Introduction to (Bayesian) Estimation
MAE 5020

Filtering and Smoothing: a Bayesian approach

Oklahoma State University

November 1, 2023

Overview

- Deviate from the textbook to present a systematic development of nonlinear filtering and smoothing
- Probabilitist models
- Generic filtering and smoothing equations
- Application to the basic state-variable model
- Reference: Bayesian filtering and smoothing, 2013, Cambridge University Press

Classification

Depending on the relative relationship of total number of available measurements up to time step N and the time point k at which we estimate $x(k)$

- $k > N$: prediction
- $k = N$: filtering
- $k < N$: smoothing.

Notation: $\hat{x}(k|j)$, $x_k = x(k)$, $x_{1:k-1}$

Probabilistic state space model

We extend the basic state-variable model to a generic model given by a sequence of conditional probability distributions:

$$x_k \sim p(x_k | x_{k-1})$$
$$z_k \sim p(z_k | x_k)$$

where

Example: the basic state-variable model

What is the probabilistic state-space model?

$$
\begin{aligned}
x(k+1) &= \Phi(k+1,k)x(k) + \Gamma(k+1,k)w(k) + \Psi(k+1,k)u(k) \\
z(k) &= H(k)x(k) + v(k), \quad k = 0, 1, \cdots .
\end{aligned}
$$

Markov properties

The states x_k, $k = 0, 1, 2, \cdots$, form a Markov sequence, which satisfies

- $p(x_k | x_{1:k-1}, z_{1:k-1}) = p(x_k | x_{k-1})$
- $p(x_{k-1} | x_{k:N}, z_{k:N}) = p(x_{k-1} | x_k)$.
- Conditional independence of measurements:

$$p(z_k | x_{1:k}, z_{1:k-1}) = p(y_k | x_k).$$

Example: Gaussian random walk

$$x_{k+1} = x_k + w_k, \quad w_{k-1} \sim \mathcal{N}(0, Q),$$
$$z_k = x_k + v_k, \quad v_k \sim \mathcal{N}(0, R).$$

Joint distribution, likelihood, and posterior

- Joint prior distribuiton:

$$p(x_{0:N}) =$$

- Joint likelihood distribution

$$p(z_{1:N}|x_{0:N}) =$$

- Posterior distribution from Bayes' rule:

$$\begin{aligned} p(x_{0:N}|y_{1:N}) &= \\ &\approx \end{aligned}$$

Challenges in real time

- The number of computations per time step increases as new observations arrive.
- Develop recursive estimation steps to compute prediction, filtering, and smoothing distributions.
- Only a constant number of computations are done at each time step.
- Start with Bayesian filtering equations (which will include prediction)
- Move to Bayesian smoothing equations

Generic Bayesian filtering equations

- Bayesian filtering: compute the marginal distribution $p(x_k|y_{1:k})$.
- Fundamental equations for recursive Bayesian filtering at time step k
  - Initialization: start with the prior distribution $p(x_0)$
  - Prediction step: Chapman-Kolmogorov equation

$$p(x_k|z_{1:k-1}) = \int p(x_k|x_{k-1})p(x_{k-1}|z_{1:k-1})dx_{k-1}$$

  - Update step: via Bayes' rule

$$p(x_k|z_{1:k-1}, z_k) = \frac{p(z_k|x_k)p(x_k|z_{1:k-1})}{Z_k}$$

$$Z_k = \int p(z_k|x_k)p(x_k|z_{1:k-1})dx_k.$$

Graphically

Proof

Joint distribution of x_k and x_{k-1}:

Marginalization of x_{k-1} yields:

Bayes rule for x_k given z_k and $z_{1:k-1}$, i.e., $z_{1:k}$:

Developing the Kalman filter

Closed-form solution to the Bayesian filtering equations for linear Gaussian dynamic and measurement models

$$x(k+1) = \Phi(k+1,k)x(k) + \Gamma(k+1,k)w(k) + \Psi(k+1,k)u(k)$$
$$z(k) = H(k)x(k) + v(k), \quad k = 0, 1, \cdots .$$

- $v(k) \sim \mathcal{N}(0, R(k))$, $w(k) \sim \mathcal{N}(0, Q(k))$, $x(0) \sim \mathcal{N}(m_x(0), P_x(0))$.
- Probabilistic models

Two useful Lemmas

Lemma 1 Suppose $x \sim \mathcal{N}(m, P)$ and $y|x \sim \mathcal{N}(Hx + u, R)$, then the joint distribution of x, y and the marginal distribution of y are given by

$$\begin{pmatrix} x \\ y \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} m \\ Hm + u \end{pmatrix}, \begin{pmatrix} P & PH^T \\ HP & HPH^T + R \end{pmatrix}\right).$$

Lemma 2: conditional distribution Suppose

$$\begin{pmatrix} x \\ y \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} a \\ b \end{pmatrix}, \begin{pmatrix} A & C \\ C^T & B \end{pmatrix}\right).$$

