# 6. The Kalman Filter

We derive the Kalman Filter (KF) as the Bayesian state estimator that keeps track of the PDF of the system state conditioned on all past measurements, $f(x(k)|z(1{:}k))$, for linear time-varying systems with Gaussian process and measurement noise. The KF algorithm follows the same formalism as the Bayesian Tracking algorithm (lecture #4); namely, it consists of a prior update step (based on the process model) and a measurement update step (based on the measurement model). In contrast to the Bayesian Tracking algorithm, the state space is now continuous and no longer finite.

Exploring the problem structure (linear system and Gaussian distributions) allows us to derive the KF as the analytic solution to the Bayesian state estimation problem. The resulting recursive equations are straightforward matrix calculations.

## 6.1 Model

We consider the linear, time-varying system:

$$x(k) = A(k{-}1)x(k{-}1) + B(k{-}1)u(k{-}1) + v(k{-}1) \tag{6.1}$$

$$z(k) = H(k)x(k) + w(k) \tag{6.2}$$

$x(k)$: state
$u(k)$: known control input
$v(k)$: process noise
$z(k)$: measurement
$w(k)$: sensor noise

for $k = 1, 2, \ldots$. The CRVs $x(0)$, $\{v(\cdot)\}$, and $\{w(\cdot)\}$ are mutually independent, Gaussian distributed:

- $x(0) \sim \mathcal{N}(x_0, P_0)$, i.e. $x(0)$ has a Gaussian distribution with mean $x_0$ and variance $P_0$,
- $v(k) \sim \mathcal{N}(0, Q(k))$, $w(k) \sim \mathcal{N}(0, R(k))$.

## 6.2 Gaussian Random Variable (GRV)

The PDF of a Gaussian distributed (also called normally distributed) vector CRV $y = (y^1, \ldots, y^D)$ is given by

$$f(y) = \frac{1}{(2\pi)^{D/2}\,\det(\Sigma)^{1/2}} \exp\left(-\frac{1}{2}(y-\mu)^T\Sigma^{-1}(y-\mu)\right),$$

where

$\mu \in \mathbb{R}^D$ is the mean vector,
$\Sigma \in \mathbb{R}^{D\times D}$ is the variance, a symmetric ($\Sigma = \Sigma^T$), positive definite ($\Sigma > 0$) matrix, and
$\det(\Sigma)$ denotes the determinant of $\Sigma$.

We often use shorthand notation and write $y \sim \mathcal{N}(\mu, \Sigma)$. Notice that a GRV is fully characterized by its mean and variance.

---

Date compiled: April 9, 2013

**Special case**

Consider the case where $\Sigma$ is a diagonal matrix:

$$\Sigma = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_D^2 \end{bmatrix}.$$

Then,

$$f(y) = \prod_{i=1,\ldots,D} \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{(y_i-\mu_i)^2}{2\sigma_i^2}\right).$$

That is, the PDF simplifies to the product of $D$ scalar GRVs. Note that the variables are independent[1] if and only if $\Sigma$ is diagonal.

**Jointly Gaussian Random Variables**

Two random variables $x$ and $y$, both of which may be vectors, are said to be jointly Gaussian if the joint vector random variable $(x, y)$ is a GRV.

Notice that, if $x$ and $y$ are independent GRVs, then this implies that they are jointly Gaussian: Let $x \sim \mathcal{N}(\mu_x, \Sigma_x)$ and $y \sim \mathcal{N}(\mu_y, \Sigma_y)$. Then,

$$\begin{aligned} f(x,y) = f(x)f(y) &\propto \exp\left(-\frac{1}{2}\Big((x-\mu_x)^T\Sigma_x^{-1}(x-\mu_x) + (y-\mu_y)^T\Sigma_y^{-1}(y-\mu_y)\Big)\right) \\ &= \exp\left(-\frac{1}{2}\begin{bmatrix}(x-\mu_x)^T & (y-\mu_y)^T\end{bmatrix}\begin{bmatrix}\Sigma_x^{-1} & 0 \\ 0 & \Sigma_y^{-1}\end{bmatrix}\begin{bmatrix}x-\mu_x \\ y-\mu_y\end{bmatrix}\right). \end{aligned}$$

### 6.2.1 Importance of the Gaussian property

The assumption of Gaussian random variables will allow us to derive the Kalman Filter as the analytic solution to the Bayesian state estimation problem. In particular, the following properties will be useful.

