Improved Unscented Kalman Filtering For a Class of Nonlinear Systems

Alexandros C. Charalampidis George P. Papavassilopoulos

National Technical University of Athens

SFHMMY 3.0, April 2009, Thessaloniki, Greece

The Filtering Problem

- General Case

$$x_{k+1} = f(x_k, w_k)$$
$$y_k = h(x_k, v_k)$$

- Known: $(y_k)_k$, the distribution of x_0, w_k, v_k.
- Unknown: $(w_k)_k$, $(v_k)_k$, $(x_k)_k$.

The Filtering Problem

- General Case

$$x_{k+1} = f(x_k, w_k)$$
$$y_k = h(x_k, v_k)$$

- Known: $(y_k)_k$, the distribution of x_0, w_k, v_k.
- Unknown: $(w_k)_k$, $(v_k)_k$, $(x_k)_k$.
- Additive Noise

$$x_{k+1} = f(x_k) + w_k$$
$$y_k = h(x_k) + v_k$$

Recursive Estimation

- Problem: Find the distribution of x_{k+1} given y_{k+1} and the distribution of x_k.
- It can be solved exactly for finite state space or for a linear system with Gaussian noise.
- Bayes rule:

$$
\begin{aligned}
p(x_{k+1}|y_{1:k}) &= \int p(x_{k+1}|x_k)p(x_k|y_{1:k})dx_k \\
p(x_{k+1}|y_{1:k+1}) &= p(y_{k+1}|x_{k+1})p(x_{k+1}|y_{1:k})/c_k, \\
c_k &= \int p(y_{k+1}|x_{k+1})p(x_{k+1}|y_{1:k})dx_k
\end{aligned}
$$

Kalman Filter (Prediction Step)

- System:

$$x_{k+1} = A_k x_k + B_k + w_k$$
$$y_k = C_k x_k + D_k + v_k$$

Kalman Filter (Prediction Step)

- System:

$$x_{k+1} = A_k x_k + B_k + w_k$$
$$y_k = C_k x_k + D_k + v_k$$

- Prediction Step:

$$\hat{x}_{k+1}^{-} = A_k \hat{x}_k + B_k$$
$$P_{x_{k+1}}^{-} = A_k P_{x_k} A_k^T + Q$$

Kalman Filter (Correction Step)

- System:

$$x_{k+1} = A_k x_k + B_k + w_k$$
$$y_k = C_k x_k + D_k + v_k$$

- Mean and Covariance Matrices:

$$\hat{y}_{k+1}^{-} = C_{k+1}\hat{x}_{k+1}^{-} + D_k$$
$$P_{y_{k+1}} = C_{k+1} P_{x_{k+1}}^{-} C_{k+1}^{T} + R$$
$$P_{x_{k+1}y_{k+1}} = P_{x_{k+1}}^{-} C_{k+1}^{T}$$

Kalman Filter (Correction Step)

- System:

$$x_{k+1} = A_k x_k + B_k + w_k$$
$$y_k = C_k x_k + D_k + v_k$$

- Mean and Covariance Matrices:

$$\hat{y}_{k+1}^- = C_{k+1}\hat{x}_{k+1}^- + D_k$$
$$P_{y_{k+1}} = C_{k+1} P_{x_{k+1}}^- C_{k+1}^T + R$$
$$P_{x_{k+1}y_{k+1}} = P_{x_{k+1}}^- C_{k+1}^T$$

- Correction Step:

$$K_{k+1} = P_{x_{k+1}y_{k+1}} P_{y_{k+1}}^{-1}$$
$$\hat{x}_{k+1} = \hat{x}_{k+1}^- + K_{k+1}\left(y_{k+1} - \hat{y}_{k+1}^-\right)$$
$$P_{x_{k+1}} = P_{x_{k+1}}^- - K_{k+1} P_{y_{k+1}} K_{k+1}^T$$

Filtering for Nonlinear Systems

- Extended Kalman Filter (EKF): Linearize, then apply KF. Satisfactory only for small noise covariance or slight nonlinearities.

Filtering for Nonlinear Systems

Alexandros C. Charalampidis,
George P. Papavassilopoulos

- Extended Kalman Filter (EKF): Linearize, then apply KF. Satisfactory only for small noise covariance or slight nonlinearities.
- Unscented Kalman Filter (UKF):

  > *"It is easier to approximate a probability distribution than it is to approximate a nonlinear function or transformation".*

  But accurate approximation in high dimensional spaces leads to computational burden.

Filtering for Nonlinear Systems

- Extended Kalman Filter (EKF): Linearize, then apply KF. Satisfactory only for small noise covariance or slight nonlinearities.
- Unscented Kalman Filter (UKF):

  > *"It is easier to approximate a probability distribution than it is to approximate a nonlinear function or transformation".*

  But accurate approximation in high dimensional spaces leads to computational burden.
- The curse of dimensionality

The Class Under Study

$$x_{k+1} = Ax_k + \sum_{i=1}^{N_c} B_i g_i(D_i^T x_k) + w_k$$

$$y_{k,i} = h_i(C_i^T x_k) + v_{k,i}, i = 1, \ldots, N_o$$

g_i and h_i are nonlinear one-variable functions, $N_c, N_o \in \mathbb{N}$, while C_i and D_i are column vectors in $\mathbb{R}^{n_x}$.

