Linear Gaussian State-Space Model Estimation

Gu, Xin*

Contents

1 Linear Gaussian State-Space Model

Consider a discrete-time linear dynamical system characterized by the following state-space representation:

$$\begin{aligned} y_t &= Hs_t + v_t\,, \quad v_t \sim \mathcal{N}(0, R)\,, && (1) \\ s_t &= As_{t-1} + w_t\,, \quad w_t \sim \mathcal{N}(0, Q)\,, && (2) \end{aligned}$$

Here, (2) denotes the **state transition equation**, governing the evolution of the latent state $s_t \in \mathbb{R}^m$ over time, where A represents the state transition dynamics, and w_t is the process noise. Equation (1) defines the **observation model**, mapping the latent state to the observed measurement $y_t \in \mathbb{R}^n$ via the observation matrix H, subject to the measurement noise v_t. Both w_t and v_t are assumed to be mutually independent, zero-mean Gaussian white noise processes with covariance matrices Q and R, respectively.

2 Recursive Estimation: The Kalman Filter

The objective of the filtering problem is to determine the posterior distribution of the state s_t given the history of observations $\mathcal{Y}_t = \{y_1, \ldots, y_t\}$. Within the Bayesian framework, this is achieved through a recursive two-step process: **prediction** and **update**.

2.1 The Prediction Phase

The prediction step propagates the previous posterior estimate into the current time step. The **a priori** state estimate, $s_{t|t-1}$, and its associated error covariance, $P_{t|t-1}$, are defined as the conditional expectations:

$$\begin{aligned} s_{t|t-1} &:= \mathbb{E}[s_t \mid y_{1:t-1}]\,, \\ P_{t|t-1} &:= \mathbb{E}[(s_t - s_{t|t-1})(s_t - s_{t|t-1})^\top]\,, \end{aligned}$$

*PhD in Economics, School of Finance (School of Zheshang Asset Management), Zhejiang Gongshang University, Hangzhou, Zhejiang, China. Email: richardgu26@zgjsu.edu.cn.

Given the state-space representation shown in Section 1, the estimated state and its covariance can be expressed as

$$
\begin{aligned}
s_{t|t-1} &= As_{t-1|t-1}\,, && (3)\\
P_{t|t-1} &= \mathbb{E}[(s_t - s_{t|t-1})(s_t - s_{t|t-1})^\top]\,,\\
&= \mathbb{E}[(As_{t-1} - As_{t-1|t-1} + w_t)(As_{t-1} - As_{t-1|t-1} + w_t)^\top]\,,\\
&= A\mathbb{E}[(s_{t-1} - s_{t-1|t-1})(s_{t-1} - s_{t-1|t-1})^\top]A^\top + \mathbb{E}[w_t w_t^\top]\,,\\
&= AP_{t-1|t-1}A^\top + Q\,, && (4)
\end{aligned}
$$

Simultaneously, we define the predicted observation $y_{t|t-1}$ and the innovation covariance $F_{t|t-1}$ as:

$$
\begin{aligned}
y_{t|t-1} &= Hs_{t|t-1}\,, && (5)\\
F_{t|t-1} &= \mathbb{E}[(y_t - y_{t|t-1})(y_t - y_{t|t-1})^\top]\,,\\
&= \mathbb{E}[(Hs_t - Hs_{t|t-1} + v_t)(Hs_t - Hs_{t|t-1} + v_t)^\top]\,,\\
&= H\mathbb{E}[(s_t - s_{t|t-1})(s_t - s_{t|t-1})^\top]H^\top + \mathbb{E}[v_t v_t^\top]\,,\\
&= HP_{t|t-1}H^\top + R\,, && (6)
\end{aligned}
$$

The log-likelihood of the observation y_t is then given by the Gaussian density:

$$
\mathcal{L} = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log(\det(F_{t|t-1})) - \frac{1}{2}\left((y_t - y_{t|t-1})^\top F_{t|t-1}^{-1}(y_t - y_{t|t-1})\right)\,, \tag{7}
$$

