Nonlinear Kalman filters

Jason Filos (Imperial College London)

February 1, 2013
Outline

1. Standard Kalman filter
2. Extended Kalman filter
3. Unscented Kalman filter
4. What's more
Standard Kalman filter

- Kalman filtering considered up to this point addressed estimation of a state vector in a **linear** model of a dynamical system:

\[
\begin{align*}
    x_t &= A_t x_{t-1} + u_t, \\
    y_t &= B_t x_t + v_t,
\end{align*}
\]

where \(A_t\) is the **transition matrix** taking the state \(x_{t-1}\) from time \(t-1\) to time \(t\) and \(B_t\) is the **measurement matrix**.

- We assume \(x_0 \sim \mathcal{N}(0, \Sigma_0)\), **process noise** \(u_t \sim \mathcal{N}(0, \Sigma_u)\) and **measurement noise** \(v_t \sim \mathcal{N}(0, \Sigma_v)\).

- Furthermore, the initial state is uncorrelated with the noise process, i.e. \(E[u_t x_0^T] = 0, \ E[v_t x_0^T] = 0\).
If the model is **nonlinear**, we may extend the use of Kalman filtering through a linearization procedure.

Resulting filter is referred to as the *extended Kalman filter* (EKF).

Such extension is feasible by virtue of the fact that the Kalman filter is described in terms of difference equations in the case of discrete-time systems.

Consider a nonlinear dynamical system described by the state-space model:

$$x_t = f_t(x_{t-1}) + u_t,$$

$$y_t = h_t(x_t) + v_t.$$

The functional $f$ denotes a *nonlinear transition matrix* function and the functional $h$ denotes a *nonlinear measurement matrix*. 
Extended Kalman filter (ctnd.)

- Basic idea of EKF is to linearize state-space model at each time instant around the most recent state estimate.
  - Function \( f \) can be used to compute the predicted state from the previous estimate.
  - Function \( h \) can be used to compute the predicted measurement from the predicted state.

- **Note**: \( f \) and \( h \) cannot be applied to the covariance directly. Instead, matrix of partial derivatives (the Jacobian) is computed.

- Two step approximation: Compute Jacobian, then employ first-order Taylor approximation.
- Once a linear model is obtained, the standard Kalman filter equations are applied.
State transition and observation matrices are defined to be the following Jacobians:

\[
F_{t-1} = \frac{\partial f}{\partial \hat{x}} \bigg|_{\hat{x}_t-1|t-1},
\]

\[
H_t = \frac{\partial h}{\partial \hat{x}} \bigg|_{\hat{x}_t|t-1}.
\]

The \( ij \)th entry of \( F_{t-1} \) is equal to the partial derivative of the \( i \)th component of \( f \) with respect to the \( j \)th component of \( \hat{x}_{t-1} \).

Likewise, the \( ij \)th entry of \( H_t \) is equal to the partial derivative of the \( i \)th component of \( h \) with respect to the \( j \)th component of \( \hat{x}_t \).
EKF Algorithm

- **Initialization**: For $t = 0$, set

  \[
  \hat{x}_0 = E [x_0], \\
  \Sigma_0 = E [(x_0 - \hat{x}_0)(x_0 - \hat{x}_0)^T].
  \]

- **Computation**: For $t = 1, 2, \ldots$ compute:

  State estimate propagation: $\hat{x}_{t|t-1} = f(\hat{x}_{t-1|t-1})$

  Error covariance propagation: $\Sigma_{t|t-1} = F_{t-1} \Sigma_{t-1|t-1} F_{t-1}^T + \Sigma_u$

  Kalman gain matrix: $K_t = \Sigma_{t|t-1} H_t^T (H_t \Sigma_{t|t-1} H_t^T + \Sigma_v)^{-1}$

  State estimate update: $\hat{x}_{t|t} = \hat{x}_{t|t-1} + K_t (y_t - h(\hat{x}_{t|t-1}))$

  Error covariance update: $\Sigma_{t|t} = (I - K_t H_t) \Sigma_{t|t-1}$
Problems with the EKF

- The EKF algorithm provides only an approximation to optimal nonlinear estimation.
- The state distribution is approximated by a Gaussian random variable (GRV), which is then propagated analytically through the first-order linearization of the nonlinear system.
- This can introduce large errors in the true posterior mean and covariance of the transformed GRV:
  - **suboptimal performance**
  - **divergence of the filter** (if the initial estimate of the state is wrong, or if the process is modelled incorrectly)
- However, the EKF can give reasonable performance, and is arguably the de facto standard in navigation systems and GPS.
Unscented Kalman filter

Solution?

