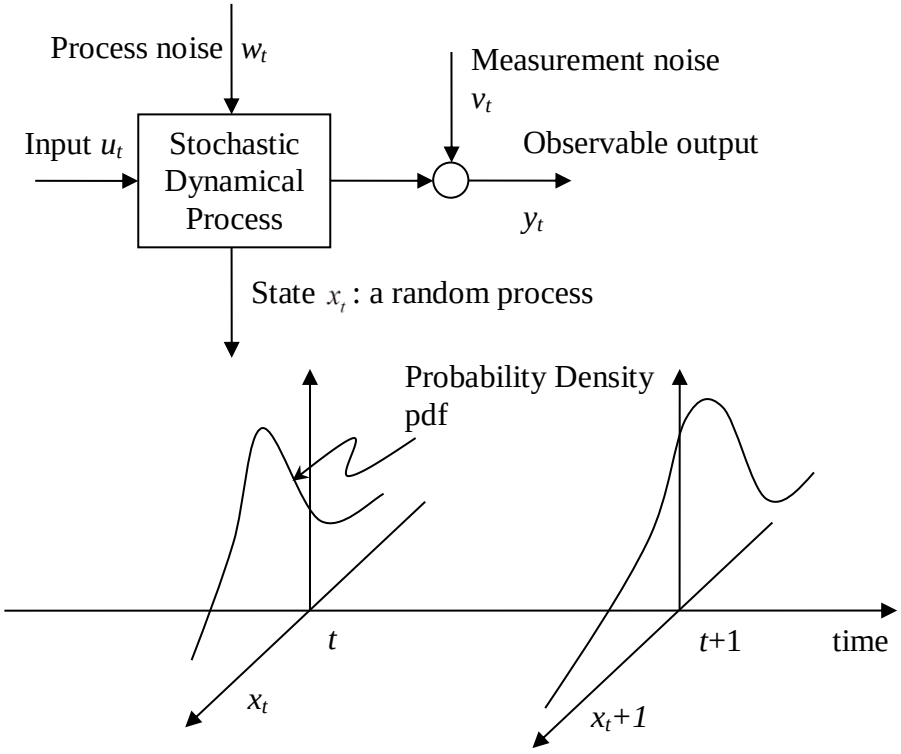
2.160 IDENTIFICATION, ESTIMATION, AND LEARNING

LECTURE NOTES NO. 8

# *8. Bayesian Filter and Gaussian Kalman Filter*

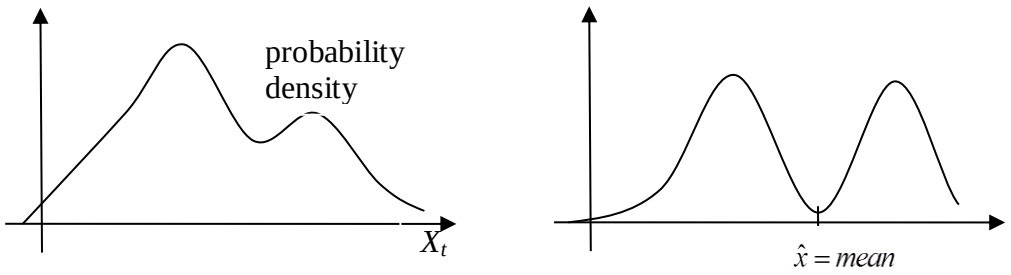
In this chapter we will look at some fundamentals of stochastic estimation. We will begin with a brief introduction to Markov Process and Chapman-Kolmogorov Equation, in which we will extend our goal from estimation of a single value, like state or parameter value estimation, to estimation of the whole distribution (pdf) of random variables. Then we will introduce Bayesian Filter, which is based on Bayes' Rule and the Chapman-Kolmogorov equation. Finally, we will revisit Kalman Filter and prove that the Kalman Filter with Gaussian noise distribution is the optimal filter among all the linear and nonlinear filters. This proof is based on the Bayesian Filter; in other words, Gaussian Kalman Filter is a special case of Bayesian Filter.

## 8.1 Estimation of Distribution/Density




Our objective of state estimation thus far is to determine a single value, $\hat{x}_t$, from the random variable having some distribution. This section addresses how to estimate the whole distribution, rather than a single value.

A single value, e.g. mean, is sometimes a poor representation. See the bi-modal case below. The mean is least likely.




