Improvements over EKF and PF

Johan Bjurgert*

2014-01-14

Abstract

We analyse how issues with the Extended Kalman Filter (EKF) and the Particle Filter (PF) are solved by employing the Iterated Extended Kalman Filter (IEKF), the Gaussian Mixture Model (GMM) and the Unscented Kalman Filter (UKF). We show examples of where the EKF and PF fails and show how the filters we have chosen improve the estimation. We conclude that the performance of the EKF and the PF can be improved by using other filters and that filter choice is a trade-off between accuracy and computational complexity.

I. Introduction

The purpose of this study is to analyse the properties of IEKF, GMM and UKF as improvements over EKF and PF. How and when is EKF estimation improved by using IEKF. What properties does IEKF have? How does GMM relate to EKF and does it offer any improvements over PF, if so when and how? What properties does UKF have and does it offer any improvements over EKF?

As pointed out by Merwe [2], the EKF solution of propagating a Gaussian through a first order linearized nonlinear system can cause large estimation errors. The IEKF has an iterative measurement update which aims at improving the EKF update procedure by linearizing around the updated state rather than around the predicted. The UKF is another way of solving the same problem. It aims at solving the linearization inaccuracy by propagating so called sigma points through the non linearity and then compute an average and covariance of the propagated sigma points.
Another shortcoming of the EKF is the inability to keep many simultaneous hypotheses about a state. For such applications, the PF usually is of interest although it too has its shortcomings. The biggest problems with the PF is particle deprivation, having too few meaningful particles, and the filters high computational cost. The PF therefore has issues keeping hypothesis over time which is something that can be addressed by using the GMM which essentially is running multiple EKFs' simultaneously.

Outline of report

The theory section briefly outlines the EKF and the PF with a focus on their drawbacks. In the method section we present the filters we study along with their properties and we also mention implementational details. In the experiments section we show examples of where the EKF and the PF fails and how the studied filters improve the solution. The conclusion part contains findings and suggested future work.

II. Theory

The theory is adapted from [3] and [4].

The Extended Kalman Filter

The main difference between the Kalman filter and the EKF is that the measurement function h and state transition function g can be non-linear for the EKF.

$$\boldsymbol{x}_t = g(\boldsymbol{u}_t, \boldsymbol{x}_{t-1}) + \boldsymbol{\epsilon}_t$$
$$z_t = h(\boldsymbol{x}_t) + \delta_t$$

The EKF utilizes a first order Taylor expansion for g and h and thereby retains the Gaussian form of the posterior. This means that the error can be big if g and h are highly non-linear. When these errors can not be neglected, the estimates get inconsistent and the EKF may diverge.

*Project carried out together with Ioannis Karagiannis and Dirk van Dooren, reports written individually.

In terms of computational complexity, the EKF is almost cubic in k, the measurement dimension, and quadratic in n the state dimension $\mathcal{O}(k^{2.4} + n^2)$. The cubic part comes from the matrix inversion in calculating the Kalman gain and the quadratic part from matrix multiplication.

The Particle Filter

The particle filter uses the non-linear state transition model and measurement function as they are. The distribution is approximated by the particles, there is no assumption about distributions. Besides having a high computational complexity, the main problem with the PF is particle deprivation. It happens because particles with large weights tend to get drawn again whereas particles with low probability may not get drawn again. The diversity of the samples may decrease during resampling.

III. Methods

The methods explained are adapted from [3] and [4].

The Iterated Extended Kalman Filter

The IEKF is an iterative approach to linearize the measurement model around the updated state estimate.

Consider the estimation problem

$$\boldsymbol{\mu}_t = \underset{\boldsymbol{x}_t}{\operatorname{argmax}}[p(\boldsymbol{x}_t|z_{1:t}, \boldsymbol{u}_{1:t})] = \underset{\boldsymbol{x}_t}{\operatorname{argmax}}[G(z_t, h(\boldsymbol{x}_t), \boldsymbol{R}_t)G(\boldsymbol{x}_t, \bar{\boldsymbol{\mu}}_t, \bar{\boldsymbol{\Sigma}}_t)] = \underset{\boldsymbol{x}_t}{\operatorname{argmin}}[(z_t - h(\boldsymbol{x}_t))^{\mathsf{T}}\boldsymbol{R}_t^{-1}(z_t - h(\boldsymbol{x}_t)) + (\boldsymbol{x}_t - \bar{\boldsymbol{\mu}}_t)^T\bar{\boldsymbol{\Sigma}}_t^{-1}(\boldsymbol{x}_t - \bar{\boldsymbol{\mu}}_t)] = \underset{\boldsymbol{x}_t}{\operatorname{argmin}} f(\boldsymbol{x}_t)$$