In the following we present an alternative filter with performance superior to that of the EKF. This algorithm, referred to as the unscented Kalman filter (UKF), was first proposed by Julier et al.

- The UKF addresses the problems of the EKF by using a deterministic sampling approach.
- State distribution again approximated by a GRV, but now represented using a minimal set of carefully chosen sample points (called *sigma points*).
- Capture true mean and covariance of the GRV, and, when propagated through the *true* nonlinear system, give accurate posterior mean and covariance up to second order (Taylor series expansion) for *any* nonlinearity.
Unscented Transformation

- The unscented transformation (UT) is a method for calculating the statistics of a random variable which undergoes a nonlinear transformation.
- Consider propagating a random variable $x$ (dimension $L$) through a nonlinear function, $y = f(x)$ with $x \sim \mathcal{N}(\mu, \Sigma_x)$.
- To calculate the statistics of $y$ we form a matrix $\mathbf{X}$ of $2L + 1$ sigma vectors $\chi^i$ according to the following:
  
  $\chi^0 = \mu,$
  
  $\chi^i = \mu + \left( \sqrt{(L + \lambda)\Sigma_x} \right)_i, \ i = 1 \ldots L,$
  
  $\chi^i = \mu - \left( \sqrt{(L + \lambda)\Sigma_x} \right)_{i-L}, \ i = L + 1, \ldots 2L,$

  where $\lambda = \alpha^2(L + \kappa) - L$ is a scaling parameter. The constants $\alpha$ and $\kappa$ control the spread of the sigma points.
Unscented Transformation (ctnd.)

- The constant $\alpha$ determines the spread of the sigma points around $\mu$, and is usually set to a small positive value (e.g., $1 \leq \alpha \leq 10^{-4}$). The constant $\kappa$ is a secondary scaling parameter, which is usually set to $3 - L$.

- $\left(\sqrt{(L + \lambda)\Sigma_x}\right)_i$ is the $i$th column of the matrix square root (e.g., lower-triangular Cholesky factorization) of $(L + \lambda)\Sigma_x$.

- The sigma points are propagated through the transition function $f$: $\mathcal{Y}^i = f(\chi^i), \quad i = 0 \ldots 2L$. 
Standard Kalman filter  Extended Kalman filter  Unscented Kalman filter  What’s more

Table 7.2. Note that no explicit calculations of Jacobians or Hessians are necessary to implement this algorithm. Furthermore, the overall number of computations is of the same order as the EKF.

Implementation Variations

For the special (but often encountered) case where the process and measurement noise are purely additive, the computational complexity of the UKF can be reduced. In such a case, the system state need not be augmented with the noise RVs. This reduces the dimension of the sigma points as well as the total number of sigma points used. The covariances of the noise source are then incorporated into the state covariance using a simple additive procedure. This implementation is given in Table 7.3. The complexity of the algorithm is of order $L^3$, where $L$ is the dimension of the state. This is the same complexity as the EKF. The most costly operation is in forming the sample prior covariance matrix $P_{y/C_0}$. Depending on the form of $F$, this may be simplified; for example, for univariate time series or with parameter estimation (see Section 7.4), the complexity reduces to order $L^2$.

Figure 7.3 Example of the UT for mean and covariance propagation: (a) actual; (b) first-order linearization (EFK); (c) UT.

