Notes on the Unscented Kalman Filter

Ioannis Mandralis

March 2023

Optimal Recursive Estimation

The basic framework for state estimation of a discrete time nonlinear dynamical system we have considered so far is:

$$X[k+1] = F(X[k], u[k], V[k]) \tag{1}$$
$$Y[k] = H(X[k], W[k]). \tag{2}$$

Our aim is to estimate the unobserved state $X[k]$ given observations $Y[k]$. Under the assumption that all densities remain Gaussian it can be shown that the recursion

$$\hat{X}[k|k] = \hat{X}[k|k-1] + L[k](Y[k] - \hat{Y}[k|k-1]) \tag{3}$$
$$P[k|k] = P[k|k-1] - L[k]P_{\tilde{Y}[k]\tilde{Y}[k]}L[k]^\top \tag{4}$$

provides the minimum mean-squared error (MMSE) estimate for $X[k]$, where $\tilde{Y}[k] = Y[k] - \hat{Y}[k|k-1]$. The optimality of this recursion does not require model linearity. It can be shown that, in general, the optimal terms of this recursion are given by:

$$\boxed{\begin{aligned}\hat{X}[k|k-1] &= \mathbb{E}[F(\hat{X}[k-1|k-1], u[k-1], V[k-1])] && (5)\\ \hat{Y}[k|k-1] &= \mathbb{E}[H(\hat{X}[k|k-1], W[k])] && (6)\\ L[k] &= P_{X[k]Y[k]}P^{-1}_{\tilde{Y}[k]\tilde{Y}[k]} && (7)\end{aligned}}$$

Equivalence to Discrete Time Kalman Filter

To see an example of how to apply the optimal recursion, consider the Discrete Time Kalman filter as was considered in class. Here the system dynamics were:

$$X[k+1] = AX[k] + Bu[k] + FV[k] \tag{8}$$
$$Y[k] = CX[k] + W[k] \tag{9}$$

where $V[k], W[k]$ are zero-mean white noise processes. Lets compute the terms in equations (22)-(23) one by one and see that they match the predictor-corrector form of the discrete time Kalman Filter seen in class. Start with (22):

$$\begin{aligned}\hat{X}[k|k-1] &= \mathbb{E}[A\hat{X}[k-1|k-1] + Bu[k-1] + FV[k-1]] \\ &= A\hat{X}[k-1|k-1] + Bu[k-1]\end{aligned} \tag{10}$$

Next compute (23):

$$\hat{Y}[k|k-1] = \mathbb{E}[C\hat{X}[k|k-1] + W[k]] = C\hat{X}[k|k-1] \tag{11}$$

Finally compute (24) by computing the two covariances. First compute $P_{X[k]Y[k]}$:

$$\begin{aligned}P_{X[k]Y[k]} &= \mathbb{E}[(X[k] - \mathbb{E}[X[k]])(Y[k] - \mathbb{E}[Y[k]])^\top] \\ &= \mathbb{E}[(X[k] - \mathbb{E}[X[k]])(Y[k] - C\mathbb{E}[X[k]])^\top] \\ &= \mathbb{E}[X[k]Y[k]^\top] - \mathbb{E}[X[k]]\mathbb{E}[X[k]]^\top C^\top \\ &= (\mathbb{E}[X[k]X[k]^\top] - \mathbb{E}[X[k]]\mathbb{E}[X[k]]^\top)C^\top \\ &= P_{X[k]X[k]}C^\top\end{aligned} \tag{12}$$

Now compute $P_{\tilde{Y}[k]\tilde{Y}[k]}$. Firt compute $\mathbb{E}[\tilde{Y}[k]]$:

$$\begin{aligned}\mathbb{E}[\tilde{Y}[k]] &= \mathbb{E}[Y[k] - \hat{Y}[k|k-1]] \\ &= \mathbb{E}[C(X[k] - \hat{X}[k|k-1]) + W[k]] \\ &= C(\mathbb{E}[X[k]] - \hat{X}[k|k-1])\end{aligned} \tag{13}$$

Now since $\tilde{Y}[k] - \mathbb{E}[\tilde{Y}[k]] = C(X[k] - \mathbb{E}[X[k]]) + W[k]$ we can compute $P_{\tilde{Y}[k]\tilde{Y}[k]}$:

$$\begin{aligned}P_{\tilde{Y}[k]\tilde{Y}[k]} &= \mathbb{E}[(\tilde{Y}[k] - \mathbb{E}[\tilde{Y}[k]])(\tilde{Y}[k] - \mathbb{E}[\tilde{Y}[k]])^\top] \\ &= R_W + CP_{X[k]X[k]}C^\top\end{aligned} \tag{14}$$

