Summary

This note is a self-contained derivation of the forward (a.k.a. causal) Kalman Filter. The derivation is based on the key observation that, Kalman Filter recursively calculates the posterior distribution for a *Linear-Gaussian* state-space model.

1 Matrix Inversion Lemma

Lemma 1: For square and invertible $\boldsymbol{X}, \boldsymbol{Y}$ and conformable $\boldsymbol{U}, \boldsymbol{V}$:

$$(\boldsymbol{X} + \boldsymbol{U}\boldsymbol{Y}\boldsymbol{V})^{-1} = \boldsymbol{X}^{-1} - \boldsymbol{X}^{-1}\boldsymbol{U}[\boldsymbol{Y}^{-1} + \boldsymbol{V}\boldsymbol{X}^{-1}\boldsymbol{U}]^{-1}\boldsymbol{V}\boldsymbol{X}^{-1} \ , \tag{1}$$

if the inverse exists.

Proof: by multiplying the left-hand side and right-hand side.

$$\begin{aligned}
&[(\boldsymbol{X} + \boldsymbol{U}\boldsymbol{Y}\boldsymbol{V})]\left[\boldsymbol{X}^{-1} - \boldsymbol{X}^{-1}\boldsymbol{U}[\boldsymbol{Y}^{-1} + \boldsymbol{V}\boldsymbol{X}^{-1}\boldsymbol{U}]^{-1}\boldsymbol{V}\boldsymbol{X}^{-1}\right] \\
&\quad = \boldsymbol{X}\boldsymbol{X}^{-1} - \boldsymbol{X}\boldsymbol{X}^{-1}\boldsymbol{U}[\boldsymbol{Y}^{-1} + \boldsymbol{V}\boldsymbol{X}^{-1}\boldsymbol{U}]^{-1}\boldsymbol{V}\boldsymbol{X}^{-1} + \boldsymbol{U}\boldsymbol{Y}\boldsymbol{V}\boldsymbol{X}^{-1} \\
&\qquad\quad - \boldsymbol{U}\boldsymbol{Y}\boldsymbol{V}\boldsymbol{X}^{-1}\boldsymbol{U}[\boldsymbol{Y}^{-1} + \boldsymbol{V}\boldsymbol{X}^{-1}\boldsymbol{U}]^{-1}\boldsymbol{V}\boldsymbol{X}^{-1} \\
&\quad = \boldsymbol{I} + \boldsymbol{U}\boldsymbol{Y}\boldsymbol{V}\boldsymbol{X}^{-1} \\
&\qquad\quad - [\boldsymbol{U} + \boldsymbol{U}\boldsymbol{Y}\boldsymbol{V}\boldsymbol{X}^{-1}\boldsymbol{U}][\boldsymbol{Y}^{-1} + \boldsymbol{V}\boldsymbol{X}^{-1}\boldsymbol{U}]^{-1}\boldsymbol{V}\boldsymbol{X}^{-1} \\
&\quad = \boldsymbol{I} + \boldsymbol{U}\boldsymbol{Y}\boldsymbol{V}\boldsymbol{X}^{-1} \\
&\qquad\quad - \boldsymbol{U}\boldsymbol{Y}[\boldsymbol{Y}^{-1} + \boldsymbol{V}\boldsymbol{X}^{-1}\boldsymbol{U}][\boldsymbol{Y}^{-1} + \boldsymbol{V}\boldsymbol{X}^{-1}\boldsymbol{U}]^{-1}\boldsymbol{V}\boldsymbol{X}^{-1} \\
&\quad = \boldsymbol{I} + \boldsymbol{U}\boldsymbol{Y}\boldsymbol{V}\boldsymbol{X}^{-1} - \boldsymbol{U}\boldsymbol{Y}\boldsymbol{V}\boldsymbol{X}^{-1} = \boldsymbol{I}
\end{aligned} \tag{2}$$

This completes the proof. □

As will be seen in the next section, Matrix Inversion Lemma is of key importance for manipulating the algebraic expressions of posterior distribution parameters, from which the familiar form of Kalman Filter equations follows.

