Lecture Notes: Noises

The problem of tracking something can be stated as a question of "where is it?" As figure 1 demonstrates, the answer to this question can be given as a scalar ("it is at 15.2"). This provides useful information for making decisions. For example, in the case of tracking an enemy plane, this provides a location at which to aim a weapon.

But the reality in a tracking problem is that the answer is rarely if ever certain. A filter makes this explicit by calculating a probability distribution for each variable of interest, rather than a scalar. As figure 2 illustrates, "you are here" on a map can be reimagined as "you are likely somewhere in this area, according to this probability curve". (It takes a lot more room to say this on a map, which could be why "you are here" is more common on mall maps. Or maybe mall-goers just don't like math.)

In the context of filtering, there are two types of noises that are modeled using probability distributions. **Dynamic noise** (also called system noise) refers to the uncertainty in predictions in the state transition equations. For example, consider the equations we used previously to model the the 1D motion of an object moving along an x axis. The state variables are $[x_t, \dot{x}_t]$, where x_t provides the position of the object and $\dot{x}_t$ provides the velocity of the object at time t. For the state transition equations we wrote:

$$x_{t+1,t} = x_{t,t} + \dot{x}_{t,t}T \quad (1)$$
$$\dot{x}_{t+1,t} = \dot{x}_{t,t} \quad (2)$$

where T is the interval of time between sensor readings. These equations assume that the object is moving at a constant velocity. However, suppose that the object can have a non-zero acceleration. This can be written into the equations as follows:

$$x_{t+1,t} = x_{t,t} + \dot{x}_{t,t}T \quad (3)$$
$$\dot{x}_{t+1,t} = \dot{x}_{t,t} + N(0, \sigma_a^2) \quad (4)$$

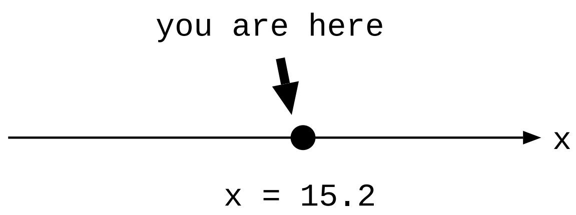

Figure 1: The scalar answer to "you are here".

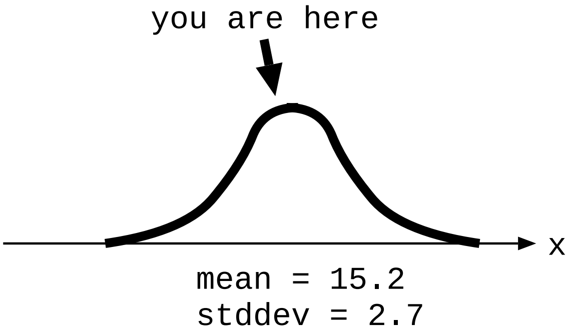

Figure 2: The probability distribution answer to "you are here". In a filtering problem, a probability distribution is associated with every variable of interest.

where $N(0, \sigma_a^2)$ denotes a random, normally distributed variable with a mean of zero and a standard deviation of σ_a. The size of σ_a defines how large of an acceleration can be expected during each prediction interval.

At least one of the predicted state variables must be affected by a dynamic noise, but this is not required of all predicted state variables. For this example, it would not make sense to have a dynamic noise on the position variable (unless the object can teleport). It is better to leave the prediction uncertainty in velocity, which has a real-world interpretation. If no state variables have dynamic noise, then there is no reason to filter, because the thing being tracked has no uncertainty in its behavior.

Measurement noise refers to the uncertainty in the sensor readings. For our 1D example, we previously wrote the measurement equation as:

$$y_t = x_{t,t} \tag{5}$$

This equation states that we observe the position of the object along the x axis with no uncertainty. However, suppose that the measurements are corrupted by noise. This can be written into the measurement equation as:

$$y_t = x_{t,t} + N(0, \sigma_n^2) \tag{6}$$

where $N(0, \sigma_n^2)$ denotes a random, normally distributed variable with a mean of zero and a standard deviation of σ_n. The size of σ_n defines the amount of expected corruption in a measurement.

As with dynamic noise, at least one measurement variable (but not necessarily all of them) should be affected by measurement noise. If there is no measurement noise, then the only reason to filter would be to model the behavior between measurements.

