

LG-SSM

$$\mathbf{z}_t = \mathbf{A}\mathbf{z}_{t-1} + \epsilon_t, \tag{1}$$
$$\mathbf{y}_t = \mathbf{C}\mathbf{z}_t + \delta_t. \tag{2}$$

- $\mathbf{z}_t$: hidden state;
  $\mathbf{y}_t$: observation;
- $\epsilon_t \sim \mathcal{N}(\mathbf{0}, \mathbf{Q})$: system noise;
  $\delta_t \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$: observation noise.

Assume a Gaussian initial state distribution

$$\mathbf{z}_1 \sim \mathcal{N}(\boldsymbol{\mu}_{1|0}, \boldsymbol{\Sigma}_{1|0}). \tag{3}$$

LG-SSM

$$\log p(\mathbf{z}_{1:T}, \mathbf{y}_{1:T}) = \log[p(\mathbf{z}_1) \prod_{t=2}^{T} p(\mathbf{z}_t|\mathbf{z}_{t-1}) \prod_{t=1}^{T} p(\mathbf{y}_t|\mathbf{z}_t)] \tag{4}$$

$$= -\frac{1}{2}(\mathbf{z}_1 - \boldsymbol{\mu}_{1|0})'\boldsymbol{\Sigma}_{1|0}^{-1}(\mathbf{z}_1 - \boldsymbol{\mu}_{1|0}) - \frac{1}{2}\log|\boldsymbol{\Sigma}_{1|0}| \tag{5}$$

$$- \frac{1}{2}\sum_{t=2}^{T}(\mathbf{z}_t - \mathbf{A}\mathbf{z}_{t-1})'\mathbf{Q}^{-1}(\mathbf{z}_t - \mathbf{A}\mathbf{z}_{t-1}) - \frac{T-1}{2}\log|\mathbf{Q}| \tag{6}$$

$$- \frac{1}{2}\sum_{t=1}^{T}(\mathbf{y}_t - \mathbf{C}\mathbf{z}_t)'\mathbf{R}^{-1}(\mathbf{y}_t - \mathbf{C}\mathbf{z}_t) - \frac{T}{2}\log|\mathbf{R}| + C \tag{7}$$

E step

Compute

$$\boldsymbol{\mu}_{t|T} \triangleq \mathbb{E}(\mathbf{z}_t|\mathbf{y}_{1:T}), \tag{8}$$

$$\mathbf{P}_t \triangleq \mathbb{E}(\mathbf{z}_t\mathbf{z}_t'|\mathbf{y}_{1:T}) = \boldsymbol{\Sigma}_{t|T} + \boldsymbol{\mu}_{t|T}\boldsymbol{\mu}_{t|T}', \tag{9}$$

$$\mathbf{P}_{t,t-1} \triangleq \mathbb{E}(\mathbf{z}_t\mathbf{z}_{t-1}'|\mathbf{y}_{1:T}) \tag{10}$$

by Kalman smoothing algorithm.

E step

$$\boldsymbol{\Sigma}_{t,t-1|T} \triangleq cov(\mathbf{z}_t, \mathbf{z}_{t-1}|\mathbf{y}_{1:T}), \tag{11}$$
$$\mathbf{P}_{t,t-1} = \boldsymbol{\Sigma}_{t,t-1|T} + \boldsymbol{\mu}_{t|T}\boldsymbol{\mu}'_{t-1|T}. \tag{12}$$

The backward recursions:

$$\mathbf{J}_t = \boldsymbol{\Sigma}_t \mathbf{A}' \boldsymbol{\Sigma}^{-1}_{t+1|t}, \tag{13}$$
$$\mathbf{K}_t = \boldsymbol{\Sigma}_{t|t-1}\mathbf{C}'(\mathbf{C}\boldsymbol{\Sigma}_{t|t-1}\mathbf{C}' + \mathbf{R})^{-1}, \tag{14}$$
$$\boldsymbol{\Sigma}_{T,T-1|T} = (\mathbf{I} - \mathbf{K}_T\mathbf{C})\mathbf{A}\boldsymbol{\Sigma}_{T-1}, \tag{15}$$
$$\boldsymbol{\Sigma}_{t,t-1|T} = \boldsymbol{\Sigma}_t\mathbf{J}'_{t-1} + \mathbf{J}_t(\boldsymbol{\Sigma}_{t+1,t|T} - \mathbf{A}\boldsymbol{\Sigma}_t)\mathbf{J}'_{t-1}. \tag{16}$$

