AERO 626 Challenge Problem

Tim Woodbury

November 29, 2015

1 Filter algorithms

Five estimation algorithms are considered for the nonlinear sequential estimation problem: the Extended Kalman Filter (EKF), the Unscented Kalman Filter (UKF), the Sequential Importance Resampling (SIR) particle filter (PF), the Ensemble Kalman Filter (ENKF), and the Gaussian Mixture Model (GMM). For completeness, the algorithm governing each filter is described in the following sections.

1.1 Extended Kalman filter

The Extended Kalman filter is, by far, the most popular sequential estimation algorithm in the field of guidance, navigation, and control. Simply put, the EKF replaces the linear process and measurement influence matrices in the linear Kalman filter algorithm with the Jacobian of nonlinear governing functions evaluated at an aposteriori state estimate. The assumed prior estimate $\hat{x}(0)$, covariance $P(0)$, and dynamic system are of the following form:

$$\hat{x}(0) = E[x_0] \tag{1}$$

$$P(0) = E[(\hat{x}(0) - x_0)(\hat{x}(0) - x_0)^T] \tag{2}$$

$$x_k = f(x_{k-1}, t_{k-1}) + G_k w_k, w_k \sim N(0, Q_k) \tag{3}$$

$$y_k = h(x_k, t_k) + v_k, v_k \sim N(0, R_k) \tag{4}$$

The discrete propagation equations for the state and covariance follow:

$$\hat{x}_k^- = f(\hat{x}_{k-1}, t_{k-1}) \tag{5}$$

$$P_k^- = F_{k-1} P_{k-1} F_{k-1}^T + G_{k-1} Q_{k-1} G_{k-1}^T \tag{6}$$

$$F_{k-1} \equiv \frac{\partial f}{\partial x}(\hat{x}_{k-1}, t_{k-1}) \tag{7}$$

The update equations for the state and covariance, given a measurement $\tilde{y}_k$, are:

$$\hat{x}_k = \hat{x}_k^- + K_k(\tilde{y}_k - h(\hat{x}_k^-)) \tag{8}$$

$$P_k = (I - K_k H_k)P_k^- \tag{9}$$

$$K_k \equiv P_k^- H_k^T (H_k P_k^- H_k^T + R_k)^{-1} \tag{10}$$

$$H_k \equiv \frac{\partial h}{\partial x}(\hat{x}_k^-, t_k) \tag{11}$$

The EKF has been shown to work well for a variety of systems, given that the estimate PDF remains well approximated by a Gaussian. Many studies have also shown that the UKF generally performs better for nonlinear systems, at the cost of slightly greater computational complexity.

1.2 Unscented Kalman filter

The UKF is based on the same idea as the EKF; extend the linear Kalman update to nonlinear systems in an approximate fashion. In the UKF, however, function Jacobians are no longer used; instead, a discrete set of "sigma points" are defined and passed through the propagation and update equations. The output points are used to evaluate the system covariances from their definitions.

Define the augmented state vector x^a as follows:

$$x_k^a = \begin{bmatrix} x_k \\ w_k \\ v_k \end{bmatrix} \equiv \begin{bmatrix} x_k^x \\ x_k^w \\ x_k^v \end{bmatrix} \tag{12}$$

In Eq. 12, x_k^x, x_k^w, and x_k^v are used as shorthand for the augmented state members associated with the state, process noise, and measurement noise. The $L \times 2L + 1$ array of sigma points is defined as follows:

$$\chi_{k-1}^a = [\hat{x}_{k-1}^a \ \hat{x}_{k-1}^a \pm \sqrt{(L+\lambda)P_{k-1}^a}] \tag{13}$$

Each sigma point is then propagated through the governing nonlinear equation as follows:

$$\chi_k^-(:,j) = f(\chi_{k-1}^x(:,j), t_{k-1}) + G_k \chi_{k-1}^w(:,j), \ \forall \ j \tag{14}$$

The state and covariance apriori estimates are given as weighted averages of the sigma points:

$$\hat{x}_k^- = \sum_{i=1}^{2L+1} W_i^m \chi_k^{x-}(:,i) \tag{15}$$

$$P_k^- = \sum_{i=1}^{2L+1} W_i^c (\chi_k^{x-}(:,i) - \hat{x}_k^-)(\chi_k^{x-}(:,i) - \hat{x}_k^-)^T \tag{16}$$