7.3 THE UNSCENTED KALMAN FILTER 231
UKF Algorithm: Predict

- The estimated state and covariance are augmented with the mean and covariance of the process noise:

\[
x_{t-1|t-1} = \begin{bmatrix} \hat{x}_{t-1|t-1}^T & E[u_t^T] \end{bmatrix}^T,
\]

\[
\Sigma_{t-1|t-1} = \begin{bmatrix} \Sigma_{t-1|t-1} & 0 \\ 0 & \Sigma_u \end{bmatrix}.
\]

- A set of $2L + 1$ sigma points is derived from the augmented state and covariance where $L$ is the dimension of the augmented state.

\[
\chi_{t-1|t-1}^0 = x_{t-1|t-1};
\]

\[
\chi_{t-1|t-1}^i = x_{t-1|t-1} + \left( \sqrt{(L + \lambda)\Sigma_{t-1|t-1}} \right)_i, \quad i = 1 \ldots L,
\]

\[
\chi_{t-1|t-1}^i = x_{t-1|t-1} - \left( \sqrt{(L + \lambda)\Sigma_{t-1|t-1}} \right)_{i-L}, \quad i = L + 1, \ldots 2L.
\]
UKF Algorithm: Update

- The predicted state and covariance are augmented as before, except now with the mean and covariance of the measurement noise:

  \[
  \mathbf{x}_{t-1|t-1} = \begin{bmatrix} \hat{x}_{t-1|t-1}^T & E[\mathbf{v}_t^T] \end{bmatrix}^T, \\
  \Sigma_{t-1|t-1} = \begin{bmatrix} \Sigma_{t-1|t-1} & 0 \\ 0 & \Sigma_v \end{bmatrix}.
  \]

- As before, a set of \(2L + 1\) sigma points is derived from the augmented state and covariance where \(L\) is the dimension of the augmented state.

  \[
  \begin{align*}
  \chi_{t-1|t-1}^0 &= \mathbf{x}_{t-1|t-1}, \\
  \chi_{t-1|t-1}^i &= \mathbf{x}_{t-1|t-1} + \left( \sqrt{(L + \lambda) \Sigma_{t-1|t-1}} \right)_i, \quad i = 1 \ldots L, \\
  \chi_{t-1|t-1}^i &= \mathbf{x}_{t-1|t-1} - \left( \sqrt{(L + \lambda) \Sigma_{t-1|t-1}} \right)_{i-L}, \quad i = L + 1, \ldots 2L.
  \end{align*}
  \]
**UKF Algorithm**

- **Initialization:** For $t = 0$, set
  
  $$\hat{x}_0 = E[\mathbf{x}_0],$$
  
  $$\Sigma_0 = E[(\mathbf{x}_0 - \hat{x}_0)(\mathbf{x}_0 - \hat{x}_0)^T].$$

- **Computation:** For $t = 1, 2, \ldots$ compute:

  State estimate propagation:  \( \hat{x}_{t|t-1} = f(\hat{x}_{t-1|t-1}) \)

  Error covariance propagation: \( \Sigma_{t|t-1} = F_{t-1} \Sigma_{t-1|t-1} F_{t-1}^T + \Sigma_u \)

  Kalman gain matrix: \( K_t = \Sigma_{t|t-1} H_t^T (H_t \Sigma_{t|t-1} H_t^T + \Sigma_v)^{-1} \)

  State estimate update: \( \hat{x}_{t|t} = \hat{x}_{t|t-1} + K_t (y_t - h(\hat{x}_{t|t-1})) \)

  Error covariance update: \( \Sigma_{t|t} = (I - K_t H_t) \Sigma_{t|t-1} \)
As has been discussed, the Kalman filter is a recursive algorithm providing the conditional expectation of the state $x_k$ given all observations $Y_{0:k}$ up to the current time $k$. In contrast, the Kalman smoother estimates the state given all observations past and future, $Y_{0:N}$, where $N$ is the final time.

Kalman smoothers are commonly used for applications such as trajectory planning, noncausal noise reduction, and the E-step in the EM algorithm.

Figure 7.6 illustrates the estimation of Mackey–Glass time series using a known model: (a) with the EKF; (b) with the UKF. (c) shows a comparison of estimation errors for the complete sequence.
Recovery of sparse signals

Recently the traditional Kalman filter has been employed for the recovery of (dynamic) sparse signals from noisy observations using notions from the theory of compressed sensing/sampling, such as the restricted isometry property and related probabilistic recovery arguments, for sequentially estimating the sparse state in intrinsically low-dimensional systems.