

Lecture Notes: Nonlinear filtering

In filtering, there are two equations that describe the model information brought to bear on a problem: the observation equation and the state transition equation. The observation equation describes how the measurements are obtained during sensing. The state transition equation describes how the system is expected to change over time.

In the Kalman filter, we assume that both these equations are linear. The measurement equation is:

$$Y_t = MX_t + N_t \quad (1)$$

where Y_t is the measurement matrix (sensor readings), X_t is the actual state (which is unknown), N_t is the random measurement noise at time t , and M is the observation matrix. For example:

$$[\text{pos}] = [1 \ 0] \begin{bmatrix} \text{pos} \\ \text{vel} \end{bmatrix} + [\text{meas noise}] \quad (2)$$

Now suppose that we want to track something in an $X - Y$ space, but that our sensor measures polar $R-\theta$ readings (like for a radar). What happens to the measurement equation?

$$\begin{bmatrix} r_t \\ \theta_t \end{bmatrix} = \begin{bmatrix} ? & ? \\ ? & ? \end{bmatrix} \begin{bmatrix} x_t \\ y_t \end{bmatrix} + \begin{bmatrix} \text{dist noise} \\ \text{dir noise} \end{bmatrix} \quad (3)$$

The equations relating polar coordinates to Cartesian coordinates are

$$r = \sqrt{x^2 + y^2} \quad (4)$$

$$\theta = \tan^{-1} \frac{y}{x} \quad (5)$$

How can these be put into linear matrix form? They cannot. Instead, we must rewrite the measurement equation in nonlinear form:

$$y_t = g(x_t, n_t) \quad (6)$$

where y_t is the set of quantities being observed (measurements), x_t is the set of variables in the state model, and n_t is the set of measurement noises. The notation g indicates a matrix of equations. For example:

$$g(x_t, n_t) = \begin{bmatrix} r_t = \sqrt{x_t^2 + y_t^2} + N(0, \sigma_{\text{dist}}^2) \\ \theta_t = \tan^{-1} \frac{y_t}{x_t} + N(0, \sigma_{\text{dir}}^2) \end{bmatrix} \quad (7)$$

Similarly, the state transition portion of the model in linear form is:

$$X_t = \Theta X_{t-1} + A_t \quad (8)$$

where X_{t-1} is the actual state at time $t-1$ (which is unknown), Θ is the linear state transition matrix, A_t is the random dynamic noise during the time from $t-1$ to t , and X_t is the actual state at time t (which is unknown). This can be rewritten in nonlinear form as:

$$x_t = f(x_{t-1}, a_t) \quad (9)$$

where f is the matrix of state transition equations.

In order to apply the Kalman filtering approach to problems with nonlinear models, the extended Kalman filter linearizes equations 6 and 9. This is done by rewriting the measurement and state transition model equations in terms of computable approximations, via a Taylor series expansion. Recall that

$$x(t + \Delta t) = x(t) + \Delta t \dot{x}(t) + \frac{(\Delta t)^2}{2!} \ddot{x}(t) + \frac{(\Delta t)^3}{3!} \ddot{\ddot{x}}(t) + \dots \quad (10)$$

For small Δt or for small $\ddot{x}(t)$ all but the first two terms are negligible. In the case of the state transition equation (equation 9), we can write a similar expansion as follows:

$$x_t \approx \tilde{x}_t + \frac{\partial f}{\partial x}(x_{t-1} - x_{t-1,t-1}) + \frac{\partial f}{\partial a} a_t \quad (11)$$

Each of these terms takes some effort to understand.

The first term \tilde{x}_t is calculated using the previous estimate of state $x_{t-1,t-1}$ assuming no dynamic noise occurred during the interval $t-1$ to t :

$$\tilde{x}_t = f(x_{t-1,t-1}, 0) \quad (12)$$

The values in $x_{t-1,t-1}$ are the estimate of the state from the previous filter iteration. This portion of equation 11 is computable.

The second term in equation 11 is the equivalent of a first derivative term in a Taylor series expansion. It considers the difference between the actual state x_{t-1} (which is unknown) and the filtered state $x_{t-1,t-1}$. This difference, which can be considered a “delta error”, is multiplied by the partial derivative $\frac{\partial f}{\partial x}$.