2

In a similar fashion, the update is performed by first passing the propagated sigma points through the measurement equation:

$$Y_k(:,j) = h(\chi_k^{x-}(:,j), t_k) + \chi_{k-1}^v(:,j) \tag{17}$$

1.3 Gaussian mixture models

The implementation of the Gaussian mixture model (GMM) is based on Ref. [2]. The algorithm is repeated here for completeness. The prior density function is approximated as a sum of multivariate Gaussians with normalizing weights α'_{ki} as follows:

$$p(x_k|y_k) = \sum_{i=1}^{\chi_k} \alpha_{ki} \eta(x_k - a_{ki}, P_{ki}) \tag{18}$$

$$\eta(a, P) \equiv \frac{1}{\sqrt{(2\pi)^n \det(P)}} e^{-\frac{1}{2} a^T P^{-1} a} \tag{19}$$

$$\sum_{i=1}^{\chi_k} \alpha_{ki} = 1 \tag{20}$$

For a discrete-time dynamic system with state influence function $x_{k+1} = f(x_k) + G_k w_k, w_k \sim N(0, Q_k)$, the individual Gaussian filters are propagated using the Extended Kalman Filter update for the means a_{ki} and the covariances P'_{ki}:

$$p(x_k|y_{k-1}) = \sum_{i=1}^{\chi'_k} \alpha'_{ki} \eta(x_k - a'_{ki}, P'_{ki}) \tag{21}$$

$$\chi'_k = \chi_{k-1} \tag{22}$$

$$\alpha'_{ki} = \alpha_{(k-1)i} \tag{23}$$

$$a'_{ki} = f_k(a_{(k-1)i}) \tag{24}$$

$$P'_{ki} = F_{ki} P_{(k-1)i} F_{ki}^T + G_k Q_k G_k^T \tag{25}$$

$$F_{ki} \equiv \frac{\partial f_k}{\partial x}(a_{(k-1)i}) \tag{26}$$

The update equations assume a nonlinear measurement of the form $y_k = h(x_k) + n_k, n_k \sim N(0, R_k)$. The aposteriori density function is approximated as follows:

$$p(x_k|y_k) = \sum_{i=1}^{\chi_k} \alpha_{ki} \eta(x_k - (a_{ki}' + K_{ki}(y_k - h_k(a_{ki}'))), P_{ki}) \tag{27}$$

$$\chi_k \equiv \chi_k' \tag{28}$$

$$P_{ki} \equiv P_{ki}' - K_{ki} H_{ki} P_{ki}' \tag{29}$$

$$K_{ki} \equiv P_{ki}' H_{ki}^T (H_{ki} P_{ki}' H_{ki}^T + R_k)^{-1} \tag{30}$$

$$\alpha_{ki} \equiv \frac{\alpha_{ki}' \beta_{ki}}{\sum_{j=1}^{\chi_k} \alpha_{kj}' \beta_{kj}} \tag{31}$$

$$\beta_{kj} \equiv \eta(y_k - h_k(a_{kj}), H_{kj} P_{kj}' H_{kj}^T + R_k) \tag{32}$$

This completes the description of the basic algorithm. Initialization of the prior is not addressed in Ref. [2]. Two schemes are employed in the current work:

1. Initialization of the means a_{ki} to the sigma points used by the Unscented Kalman Filter

2. Monte Carlo selection of the initial means

In each case, the initial covariances P_{0i} are selected to satisfy a prior Gaussian proposal with mean $\mu(0)$ and covariance $P(0)$. The mean and covariance of the initial distribution are computed as follows:

$$\mu \equiv E(x_k) = \sum_{i=1}^{\chi_k} \alpha_{ki} a_{ki} \tag{33}$$

$$cov(x_k) = \sum_{i=1}^{\chi_k} \alpha_{ki}(P_{ki} + (a_{ki} - \mu)(a_{ki} - \mu)^T) \tag{34}$$

Eq. 34 represents an underconstrained system. By specifying that the initial distributions have equal covariance, the expression can be made tractable and solved for P_0:

$$P_0 = P(0) - \sum_{i=1}^{\chi_k} (\alpha_{ki}(a_{ki} - \mu)(a_{ki} - \mu)^T) \tag{35}$$

One practical note: astute readers will observe that the sum in Eq. 35 with equal weights matches the statistical definition of the covariance (to within a small bias in the weights). Depending on the number of initial points and their distribution, the value of P_0 may not satisfy the requirement of positive definiteness for the covariance matrix. To ensure positive definiteness, P_0 is processed additionally by taking the absolute value of all members, and setting the diagonal elements to zero, to ensure the determinant remains positive.