Important Subclasses

- Systems with linear dynamics and nonlinear output.
- SISO linear systems with nonlinear feedback.
- MIMO linear systems with nonlinear decoupled feedback.
- Cascades of linear systems with nonlinear characteristics.

Important Subclasses

- Systems with linear dynamics and nonlinear output.
- SISO linear systems with nonlinear feedback.
- MIMO linear systems with nonlinear decoupled feedback.
- Cascades of linear systems with nonlinear characteristics.
- Arbitrary networks of linear systems interconnected with nonlinear characteristics.

Prediction Step

- Mean: $\mathbb{E}[x_{k+1}] = A\hat{x}_k + \sum_{i=1}^{N_c} B_i \mathbb{E}[g_i(D_i^T x_k)]$
- Covariance:

$$P^{-}_{x_{k+1}} = V[Ax_k] + V[\sum_{i=1}^{N_c} B_i g_i(D_i^T x_k)] +$$

$$Cov(Ax_k, \sum_{i=1}^{N_c} B_i g_i(D_i^T x_k)) + Cov(\sum_{i=1}^{N_c} B_i g_i(D_i^T x_k), Ax_k) + V[w_k]$$

Prediction Step

- Mean: $\mathbb{E}[x_{k+1}] = A\hat{x}_k + \sum_{i=1}^{N_c} B_i \mathbb{E}[g_i(D_i^T x_k)]$
- Covariance:

$$P_{x_{k+1}}^{-} = V[Ax_k] + V[\sum_{i=1}^{N_c} B_i g_i(D_i^T x_k)] +$$

$$Cov(Ax_k, \sum_{i=1}^{N_c} B_i g_i(D_i^T x_k)) + Cov(\sum_{i=1}^{N_c} B_i g_i(D_i^T x_k), Ax_k) + V[w_k]$$

- Terms of the following form appear:
  - $\mathbb{E}[g_i(D_i^T x_k)]$
  - $\mathbb{E}[g_i(D_i^T x_k) g_j(D_j^T x_k)]$
  - $\mathbb{E}[x_k g_i(D_i^T x_k)]$

Prediction Step

- Mean: $\mathbb{E}[x_{k+1}] = A\hat{x}_k + \sum_{i=1}^{N_c} B_i \mathbb{E}[g_i(D_i^T x_k)]$
- Covariance:

$$P^-_{x_{k+1}} = V[Ax_k] + V[\sum_{i=1}^{N_c} B_i g_i(D_i^T x_k)] +$$

$$Cov(Ax_k, \sum_{i=1}^{N_c} B_i g_i(D_i^T x_k)) + Cov(\sum_{i=1}^{N_c} B_i g_i(D_i^T x_k), Ax_k) + V[w_k]$$

- Terms of the following form appear:
  - $\mathbb{E}[g_i(D_i^T x_k)]$
  - $\mathbb{E}[g_i(D_i^T x_k) g_j(D_j^T x_k)]$
  - $\mathbb{E}[x_k g_i(D_i^T x_k)]$
- But the real valued $D_i^T x_k \sim N(D_i^T \hat{x}_k, D_i^T P_{x_k} D_i)$!

Correction Step

- Terms of the same forms are needed in order to calculate $\hat{y}_{k+1}^{-}$, $P_{y_{k+1}}$, $P_{x_{k+1},y_{k+1}}$. Then Kalman Filter Eqs. are used, as in standard UKF.

Correction Step

- Terms of the same forms are needed in order to calculate $\hat{y}_{k+1}^{-}$, $P_{y_{k+1}}$, $P_{x_{k+1},y_{k+1}}$. Then Kalman Filter Eqs. are used, as in standard UKF.
- When x is a normally distributed random variable, $\mathbb{E}[g_1(D_1^T x)]$ and $\mathbb{E}[g_1(D_1^T x)g_2(D_2^T x)]$ are integrals in $\mathbb{R}$ and $\mathbb{R}^2$ respectively, thus not very demanding to compute.

Correction Step

- Terms of the same forms are needed in order to calculate $\hat{y}_{k+1}^{-}$, $P_{y_{k+1}}$, $P_{x_{k+1},y_{k+1}}$. Then Kalman Filter Eqs. are used, as in standard UKF.
- When x is a normally distributed random variable, $\mathbb{E}[g_1(D_1^T x)]$ and $\mathbb{E}[g_1(D_1^T x)g_2(D_2^T x)]$ are integrals in $\mathbb{R}$ and $\mathbb{R}^2$ respectively, thus not very demanding to compute.
- Thus it remains to show how $S = \mathbb{E}[xg(C^T x)]$ can be computed effectively, when $x \sim N(M, P)$.