2.2 The Update Phase

Under the Gaussian assumption, the joint distribution of the state s_t and the observation y_t conditional on $\mathcal{Y}_{t-1}$ is expressed as:

$$
\begin{pmatrix} y_t \\ s_t \end{pmatrix} \sim \mathcal{N}\left(\begin{bmatrix} y_{t|t-1} \\ s_{t|t-1} \end{bmatrix}, \begin{bmatrix} F_{t|t-1} & HP_{t|t-1} \\ P_{t|t-1}H^\top & P_{t|t-1} \end{bmatrix}\right)\,, \tag{8}
$$

where

$$
\begin{aligned}
\mathrm{cov}(s_t, y_t) &= \mathbb{E}[(s_t - s_{t|t-1})(y_t - y_{t|t-1})^\top]\,,\\
&= \mathbb{E}[(s_t - s_{t|t-1})(Hs_t - Hs_{t|t-1} + v_t)^\top]\,,\\
&= \mathbb{E}[(s_t - s_{t|t-1})(s_t - s_{t|t-1})^\top]H^\top\,,\\
&= P_{t|t-1}H^\top\,,
\end{aligned}
$$

By applying the properties of conditional Gaussian distributions, the **a posteriori** state estimate $s_{t|t}$ and covariance $P_{t|t}$ are derived as:

$$
\begin{aligned}
s_{t|t} &= s_{t|t-1} + P_{t|t-1}H^\top F_{t|t-1}^{-1}(y_t - y_{t|t-1})\,,\\
P_{t|t} &= P_{t|t-1} - P_{t|t-1}H^\top F_{t|t-1}^{-1}HP_{t|t-1}\,,
\end{aligned}
$$

where **Kalman gain** $K = P_{t|t-1}H^\top F_{t|t-1}^{-1}$, minimizing the mean square error (MSE) of the estimate.

We rewrite the updated state estimate and its covariance (related to Kalman gain) as follows.

$$
s_{t|t} = s_{t|t-1} + K(y_t - y_{t|t-1})\,, \tag{9}
$$

and

$$
P_{t|t} = P_{t|t-1}(I - KH)\,, \tag{10}
$$

Numerical Stability: The Joseph Form To guarantee $P_{t|t}$ symmetric, we utilize the Joseph form of covariance update.

$$
\begin{aligned}
s_{t|t} &= s_{t|t-1} + K(y_t - y_{t|t-1})\,, \\
&= s_{t|t-1} + K(Hs_t - Hs_{t|t-1} + v_t)\,, \\
&= s_{t|t-1} + KH(s_t - s_{t|t-1}) + Kv_t\,,
\end{aligned}
$$

Define the estimate error as $e_t = s_t - s_{t|t}$.

$$
\begin{aligned}
s_t - s_{t|t} &= s_t - s_{t|t-1} - KH(s_t - s_{t|t-1}) - Kv_t\,, \\
&= (I - KH)(s_t - s_{t|t-1}) - Kv_t\,,
\end{aligned}
$$

Updated covariance $P_{t|t} = \mathbb{E}[e_t e_t^\top]$.

$$
\begin{aligned}
P_{t|t} &= \mathbb{E}[((I-KH)(s_t - s_{t|t-1}) - Kv_t)((I-KH)(s_t - s_{t|t-1}) - Kv_t)^\top]\,, \\
&= (I-KH)\mathbb{E}[(s_t - s_{t|t-1})(s_t - s_{t|t-1})^\top](I-KH)^\top + K\mathbb{E}[v_t v_t^\top]K^\top\,, \\
&= (I-KH)P_{t|t-1}(I-KH)^\top + KRK^\top\,, \qquad (11)
\end{aligned}
$$

(11) is robust to rounding errors and is preferred for practical implementation.

3 Kalman Smoother

$$
\begin{pmatrix} s_t \\ s_{t+1} \end{pmatrix} \mid Y_t \sim \mathcal{N}\left(\begin{bmatrix} s_{t|t} \\ s_{t+1|t} \end{bmatrix}, \begin{bmatrix} P_{t|t} & P_{t|t}A^\top \\ AP_{t|t} & P_{t+1|t} \end{bmatrix} \right), \qquad (12)
$$

where the covariance is

$$
\begin{aligned}
\mathrm{cov}(s_t, s_{t+1}) &= \mathbb{E}[(s_t - s_{t|t})(s_{t+1} - s_{t+1|t})^\top]\,, \\
&= \mathbb{E}[(s_t - s_{t|t})(As_t - As_{t|t} + w_t)^\top]\,, \\
&= \mathbb{E}[(s_t - s_{t|t})(s_t - s_{t|t})^\top]A^\top\,, \\
&= P_{t|t}A^\top\,,
\end{aligned}
$$