Combining the two covariances (12),(14) gives us the optimal gain (using (24)):

$$L[k] = P_{X[k]X[k]}C^\top(R_W + CP_{X[k]X[k]}C^\top)^{-1} \tag{15}$$

Now define $P[k|k-1] := P_{X[k]X[k]}$ and recall from OBC that we can compute the covariance of the state of discrete time system recursively as:

$$P[k|k-1] = AP[k-1|k-1]A^\top + FR_VF^\top \tag{16}$$

Combining all the above equations we retrieve the predictor corrector form of the discrete time Kalman Filter.
Prediction:

$$\hat{X}[k|k-1] = A\hat{X}[k-1|k-1] + Bu[k-1] \tag{17}$$

$$P[k|k-1] = AP[k-1|k-1]A^\top + FR_VF^\top \tag{18}$$

Correction:

$$L[k] = P[k|k-1]C^\top(R_W + CP[k|k-1]C^\top)^{-1} \tag{19}$$

$$\hat{X}[k|k] = \hat{X}[k|k-1] + L[k](Y[k] - \hat{Y}[k|k-1]) \tag{20}$$

$$P[k|k] = P[k|k-1] - L[k]CP[k|k-1] \tag{21}$$

The problem with the Extended Kalman Filter

In the Extended Kalman Filter instead of applying Equations (22)-(24), the following approximation is made:

$$\hat{X}[k|k-1] = F(\hat{X}[k-1|k-1], u[k-1], 0) \tag{22}$$

$$\hat{Y}[k|k-1] = H(\hat{X}[k|k-1], 0) \tag{23}$$

$$L[k] = \hat{P}_{X[k]Y[k]} \hat{P}^{-1}_{\tilde{Y}[k]\tilde{Y}[k]} \tag{24}$$

i.e. no expectations are taken and the process and disturbance noise are simply set to zero in the prediction steps. Furthermore, the covariance is estimated by linearizing the dynamics equations and then determining the posterior covariance matrices analytically for the linearized system. Thus, in the EKF the state distribution is viewed as a Gaussian Random Variable that is propagated analytically though the "first order" linearization of the nonlinear system. In essence, the EKF is providing " first-order" approximations to the optimal estimation terms, which can introduce large errors in the true posterior mean and covariance of the transformed Gaussian Random Variable. This can lead to suboptimal performance and/or divergence of the filter.

The Unscented Kalman Filter

The Unscented Kalman Filter improves on the approximation issues of the EKF by propagating special sample points called *sigma points* through the true *nonlinear* system capturing the posterior mean and covariance accurately to the 3rd order for any nonlinearity. The main machinery used in the UKF is the *Unscented Transformation*.

Unscented Transform (UT)

The UT estimates the mean and covariance of a random variable which goes through a nonlinear transformation. Consider propagating a random variable X of dimension L through a nonlinear function $Y = g(X)$. Assume X has mean μ_X and covariance P_X. To calculate the statistics of Y we form a matrix $\mathcal{X}$ as follows:

$$\mathcal{X}_0 = \mu_X \tag{25}$$

$$\mathcal{X}_i = \mu_X + (\sqrt{(L+\lambda)P_X})_i, \quad i = 1, \ldots, L \tag{26}$$

$$\mathcal{X}_i = \mu_X - (\sqrt{(L+\lambda)P_X})_i, \quad i = L+1, \ldots, 2L \tag{27}$$

$$W_0^{(m)} = \lambda/(L+\lambda) \tag{28}$$

$$W_0^{(c)} = \lambda/(L+\lambda) + (1 - \alpha^2 + \beta) \tag{29}$$

$$W_i^{(m)} = W_i^{(c)} = 1/2(L+\lambda), \quad i = 1, \ldots, 2L \tag{30}$$

where $\lambda = \alpha^2(L+\kappa) - L$ is a scaling parameter, α determines the spread of sigma points about the mean and κ is usually set to zero. $\beta = 2$ is optimal for a state which is Gaussian distributed. Finally, $(\sqrt{(L+\lambda)P_X})_i$ denotes the i-th row of the matrix square root. Once the sigma vectors are computed, they are propagated through the nonlinear transformation:

$$\mathcal{Y}_i = g(\mathcal{X}_i) \tag{31}$$

Finally, the mean and covariance of the transformed random variable are estimated using:

$$\mu_Y = \sum_0^{2L} W_i^{(m)} \mathcal{Y}_i \tag{32}$$

$$P_Y = \sum_0^{2L} W_i^{(c)} [\mathcal{Y}_i - \mu_Y][\mathcal{Y}_i - \mu_Y]^\top \tag{33}$$

This approach results in an approximation of the posterior mean and covariance which is accurate to the third order for Gaussian inputs for all nonlinearities.

The Unscented Kalman Filter (UKF)

In order to apply the Unscented Transform to the recursive estimation framework detailed above we need to define a new state random variable as the concatenation $X^a[k] = [X[k]^\top \ V[k]^\top \ W[k]^\top]^\top$. Then we can directly apply the UT as follows:
Prediction:

$$\hat{X}[k|k-1], P[k|k-1], \hat{Y}[k|k-1], \hat{P}_{\tilde{Y}[k]\tilde{Y}[k]}, \hat{P}_{X[k]Y[k]} \tag{34}$$

$$= \mathbf{UT}(\hat{X}[k-1|k-1], P[k-1|k-1], u[k-1], F, H) \tag{35}$$

Correction:

$$L[k] = \hat{P}_{X[k]Y[k]} \hat{P}^{-1}_{\tilde{Y}[k]\tilde{Y}[k]} \tag{36}$$

$$\hat{X}[k|k] = \hat{X}[k|k-1] + L[k](Y[k] - \hat{Y}[k|k-1]) \tag{37}$$

$$P[k|k] = P[k|k-1] - L[k]\hat{P}_{\tilde{Y}[k]\tilde{Y}[k]} L[k]^\top \tag{38}$$

The full equations of the Unscented Kalman Filter can be found in the original paper or the python implementation.