**Property 1: An (affine) linear transformation of a GRV is a GRV**

Let $y$ be a GRV, and let $x$ be defined by $x = My + b$ with $M$ a constant matrix and $b$ a constant vector of appropriate dimensions. Then, $x$ is a GRV.

We will show this for scalar random variables.

- Let $y \sim \mathcal{N}(\mu, \sigma^2)$. That is,

  $$f(y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right).$$

  Let $x = my + b$, with $m \neq 0$ and $b$ constants. What is $f_x(x)$?
- Using change of variables, we get

  $$\begin{aligned} f_x(x) = \frac{f_y(y)}{\left|\frac{dx}{dy}\right|} &= \frac{1}{|m|\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}\left(\frac{x-b}{m} - \mu\right)^2\right) \\ &= \frac{1}{\sqrt{2\pi m^2\sigma^2}} \exp\left(-\frac{1}{2m^2\sigma^2}(x-b-m\mu)^2\right). \end{aligned}$$

  That is, $x \sim \mathcal{N}(b + m\mu, m^2\sigma^2)$.

[1] Spatially independent, do not confuse with temporally independent. For a vector random variable $y(k)$, spatial independence means that all elements at a fixed time $k$ (i.e. $y^1(k), \ldots, y^D(k)$) are independent; whereas temporal independence refers to $y(1), \ldots, y(k)$ being independent.

Notice that the derivation above is *not* the same as showing that the mean is $b + m\mu$ and the variance is $m^2\sigma^2$, which is easy to do, and applies to non-Gaussian PDFs as well:

$$\begin{aligned}
\mathrm{E}[x] &= m\mathrm{E}[y] + b = m\mu + b \\
\mathrm{Var}[x] &= \mathrm{E}[(my + b - (m\mu + b))^2] = \mathrm{E}[m^2(y-\mu)^2] = m^2\mathrm{E}[(y - \mathrm{E}[y])^2] = m^2\sigma^2.
\end{aligned}$$

**Property 2: A linear combination of two jointly GRVs is a GRV**

Let $x$ and $y$ be two jointly Gaussian random variables, and let $z$ be defined by $z = M_1x + M_2y$, where $M_1$ and $M_2$ are constant matrices of appropriate dimensions. Then, $z$ is a GRV.

This property follows directly from Property 1 above:

$$z = M\xi \qquad \text{with } M = [M_1\, M_2] \text{ and } \xi = (x, y) \text{ a GRV.}$$

Hence, $z \sim \mathcal{N}(\mu_z, \Sigma_z)$. As before, we can calculate $\mu_z$ and $\Sigma_z$ directly. Let $x \sim \mathcal{N}(\mu_x, \Sigma_x)$ and $y \sim \mathcal{N}(\mu_y, \Sigma_y)$, and we assume that $x$ and $y$ are independent (for $x$ and $y$ jointly Gaussian, but not necessarily independent, see (**PSET 4: P1**)). Then,

$$\begin{aligned}
\mu_z &= \mathrm{E}[z] = M_1\mathrm{E}[x] + M_2\mathrm{E}[y] = M_1\mu_x + M_2\mu_y \\
\Sigma_z &= \mathrm{E}[(z-\mu_z)(z-\mu_z)^T] \\
&= \mathrm{E}[(M_1x - M_1\mu_x + M_2y - M_2\mu_y)(M_1x - M_1\mu_x + M_2y - M_2\mu_y)^T] \\
&= M_1\mathrm{E}[(x-\mu_x)(x-\mu_x)^T]M_1^T + M_2\mathrm{E}[(y-\mu_y)(y-\mu_y)^T]M_2^T \\
&\quad + M_1 \underbrace{\mathrm{E}[(x-\mu_x)(y-\mu_y)^T]}_{=0 \text{ (by independence)}} M_2^T + M_2 \underbrace{\mathrm{E}[(y-\mu_y)(x-\mu_x)^T]}_{=0 \text{ (by independence)}} M_1^T \\
&= M_1\Sigma_x M_1^T + M_2\Sigma_y M_2^T.
\end{aligned}$$