The initial weights are initially set equal to one another, $\alpha_{ki} = \chi_k^{-1}$. It should be observed that for a unimodal near-Gaussian system, the algorithm will collapse to the Extended Kalman filter after a few iterations, with the means all concentrated near a single point. The GMM is intended to address the case of bimodality in the propagated density function, and is of most interest as an alternative to the PF and ENKF for handling multi-modality of the density.

2 Governing equation and bifurcation time

The system dynamics are governed by Eq. 36, given below. The system consists of a cubic nonlinearity with white-noise and cosine forcing terms. For the purposes of estimation, the cosine forcing term is considered "unknown"; that is, the magnitude may be known but the exact form cannot be used in the system propagation equation. Therefore, in practice both the white-noise term and the cosine forcing terms are lumped into a single process noise term that must encapsulate both sources.

$$\ddot{x} = x - \epsilon x^3 + a_0 \cos \omega_t t + w(t) \tag{36}$$

$$w(t) \sim N(0, q) \tag{37}$$

2.1 System parameters and measurement model

The system measurement model is given as follows:

$$y_k = x_k + n_k, n_k \sim N(0, r) \tag{38}$$

System measurements occur at discrete times $t_k = kT$. The total set of parameters that govern system behavior are defined below:

- $\epsilon = 0.01$

- $a_0 = 2.0$

- $\omega_t = 1.25$

- $q = 1.0$

- $r = 1.0$

- $T = 0.1$

In addition, for numerical convenience in simulating the system dynamics, the process noise $w(t)$ is discretized at a fixed rate and held piecewise constant between samples. This effectively transforms as follows:

$$w(t) \to w_k \sim N(0, qT_s), t_k \leq t \leq t_{k+1} \tag{39}$$

The dynamic uncertainty discretization time is taken as $T_s = 0.001$, which is an order of magnitude smaller than any of the measurement sample times used later.

2.2 Bifurcation

One feature of interest for the system is its equilibria: the unforced system has roots at $x = 0, \pm\sqrt{\frac{1}{\epsilon}}$. The roots at $\pm\sqrt{\frac{1}{\epsilon}}$ are attractive; the root at the origin is repulsive. The net effect is that, when propagated forward for a long enough time, a probability density initially centered at the origin will bifurcate into two distributions, roughly centered around the attractive equilibria. It is of interest to determine the bifurcation time, in examining the effectiveness of various filters.

Bifurcation time is analyzed with no forcing ($a_0 = 0$) by performing simulating a number of points from a normal distribution about the origin, then propagating forward in time. Results are plotted, and the bifurcation time estimated qualitatively. Fig. 1 shows the development of the system bifurcation from its initial condition. The unforced motion is also shown by plotting the system after 9.5 seconds of development. From these plots, it is clear that bifurcation appears qualitatively on the order of a second or so. Therefore, the sample time of 10.0 seconds is used to consider the bifurcated system in later study.

2.3 Test for multimodality

The test for multimodality is Silverman's[3]; for implementation, reference is also made to [1]. The KDE is a method of nonparametric estimation of a density function. The KDE for a particular kernel $K(.)$ is defined for samples $\chi = \{X_1, X_2, ...X_n\}$ as follows:

$$\hat{f}(x; h) = \frac{1}{nh} \sum_{i=1}^{n} K(\frac{x - X_i}{h})$$

(40)

In Eq. 40, h is a smoothing parameter. The kernel function in general may take different forms, but the Silverman test uses a univariate normal distribution. For a given data set χ, the number of modes in the data (determined by the number of local extrema) is monotonically decreasing with the value of the smoothing parameter, according to Silverman[3]. So, a simple binary search can be used to determine the critical smoothing parameter h_{crit} at which the KDE has a particular number of modes. The objective of the test is to if the distribution of the position and velocity states at a fixed time is unimodal; therefore, the smoothing parameter for which the KDE has only one extrema is sought.