Then,

$$x \sim \mathcal{N}(a, A), \quad y \sim \mathcal{N}(b, B)$$

$$x|y \sim \mathcal{N}(a + CB^{-1}(y - b), A - CB^{-1}C^T)$$

$$y|x \sim \mathcal{N}(b + C^T B^{-1}(x - a), B - C^T A^{-1}C).$$

The KF equations from the Bayesian filtering equation

Assume that $p(x_{k-1}|z_{1:k-1}) \sim \mathcal{N}(m_{k-1}, P_{k-1})$

Prediction step: Calculate $p(x_k|z_{1:k-1})$ from

$p(x_k, x_{k-1}|z_{1:k-1}) = p(x_k|x_{k-1})p(x_{k-1}|z_{1:k-1})$

Update step: $p(x_k|z_{1:k}) \propto p(z_k|x_k)p(x_k|z_{1:k-1})$ calculated from $p(x_k, z_k|z_{1:k-1})$.

Final equations adapted to the basic state-variable model

$$x(k+1) = \Phi(k+1,k)x(k) + \Gamma(k+1,k)w(k) + \Psi(k+1,k)u(k)$$
$$z(k) = H(k)x(k) + v(k), \quad k = 0, 1, \cdots .$$

Prediction:

$$\hat{x}(k+1|k) = \Phi(k+1,k)\hat{x}(k|k) + \Psi(k+1,k)u(k)$$
$$P(k+1|k) = \Phi(k+1,k)P(k|k)\Phi^T(k+1,k) + \Gamma(k+1,k)Q(k)\Gamma^T(k+1,k)$$

Update: $\tilde{z}(k) = z(k) - H(k)\hat{x}(k+1|k)$

$$\hat{x}(k+1|k+1) = \hat{x}(k+1|k) + K(k+1)\tilde{z}(k+1|k)$$
$$K(k+1) = P(k+1|k)H^T(k+1)\left(H(k+1)P(k+1|k)H^T(k+1) + R(k+1)\right)^{-1}$$
$$P(k+1|k+1) = (I - K(k+1)H(k+1))P(k+1|k).$$

Comments

Example: Gaussian random walk

Bayesian smoothing: Fixed-interval

- Compute the marginal posterior distribution $p(x_k|z_{1:N})$, where $N > k$.
- Smoothing also uses future measurements: Fixed-interval smoothing cannot be implemented online.
- Other types of smoothing:
  - Fixed-point smoother: $p(x_k|z_{1:j})$, $j = k+1, \cdots$, with a fixed k. In this case, we improve our estimate of a state at a particular time by using future measurements. It can be calculated online, but subject to a delay of $(j-k)$ steps.
  - fixed-lag smoother: $p(x_k|x_{k:k+L})$, $k = 0, 1, \cdots$, with a fixed L (positive integer). It can be used online where a constant lag between measurements and state estimates is permissible, subject to L steps of delay.

Bayesian smoothing equations

Forward computation: obtain the filtering posterior state distributions, $p(x_k|z_{1:k})$, $k = 1, \cdots, N$, via a filter.

Backward computation:

$$p(x_k|z_{1:N}) = p(x_k|z_{1:k}) \int \frac{p(x_{k+1}|x_k)p(x_{k+1}|z_{1:T})}{p(x_{k+1}|z_{1:k})} dx_{k+1}$$

- $p(x_k|z_{1:k})$:
- $p(x_{k+1}|x_k)$:
- $p(x_{k+1}|z_{1:k})$:

- $p(x_{k+1}|z_{1:T})$:

Why?

Markov property: $p(x_k|x_{k+1}, z_{1:N}) =$

Bayes' rule: $p(x_k|x_{k+1}, z_{1:N}) =$

Joint distribution $p(x_k, x_{k+1}|z_{1:N}) =$

Marginal distribution $p(x_k|z_{1:N}) = \int p(x_k, x_{k+1}|z_{1:N}) dx_{k+1}$

Rauch-Tung-Striebel (RTS) smoother: Kalman smoother

$$x(k+1) = \Phi(k+1,k)x(k) + \Gamma(k+1,k)w(k) + \Psi(k+1,k)u(k)$$
$$z(k) = H(k)x(k) + v(k), \quad k = 0, 1, \cdots .$$

- Closed-form smoother equations for linear Gaussian dynamic and measurement models
- Look for smoothed distributions $p(x_k|z_{1:T}) = \mathcal{N}(m_k^s, P_k^s)$.
- Consists of a forward pass and a backward pass
  - Forward pass: the Kalman filter
  - Backward pass: backward-running recursive predictor