## 6.3 Problem Formulation and Auxiliary Variables

We want to calculate

$$f(x(k)|z(1{:}k)).$$

We have already done this for Bayesian Tracking: we satisfy all the assumptions on model structure and noise independence. The only difference is that we are now working with CRVs instead of DRVs, which can readily be accommodated for by replacing sums with integrals. In particular, we can immediately write down the solution:

**Step 1 (S1):** Prior update (yields a priori state estimate).

$$f(x(k)|z(1{:}k{-}1)) = \int \overbrace{f(x(k)|x(k{-}1))}^{\text{process model}} \; \overbrace{f(x(k{-}1)|z(1{:}k{-}1))}^{\text{previous iteration (S2)}} \; dx(k{-}1) \tag{6.3}$$

**Step 2 (S2):** Measurement update (yields a posteriori state estimate).

$$f(x(k)|z(1{:}k)) = \frac{\overbrace{f(z(k)|x(k))}^{\text{measurement model}} \; \overbrace{f(x(k)|z(1{:}k{-}1))}^{\text{prior (S1)}}}{\underbrace{\int f(z(k)|x(k))\, f(x(k)|z(1{:}k{-}1))\; dx(k)}_{\text{normalization } = f(z(k)|z(1:k-1))}} \tag{6.4}$$

So what is left to do? Exploit the structure:

1. Linearity
2. Gaussian random variables

in order to convert the problem to pure matrix manipulations. (All our problem data consists of matrices, so this is a reasonable goal.)

### 6.3.1 Auxiliary variables

So far in the notes, we have mostly used simplified notation where we do not explicitly distinguish between a random variable and the value that it takes. In this section, however, we will use explicit notation in order to avoid confusion. In particular, we use $z(k)$ to denote the random variable according to (6.2) and $\mathbf{z}(k)$ to denote the value that it takes; that is, the actual measurement at time $k$.

We define new random variables to simplify the development: $x_p(k)$, $x_m(k)$, and $z_m(k)$, where the subscript $p$ denotes "prediction" or "prior update", and subscript $m$ denotes "measurement" or "measurement update":

$$
\begin{array}{ll}
\textbf{Init:} & x_m(0) := x(0) \\[6pt]
\textbf{S1:} & x_p(k) := A(k{-}1)x_m(k{-}1) + B(k{-}1)u(k{-}1) + v(k{-}1) \\
\textbf{S2:} & z_m(k) := H(k)x_p(k) + w(k) \\
 & x_m(k) \text{ defined via its PDF} \\
 & \quad f_{x_m(k)}(\xi) := f_{x_p(k)|z_m(k)}(\xi|\mathbf{z}(k)) \quad \forall \xi
\end{array}
\left.\vphantom{\begin{array}{l} a\\ a\\ a\\ a \end{array}}\right\} \quad k = 1, 2, \dots
$$

We now claim that: for all $\xi$ and $k = 1, 2, \dots,$

$$f_{x_p(k)}(\xi) = f_{x(k)|z(1:k-1)}(\xi|\mathbf{z}(1:k{-}1)) \tag{6.5}$$

$$f_{x_m(k)}(\xi) = f_{x(k)|z(1:k)}(\xi|\mathbf{z}(1:k))\,. \tag{6.6}$$

That is, $x_p(k)$ is the random variable $x(k)$ conditioned on $z(1:k{-}1)$, and $x_m(k)$ is the random variable $x(k)$ conditioned on $z(1:k)$.