M step

$$
\frac{\partial L}{\partial \boldsymbol{\mu}_{1|0}} = (\boldsymbol{\mu}_{1|T} - \boldsymbol{\mu}_{1|0})' \boldsymbol{\Sigma}_{1|0}^{-1}, \tag{17}
$$

$$
\boldsymbol{\mu}_{1|0} = \boldsymbol{\mu}_{1|T}. \tag{18}
$$

$$
\frac{\partial L}{\partial \boldsymbol{\Sigma}_{1|0}^{-1}} = \frac{1}{2}\boldsymbol{\Sigma}_{1|0} - \frac{1}{2}(\mathbf{P}_1 - \boldsymbol{\mu}_{1|T}\boldsymbol{\mu}_{1|0}' - \boldsymbol{\mu}_{1|0}\boldsymbol{\mu}_{1|T}' + \boldsymbol{\mu}_{1|0}\boldsymbol{\mu}_{1|0}'), \tag{19}
$$

$$
\boldsymbol{\Sigma}_{1|0} = \mathbf{P}_1 - \boldsymbol{\mu}_{1|T}\boldsymbol{\mu}_{1|T}'. \tag{20}
$$

M step

$$\frac{\partial L}{\partial \mathbf{C}} = \mathbf{R}^{-1} \sum_{t=1}^{T} \mathbf{y}_t \mu'_{t|T} - \mathbf{R}^{-1}\mathbf{C} \sum_{t=1}^{T} \mathbf{P}_t, \tag{21}$$

$$\mathbf{C} = (\sum_{t=1}^{T} \mathbf{y}_t \mu'_{t|T})(\sum_{t=1}^{T} \mathbf{P}_t)^{-1}. \tag{22}$$

$$\frac{\partial L}{\partial \mathbf{R}^{-1}} = \frac{T}{2}\mathbf{R} - \frac{1}{2} \sum_{t=1}^{T} (\mathbf{y}_t \mathbf{y}'_t - \mathbf{C}\mu_{t|T}\mathbf{y}'_t - \mathbf{y}_t \mu'_{t|T}\mathbf{C}' + \mathbf{C}\mathbf{P}_t\mathbf{C}'), \tag{23}$$

$$\mathbf{R} = \frac{1}{T} \sum_{t=1}^{T} (\mathbf{y}_t \mathbf{y}'_t - \mathbf{C}\mu_{t|T}\mathbf{y}'_t). \tag{24}$$

M step

$$\frac{\partial L}{\partial \mathbf{A}} = \mathbf{Q}^{-1} \sum_{t=2}^{T} \mathbf{P}_{t,t-1} - \mathbf{Q}^{-1} \mathbf{A} \sum_{t=2}^{T} \mathbf{P}_{t-1}, \tag{25}$$

$$\mathbf{A} = (\sum_{t=2}^{T} \mathbf{P}_{t,t-1})(\sum_{t=2}^{T} \mathbf{P}_{t-1})^{-1}. \tag{26}$$

$$\frac{\partial L}{\partial \mathbf{Q}^{-1}} = \frac{T-1}{2}\mathbf{Q} - \frac{1}{2} \sum_{t=2}^{T} (\mathbf{P}_t - \mathbf{A}\mathbf{P}_{t-1,t} - \mathbf{P}_{t,t-1}\mathbf{A}' + \mathbf{A}\mathbf{P}_{t-1}\mathbf{A}'), \tag{27}$$

$$\mathbf{Q} = \frac{1}{T-1} \sum_{t=2}^{T} (\mathbf{P}_t - \mathbf{A}\mathbf{P}_{t-1,t}). \tag{28}$$