Representation in terms of:

- Pdf, pmf
- Parametric, e.g. σ, μ, $P_t$
- Non-parametric, e.g. samples, particles.

## 8.2 Bayes' Rule

Consider a random variable *X* and its observation *Y*. Suppose we know the conditional probability:

$$p(y|x): X = x \rightarrow Y = y \tag{1}$$

Then, can we estimate *x* by observing $Y = y$

$$p(x|y): Y = y \rightarrow X = x \text{ (Infer } x \text{ from } y\text{)}$$

Recall joint probability and Bayes Rule:

$$\begin{aligned} p(x,y) &= p(x|y)\,p(y) = p(y|x)\,p(x) \\ \therefore\ p(x|y) &= \frac{p(y|x)\,p(x)}{p(y)} \end{aligned} \tag{2}$$

Remarks:

- *p(y)* does not affect the estimation of *x*; it is merely a scaling factor.
- For *p*(*x*|*y*)to be a pdf, it must integrate to 1.
- Since $\int_{-\infty}^{\infty} p(y|x)\,p(x)\,dx = p(y)$, dividing *p*(*y*|*x*)*p*(*x*) by *p*(*y*) makes *p*(*x*|*y*) a pdf.

Replacing $\eta \triangleq \frac{1}{p(y)}$, a scaling factor, we obtain

$$p(x|y) = \eta\, p(y|x)\,p(x) \tag{3}$$

Important terminology:

- $p(x|y) \rightarrow$ posterior probability
- $p(x) \rightarrow$ prior probability
- $y \rightarrow$ data
- $p(y|x) \rightarrow$ generative model

## 8.3 Markov Process and Recursive Bayes Filter

Our interest is to estimate a random process governed by a state transition equation:

$$x_t = \mathbf{f}(x_{t-1}, u_{t-1}) + w_{t-1} \tag{4}$$

where $w_{t-1}$ is uncorrelated process noise with pdf $f_W(w_t)$, and an observation equation

$$y_t = \mathbf{h}(x_t) + v_t \tag{5}$$

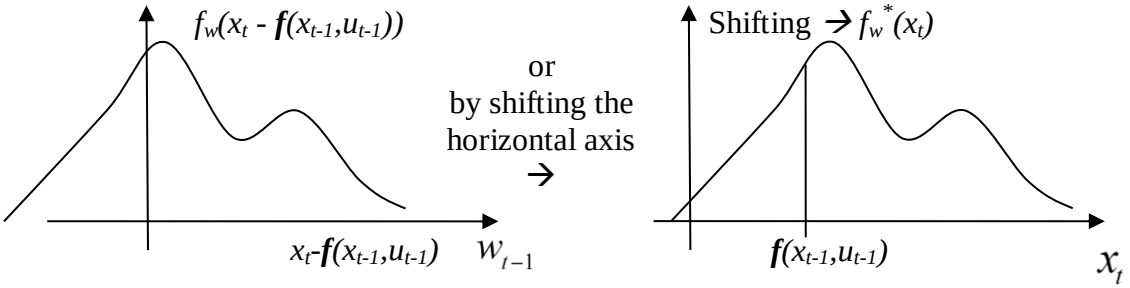
where $v_t$ is uncorrelated measurement noise with pdf $f_v(v_t)$.

A random process is called a Markov process if the probability of $X_t = x_t$ depends on $x_{t-1}$ and $u_{t-1}$ alone, and not on past states and input:

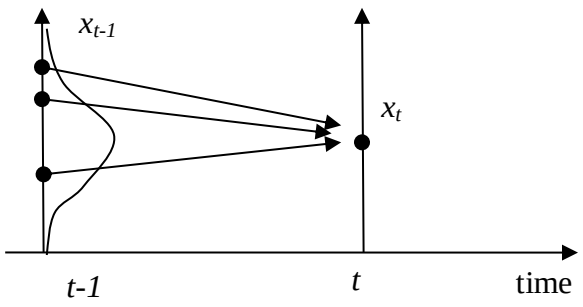
$$\Pr(x_t | x_0, x_1, \cdots, x_{t-1}, u_0, \cdots, u_{t-1}) = \Pr(x_t | x_{t-1}, u_{t-1}) \tag{6}$$

This is the case for the above state equation. Given $x_{t-1}$ and $u_{t-1}$, the randomness of $x_t$ comes only from $w_{t-1}$. Therefore, replacing $w_{t-1}$ by $x_t$ - $\mathbf{f}(x_{t-1}, u_{t-1})$, we obtain

$$\Pr(x_t | x_{t-1}, u_{t-1}) = f_W(x_t - \mathbf{f}(x_{t-1}, u_{t-1})) \tag{7}$$




Let $g_{t|t-1}(x_t)$ be the probability density of $x_t$ propagated through the state equation. State $x_t$ can be reached from various states in one time step earlier, $x_{t-1}$, which has probability density $g_{t-1}(x_{t-1})$.