One additional procedure is used in the test; smoothed bootstrap sampling. The test works by evaluating the significance of the null hypothesis that the underlying distribution is unimodal. If the significance is too small, the null hypothesis is rejected, and the distribution is assumed to be multimodal. Smoothed bootstrap sampling is used to evaluate the significance level. Bootstrap samples X_i^* are drawn from the original data set $\chi = \{X_1, X_2, ...X_n\}$ as follows:

$$X_i^* = \frac{X_{j(i)} + h\epsilon_i}{\sqrt{1 + h^2/\sigma^2}}, i \in [1, \ldots, n] \tag{41}$$

In Eq. 41, the sample $X_{j(i)}$ is selected uniformly with replacement. σ is the sample standard deviation of the data. ϵ_i is a normally distributed random variable with uniform variance. The smoothed bootstrap samples are used to evaluate the null hypothesis that the distribution is unimodal with a critical KDE parameter h_{crit}; N_b sets are drawn using smoothed bootstrap sampling, and the number of maxima of the resulting KDE with h_{crit} are tabulated. The fraction of sets with only one maximum is taken to be the significance level of the null hypothesis.

The test procedure is summarized as follows:

1. Simulate N_p Monte Carlo simulations for a fixed time

2. For each time t^* in the simulation outputs:

    (a) Collect the N_p state values at $t = t^*$

    (b) Compute the critical smoothing parameter h_{crit} for which the KDE is unimodal

    (c) Evaluate the probability that the distribution has one mode

        i. Perform smoothed bootstrap sampling from the critical KDE N_b times

        ii. For each of the N_b samples:

            A. Evaluate the KDE with the critical smoothing parameter for the current sample. Record the number of modes.

        iii. The fraction of unimodal KDEs is taken as the P-value for the test.

References

[1] M. Adereth. Silverman's mode estimation method explained. Online, October 2014. http://adereth.github.io/blog/2014/10/12/silvermans-mode-detection-method-explained/.

[2] Daniel L. Aspach and Harold W. Sorenson. Nonlinear bayesian estimation using gaussian sum approximations. *IEEE Transactions on Automatic Control*, AC-17(4), 1972.

[3] B.W. Silverman. Using kernel density estimates to investigate multimodality. *Journal of the royal statistical society*, 43(1):97–99, 1981. http://www.stat.washington.edu/wxs/Stat593-s03/Literature/silverman-81a.pdf.
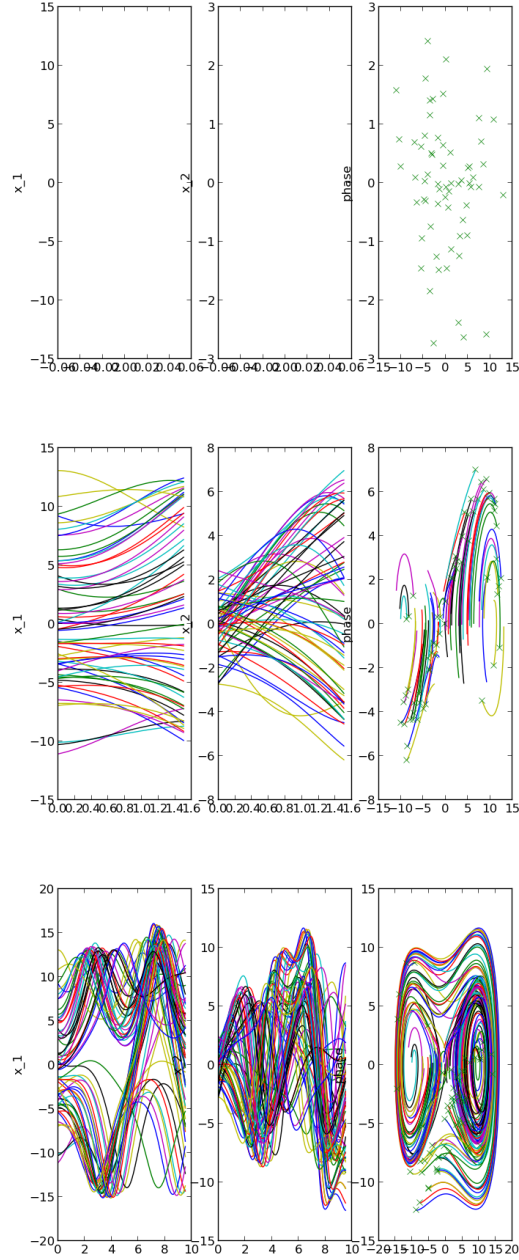
Figure 1: Figures showing the development of the distribution. Left: time hsitory of position. Center: time history of speed. Right: phase portrait of system, with 'x' used to indicate the current condition of the state. Top: Initially Gaussian points with position variance 0.1 and speed variance 1.0. Middle: Condition of the sustem after 1.5 seconds; clear separation between points in the phase portratit is visible. Bottom: Long-term development of the system after 9.5 seconds, showing the unforced motion of points.