2 Forward Kalman Filter

The Linear-Gaussian state-space model has the form

$$\begin{aligned}
\boldsymbol{x}_t &= \boldsymbol{F}_t\boldsymbol{x}_{t-1} + \boldsymbol{w}_t \ , \ \boldsymbol{w}_t \sim N(\boldsymbol{0}, \boldsymbol{Q}_t) \\
\boldsymbol{y}_t &= \boldsymbol{H}_t\boldsymbol{x}_t + \boldsymbol{v}_t \ , \ \boldsymbol{v}_t \sim N(\boldsymbol{0}, \boldsymbol{R}_t)
\end{aligned} \tag{3}$$

where the observation $\boldsymbol{y}_t$ is linear in terms of the hidden state $\boldsymbol{x}_t$. The Kalman filter studied here computes the posterior recursively for a finite time horizon $1, \dots, T$. The key to the recursive computation is the *Markov* property of the stochastic process for the hidden state $\boldsymbol{x}_t$. That is, the previous state $\boldsymbol{x}_{t-1}$ completely characterizes $\boldsymbol{x}_t$. This is put into use as follows: At time t it is of interest to calculate the posterior $p(\boldsymbol{x}_{1:t}|\boldsymbol{y}_{1:t})$ where the shorthand $1:t$ denotes all the observations up to time t. Applying Bayes rule

$$
\begin{aligned}
p(\boldsymbol{x}_t|\boldsymbol{y}_{1:t}) &= \frac{p(\boldsymbol{y}_{1:t}|\boldsymbol{x}_{1:t})\ p(\boldsymbol{x}_{1:t})}{p(\boldsymbol{y}_{1:t})} && (4)\\
&= p(\boldsymbol{y}_t|\boldsymbol{x}_t)\ p(\boldsymbol{x}_t|\boldsymbol{x}_{1:t-1})\ \frac{p(\boldsymbol{y}_{1:t-1}|\boldsymbol{x}_{1:t-1})p(\boldsymbol{x}_{1:t-1})}{p(\boldsymbol{y}_{1:t})}\ . && (5)
\end{aligned}
$$

It is easy to overlook the transition from Eq. (4) to Eq. (5). However note that this recursive split would not have been possible had not the stochastic process model of $\boldsymbol{x}_t$ been Markov. The last fractional term is independent of $\boldsymbol{x}_t$, so as far as posterior estimation is concerned, it is irrelevant. The posterior of interest is proportional to two terms: the prior distribution $p(\boldsymbol{x}_t|\boldsymbol{x}_{1:t-1})$ and the likelihood $p(\boldsymbol{y}_t|\boldsymbol{x}_t)$.

Using linear transformation property of a random variable, it is easy to see that the prior distribution is of form

$$
\begin{aligned}
p(\boldsymbol{x}_t|\boldsymbol{x}_{1:t-1}) &= N(\boldsymbol{x}_{t|t-1}, \boldsymbol{P}_{t|t-1})\\
\boldsymbol{x}_{t|t-1} &= \boldsymbol{F}_t \boldsymbol{x}_{t-1|t-1}\\
\boldsymbol{P}_{t|t-1} &= \boldsymbol{F}_t \boldsymbol{P}_{t-1|t-1} \boldsymbol{F}_t^\top + \boldsymbol{Q}_t\ . && (6)
\end{aligned}
$$

Here $(\boldsymbol{x}_{t|t-1}, \boldsymbol{P}_{t|t-1})$ parametrize the prior distribution of $\boldsymbol{x}$ at time t. Similarly $(\boldsymbol{x}_{t-1|t-1}, \boldsymbol{P}_{t-1|t-1})$ parametrize the posterior distribution at time $t-1$. This notation is choosen to be consistent with the literature. Next, the posterior distribution follows Eq. (5)

$$
\begin{aligned}
p(\boldsymbol{x}_t|\boldsymbol{y}_t) &= p(\boldsymbol{y}_t|\boldsymbol{x}_t) \qquad\qquad \times\ p(\boldsymbol{x}_t|\boldsymbol{x}_{1:t-1}) \qquad\qquad \times\ C\\
&= \exp\left\{-\frac{1}{2}(\boldsymbol{y}_t - \boldsymbol{H}_t\boldsymbol{x}_t)^\top \boldsymbol{R}_t^{-1}(\boldsymbol{y}_t - \boldsymbol{H}_t\boldsymbol{x}_t)\right\} \times \exp\left\{-\frac{1}{2}(\boldsymbol{x}_t - \boldsymbol{x}_{t|t-1})^\top \boldsymbol{R}_t^{-1}(\boldsymbol{x}_t - \boldsymbol{x}_{t|t-1})\right\} \times C && (7)
\end{aligned}
$$

where the constant terms are absorbed into C. The expressions can be completed to square from which it follows that

$$
\begin{aligned}
p(\boldsymbol{x}_t|\boldsymbol{y}_t) &= N(\boldsymbol{x}_{t|t}, \boldsymbol{P}_{t|t})\\
\boldsymbol{x}_{t|t} &= \boldsymbol{P}_{t|t}\ [\boldsymbol{P}_{t|t-1}^{-1}\boldsymbol{x}_{t|t-1} + \boldsymbol{H}_t^\top \boldsymbol{R}_t^{-1}\boldsymbol{y}_t]\\
\boldsymbol{P}_{t|t} &= [\boldsymbol{P}_{t|t-1}^{-1} + \boldsymbol{H}_t^\top \boldsymbol{R}_t^{-1}\boldsymbol{H}_t]^{-1}\ . && (8)
\end{aligned}
$$

Now, while Eqs. (6) and (8) completely describe the Forward Kalman Filter, they are not in the traditional Kalman gain/innovation form. Upcoming is the derivation of the standard equations.