Therefore, (recall $p(x) = \int_{-\infty}^{\infty} p(x|y)\,p(y)\,dy$ )

$$g_{t|t-1}(x_t) = \int_{-\infty}^{\infty} f_W\left(x_t - \mathbf{f}\left(x_{t-1}, u_{t-1}\right)\right) g_{t-1}(x_{t-1})\,dx_{t-1} \tag{8}$$

Given the (posterior) density $g_{t-1}(x_{t-1})$ at time $t$-1 and input $u_{t-1}$, the probability density of $x_t$ propagated through the state equation can be computed with the above equation. . . . state propagation.
This is a type of **Chapman-Kolmogorov Equation**.

Next, we update the probability density $g_{t|t-1}(x_t)$ by assimilating data $y_t$.
. . . state update/correction.

Recall Bayes Rule

$$p(x|y) = \eta\, p(y|x)\, p(x) \tag{9}$$

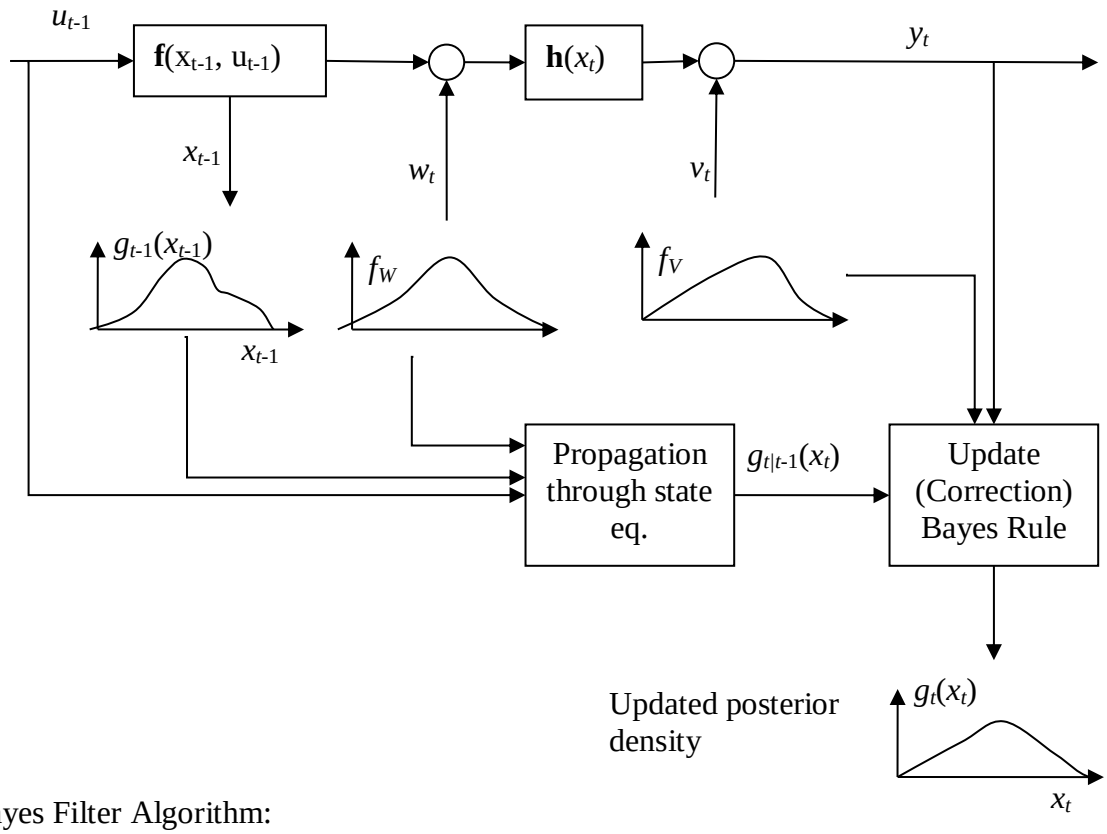
Where $y$ is new data, $p(x)$ is $g_{t|t-1}(x_t)$, $p(x|y)$ corresponds to $g_t(x_t)$, the posterior density after assimilating $y_t$, and $p(y|x)$ is the generative model obtained from equation (5):

$$p(y|x) = f_V\left(y_t - \mathbf{h}(x_t)\right) \tag{10}$$

Therefore,

$$g_t(x_t) = \eta\, f_V\left(y_t - \mathbf{h}(x_t)\right) g_{t|t-1}(x_t) \tag{11}$$