The partial derivative is known as a Jacobian. A Jacobian of a set of functions of several variables is the differential of every equation with respect to every variable, with the result in matrix form. For example:

$$\frac{\partial f}{\partial x} = \frac{\partial(f_1, f_2, \dots, f_i)}{\partial(x_1, x_2, \dots, x_j)} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \dots & \frac{\partial f_1}{\partial x_j} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_i}{\partial x_1} & \dots & \frac{\partial f_i}{\partial x_j} \end{bmatrix} \quad (13)$$

A Jacobian can be written in several different notations, including:

$$\frac{\partial f}{\partial x} = J_x(f(x_{t-1}, a_t)) = J_t^{f,x} \quad (14)$$

In my notes, I will use only the first notation in order to emphasize it as a derivative. However, it is important to note that it is in general time-dependent. The shape of the derivative naturally changes depending on the position on the nonlinear curve.

The third term in equation 11 is also a first derivative term but considers the effect of the dynamic noise. The value a_t can be considered as another “delta error” and is multiplied by its respective Jacobian $\frac{\partial f}{\partial a}$, which can also be written as:

$$\frac{\partial f}{\partial a} = J_a(f(x_{t-1}, a_t)) = J_t^{f,a} \quad (15)$$

To summarize equation 11, we have expanded it about its approximation $x_{t-1,t-1}$ (the filtered state) using a Taylor series expansion to project to time t . The second and third terms in the equation are the first order derivatives of the expansion.

Similarly, the measurement equation (equation 6) can be rewritten using a Taylor series expansion as follows:

$$y_t \approx \tilde{y}_t + \frac{\partial g}{\partial x}(x_t - \tilde{x}_t) + \frac{\partial g}{\partial n} n_t \quad (16)$$

We have again ignored all terms involving second and higher order derivatives. The first term \tilde{y}_t is calculated as

$$\tilde{y}_t = g(\tilde{x}_t, 0) \quad (17)$$

The second and third terms cannot be calculated, as they involve the true state and the actual (unknown) measurement noise. However, the Jacobians can be calculated and can be written in shorthand as:

$$\frac{\partial g}{\partial x} = J_x(g(x_t, n_t)) = J_t^{g,x} \quad (18)$$

$$\frac{\partial g}{\partial n} = J_n(g(x_t, n_t)) = J_t^{g,n} \quad (19)$$

We can now use equations 11 and 16 to derive an equation for the filtered state estimate. To do so, we imagine a hypothetical Kalman filter that tracks the difference between the actual state and measurement, as compared to the approximations calculated from equations 12 and 17. We use the difference between the actual measurement and the approximated measurement (both of which are known) to update our approximated estimate of state (also known) towards the actual state (unknown).

We define the “error state” as the difference between the actual state and the calculated approximate state:

$$\tilde{e}_{x_t} = x_t - \tilde{x}_t \approx \frac{\partial f}{\partial x}(x_{t-1} - x_{t-1,t-1}) + \frac{\partial f}{\partial a} a_t \quad (20)$$

We define the “error measurement” as the difference between the actual measurement and the calculated approximate measurement:

$$\tilde{e}_{y_t} = y_t - \tilde{y}_t \approx \frac{\partial g}{\partial x}(x_t - \tilde{x}_t) + \frac{\partial g}{\partial n} n_t \quad (21)$$

In this hypothetical Kalman filter we can think of \tilde{e}_{x_t} as the actual error (unknown), and \tilde{e}_{y_t} as the observation of error. We can imagine calculating an estimated error e_t using a state update equation:

$$e_t = e_{t-1} + K_t(\tilde{e}_{y_t} - e_{t-1}) \quad (22)$$

However, we hope that the actual error will be zero. Therefore we will predict (in this hypothetical Kalman filter) that the error is always zero, simplifying equation 22 as

$$e_t = 0 + K_t(\tilde{e}_{y_t} - 0) = K_t\tilde{e}_{y_t} \quad (23)$$

Finally, using this estimated state of error, we can write a filtered state estimate as

$$x_{t,t} = \tilde{x}_t + e_t = \tilde{x}_t + K_t\tilde{e}_{y_t} = \tilde{x}_t + K_t(y_t - \tilde{y}_t) \quad (24)$$

We know the values of the actual measurements y_t , and we can calculate the values of the approximated state \tilde{x}_t and approximated measurement \tilde{y}_t . We have now finished linearizing the equations.

It is important to note that most of the equations presented in this set of notes are not things that are actually implemented. They are all presented here only to explain the derivation of the extended Kalman filter. It is also important to note that I left out a detail concerning the covariances. These need to be updated to reflect the effect of the Jacobians. In the next set of notes, these details will be discussed as the actual implementation is presented.