The expression can be minimized through Newton-Rhapson iterations. Starting from $x_t^0 = \bar{\mu}_t$ and setting a stopping criterion, e.g $i = i_{max}$, or $|\bigtriangledown f(x_t)| < \lambda_\psi$ or by requiring that the mean suare error of the two consecutive estimates are sufficiently small, the following change in the update step of the extended Kalman filter is taking place

$$\boldsymbol{x}_t^{i+1} = \boldsymbol{x}_t^i + \boldsymbol{P}_i[\boldsymbol{H}^T(\boldsymbol{x}_t^i)\boldsymbol{R}_t^{-1}(z_t - h(\boldsymbol{x}_t^i)) + \boldsymbol{\Sigma}_t^{-1}(\bar{\boldsymbol{\mu}}_t - \boldsymbol{x}_t^i)]$$

where $\boldsymbol{P}_i = [\boldsymbol{H}^T(\boldsymbol{x}_t^i)\boldsymbol{R}_t^{-1}\boldsymbol{H}(\boldsymbol{x}_t^i) + \bar{\boldsymbol{\Sigma}}_t^{-1}]^{-1}$ is the Kalman update equation for the covariance matrix in the Kalman filter.
The method is expected to be more accurate but slower than the EKF depending on the stopping criteria and the amount of non-linearity.

Implementational details IEKF

• If there is no x_t that better minimizes f_t than the EKF solution then the IEKF estimation is the same as the EKF estimation. Also, if the solver gets stuck between two solutions, we take the EKF solution instead.
• Choose stopping criterion $|\bigtriangledown f(x_t)| < \lambda_\psi$ or iterations $> n$ or that the solution has converged or a combination.

The Gaussian Mixture Model

With a mixture of Gaussians, several EKF filters run simultaneously to account for the multiple hypotheses. This leads to a new form for the belief given by $bel(\mathbf{x}_t) = p(\mathbf{x}_t|\mathbf{z}_{1:t}, \mathbf{u}_{1:t}) = \frac{1}{\Sigma_i \psi_{t,i}} \Sigma_i \psi_{t,i} G(\mathbf{x}_t, \mathbf{z}_t, \Sigma_t)$, where $\psi_{t,i}$ is the mixture weight for hypothesis i at time t.

Each hypothesis i has different data association $\mathbf{c}_{1:t}$ and the belief of a particular hypothesis i is given by $bel_i(\mathbf{x}_t) = p(\mathbf{x}_t|\mathbf{z}_{1:t}, \mathbf{u}_{1:t}, \mathbf{c}_{1:t,i}) = G(\mathbf{x}_t, \mu_{t,i}, \Sigma_{t,i})$.

To derive the weight update, consider the posteriori over $\mathbf{x}_t$ and $\mathbf{c}_{t,i}$, $p(\mathbf{x}_t|\mathbf{z}_{1:t}, \mathbf{u}_{1:t}) = \Sigma_i p(\mathbf{x}_t|\mathbf{z}_{1:t}, \mathbf{u}_{1:t}, \mathbf{c}_{1:t,i}) p(\mathbf{c}_{1:t,i}|\mathbf{z}_{1:t}, \mathbf{u}_{1:t})$ where $\psi_{t,i} \propto p(\mathbf{c}_{1:t,i}|\mathbf{z}_{1:t}, \mathbf{u}_{1:t})$. Apply Bayes' rule and condition on $\mathbf{z}_t \implies \psi_{t,i} \propto \frac{p(\mathbf{z}_t|\mathbf{z}_{1:t-1}, \mathbf{u}_{1:t}, \mathbf{c}_{1:t,i}) p(\mathbf{c}_{1:t,i}|\mathbf{z}_{1:t-1}, \mathbf{u}_{1:t})}{p(\mathbf{z}_t|\mathbf{z}_{1:t-1}, \mathbf{u}_{1:t})} \propto p(\mathbf{z}_t|\mathbf{z}_{1:t-1}, \mathbf{u}_{1:t}, \mathbf{c}_{1:t,i})\psi_{t-1,i}$

Now that the mixture weights can be calculated it is possible to define a pruning rule to exclude hypotheses, $\frac{1}{\Sigma_i \psi_{t,i}} \leq \psi_{min}$.