Conditional distribution of s_t given s_{t+1}

$$
\mathbb{E}[s_t \mid s_{t+1}, Y_t] = s_{t|t} + \underbrace{P_{t|t}A^\top P_{t+1|t}^{-1}}_{J_t}(s_{t+1} - s_{t+1|t})\,, \qquad (13)
$$

$$
\begin{aligned}
\mathrm{cov}[s_t \mid s_{t+1}, Y_t] &= P_{t|t} - P_{t|t}A^\top P_{t+1|t}^{-1}AP_{t|t}\,, \\
&= P_{t|t} - \underbrace{P_{t|t}A^\top P_{t+1|t}^{-1}}_{J_t} P_{t+1|t} \underbrace{P_{t+1|t}^{-1}AP_{t|t}}_{J_t^\top}\,, \\
&= P_{t|t} - J_t P_{t+1|t} J_t^\top\,, \qquad (14)
\end{aligned}
$$

State update (incorporating the future data)

$$
\begin{aligned}
s_{t|T} &= \mathbb{E}[s_t \mid Y_T] = \mathbb{E}[\mathbb{E}(s_t \mid s_{t+1}, Y_t) \mid Y_T]\,, \\
&= \mathbb{E}[s_{t|t} + J_t(s_{t+1} - s_{t+1|t}) \mid Y_T]\,, \\
&= s_{t|t} + J_t(s_{t+1|T} - s_{t+1|t})\,, \qquad (15)
\end{aligned}
$$

Covariance update[1]

$$
\begin{aligned}
P_{t|T} &= \mathbb{E}[\mathrm{var}(s_t \mid s_{t+1}, Y_t)] + \mathrm{var}[\mathbb{E}(s_t \mid s_{t+1}, Y_t)]\,, \\
&= \mathbb{E}[P_{t|t} - J_t P_{t+1|t} J_t^\top \mid Y_T] + \mathrm{var}[s_{t|t} + J_t(s_{t+1} - s_{t+1|t}) \mid Y_T]\,, \\
&= P_{t|t} - J_t P_{t+1|t} J_t^\top + J_t P_{t+1|T} J_t^\top\,, \\
&= P_{t|t} + J_t \left(P_{t+1|T} - P_{t+1|t} \right) J_t^\top\,, \qquad (16)
\end{aligned}
$$

[1]The update of covariance is based on **the law of total variance**, which can be derived as follows. For a random vector $Y \in \mathbb{R}^n$,

4 Durbin and Koopman Smoother

In academic literature (e.g., Koopman 1993, Durbin & Koopman 2012), r_t is formally defined as the **score vector** of the future observations. Specifically, it is the derivative of the log-likelihood of the future observations conditional on the state s_{t+1}:

$$r_t := \frac{\partial \log p(y_{t+1}, \ldots, y_T \mid s_{t+1})}{\partial s_{t+1}} \,,$$

From a geometric perspective, smoothing can be viewed as an orthogonal projection in a **Hilbert space** of random variables. While the Kalman filter $s_{t|t}$ provides the projection of s_t onto the subspace spanned by $\{y_1, \ldots, y_t\}$, the adjoint variable r_t represents the **additional information** provided by the future subspaces $\{y_{t+1}, \ldots, y_T\}$ that is orthogonal to the past.

In the context of the Optimal Control, r_t is the costate. It measures the sensitivity of the total objective function (the log-likelihood) to a small perturbation in the state s_{t+1}. It essentially tells us :"If I nudge the state s_{t+1} by a tiny amount, how much does the total likelihood of seeing the future data change?"

Let $\ell = \log p(y_{1:T})$ be the joint log-likelihood of all observations. In a Gaussian system, the posterior mean $\hat{s}_{t|T}$ is the value that maximizes the conditional density. The log-likelihood can be split into two independent parts relative to the state s_t:

1. **the past**: $\log p(s_t \mid y_{1:t-1})$, which is the prior (predicted) density from the Kalman filter.
2. **the future**: $\log p(y_t, \ldots, y_T \mid s_t)$, which contains the current and future information.