RTS Smoother Algorithm adapted to the basic model

Forward pass: Perform the KF to obtain $\hat{x}(k|k)$, $\hat{x}(k|k-1)$, $P(k|k-1)$, and $P(k|k)$, $k = 1, \cdots, N$.
Backward pass: for $k = N-1, N-2, \cdots, 0$ as

$$\begin{aligned} x(k|N) &= \hat{x}(k|k) + A(k)[\hat{x}(k+1|N) - \hat{x}(k+1|k)] \\ A(k) &= P(k|k)\Phi^T(k+1,k)P^{-1}(k+1|k) \\ P(k|N) &= P(k|k) + A(k)[P(k+1|N) - P(k+1|k)]A^T(k). \end{aligned}$$

- $P(k|N)$:
- Availability of $\hat{x}(k|k)$, $A(k)$, $\hat{x}(k+1|N)$, $\hat{x}(k+1|k)$, $P(k+1|N)$, $P(k|k)$, $P(k+1|k)$:

Example

Nonlinear filters and smoothers

- Practical applications involve nonlinear dynamic and measurement models.
- Posterior distributions are no longer Gaussian even if the added noise is Gaussian.
- A large class of nonlinear filters aim at approximating the posterior distributions as Gaussian distributions
  - Extended Kalman filter (EKF)
  - Unscented Kalman filter (UKF)
  - Gaussian filter, ...
- Another type of popular nonlinear filters is Particle filters (PF), using a similar concept to MCMC.
- For each nonlinear filter, there is a corresponding nonlinear smoother, e.g., EKS, Particle smoother.

Taylor series expansion in linearization

Consider $x \sim \mathcal{N}(m, P)$ and $y = g(x)$.

- $p(y) = |J(y)|\mathcal{N}(g^{-1}(y)|m, P)$, where $J(y)$ is the Jacobian matrix of the inverse transform $g^{-1}(y)$.
- Except linear g, $p(y)$ is difficult to obtain and handle.

First-order Taylor series expansion (linearization):

- Let $x = m + \delta x$, where $\delta x \sim \mathcal{N}(0, P)$.

$$g(x) = g(m + \delta x) \approx g(m) + G_x(m)\delta x + H.O.T.$$

$G_x(m)$ is the Jacobian matrix of $g(x)$ evaluated at $x = m$.

Expectation, covariance, joint distribution with linearization

A General Case: Mean, covariance, joint distribution

$$x \sim \mathcal{N}(m, P), w \sim \mathcal{N}(0, Q), y = g(x, q)$$

Extended Kalman filter

- Key idea: linearization, assume Gaussian distributions for posterior
- Generic model:

$$
\begin{aligned}
x_k &= f(x_{k-1}, u_{k-1}, w_{k-1}), \quad w_{k-1} \sim \mathcal{N}(0, Q_{k-1}) \\
z_k &= h(x_k, v_k), \qquad v_k \sim \mathcal{N}(0, R_k).
\end{aligned}
$$

- Linearization

$$f(x_{k-1}, u_{k-1}, w_{k-1}) \approx f(m_{k-1}, u_{k-1}, 0) + F_{x,k}\delta x_{k-1} + F_{w,k} w_{k-1}$$

$$h(x_k, v_k) \approx h(m_k, 0) + H_{x,k}\delta x_k + H_{v,k} v(k)$$

EKF algorithm

Prediction:

Update:

Comments

- Advantage of EKF: simplicity compared to its performance.
- Disadvantages of EKF:
  - Local linear approximation may not be sufficient for considerable nonlinearity.
  - Also not applicable to non-Gaussian distributions.
  - Requires differentiable dynamics and measurement models.
  - Computation of Jacobian matrices can be error prone.
- Despite all these disadvantages, its simplicity makes EKF a first try and a common engineering approach to nonlinear filtering.

Example

Gaussian filters

- Unify various Gaussian approximations to nonlinear transforms: approximate computation of the expectation

$$E_x[g(x)] = \int g(x)\mathcal{N}(x|m, P)dx.$$

- If we can compute the above integral, we can then match the first/second order moments of the Gaussian approximation of (x, y) to the true moments.

Moment matching for $y = g(x, w)$, $x \sim \mathcal{N}(m, P)$, $w \sim \mathcal{N}(0, Q)$

Approximate joint distribution of (x, y)

$$\begin{pmatrix} x \\ y \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} m \\ \mu_M \end{pmatrix}, \begin{pmatrix} P & C_M \\ C_M^T & S_M \end{pmatrix} \right)$$

Gaussian filter

- Assume we can compute the integrals numerically and $p(x_k|z_{1:k}) = \mathcal{N}(m_k, P_k)$.
- Generic model:

$$
\begin{aligned}
x_k &= f(x_{k-1}, u_{k-1}, w_{k-1}), \quad w_{k-1} \sim \mathcal{N}(0, Q_{k-1}) \\
z_k &= h(x_k, v_k), \qquad v_k \sim \mathcal{N}(0, R_k).
\end{aligned}
$$

- Prediction:

- Update:

Computation of the integral

- Linearization leads to EKF
- Gauss-Hermite quadratures
- Cubature rules
- Unscented Transforms
- Bayes-Hermite quadrature
- Monte-Carlo integration