Implementational details GMM

• We extended the EKF code to run several filters by looping over the existing code. This way it would be possible to combine the IEKF and the GMM.
• The mixture weights are started with a uniform distribution.
• The visual representation of the hypotheses was extended.
• To do global localization, the starting positions need to be spread out. We developed two procedures for this, either distribute the Gaussians evenly over the map or place them randomly within a circle around the landmarks.
• The pruning excludes unlikely hypotheses and additionally it is not allowed for hypotheses to be too similar. This makes it possible to start with lots of hypotheses without suffering from excessive slowdown. The complexity of the filter has a constant overhead of the EKF.

The Unscented Kalman Filter

The idea of the UKF is to pick a set of sample points called sigma points and propagate these through the non-linearities. $\{\mathcal{X}_t^i\} = \{\boldsymbol{\mu}_t, \boldsymbol{\mu}_t \pm (\sqrt{(n+\lambda)\boldsymbol{\Sigma}_t})_i\}$ where the i indexation denotes the i^{th} column of the matrix $\boldsymbol{\Sigma}$ and $\lambda = \alpha^2(n+\kappa) - n$, with α and κ being scale parameters determining how far the sigma points are spread from the mean.

Each sigma point has two sets of weights associated with it, w_m and w_c.

$$
\begin{aligned}
w_m^0 &= \lambda/(n+\lambda) \\
w_c^0 &= \lambda/(n+\lambda) + (1 - \alpha^2 + \beta) \\
w_m^i &= w_c^i = 1/2(n+\lambda) \quad i = 1 \ldots 2n
\end{aligned}
$$

The sigma points $\mathcal{X}$ are propagated through the non-linear function g and the resulting $\boldsymbol{\mu}'$ and $\boldsymbol{\Sigma}'$ are retrieved as

$$
\begin{aligned}
\mathcal{Y}^i &= g(\mathcal{X}^i) \\
\boldsymbol{\mu}' &= \sum_{i=0}^{2n} w_m^i \mathcal{Y}^i \\
\boldsymbol{\Sigma}' &= \sum_{i=0}^{2n} w_c^i (\mathcal{Y}^i - \boldsymbol{\mu}')(\mathcal{Y}^i - \boldsymbol{\mu}')^\top
\end{aligned}
$$

The prediction is done as described and g is the process model. For the covariance estimate, the process covariance $\mathbf{R}_t$ is added to $\boldsymbol{\Sigma}'$. Similarly for the update, the sigma points are propagated through the non-linear observation model $\mathcal{Z}_t = h(\mathcal{X}_t)$ and additionally, the covariance $\boldsymbol{S}_t$ and cross-covariance $\boldsymbol{\Sigma}^{x,z}$ are computed

$$
\begin{aligned}
\hat{z}_t &= \sum_{i=0}^{2n} w_m^i \mathcal{Z}_t^i \\
\boldsymbol{S}_t &= \sum_{i=0}^{2n} w_c^i (\mathcal{Z}_t^i - \hat{z}_t)(\mathcal{Z}_t^i - \hat{z}_t)^\top \\
\boldsymbol{\Sigma}^{x,z} &= \sum_{i=0}^{2n} w_c^i (\mathcal{X}^i - \boldsymbol{\mu}')(\mathcal{Z}^i - \hat{z}_t)^\top
\end{aligned}
$$

The Kalman gain $\boldsymbol{K}$ and the posterior estimates $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are then calculated as

$$
\begin{aligned}
&\boldsymbol{K}_t = \boldsymbol{\Sigma}_t^{x,z} \boldsymbol{S}_t^{-1} \\
&\boldsymbol{\mu}_t = \bar{\boldsymbol{\mu}}_t + \boldsymbol{K}_t (z_t - \hat{z}_t) \\
&\boldsymbol{\Sigma}_t = \bar{\boldsymbol{\Sigma}}_t - \boldsymbol{K}_t \boldsymbol{S}_t \boldsymbol{K}_t^\top
\end{aligned}
$$

The computational complexity of the UKF is almost the same as the EKF, the UKF has a constant overhead over the EKF.