Therefore:

$$\ell(s_t) = \log p(s_t \mid y_{1:t-1}) + \log p(y_{t:T} \mid s_t) \,,$$

Now, taking derivative with respect to s_t, start with the first term.

$$\frac{\partial}{\partial s_t}\left[-\frac{1}{2}(s_t - s_{t|t-1})^\top P_{t|t-1}^{-1}(s_t - s_{t|t-1})\right] = -P_{t|t-1}^{-1}(s_t - s_{t|t-1}) \,,$$

Now, the second term (the definition of r_{t-1})

$$r_{t-1} = \frac{\partial}{\partial s_t} \log p(y_{t:T} \mid s_t) \,,$$

Combine with the first term and obtain the first-order condition:

$$s_t - s_{t|t-1} = P_{t|t-1} r_{t-1} \,,$$

If we expand r_{t-1} using the chain rule:

$$\begin{aligned}
r_{t-1} &= \frac{\partial \log p(y_t \mid s_t)}{\partial s_t} + \frac{\partial \log p(y_{t+1:T} \mid s_t)}{\partial s_t} \,, \\
&= H^\top R^{-1}(y_t - Hs_t) + \left(\frac{\partial s_{t+1}}{\partial s_t}\right)^\top \frac{\partial}{\partial s_t} \log p(y_{t+1:T} \mid s_t) \,, \\
&= H^\top R^{-1}(y_t - Hs_t) + A^\top r_t \,,
\end{aligned}$$

---

the covariance matrix is defined as the expectation of the outer product of its centered values.

$$\begin{aligned}
\mathrm{var}[Y] &= \mathbb{E}\left[(Y - \mathbb{E}[Y])(Y - \mathbb{E}[Y])^\top\right] \\
&= \mathbb{E}\left[YY^\top - Y(\mathbb{E}[Y])^\top - \mathbb{E}[Y]Y^\top + \mathbb{E}[Y]\mathbb{E}[Y]^\top\right] \\
&= \mathbb{E}[YY^\top] - \mathbb{E}[Y]\mathbb{E}[Y]^\top
\end{aligned}$$

This definition extends naturally to the conditional case. For a random vector Y given an observation X, the conditional covariance is expressed as:

$$\mathrm{var}[Y \mid X] = \mathbb{E}[YY^\top \mid X] - \mathbb{E}[Y \mid X]\mathbb{E}[Y \mid X]^\top \,,$$

The Law of Total Variance decomposes the total uncertainty of a latent state into the mean of the conditional variance and the variance of the conditional mean:

$$\mathrm{var}[Y] = \mathbb{E}[\mathrm{var}(Y \mid X)] + \mathrm{var}(\mathbb{E}[Y \mid X]) \,,$$

For a linear Gaussian model, the smoothed estimates of the disturbances (w_t and v_t) are those that minimize the following quadratic cost function (which is the negative log-likelihood of the joint distribution). The **cost function** $\mathcal{J}$ is defined as

$$\mathcal{J} = \frac{1}{2} \sum_{t=1}^{T} \left(v_t^\top R^{-1} v_t + w_t^\top Q^{-1} w_t \right) , \tag{17}$$

It is subject to the constraints of the system equations

$$\begin{aligned} v_t &= y_t - H s_t \,, \\ w_t &= s_{t+1} - A s_t \,, \end{aligned}$$

To solve this constrained optimization, we introduce a vector of **Lagrangian Multipliers (Adjoint Variables)**, which we call r_t. We define the Lagrangian as:

$$\mathcal{L} = \frac{1}{2} \sum_{t=1}^{T} \left((y_t - H s_t)^\top R^{-1} (y_t - H s_t) + w_t^\top Q^{-1} w_t \right) + \sum_{t=1}^{T} r_t^\top \left(s_{t+1} - A s_t - w_t \right) ,$$

The first-order conditions are

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial s_t} &= -H^\top R^{-1} v_t - A^\top r_t + r_{t-1} = 0 \,, \\ \frac{\partial \mathcal{L}}{\partial w_t} &= Q^{-1} w_t - r_t = 0 \,, \end{aligned}$$

Rearrange and obtain

$$r_{t-1} = H^\top R^{-1} v_t + A^\top r_t \,, \tag{18}$$

$$r_t = Q^{-1} w_t \,, \tag{19}$$

We have the state variable s_t distributed as Gaussian, and its log-density function is

$$\log p(s_t \mid y_{1:t-1}) = -\frac{1}{2} (s_t - s_{t|t-1})^\top P_{t|t-1}^{-1} (s_t - s_{t|t-1}) + \text{const} \,,$$

We take the derivative (gradient) of this log-density with respect to s_t, and obtain:

$$\nabla_{s_t} \log p(s_t \mid y_{1:t-1}) = -P_{t|t-1}^{-1} (s_t - s_{t|t-1}) \,,$$