We now apply Matrix Inversion Lemma to the expression of $\boldsymbol{P}_{t|t}$

$$\begin{aligned}
\boldsymbol{P}_{t|t} &= [\boldsymbol{P}_{t|t-1}^{-1} + \boldsymbol{H}_t^\top \boldsymbol{R}_t^{-1} \boldsymbol{H}_t]^{-1} \\
&= \boldsymbol{P}_{t|t-1} - \boldsymbol{P}_{t|t-1}\boldsymbol{H}_t^\top [\boldsymbol{R}_t + \boldsymbol{H}_t \boldsymbol{P}_{t|t-1} \boldsymbol{H}_t^\top]^{-1} \boldsymbol{H}_t \boldsymbol{P}_{t|t-1} \\
&= \boldsymbol{P}_{t|t-1} - [\boldsymbol{P}_{t|t-1}\boldsymbol{H}_t^\top \boldsymbol{S}_t^{-1}]\ \boldsymbol{S}_t\ [\boldsymbol{S}_t^{-1}\boldsymbol{H}_t \boldsymbol{P}_{t|t-1}] \\
&= \boldsymbol{P}_{t|t-1} - \boldsymbol{K}_t \boldsymbol{S}_t \boldsymbol{K}_t^\top
\end{aligned} \tag{9}$$

where two new quantities are defined. $\boldsymbol{S}_t$ is known as the measurement covariance, as it is the covariance of $\boldsymbol{y}_t$ given the prior distribution of $\boldsymbol{x}_t$. $\boldsymbol{K}_t$ is known as the Kalman gain, a standard quantity in Kalman Filter derivation. To summarize

$$\boldsymbol{S}_t = [\boldsymbol{R}_t + \boldsymbol{H}_t \boldsymbol{P}_{t|t-1} \boldsymbol{H}_t^\top]^{-1} \quad , \quad \boldsymbol{K}_t = \boldsymbol{P}_{t|t-1}\boldsymbol{H}_t^\top \boldsymbol{S}_t^{-1} \ . \tag{10}$$

Obtaining the expression for the mean is tricky. The first thing to note is the equivalence

$$\boldsymbol{P}_{t|t} = \boldsymbol{P}_{t|t-1} - \boldsymbol{K}_t \boldsymbol{S}_t \boldsymbol{K}_t^\top = \boldsymbol{P}_{t|t-1} - \boldsymbol{K}_t \boldsymbol{H}_t \boldsymbol{P}_{t|t-1} \tag{11}$$

and using the latter expression in Eq. (8)

$$\begin{aligned}
\boldsymbol{x}_{t|t} &= [\boldsymbol{P}_{t|t-1} - \boldsymbol{K}_t \boldsymbol{H}_t \boldsymbol{P}_{t|t-1}] \times [\boldsymbol{P}_{t|t-1}\boldsymbol{x}_{t|t-1} + \boldsymbol{H}_t \boldsymbol{R}_t^{-1} \boldsymbol{y}_t] \\
&= \boldsymbol{x}_{t|t-1} - \boldsymbol{K}_t \boldsymbol{H}_t \boldsymbol{x}_{t|t-1} + (\boldsymbol{I} - \boldsymbol{K}_t \boldsymbol{H}_t)\boldsymbol{P}_{t|t-1}\boldsymbol{H}_t^\top \boldsymbol{R}_t^{-1} \boldsymbol{y}_t \\
&= \boldsymbol{x}_{t|t-1} - \boldsymbol{K}_t \hat{\boldsymbol{y}}_t + \boldsymbol{K}_t \boldsymbol{y}_t \\
&= \boldsymbol{x}_{t|t-1} + \boldsymbol{K}_t(\boldsymbol{y}_t - \hat{\boldsymbol{y}}_t)
\end{aligned} \tag{12}$$

where $\hat{\boldsymbol{y}}_t = \boldsymbol{H}_t \boldsymbol{x}_{t|t-1}$ is the measurement prediction, and $\boldsymbol{y}_t - \hat{\boldsymbol{y}}_t$ is called the innovation. All that remains is to verify