This recursive Bayes Algorithm is called Bayes Filter. See the block diagram below.




Bayes Filter Algorithm:

Given $g_{t-1}(x_{t-1})$, $u_{t-1}$, and $y_t$

Compute:

$$g_{t|t-1}(x_t) = \int_{-\infty}^{\infty} f_W\left(x_t - \mathbf{f}\left(x_{t-1}, u_{t-1}\right)\right) g_{t-1}\left(x_{t-1}\right) dx_{t-1}$$
$$g_t(x_t) = \eta f_V\left(y_t - \mathbf{h}\left(x_t\right)\right) g_{t|t-1}\left(x_t\right) \tag{12}$$

Return $g_t(x_t)$.

## 8.4 Gaussian Kalman Filter

The Kalman Filter with Gaussian noise can be derived from the Bayes Filter. Namely, Gaussian Kalman Filter is a special case of Bayes Filter. Consider a linear, time-varying system,

$$x_{t+1} = A_t x_t + B_t u_t + w_t$$
$$y_t = H_t x_t + v_t \tag{13}$$

where $w_t$ and $v_t$ are zero-mean, uncorrelated process and measurement noise with Gaussian densities:

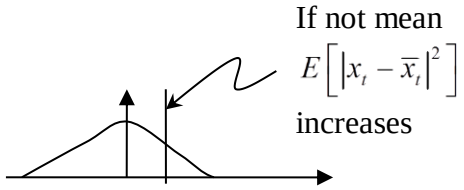
$$f_W(w_t) = \frac{1}{\sqrt{\det(2\pi Q_t)}} \exp\left(-\frac{1}{2} w_t^T Q_t^{-1} w_t\right)$$
$$f_V(v_t) = \frac{1}{\sqrt{\det(2\pi R_t)}} \exp\left(-\frac{1}{2} v_t^T R_t^{-1} v_t\right) \quad (14)$$

The problem is to find the optimal estimate $\hat{x}_t$:

$$\hat{x}_t = \arg\min_{\bar{x}_t} E\left[|x_t - \bar{x}_t|^2 | u_0, \cdots, u_{t-1}, y_1, \cdots, y_t\right] \quad (15)$$

The solution is the conditional <u>mean</u>.

$$\hat{x}_t = E\left[x_t | u_0, \cdots, u_{t-1}, y_1, \cdots, y_t\right] \quad (16)$$




which comes with

$$\hat{x}_t = \hat{x}_{t|t-1} + P_t H_t^T R_t^{-1}\left(y_t - H_t \hat{x}_{t|t-1}\right),$$
$$\text{Kalman Gain } K_t = P_t H_t^T R_t^{-1} \quad (17)$$

This is the same linear update law as the one we obtained previously, but it is the optimal among linear and non-linear filters.

Initial Conditions: $X_0$ is a Gaussian random variable with mean $x_0$ and covariance $P_0>0$ positive definite. You can skip the first step of the following proof, which is rather technical. Step 2 shows how the linear recursive update law, (17), is obtained.

**<u>Proof</u>:**

**Step 1.** Show that the propagated probability density $g_{t|t-1}(x_t)$ is Gaussian.

$$g_{t|t-1}(x_t) = \int_{-\infty}^{\infty} f_W\left(x_t - A_{t-1}x_{t-1} - B_{t-1}u_{t-1}\right) g_{t-1}(x_{t-1})\, dx_{t-1} \quad (18)$$

We use Induction: Assuming that $g_{t-1}(x_t)$ is Gaussian with mean $\hat{x}_{t-1}$ and covariance $P_{t-1}$, show that $g_{t|t-1}(x_t)$ is also Gaussian.