*Proof:*
We will prove this by induction. The statement (6.6) is true for $k = 0$ by definition of $x_m(0)$. Assume now that (6.6) is true for $k - 1$, then prove (6.5) and (6.6) are true for $k$.

**S1:** By the total probability theorem:

$$f_{x_p(k)}(\xi) = \int f_{x_p(k)|x_m(k-1)}(\xi|\lambda)\, f_{x_m(k-1)}(\lambda)\, d\lambda \quad \forall \xi.$$

We want to show that $f_{x_p(k)}$ is equal to $f_{x(k)|z(1:k-1)}$ in (6.3), which can be rewritten in explicit notation as

$$f_{x(k)|z(1:k-1)}(\xi|\mathbf{z}(1:k{-}1)) = \int f_{x(k)|x(k-1)}(\xi|\lambda)\; f_{x(k-1)|z(1:k-1)}(\lambda|\mathbf{z}(1:k{-}1))\; d\lambda.$$

First note that $f_{x_m(k-1)}(\lambda) = f_{x(k-1)|z(1:k-1)}(\lambda|\mathbf{z}(1:k{-}1))$ for all $\lambda$ by induction assumption. Second, note that, for all $\xi$,

$$
\begin{aligned}
f_{x_p(k)|x_m(k-1)}(\xi|\lambda) &= f_{v(k-1)}(\xi - A(k{-}1)\lambda - B(k{-}1)u(k{-}1)) \\
f_{x(k)|x(k-1)}(\xi|\lambda) &= f_{v(k-1)}(\xi - A(k{-}1)\lambda - B(k{-}1)u(k{-}1))\,.
\end{aligned}
$$

Hence, the PDFs $f_{x_p(k)}$ and $f_{x(k)|z(1:k-1)}$ are identical, as required.

**S2:** By Bayes' rule

$$f_{x_p(k)|z_m(k)}(\xi|\mathbf{z}(k)) = \frac{f_{z_m(k)|x_p(k)}(\mathbf{z}(k)|\xi)\; f_{x_p(k)}(\xi)}{\displaystyle\int f_{z_m(k)|x_p(k)}(\mathbf{z}(k)|\lambda)\; f_{x_p(k)}(\lambda)\; d\lambda} \quad \forall \xi.$$

We want to show that $f_{x_p(k)|z_m(k)}$ (and hence $f_{x_m(k)}$) is equal to $f_{x(k)|z(1:k)}$ in (6.4), which is rewritten in explicit notation,

$$f_{x(k)|z(1:k)}(\xi|\mathbf{z}(1:k)) = \frac{f_{z(k)|x(k)}(\mathbf{z}(k)|\xi)\; f_{x(k)|z(1:k-1)}(\xi|\mathbf{z}(1:k{-}1))}{\displaystyle\int f_{z(k)|x(k)}(\mathbf{z}(k)|\lambda)\; f_{x(k)|z(1:k-1)}(\lambda|\mathbf{z}(1:k{-}1))\; d\lambda}.$$

First note that $f_{x_p(k)}(\xi) = f_{x(k)|z(1:k-1)}(\xi|\mathbf{z}(1:k-1))$ for all $\xi$ by the proof in **S1** above. Second, for all $\xi$,

$$\begin{aligned} f_{z_m(k)|x_p(k)}(\mathbf{z}(k)|\xi) &= f_{w(k)}(\mathbf{z}(k) - H(k)\xi) \\ f_{z(k)|x(k)}(\mathbf{z}(k)|\xi) &= f_{w(k)}(\mathbf{z}(k) - H(k)\xi)\,. \end{aligned}$$

Therefore the PDFs $f_{x(k)|z(1:k)}$ and $f_{x_p(k)|z_m(k)}$ (and hence $f_{x_m(k)}$) are the same. This completes the proof.

What comes next: efficiently calculating the mean and variance of $x_p(k)$ and $x_m(k)$. If we can additionally show that $x_p(k)$ and $x_m(k)$ are GRVs, we have a full characterization of their PDFs, completing our task.