$$(\boldsymbol{I} - \boldsymbol{K}_t \boldsymbol{H}_t)\boldsymbol{P}_{t|t-1}\boldsymbol{H}_t^\top \boldsymbol{R}_t^{-1} = \boldsymbol{K}_t \ , \tag{13}$$

which is done as follows

$$\begin{aligned}
& (\boldsymbol{I} - \boldsymbol{K}_t \boldsymbol{H}_t)\boldsymbol{P}_{t|t-1}\boldsymbol{H}_t^\top \boldsymbol{R}_t^{-1} && = && \boldsymbol{K}_t \\
(\iff)\ & \boldsymbol{P}_{t|t-1}\boldsymbol{H}_t^\top \boldsymbol{R}_t^{-1} && = && \boldsymbol{K}_t \boldsymbol{H}_t \boldsymbol{P}_{t|t-1} \boldsymbol{H}_t^\top \boldsymbol{R}_t^{-1} + \boldsymbol{K}_t \\
(\iff)\ & \boldsymbol{P}_{t|t-1}\boldsymbol{H}_t^\top \boldsymbol{R}_t^{-1} && = && \boldsymbol{K}_t[\boldsymbol{H}_t \boldsymbol{P}_{t|t-1} \boldsymbol{H}_t^\top \boldsymbol{R}_t^{-1} + \boldsymbol{I}] \\
(\iff)\ & \boldsymbol{P}_{t|t-1}\boldsymbol{H}_t^\top \boldsymbol{R}_t^{-1} && = && \boldsymbol{K}_t \boldsymbol{S}_t \boldsymbol{R}_t^{-1} \\
(\iff)\ & \boldsymbol{P}_{t|t-1}\boldsymbol{H}_t^\top \boldsymbol{R}_t^{-1} && = && \boldsymbol{P}_{t|t-1}\boldsymbol{H}_t^\top \boldsymbol{S}_t^{-1} \boldsymbol{S}_t \boldsymbol{R}_t^{-1} \ .
\end{aligned}$$

This finishes the derivation of Kalman filter. The equations are summarized in the next page.

3 Forward Kalman Filter: Summary

3.1 State Space Model

$$\begin{aligned} \boldsymbol{x}_t &= \boldsymbol{F}_t \boldsymbol{x}_{t-1} + \boldsymbol{w}_t \ , \ \boldsymbol{w}_t \sim N(\boldsymbol{0}, \boldsymbol{Q}_t) \\ \boldsymbol{y}_t &= \boldsymbol{H}_t \boldsymbol{x}_t + \boldsymbol{v}_t \ , \ \boldsymbol{v}_t \sim N(\boldsymbol{0}, \boldsymbol{R}_t) \end{aligned}$$

3.2 Prior Computation

$$\begin{aligned} p(\boldsymbol{x}_t | \boldsymbol{x}_{1:t-1}) &= N(\boldsymbol{x}_{t|t-1}, \boldsymbol{P}_{t|t-1}) \\ \boldsymbol{x}_{t|t-1} &= \boldsymbol{F}_t \boldsymbol{x}_{t-1|t-1} \\ \boldsymbol{P}_{t|t-1} &= \boldsymbol{F}_t \boldsymbol{P}_{t-1|t-1} \boldsymbol{F}_t^\top + \boldsymbol{Q}_t \end{aligned}$$

3.3 Posterior Computation

$$\begin{aligned} p(\boldsymbol{x}_t | \boldsymbol{y}_t) &= N(\boldsymbol{x}_{t|t}, \boldsymbol{P}_{t|t}) \\ \boldsymbol{x}_{t|t} &= \boldsymbol{x}_{t|t-1} + \boldsymbol{K}_t (\boldsymbol{y}_t - \hat{\boldsymbol{y}}_t) \\ \boldsymbol{P}_{t|t} &= \boldsymbol{P}_{t|t-1} - \boldsymbol{K}_t \boldsymbol{S}_t \boldsymbol{K}_t^\top \\ \boldsymbol{S}_t &= \boldsymbol{H}_t \boldsymbol{P}_{t|t-1} \boldsymbol{H}_t^\top + \boldsymbol{R}_t \\ \boldsymbol{K}_t &= \boldsymbol{P}_{t|t-1} \boldsymbol{H}_t^\top \boldsymbol{S}_t^{-1} \\ \hat{\boldsymbol{y}}_t &= \boldsymbol{H}_t \boldsymbol{x}_{t|t-1} \end{aligned}$$