Since both $f_w()$ and $g_{t-1}(x_t)$ are Gaussian, we can combine their exp() terms together

$$g_{t|t-1}(x_t) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{\det(2\pi Q_{t-1})\det(2\pi P_{t-1})}} \exp\left(-L(x_t, x_{t-1})\right) dx_{t-1} \quad (19)$$

where $L$ is a quadratic function given by

$$L(x_t, x_{t-1}) = \frac{1}{2}(x_t - A_{t-1}x_{t-1} - B_{t-1}u_{t-1})^T Q_{t-1}^{-1}(x_t - A_{t-1}x_{t-1} - B_{t-1}u_{t-1}) + \frac{1}{2}(x_{t-1} - \hat{x}_{t-1})^T P_{t-1}^{-1}(x_{t-1} - \hat{x}_{t-1}) \quad (20)$$

Let $P_{t|t\text{-}1}$ and $\bar{x}_{t-1}$ be defined as

$$\begin{aligned} P_{t|t-1} &\triangleq \left(A_{t-1}^T Q_{t-1}^{-1} A_{t-1} + P_{t-1}^{-1}\right)^{-1} \\ \bar{x}_{t-1} &\triangleq P_{t|t-1}\left[A_{t-1}^T Q_{t-1}^{-1}(x_t - B_{t-1}u_{t-1}) + P_{t-1}^{-1}\hat{x}_{t-1}\right] \end{aligned} \quad (21)$$

Construct another quadratic function,

$$M(x_t, x_{t-1}) = \frac{1}{2}(x_{t-1} - \bar{x}_{t-1})^T P_{t|t-1}^{-1}(x_{t-1} - \bar{x}_{t-1}) \quad (22)$$

We can show that $L_0 = L(x_t, x_{t\text{-}1}) - M(x_t, x_{t\text{-}1})$ is independent of $x_{t\text{-}1}$; therefore $L_0$ can be factored out from the integral.

$$g_{t|t-1}(x_t) = \frac{1}{\sqrt{\det()}} \exp\left(-L_0(x_t)\right) \int_{-\infty}^{\infty} \exp\left(-M(x_t, x_{t-1})\right) dx_{t-1} \quad (23)$$

Since $M(x_t, x_{t\text{-}1})$ is a quadratic function, it forms another Gaussian distribution, which integrates to a constant.

$$\begin{aligned} &\frac{1}{\sqrt{\det\left(2\pi P_{t|t-1}\right)}} \int_{-\infty}^{\infty} \exp\left(\left(x_{t-1} - \bar{x}_{t-1}\right)^T P_{t|t-1}^{-1}\left(x_{t-1} - \bar{x}_{t-1}\right)\right) dx_{t-1} = 1 \\ &\rightarrow \int_{-\infty}^{\infty} \exp\left(-M(x_t, x_{t-1})\right) dx_{t-1} = \sqrt{\det\left(2\pi P_{t|t-1}\right)}: \text{ constant} \end{aligned} \quad (24)$$

Using this in (23) yields

$$\therefore g_{t|t-1}(x_t) = \eta \exp\left(-L_0(x_t)\right), \eta = \sqrt{\frac{\det\left(P_{t|t-1}\right)}{\det\left(2\pi Q_{t-1} P_{t-1}\right)}} \quad (25)$$

where

$$L_0 = L - M = \frac{1}{2}\left[x_t - \left(A_{t-1}\hat{x}_{t-1} + B_{t-1}u_{t-1}\right)\right]^T \left(A_{t-1}P_{t-1}A_{t-1}^T + Q_{t-1}\right)^{-1} \left[x_t - \left(A_{t-1}\hat{x}_{t-1} + B_{t-1}u_{t-1}\right)\right] \quad (26)$$

is a quadratic function of $x_t$.

$$\begin{aligned} &\therefore g_{t|t-1}(x_t) \text{ is Gaussian} \\ &\text{with mean } A_{t-1}\hat{x}_{t-1} + B_{t-1}u_{t-1} = \hat{x}_{t|t-1} \\ &\text{and covariance } P_{t|t-1} \triangleq \left(A_{t-1}P_{t-1}A_{t-1}^T + Q_{t-1}\right) \end{aligned}$$

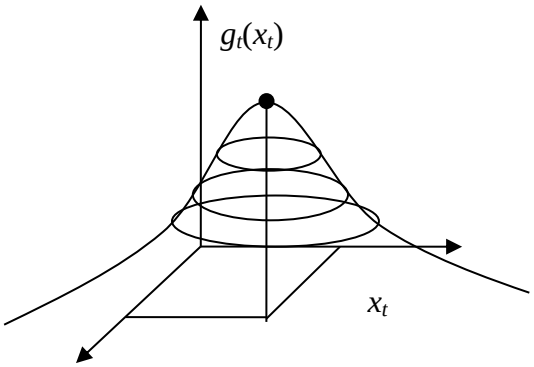
**Step 2.** State update by assimilating new measurement $y_t$.

