Sensor Fusion as Weighted Averaging

Frank Dellaert
Center for Robotics and Intelligent Machines
Georgia Institute of Technology

1 Common Sense Information Fusion

When we have two (scalar) measurements of equal weight, we can just average them to form a point estimate of the (scalar) unknown x:

$$\hat{x} = \frac{z_1 + z_2}{2}$$

When we trust z_1 more, we could give it more weight and form a weighted average, with $w_1 > w_2$:

$$\hat{x} = \frac{w_1 z_1 + w_2 z_2}{w_1 + w_2} \tag{1.1}$$

2 Bayes Law

Can we know more about x than just a point estimate? A principled way to do sensor fusion is to use Bayes law:

$$P(x|z_1, z_2) \propto P(x)L(x; z_1)L(x; z_2)$$

where $P(x)$ is a prior density over x, and $L(x; z)$ is the **likelihood of x given the measurement z.** The likelihood is defined as

$$L(x; z) \propto P(x|z)$$

i.e., proportional to the conditional density $P(x|z)$ that tells us how probable it is to receive a measurement z given a value x. However, when not x but z is given, we call the resulting function of x a likelihood function.

3 Discrete Case

If x is **discrete**, i.e., $x \in \{1..K\}$, then both $P(x)$ and $L(x;z)$ are vectors, and we can just multiply them point-wise and normalize to obtain the posterior:

$$P(x=k|z_1,z_2) = \frac{P(x=k)L(x=k;z_1)L(x=k;z_2)}{\sum_{k=1}^{K} P(x=k)L(x=k;z_1)L(x=k;z_2)}$$

Note it does not matter if z is discrete or continuous: we assume the likelihood vectors $L(x;z)$ can be computed given any value of z.

4 Scalar Case

If x is **scalar**, then things are not so simple in general: both the prior $P(x)$ and the likelihood functions $L(x;z)$ can be arbitrarily complex functions, so how do we even represent those?

4.1 Gaussian Measurement Noise

If both x and the measurements z are scalar, a very common situation is when the measurement z is simply a corrupted version of the unknown value x,

$$z = x + n$$

where n is some additive noise with some density $P(n)$. A convenient modeling choice is to assume the noise is Gaussian, i.e., drawn from a normal distribution, which means the conditional density $P(z|x)$ is also a Gaussian density on z:

$$P(z|x) = k \exp \frac{1}{2} \left(\frac{x-z}{\sigma} \right)^2$$

When z is given, the corresponding likelihood function is

$$L(x;z) = \exp \frac{1}{2} \left(\frac{x-z}{\sigma} \right)^2$$

By a happy coincidence, *this is also a Gaussian*, but please note that this is a very special case: in general, when the measurement is not simply a linear function of x, we will *not* be so lucky. Note above we also dropped the constant k: a likelihood function just has to be proportional to $P(z|x)$.

4.2 Fusing two Gaussian Measurements

Now we are in a position to apply Bayes law:

$$\begin{aligned} P(x|z_1, z_2) &\propto P(x)L(x; z_1)L(x; z_2) \\ &= 1 \times \exp \frac{1}{2} \left(\frac{x - z_1}{\sigma_1} \right)^2 \times \exp \frac{1}{2} \left(\frac{x - z_2}{\sigma_2} \right)^2 \end{aligned}$$

where we assumed the prior $P(x)$ was non-informative. As it happens, multiplying two Gaussians will yield a Gaussian again. This is true because a Gaussian is just the exponential of a quadratic, and adding two quadratics is easily seen to be a quadratic again. But what is its mean? To find that, take the negative log and minimize:

$$\begin{aligned} \hat{x} &= \arg\max_x P(x|z_1, z_2) \\ &= \arg\min_x - \log P(x|z_1, z_2) \\ &= \arg\min_x \left\{ \frac{1}{2} \left(\frac{x - z_1}{\sigma_1} \right)^2 + \frac{1}{2} \left(\frac{x - z_2}{\sigma_2} \right)^2 \right\} \end{aligned}$$

Taking the derivative of the argument and setting to zero we obtain

$$\frac{\hat{x} - z_1}{\sigma_1^2} + \frac{\hat{x} - z_2}{\sigma_2^2} = 0$$

or

$$\hat{x} = \frac{w_1 z_1 + w_2 z_2}{w_1 + w_2} \tag{4.1}$$

with $w_1 = 1/\sigma_1^2$ and $w_2 = 1/\sigma_2^2$. The weights here indicate the **information** each measurement carries, and it is equal to the inverse variance. The information of a measurement is equal to the inverse of its variance.