Previously, we wrote the filtering update equations for the 1D example as follows:

$$x_{t,t} = x_{t,t-1} + g_t(y_t - x_{t,t-1}) \tag{7}$$

$$\dot{x}_{t,t} = \dot{x}_{t,t-1} + h_t \frac{y_t - x_{t,t-1}}{T} \tag{8}$$

In an example iteration, we obtained a sensor reading that differed from the prediction. The values g_t and h_t were introduced as weights to control how to combine the estimates. Using

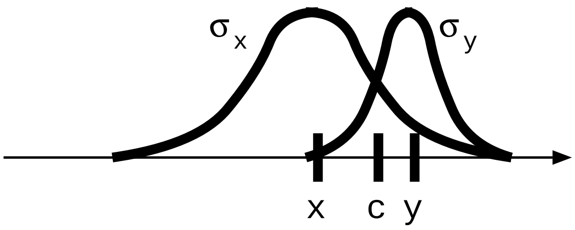

Figure 3: Given two estimates x and y of an unknown c, they can be combined by weighting according to their variances.

the concepts of dynamic noise and measurement noise, we can now develop a mathematical formula for how to calculate values for these weights. The basic idea is to balance the combination depending on the relative magnitude of the noises. For example, if the measurement noise is much higher than the dynamic noise, we would use relatively low values for g_t and h_t, relying more upon the predictions. Conversely, if the dynamic noise was much higher than the measurement noise, we would use higher values for g_t and h_t, which puts more weight on the measurements.

The Kalman filter takes the following approach to the problem. Assume that we have two estimates of a quantity, and we are seeking the best linear combination of those estimates:

$$c = K_1 x + K_2 y \tag{9}$$

where x and y are the estimates and c is the combined estimate. How should the constants K_1 and K_2 be chosen? Assume that each estimate has a known variance, as illustrated in figure 3. If we believe the estimates according to the inverse of the size of these variances, we can define the error of the combined estimate as:

$$E = \frac{(x-c)^2}{\sigma_x^2} + \frac{(y-c)^2}{\sigma_y^2} \tag{10}$$

This makes intuitive sense. If the variance of the estimate x is small, then we require x to be very close to the actual value c in order to keep the error down. The same holds for y.

We can minimize the error by taking the partial derivative:

$$\frac{\partial E}{\partial c} = \frac{-2(x-c)}{\sigma_x^2} + \frac{-2(y-c)}{\sigma_y^2} \tag{11}$$

Setting this equation equal to zero and solving for c:

$$\frac{-2(x-c)}{\sigma_x^2} + \frac{-2(y-c)}{\sigma_y^2} = 0 \tag{12}$$

$$\frac{x}{\sigma_x^2} - \frac{c}{\sigma_x^2} + \frac{y}{\sigma_y^2} - \frac{c}{\sigma_y^2} = 0 \tag{13}$$

$$c\left(\frac{1}{\sigma_x^2} + \frac{1}{\sigma_y^2}\right) = \frac{x}{\sigma_x^2} + \frac{y}{\sigma_y^2} \tag{14}$$

$$c = \frac{\frac{x}{\sigma_x^2} + \frac{y}{\sigma_y^2}}{\frac{1}{\sigma_x^2} + \frac{1}{\sigma_y^2}} \tag{15}$$

This equation again makes intuitive sense. Suppose that the variance of x is much smaller than the variance of y. Then the combined estimate c is equal to x. The same is true in the other direction. Suppose that the variances of x and y were equal, say to S. Then the combined estimate c is the mean (average) of x and y.

Now we will manipulate the equation for c algebraically. If we scale the last equation for common denominators, we obtain:

$$c = \frac{\frac{\sigma_y^2 x}{\sigma_x^2 \sigma_y^2} + \frac{\sigma_x^2 y}{\sigma_x^2 \sigma_y^2}}{\frac{\sigma_y^2}{\sigma_x^2 \sigma_y^2} + \frac{\sigma_x^2}{\sigma_x^2 \sigma_y^2}} \tag{16}$$

Eliminating common denominators gives:

$$c = \frac{\sigma_y^2 x + \sigma_x^2 y}{\sigma_y^2 + \sigma_x^2} \tag{17}$$

Expanding in terms of x and y gives:

$$c = \frac{\sigma_y^2}{\sigma_y^2 + \sigma_x^2} x + \frac{\sigma_x^2}{\sigma_y^2 + \sigma_x^2} y \tag{18}$$

Now comes a little trick. How far away is the x term from a whole $\frac{1}{1}x$? Rewriting just that part gives:

$$c = x - \frac{\sigma_x^2}{\sigma_y^2 + \sigma_x^2} x + \frac{\sigma_x^2}{\sigma_y^2 + \sigma_x^2} y \tag{19}$$

Combining terms with common fractions gives:

$$c = x + \frac{\sigma_x^2}{\sigma_y^2 + \sigma_x^2} (y - x) \tag{20}$$

Looking back at equation 7 we see that equation 20 has the same form, but the weight has been calculated as a function of the variances. In the Kalman filter, the two estimates of an unknown quantity come from sensor readings and predictions from state transition equations. They are combined in the update part of the filter according to a function of their variances.