Non-linear, Non-Gaussian SSMs

$$\mathbf{z}_t = g(\mathbf{u}_t, \mathbf{z}_{t-1}, \epsilon_t), \tag{29}$$

$$\mathbf{y}_t = h(\mathbf{z}_t, \mathbf{u}_t, \delta_t). \tag{30}$$

- $\mathbf{z}_t$: hidden state;
  $\mathbf{u}_t$: input/control signal (optional);
  $\mathbf{y}_t$: observation;
- g: transition model;
  h: observation model;
- ϵ_t: system noise;
  δ_t: observation noise.

Approximate Inference

$Y = f(X)$:

- X: random variable with a Gaussian distribution;
- f: non-linear function.

Approximate $p(Y)$ by a Gaussian:

- Use a first-order approximation of f;
- Project $f(X)$ onto the space of Gaussians by moment matching.

Extended Kalman Filter (EKF)

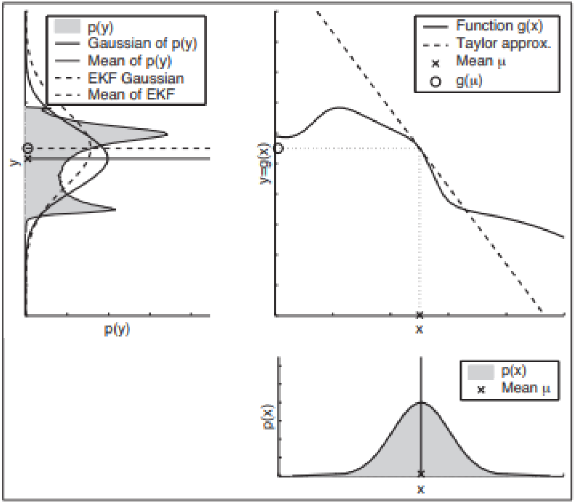
Assume the noise is Gaussian:

$$\mathbf{z}_t = g(\mathbf{u}_t, \mathbf{z}_{t-1}) + \mathcal{N}(\mathbf{0}, \mathbf{Q}_t), \tag{31}$$

$$\mathbf{y}_t = h(\mathbf{z}_t) + \mathcal{N}(\mathbf{0}, \mathbf{R}_t). \tag{32}$$

- g, h: non-linear but differentiable functions.
- Basic idea: linearize g, h about the previous state estimate using a first order Taylor series expansion

Extended Kalman Filter (EKF)



Extended Kalman Filter (EKF)

1.Approximate the measurement model:

$$p(\mathbf{y}_t|\mathbf{z}_t) \approx \mathcal{N}(\mathbf{y}_t|\mathbf{h}(\boldsymbol{\mu}_{t|t-1}) + \mathbf{H}_t(\mathbf{y}_t - \boldsymbol{\mu}_{t|t-1}), \mathbf{R}_t), \tag{33}$$

$$H_{ij} \triangleq \frac{\partial h_i(\mathbf{z})}{\partial z_j}, \tag{34}$$

$$\mathbf{H}_t \triangleq \mathbf{H}|_{\mathbf{z}=\boldsymbol{\mu}_{t|t-1}}. \tag{35}$$

Extended Kalman Filter (EKF)

2.Approximate the system model:

$$p(\mathbf{z}_t|\mathbf{z}_{t-1}, \mathbf{u}_t) \approx \mathcal{N}(\mathbf{z}_t|\mathbf{g}(\mathbf{u}_t, \boldsymbol{\mu}_{t-1}) + \mathbf{G}_t(\mathbf{z}_{t-1} - \boldsymbol{\mu}_{t-1}), \mathbf{Q}_t), \tag{36}$$

$$G_{ij}(\mathbf{u}) \triangleq \frac{\partial g_i(\mathbf{u}, \mathbf{z})}{\partial z_j}, \tag{37}$$

$$\mathbf{G}_t \triangleq \mathbf{G}(\mathbf{u}_t)|_{\mathbf{z}=\boldsymbol{\mu}_{t-1}}. \tag{38}$$

3.Apply the Kalman filter.