In Durbin and Koopman derivation, r_{t-1} is defined as the derivative of the **entire** log-likelihood of all observations $(y_1, \ldots, y_T)$ with respect to the predicted state $s_{t|t-1}$. Formally,

$$r_{t-1} = \nabla_{s_{t|t-1}} \log p(y_1, \ldots, y_T) \,,$$

Take expectation conditional on the all future observation, and obtain

$$s_{t|T} - s_{t|t-1} = P_{t|t-1} r_{t-1} \,,$$

We define the innovation at time t as $e_t = y_t - H s_{t|t-1}$, and by substituting the observation equation $y_t = H s_t + v_t$, we can rewrite the innovation in terms of the estimation error $(s_t - s_{t|t-1})$:

$$e_t = H \left(s_t - s_{t|t-1} \right) + v_t \,,$$

Rearrange for the measurement noise v_t:

$$v_t = e_t - H \left(s_t - s_{t|t-1} \right) ,$$

Multiply by R^{-1}:

$$R^{-1} v_t = R^{-1} e_t - R^{-1} H P_{t|t-1} r_{t-1} \,,$$

and $r_{t-1} = H^\top R^{-1} v_t + A^\top r_t$, then we have

$$\begin{aligned} R^{-1} v_t &= R^{-1} e_t - R^{-1} H P_{t|t-1} \left(H^\top R^{-1} v_t + A^\top r_t \right) , \\ \left(I + R^{-1} H P_{t|t-1} H^\top \right) R^{-1} v_t &= R^{-1} e_t - R^{-1} H P_{t|t-1} A^\top r_t \, , \end{aligned}$$

In the first term

$$\begin{aligned} \left(I + R^{-1} H P_{t|t-1} H^\top \right) &= R^{-1} \underbrace{\left(R + H P_{t|t-1} H^\top \right)}_{=F_{t|t-1}} , \\ \left(I + R^{-1} H P_{t|t-1} H^\top \right) R^{-1} &= \left(R^{-1} F_{t|t-1} \right)^{-1} R^{-1} , \\ &= F_{t|t-1}^{-1} R R^{-1} = F_{t|t-1}^{-1} , \end{aligned}$$

In the second term

$$\begin{aligned} \left(I + R^{-1} H P_{t|t-1} H^\top \right) R^{-1} &= F_{t|t-1}^{-1} , \\ \left(I + R^{-1} H P_{t|t-1} H^\top \right) R^{-1} H P_{t|t-1} &= F_{t|t-1}^{-1} H P_{t|t-1} = \left(P_{t|t-1} H^\top F_{t|t-1}^{-1} \right)^\top = K^\top , \end{aligned}$$

Combined, we have

$$R^{-1} v_t = F_{t|t-1}^{-1} e_t - K^\top A^\top r_t \, , \tag{20}$$

Substitute into (18) and obtain

$$\begin{aligned} r_{t-1} &= H^\top \left(F_{t|t-1}^{-1} e_t - K^\top A^\top r_t \right) + A^\top r_t \, , \\ &= H^\top F_{t|t-1}^{-1} e_t + \left(A^\top - H^\top K^\top A^\top \right) r_t \, , \\ &= H^\top F_{t|t-1}^{-1} e_t + \left(\underbrace{A - AKH}_{=L} \right)^\top r_t \, , \\ &= H^\top F_{t|t-1} e_t + L^\top r_t \end{aligned}$$

The **Durbin-Koopman Simulation Smoother Algorithm** is sketched as:

1. **Simulate synthetic data**: generate random noise $w_t^+ \sim \mathcal{N}(0, Q)$ and $v_t^+ \sim \mathcal{N}(0, R)$. Use these to create a "fake" state sequence s_t^+ and observations y_t^+.
2. **Compute residuals**: define a "diff" observation: $y_t^* = y_t - y_t^+$.
3. **Run Kalman filter**: run the filter on y_t^* to obtain the innovations, $e_t^*, F_{t|t-1}$ and K.
4. **Run r_t recursion**: use the recursive form on e_t^* to get the sequence $r_0, \ldots, r_T$.
5. **Construct the sample**:
   - The simulated sample of the state is $\tilde{s}_t = s_t^+ + P_{t|t-1} r_{t-1}$.
   - $\tilde{w}_t = w_t^+ + Q A^\top r_t$.