Statistical Techniques in Robotics (16-831, F08) Lecture #20 (Oct. 30)

Extensions to Kalman Filtering

Lecturer: Drew Bagnell *Scribe: Alex Grubb*

1 The Extended Kalman Filter

Previously, we used a Kalman Filter which could only model linear transformations:

$$x_{t+1} = Ax_t + \varepsilon, \quad \text{where } \varepsilon \sim \mathcal{N}(0, Q) \tag{1}$$
$$y_{t+1} = Ax_{t+1} + \delta, \quad \text{where } \delta \sim \mathcal{N}(0, R) \tag{2}$$

Now we would like to be able to model non-linear transformations with our filter:

$$x_{t+1} = f(x_t) + \varepsilon, \quad \text{where } \varepsilon \sim \mathcal{N}(0, Q) \tag{3}$$
$$y_{t+1} = g(x_{t+1}) + \delta, \quad \text{where } \delta \sim \mathcal{N}(0, R) \tag{4}$$

One way to deal with this non-linearity is to use the taylor expansion. In this case we approximate the transformations using the first order taylor expansion evaluated at the mean estimate of x at time t, μ_t:

$$x_{t+1} \approx f(\mu_t) + \frac{\partial f}{\partial x}(\mu_t)(x_t - \mu_t) + \varepsilon \tag{5}$$
$$y_{t+1} \approx g(\mu_{t+1}) + \frac{\partial g}{\partial x}(\mu_{t+1})(x_{t+1} - \mu_{t+1}) + \delta \tag{6}$$

The first order derivative term $\frac{\partial f}{\partial x}(\mu_t)$, also referred to as the Jacobian, is used here as the linear transformation, while the $f(\mu_t)$ term just serves to shift the mean of the transformation.

1.1 New Update Rules

Using the non-linear transformation above, we get the following new update equations for our mean and variance estimates for the motion model:

$$\mu_{t+1}^- = f(\mu_t^+) \tag{7}$$
$$\Sigma_{t+1}^- = J_f \Sigma_t^+ J_f^T + Q \tag{8}$$

And for the sensor model:

$$\mu_{t+1}^+ = \mu_{t+1}^- + \Sigma_{t+1}^- J_g^T (J_g \Sigma_{t+1}^- J_g^T + R)^{-1} [y_{t+1} - g(\mu_{t+1}^-)] \tag{9}$$
$$\Sigma_{t+1}^+ = \Sigma_{t+1}^- - \Sigma_{t+1}^- J_g^T (J_g \Sigma_{t+1}^- J_g^T + R)^{-1} J_g \Sigma_{t+1}^- \tag{10}$$

where $J_f = \frac{\partial f}{\partial x}(\mu_t^+)$ and $J_g = \frac{\partial g}{\partial x}(\mu_{t+1}^-)$.

1.2 Problems with Extended Kalman Filtering

- Taylor expansion is a poor approximation of most non-linear functions.
- Very difficult to implement correctly: you will most likely get the Jacobian wrong on the first try. Even when the Jacobian is wrong, it can be hard to tell whether your implementation is incorrect or the poor filtering results are due to the Taylor expansion.

1.3 Example: Non-Linear Regression

In this example we want to use an EKF for a non-linear regression problem. In this case our state vector is a set of weights w and we wish to estimate the mean and variance μ_w, Σ_w with our EKF, which we will just refer to as μ and Σ from here on.

The motion model is for stationary weights:

$$\mu_{t+1}^- = \mu_t^+ \tag{11}$$
$$\Sigma_{t+1}^- = \Sigma_t^+ \tag{12}$$

For our observation model we want to use the non-linear sigmoid function $\sigma(x) = \frac{1}{1+\exp(x)}$:

$$y_t = \sigma(w^T x) + \delta, \qquad \text{where } \delta \sim \mathcal{N}(0, \gamma^2) \tag{13}$$
$$\tag{14}$$

Note that our output in this case is a single value, so our Jacobian will be a vector. We can find the Jacobian J_σ using the chain rule and the fact that $\frac{\partial \sigma}{\partial x} = \sigma(x)(1-\sigma(x))$. Here we actually want to find the derivative of $g(w) = \sigma(w^T x)$, so we use the chain rule with $z = w^T x$.

$$\frac{\partial g}{\partial w} = \frac{\partial g}{\partial z}\frac{\partial z}{\partial w} \tag{15}$$
$$= \sigma(w^T x)(1 - \sigma(w^T x))x^T \tag{16}$$
$$J_\sigma = \frac{\partial g}{\partial w}(\mu) \tag{17}$$
$$= \sigma(\mu^T x)(1 - \sigma(\mu^T x))x^T \tag{18}$$

This gives update equations:

$$\mu_{t+1}^+ = \mu_{t+1}^- + \Sigma_{t+1}^- J_\sigma^T (J_\sigma \Sigma_{t+1}^- J_\sigma^T + \gamma^2)^{-1}[y_{t+1} - \sigma({\mu_{t+1}^-}^T x)] \tag{19}$$
$$\Sigma_{t+1}^+ = \Sigma_{t+1}^- - \Sigma_{t+1}^- J_\sigma^T (J_\sigma \Sigma_{t+1}^- J_\sigma^T + \gamma^2)^{-1} J_\sigma \Sigma_{t+1}^- \tag{20}$$

2 Other Non-linear Techniques

2.1 Statistical Linearization

One initial improvement on the linearization above is to use a method other than the Taylor expansion to find a linear approximation for f and g. In statistical linearization we sample points

around the mean estimate and then fit a line, plane, etc. to the sampled points and use that linear approximation to run the Kalman filter.