Weaknesses

- When the prior covariance is large;
- When the function is highly nonlinear near the current mean.

Unscented Transform

- Key intuition: it is easier to approximate a Gaussian than to approximate a function;
- Pass a deterministically chosen set of points (sigma points) through the function, and fit a Gaussian to the resulting transformed points;
- Advantages:
  1.Be accurate to at least second order;
  2.Not require the analytic evaluation of any derivatives or Jacobians (derivative free).

Unscented Transform

Assume

$$p(\mathbf{x}) = \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}), \tag{39}$$

$$\mathbf{y} = \mathbf{f}(\mathbf{x}), \tag{40}$$

and estimate $p(\mathbf{y})$.
Create $2d+1$ sigma points

$$\mathbf{x} = (\boldsymbol{\mu}, \{\boldsymbol{\mu} + (\sqrt{(d+\lambda)\boldsymbol{\Sigma}})_{:i}\}_{i=1}^{d}, \{\boldsymbol{\mu} - (\sqrt{(d+\lambda)\boldsymbol{\Sigma}})_{:i}\}_{i=1}^{d}). \tag{41}$$

- $\lambda = \alpha^2(d+\kappa) - d$;
- $\mathbf{M}_{:i}$: the i'th column of matrix $\mathbf{M}$.

Unscented Transform

$$\mathbf{y}_i = \mathbf{f}(\mathbf{x}_i), \tag{42}$$

$$\boldsymbol{\mu}_y = \sum_{i=0}^{2d} w_m^i \mathbf{y}_i, \tag{43}$$

$$\boldsymbol{\Sigma}_y = \sum_{i=0}^{2d} w_c^i (\mathbf{y}_i - \boldsymbol{\mu}_y)(\mathbf{y}_i - \boldsymbol{\mu}_y)^T. \tag{44}$$

- $w_m^0 = \frac{\lambda}{d+\lambda}$;
  $w_c^0 = \frac{\lambda}{d+\lambda} + (1 - \alpha^2 + \beta)$;
- $w_m^i = w_c^i = \frac{1}{2(d+\lambda)}$;
- $d = 1 : \alpha = 1, \beta = 0, \kappa = 2, \lambda = 2.$

Unscented Kalman Filter (UKF)

1.Approximate the predictive density

$$p(\mathbf{z}_t|\mathbf{y}_{1:(t-1)}, \mathbf{u}_{1:t}) \approx \mathcal{N}(\mathbf{z}_t|\bar{\mu}_t, \bar{\mathbf{\Sigma}}_t). \tag{45}$$

Pass the old belief state

$$p(\mathbf{z}_{t-1}|\mathbf{y}_{1:(t-1)}, \mathbf{u}_{1:(t-1)}) = \mathcal{N}(\mathbf{z}_{t-1}|\mu_{t-1}, \mathbf{\Sigma}_{t-1}) \tag{46}$$

through the system model **g**:

Unscented Kalman Filter (UKF)

$$
\begin{aligned}
\mathbf{z}^0_{t-1} &= (\boldsymbol{\mu}_{t-1}, \{\boldsymbol{\mu}_{t-1} + \gamma(\sqrt{\boldsymbol{\Sigma}_{t-1}})_{:i}\}_{i=1}^d, \{\boldsymbol{\mu}_{t-1} - \gamma(\sqrt{\boldsymbol{\Sigma}_{t-1}})_{:i}\}_{i=1}^d), && (47)\\
\bar{\mathbf{z}}_t^{*i} &= \mathbf{g}(\mathbf{u}_t, \mathbf{z}_{t-1}^{0i}), && (48)\\
\bar{\boldsymbol{\mu}}_t &= \sum_{i=0}^{2d} w_m^i \bar{\mathbf{z}}_t^{*i}, && (49)\\
\bar{\boldsymbol{\Sigma}}_t &= \sum_{i=0}^{2d} w_c^i (\bar{\mathbf{z}}_t^{*i} - \bar{\boldsymbol{\mu}}_t)(\bar{\mathbf{z}}_t^{*i} - \bar{\boldsymbol{\mu}}_t)^T + \mathbf{Q}_t. && (50)
\end{aligned}
$$

- $\gamma = \sqrt{d + \lambda}$.

