6. The Kalman Filter: Preliminaries

6.1 Model

$$x(k) = A(k)x(k-1) + B(k)u(k) + v(k) \tag{6.1}$$
$$z(k) = H(k)x(k) + w(k) \tag{6.2}$$

$x(k)$: state

$u(k)$: known control input

$v(k)$: process noise

$z(k)$: measurement

$w(k)$: sensor noise

- $\{x(0), v(1), \ldots, v(k), w(1), \ldots, w(k)\}$ are independent random variables.
- $x(0) \sim \mathcal{N}(x_0, P_0)$ has a normal (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 The Gaussian random variable (GRV)

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

$\mu \in \mathbb{R}^D$ is the mean vector.

$\Sigma \in \mathbb{R}^{D \times D}$ is a symmetric, positive definite matrix, the variance.

$|\Sigma|$ is the determinant of Σ.

Special case

Consider the case where Σ is a diagonal matrix:

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

$$f(y) = \prod_i \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{1}{2\sigma_i^2}(y_i - \mu_i)^2\right)$$

The PDF simplifies to the product of D scalar GRVs. Note that the variables are independent[1] if and only if Σ is diagonal.

Date compiled: April 16, 2012

[1] Spatially independent, do not confuse with temporally independent.

6.3 Importance of the Gaussian assumption

6.3.1 A linear function of a GRV is a GRV

- We will show this for scalar random variables. Let $y \sim \mathcal{N}(\mu, \sigma^2)$. That is,

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

Let $x = my + b$, with m and b constants. What is $f_x(x)$?

$$f_x(x) = \frac{f_y(y)}{\left|\frac{dx}{dy}\right|}$$

$$\therefore f_x(x) = \frac{1}{|m|\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}\left(\frac{x-b}{m}-\mu\right)^2\right)$$
$$= \frac{1}{\sqrt{2\pi m^2\sigma^2}} \exp\left(-\frac{1}{2m^2\sigma^2}(x-b-m\mu)^2\right)$$

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

- This 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:

$$\mathrm{E}[x] = m\mathrm{E}[y] + b = m\mu + b$$

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

6.3.2 The sum of two GRVs is a GRV

- We will show this for independent, scalar GRVs. Let $x \sim \mathcal{N}(\mu_x, \sigma_x^2)$, $y \sim \mathcal{N}(\mu_y, \sigma_y^2)$, and let $z = x + y$. Then:

$$f_z(z) = \int_{-\infty}^{\infty} f_x(x)\, f_y(z-x)\, dx \text{ (\textbf{PSET 1: P11})}$$

$$f_x(x)\, f_y(z-x) \propto \exp\left[-\frac{1}{2}\left(\frac{1}{\sigma_x^2}(x-\mu_x)^2 + \frac{1}{\sigma_y^2}(z-x-\mu_y)^2\right)\right]$$

- Consider the term in round brackets. It is a multinomial with the following terms: x^2, z^2, xz, x, z, and 1. We claim that we can rewrite it as $a\,(x-(b+cz))^2 + d\,(z-e)^2 + f$, 6 constants for 6 coefficients (**PSET 4: P1**):

  $a:$ used to match x^2 coefficient, $a > 0$ (easy to see)

  $b:$ used to match x coefficient

  $c:$ used to match xz coefficient

  $d:$ used to match z^2 coefficient, $d > 0$ (only tricky part)

  $e:$ used to match z coefficient

  $f:$ used to match 1 coefficient

  So

$$f_x(x)\, f_y(z-x) \propto \exp\left(-\frac{a}{2}\,(x-(b+cz))^2\right) \exp\left(-\frac{d}{2}\,(z-e)^2\right)$$

Note that $\int_{-\infty}^{\infty} \exp\left(-\frac{a}{2}(x-(b+cz))^2\right) dx$ is independent of z (**PSET 4: P2**). Therefore

$$f_z(z) \propto \exp\left(-\frac{d}{2}(z-e)^2\right), \quad \text{a GRV with } \mu_z = e,\, \sigma_z = \frac{1}{\sqrt{d}}$$

As before, we can calculate μ_z and σ_z directly:

$$\mu_z = \mu_x + \mu_y, \quad \sigma_z^2 = \sigma_x^2 + \sigma_y^2 \quad (\textbf{PSET 4: P3})$$

Summary

1. A linear function of a GRV is a GRV. This applies to vector valued variables as well.

2. The sum of two GRVs is a GRV, which also applies to vector valued, not necessarily independent, variables.

6.4 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) A priori state estimate.

$$f(x(k)|z(1{:}k{-}1)) = \int \overbrace{f(x(k)|x(k{-}1))}^{\text{process model}} \; \overbrace{f(x(k{-}1)|z(1{:}k{-}1))}^{\text{recursion, step 2}} \; dx(k{-}1) \tag{6.3}$$

Step 2 (S2) A posteriori state estimate.

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

So what is left to do? Exploit the structure:

1. Linearity

2. Gaussian random variables

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

6.4.1 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 development, $x_p(k)$, $x_m(k)$, and $z_m(k)$, where subscript p denotes "prediction" or "process update" and subscript m denotes "measurement" or "measurement update":

$$x_m(0) := x(0)$$

$$\left.\begin{array}{ll} \textbf{S1:} & x_p(k) := A(k)x_m(k\text{–}1) + B(k)u(k) + v(k) \\ \textbf{S2:} & z_m(k) := H(k)x_p(k) + w(k) \\ & x_m(k) \text{ defined via its PDF} \\ & \quad f_{x_m(k)}(\xi) := f_{x_p(k)|z_m(k)}(\xi|\mathbf{z}(k)) \quad \forall \xi \end{array}\right\} \quad k = 1, 2, \ldots$$

We now claim that: For all ξ and $k = 1, 2, \ldots$,

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

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

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

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

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

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

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

S2: By Bayes rule

$$f_{x_p(k)|z_m(k)}(\xi|\mathbf{z}(k)) = \frac{f_{z_m(k)|x_p(k)}(\mathbf{z}(k)|\xi)\ f_{x_p(k)}(\xi)}{\int f_{z_m(k)|x_p(k)}(\mathbf{z}(k)|\xi)\ f_{x_p(k)}(\xi)\ d\xi}.$$

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,

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

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

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

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, completing 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.