2.2 Unscented Kalman Filter

Another sampling-based method for dealing with non-linearities is to use the unscented transform. Here we simply sample points from the current mean and covariance and transform them according to the non-linear function f in our model, and then fit a new Gaussian to these transformed points, giving an updated mean and covariance matrix. This transformation is called the unscented transformation, from which the UKF derives it's name. For a pictorial comparison, see Figure 1.

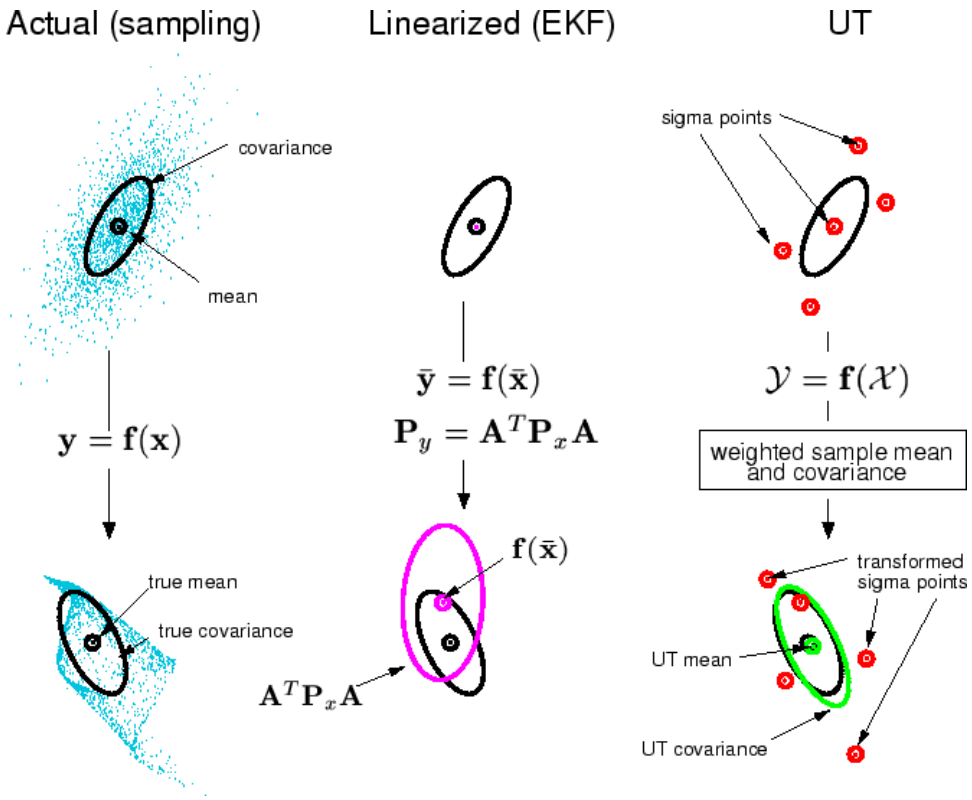

Figure 1: A comparison of the actual transformation, and the approximations given by the linear approximation and the unscented transformation. (Source: http://cslu.cse.ogi.edu/nsel/ukf/node6.html)

Typically the points used are the mean, and then points sampled at set distances along the principal axes of the Gaussian (the eigenvectors of the covariance matrix). The transformed points are also usually weighted based on the original mean and covariance and selection of these points, and a new distribution is fit to the weighted points.

The above method works for the motion model, but for the sensor model we need to operate a little differently. Here we transform the sampled points according to our sensor transform g, and then use the sampled points and the original points to estimate the unknown portions of the mean and

covariance in block form:

$$\mathcal{N}(\begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}, \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma yx & \Sigma yy \end{bmatrix}) \tag{21}$$

For example, the matrix Σ_{yy} can be found by simply taking the covariance of the transformed points. Once we have estimated all quantities above, the resulting conditional distribution is easily found using the standard equations, E.G. the conditional mean is given as:

$$\mu_{x|y} = \mu_x + \Sigma_{xy}(\Sigma_{yy})^{-1}[y - \mu_y] \tag{22}$$

There are a number of different filters that use this technique, but differ in how the select the points used and how the points are weighted. This includes: Sigma Point, Unscented, and Central Difference Filters.

2.2.1 Advantages / Disadvantages

- Advantage: You only need to implement the UKF once, as the implementation is independent of the models (I.E. f, g) used. It's also easier to implement correctly.
- Disadvantage: It will be slower than the regular Kalman filtering, and you don't gain anything if your transformation is linear.