Unscented Kalman Filter (UKF)

2.Approximate the likelihood

$$p(\mathbf{y}_t|\mathbf{z}_t) \approx \mathcal{N}(\mathbf{y}_t|\hat{\mathbf{y}}_t, \mathbf{S}_t). \tag{51}$$

Pass the prior

$$\mathcal{N}(\mathbf{z}_t|\bar{\mu}_t, \bar{\mathbf{\Sigma}}_t) \tag{52}$$

through the observation model **h**:

Unscented Kalman Filter (UKF)

$$\bar{\mathbf{z}}_t^0 = (\bar{\boldsymbol{\mu}}_t, \{\bar{\boldsymbol{\mu}}_t + \gamma(\sqrt{\bar{\boldsymbol{\Sigma}}_t})_{:i}\}_{i=1}^d, \{\bar{\boldsymbol{\mu}}_t - \gamma(\sqrt{\bar{\boldsymbol{\Sigma}}_t})_{:i}\}_{i=1}^d), \tag{53}$$

$$\bar{\mathbf{y}}_t^{*i} = \mathbf{h}(\bar{\mathbf{z}}_t^{0i}), \tag{54}$$

$$\hat{\mathbf{y}}_t = \sum_{i=0}^{2d} w_m^i \bar{\mathbf{y}}_t^{*i}, \tag{55}$$

$$\mathbf{S}_t = \sum_{i=0}^{2d} w_c^i (\bar{\mathbf{y}}_t^{*i} - \hat{\mathbf{y}}_t)(\bar{\mathbf{y}}_t^{*i} - \hat{\mathbf{y}}_t)^T + \mathbf{R}_t. \tag{56}$$

Unscented Kalman Filter (UKF)

3.Compute the posterior

$$p(\mathbf{z}_t|\mathbf{y}_{1:t},\mathbf{u}_{1:t}) \approx \mathcal{N}(\mathbf{z}_t|\mu_t,\mathbf{\Sigma}_t). \tag{57}$$

$$\bar{\mathbf{\Sigma}}_t^{z,y} = \sum_{i=0}^{2d} w_c^i(\bar{\mathbf{z}}_t^{*i} - \bar{\mu}_t)(\bar{\mathbf{y}}_t^{*i} - \hat{\mathbf{y}}_t)^T, \tag{58}$$

$$\mathbf{K}_t = \bar{\mathbf{\Sigma}}_t^{z,y}\mathbf{S}_t^{-1}. \tag{59}$$

Assumed Density Filtering (ADF)

- Idea: perform an exact update step, and then approximate the posterior by a distribution of a certain convenient form.
- $\mathcal{Q}$: a set of tractable distributions;
- θ_t: the unknowns we want to infer.

Assumed Density Filtering (ADF)

Suppose we have an approximate prior

$$q_{t-1}(\boldsymbol{\theta}_{t-1}) \approx p(\boldsymbol{\theta}_{t-1}|\mathbf{y}_{1:(t-1)})(q_{t-1} \in \mathcal{Q}). \tag{60}$$

$$q_{t|t-1}(\boldsymbol{\theta}_t) = \int p(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1})q_{t-1}(\boldsymbol{\theta}_{t-1})d\boldsymbol{\theta}_{t-1}, \tag{61}$$

$$\hat{p}(\boldsymbol{\theta}_t) = \frac{1}{Z_t}p(\mathbf{y}_t|\boldsymbol{\theta}_t)q_{t|t-1}(\boldsymbol{\theta}_t), \tag{62}$$

$$q_t(\boldsymbol{\theta}_t) = argmin_{q\in\mathcal{Q}}\mathbb{KL}(\hat{p}(\boldsymbol{\theta}_t)||q(\boldsymbol{\theta}_t)). \tag{63}$$

Assumed Density Filtering (ADF)