4.3 The shape of the Posterior

Information is additive: the information of the fused estimate $\hat{x}$ is equal to $w_1 + w_2$, and hence the variance of the resulting Gaussian $P(x|z_1, z_2)$ (with mean $\hat{x}$) is

$$\sigma^2 = \frac{1}{w_1 + w_2}$$

5 Multivariate Case

In the n-dimensional **multivariate** case, this generalizes to the information matrix, defined as the inverse of the covariance matrix R:

$$\Lambda \triangleq R^{-1}$$

and Equation 4.1 generalizes to

$$\hat{x} = \frac{\Lambda_1 z_1 + \Lambda_2 z_2}{\Lambda_1 + \Lambda_2} = (\Lambda_1 + \Lambda_2)^{-1} (\Lambda_1 z_1 + \Lambda_2 z_2)$$

Here again, information is additive**:** the information matrix of the fused estimate $\hat{x}$ is equal to $\Lambda_1 + \Lambda_2$, and hence the covariance matrix of the resulting multivariate Gaussian is

$$R = (\Lambda_1 + \Lambda_2)^{-1}$$

6 Kalman Filter

The general Bayes filter is

$$\begin{aligned} P(x_t|Z^t) &\propto P(z_t|x_t)P(x_t|Z^{t-1}) \\ &= P(z_t|x_t)\int_{x_{t-1}} P(x_t|x_{t-1})P(x_{t-1}|Z^{t-1}) \end{aligned}$$

In a Kalman filter, we assume that the likelihood, the motion model, and the posterior are all Gaussian. If that is the case, the predictive density will also be Gaussian. We see this by realizing that

$$P(x_t, x_{t-1}|Z^{t-1}) = P(x_t|x_{t-1})P(x_{t-1}|Z^{t-1})$$

The negative log is, for the scalar case

$$\begin{aligned} -\log P(x_t, x_{t-1}|Z^{t-1}) &= \frac{1}{2\sigma_u^2}(x_t - x_{t-1} - u_{t-1})^2 + \frac{1}{2\sigma_{t-1}^2}(x_{t-1} - \mu_{t-1})^2 \\ &= 0.5w_u(x_t - x_{t-1} - u_{t-1})^2 + 0.5w_{t-1}(x_{t-1} - \mu_{t-1})^2 \end{aligned}$$

Hence, the mean of the joint $P(x_t, x_{t-1}|Z^{t-1})$ is simply $(x_{t-1}, x_t) = (\mu_{t-1}, \mu_{t-1} + u_{t-1})$, and the corresponding information matrix is the curvature of this 2D Gaussian, obtained by taking the second order derivatives and arranging into a matrix:

$$\Lambda = \begin{bmatrix} w_u + w_{t-1} & -w_u \\ -w_u & w_u \end{bmatrix}$$

However, we are interested in predictive density $P(x_t|Z^{t-1})$ which is the marginal of the joint:

$$P(x_t|Z^{t-1}) = \int_{x_{t-1}} P(x_t, x_{t-1}|Z^{t-1})$$

Taking a marginal is really easy if you have the covariance matrix, as we can just take the sub-matrix. The covariance matrix can be calculated analytically and is

$$\Sigma = \Lambda^{-1} = \begin{bmatrix} 1/w_{t-1} & 1/w_{t-1} \\ 1/w_{t-1} & 1/w_{t-1} + 1/w_u \end{bmatrix} = \begin{bmatrix} \sigma^2_{t-1} & \sigma^2_{t-1} \\ \sigma^2_{t-1} & \sigma^2_{t-1} + \sigma^2_u \end{bmatrix}$$

and hence we have the very satisfying result that, for the predictive step, the variances add up. In other words, the Kalman filter, in the scalar case, is simply iterating the following two steps, starting from a prior density $\mathcal{N}(x_0; \mu_0, \sigma_0^2)$:

1. Prediction:

$$P(x_t|Z^{t-1}) = \mathcal{N}(x_t; \mu_{t|t-1}, \sigma^2_{t|t-1})$$
$$\mu_{t|t-1} = \mu_{t-1} + u$$
$$\sigma^2_{t|t-1} = \sigma^2_{t-1} + \sigma^2_u$$

2. Update

$$P(x_t|Z^t) = \mathcal{N}(x_t; \mu_t, \sigma_t^2) = \mathcal{N}(x_t; \mu_t, \frac{1}{w_t})$$
$$\mu_t = \frac{w_{t|t-1}\mu_{t|t-1} + w_z z_t}{w_{t|t-1} + w_z}$$
$$w_t = w_{t|t-1} + w_z$$

Of course, it will be very similar for the multivariate case. starting from a prior density $\mathcal{N}(x_0; \mu_0, \sigma_0^2)$:

1. Prediction:

$$P(x_t|Z^{t-1}) = \mathcal{N}(x_t; \mu_{t|t-1}, \Sigma_{t|t-1})$$
$$\mu_{t|t-1} = \mu_{t-1} + u$$
$$\Sigma_{t|t-1} = \Sigma_{t-1} + \Sigma_u$$

2. Update

$$P(x_t|Z^t) = \mathcal{N}(x_t; \mu_t, \Sigma_t) = \mathcal{N}(x_t; \mu_t, \Lambda_t^{-1})$$
$$\mu_t = \left(\Lambda_{t|t-1} + \Lambda_z\right)^{-1} \left(\Lambda_{t|t-1}\mu_{t|t-1} + \Lambda_z z_t\right)$$
$$\Lambda_t = \Lambda_{t|t-1} + \Lambda_z$$

If the measurement z_t is not a straight measurement but a linear or non-linear function of x, there is some more subtlety.