Recall

$$
\begin{aligned}
& g_t(x_t) = \eta\, p\left(y_t \mid x_t\right) g_{t|t-1}(x_t) \\
& \eta = \text{ scaling factor} \\
& g_{t|t-1}(x_t) \text{ Gaussian with mean } \hat{x}_{t|t-1}, \text{ covariance } P_{t|t-1} \\
& p\left(y_t \mid x_t\right) = f_V\left(y_t - H_t x_t\right) \text{ Gaussian with mean 0, covariance } R_t \\
& g_t(x_t) = \eta \exp\left(-N(x_t)\right) \\
& N(x_t) = \frac{1}{2}\left(y_t - H_t x_t\right)^T R_t^{-1}\left(y_t - H_t x_t\right) + \frac{1}{2}\left(x_t - \hat{x}_{t|t-1}\right)^T P_{t|t-1}^{-1}\left(x_t - \hat{x}_{t|t-1}\right)
\end{aligned}
\tag{27}
$$

$N(x_t)$ is a quadratic function of $x_t$. The optimal estimate of $\hat{x}_t$ is the mean:

$$
\hat{x}_t = E\left[x_t \mid u_0, \cdots, u_{t-1}, y_t, \cdots, y_t\right] = \int_{-\infty}^{\infty} x_t g_t(x_t) dx_t \tag{28}
$$




At the mean $\hat{x}_t = E\left[x_t \mid u, y\right]$,

$$
\begin{aligned}
& \frac{dg_t(x_t)}{dx_t} = 0 \\
& \frac{dg_t}{dx_t} = \eta \exp\left(-N(x_t)\right) \frac{d}{dx_t}\left(-N(x_t)\right) = 0 \\
& \rightarrow \frac{dN(x_t)}{dx_t} = 0 \\
& \rightarrow -H_t^T R_t^{-1}\left(y_t - H_t x_t\right) + P_{t|t-1}^{-1}\left(x_t - \hat{x}_{t|t-1}\right) = 0
\end{aligned}
\tag{29}
$$

Replacing $x_t$ satisfying this condition by $\hat{x}_t$ yields:

$$
\begin{aligned}
P_{t|t-1}^{-1}\left(\hat{x}_t - \hat{x}_{t|t-1}\right) &= H_t^T R_t^{-1}\left(y_t - H_t \hat{x}_{t|t-1} + H_t \hat{x}_{t|t-1} - H_t \hat{x}_t\right) \\
&= H_t^T R_t^{-1}\left(y_t - H_t \hat{x}_{t|t-1}\right) + H_t^T R_t^{-1} H_t\left(\hat{x}_{t|t-1} - \hat{x}_t\right) \\
\therefore \left[P_{t|t-1}^{-1} + H_t^T R_t^{-1} H_t\right]\left(\hat{x}_t - \hat{x}_{t|t-1}\right) &= H_t^T R_t^{-1}\left(y_t - H_t \hat{x}_{t|t-1}\right)
\end{aligned}
\tag{30}
$$

Recall the covariance update law of Discrete Kalman Filter, eq.(5-41). We can find that

$$P_t^{-1} = P_{t|t-1}^{-1} + H_t^T R_t^{-1} H_t \tag{31}$$

Therefore,

$$\begin{aligned} &\therefore \hat{x}_t = \hat{x}_{t|t-1} + P_t H_t^T R_t^{-1} \left( y_t - H_t \hat{x}_{t|t-1} \right) \\ &\left( K_t = P_t H_t^T R_t^{-1} \right) \text{ Kalman Gain} \end{aligned} \tag{32}$$

We have arrived at the familiar linear filter. In finding this optimal state estimate, we never assumed that the filter is linear. This implies that the Gaussian Kalman filter is the optimal among linear and nonlinear filters. Remember that we assumed the linear update law and obtained the optimal gain $K_t$ previously. Now we assumed Gaussian distributions, but did not assume the linear form. It is a consequence.