Dimension Reduction (1)

- It is easy to find $n-1$ linearly indepenendent vectors $\{v_i, i = 1, \ldots, n-1\}$ such that $v_i^T PC = 0$.
- Using the fact that $P > 0$, it can be proved that $\{v_1, \ldots, v_{n-1}, c\}$ are linearly indepenendent, too.
- Then $Cov(C^T x, v_i^T x) = C^T P v_i = 0$.
- Thus $v_i^T S = \mathbb{E}[v_i^T x g(C^T x)] = \mathbb{E}[v_i^T x]\mathbb{E}[g(C^T x)] = v_i^T M \mathbb{E}[g(C^T x)]$.

Dimension Reduction (2)

Thus S is the solution of

$$\begin{bmatrix} C^T \\ v_1^T \\ \vdots \\ v_{n-1}^T \end{bmatrix} S = \begin{bmatrix} \mathbb{E}[C^T x g(C^T x)] \\ v_1^T M \mathbb{E}[g(C^T x)] \\ \vdots \\ v_{n-1}^T M \mathbb{E}[g(C^T x)] \end{bmatrix}$$

The coefficient matrix is non-singular, while the right hand side terms need only one-dimensional integration to be computed.

Example 1 (Description)

A linear system with matrices equal to

$$A = \begin{bmatrix} 0.9 & 1 & 0 \\ 0 & 0.7794 & 1 \\ 0 & -0.2025 & 0.7794 \end{bmatrix}, \; b = \begin{bmatrix} 0 \\ 0 \\ 0.25 \end{bmatrix},$$

$$c = \begin{bmatrix} 0.3730 & 0 & 0 \end{bmatrix}$$

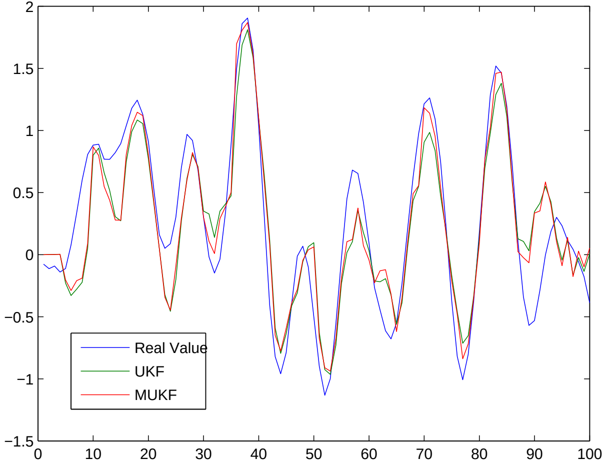
and whose transfer function is

$$G_{sys}(z) = \frac{0.093258}{(z - 0.9)(z^2 - 1.559z + 0.81)}$$

is driven by GWN following $N(0, 1)$. The sensor suffers from nonlinearity and noise so that $y(k) = s(k)^3 + v(k)$, where v is GWN following $N(0, 0.09)$. The goal is to estimate the output of the linear system.

Example 1 (Results)

One typical run



Example 1 (Results)

RMS error statistics from 100 runs

Estim. Techn.	Mean Value	Standard Deviation	Worst Case
EKF	0.9301	0.1735	1.3363
UKF	0.2992	0.0390	0.3970
MUKF	0.2844	0.0371	0.3643

Example 2

Suppose now that $y(k) = s(k)^3 \left(1 + 0.25\cos(20s(k))\right) + v(k)$.
The RMS error statistics from 100 runs are:

Estim. Techn.	Mean Value	Standard Deviation	Worst Case
EKF	0.9790	0.1794	1.5363
UKF	0.3265	0.0677	0.7334
MUKF	0.2896	0.0331	0.3757

Conclusion

- Reducing the integration in the n-dimensions to a number of n-dimensional linear systems and integration problems in one and two dimensions permits more accurate computations.
- Future research:
  - Exploiting the special structure of other classes.
  - Accounting for non-gaussianity.

Bibliography

- R. E. Kalman, "A new approach to linear filtering and prediction problem," in *ASME Journal of Basic Engineering*, Vol. 82, Issue 1, pp. 35-45, Mar. 1960.
- B. D. O. Anderson and J. B. Moore, *Optimal Filtering*. Englewood Cliffs, NJ: Prentice-Hall, 1979.
- S. Julier and J.K. Uhlmann, "Unscented filtering and nonlinear estimation," in *Proceedings of the IEEE*, Vol. 92, Issue 3, pp. 401-422, Mar. 2004.
- S. Julier, J. Uhlmann and H.F. Durrant-Whyte, "A New Method for the Nonlinear Transformation of Means and Covariances in Filters and Estimators," in *IEEE Transactions on Automatic Control*, Vol. 45, Issue 3, pp. 477-482, Mar. 2000.