Implementational details UKF

• The sigma points were computed using Cholesky decomposition, which gives numerical robustness.
• In the batch association we used known correspondences.

Segmentation of hypotheses for the PF

In order to compare the GMM with the PF, we also needed to group hypotheses of the PF.

1	Initialize grid, define coordinate system
2	Update grid for each particle
3	Find local maxima, return coordinates
4	Compute distance between local maxima
5	Assign particles to local maxima

In step 1, the grid has to be small (20x20) otherwise it is likely that multiple local maxima are detected. For step 2, particles fall into bins, so the grid is basically a histogram in 2D. In step 3, the local maxima are computed using an 8 connected kernel. In step 4 we want to include all particles within a certain distance from a maximum. The maximum distance is half the distance between the closest pairs, so that the circles do not overlap. In table 1 we see groups of particles that represent the hypotheses. For our application we also need to track the hypotheses which is straightforward to do, now that we can group the hypotheses at each time step.

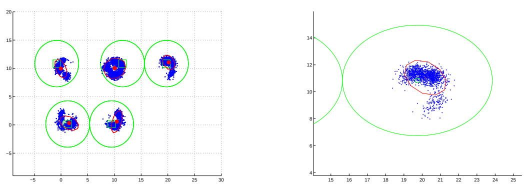

Table 1: Segmentation of hypotheses for PF

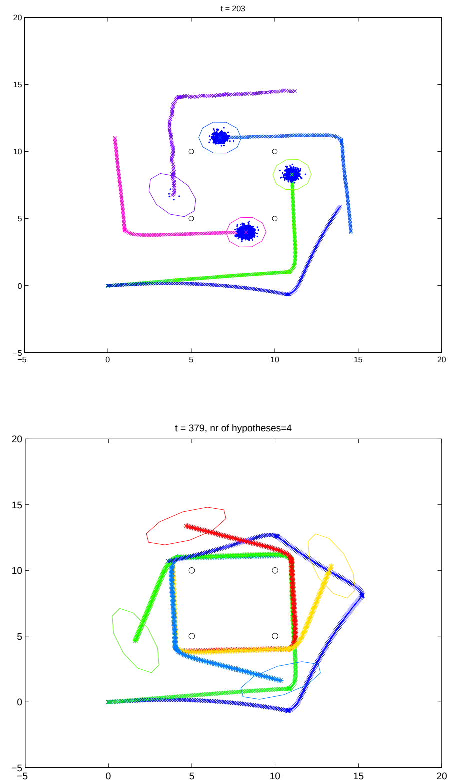

Table 2: Losing versus keeping hypotheses

IV. Experiments

In this section we investigate how the studied filters handle the situation where the EKF and the PF have problems. We use the maps and datasets from previous labs.

I. Handling multiple hypotheses

As previously addressed, the PF may lose a hypothesis due to particle deprivation, see table 2 for an example where the above plot is the PF and the below plot is the GMM. In this run we use a symmetric map with four possible hypotheses. In the example we see the particle deprivation of the PF and that the GMM reliably keeps all hypotheses. Another PF example is shown in table 3. The PF does not see a landmark for some time, particles are spread out and new hypotheses cluster form. This can not happen for the GMM. It is worth mentioning that in this example, the PF converges faster than the GMM as it sees the fifth landmark. This is due to the pruning step of the GMM. If set too high, the initially uniformly distributed probabilities would not get accepted.

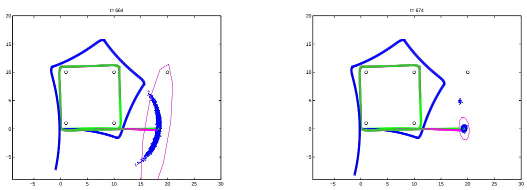

Table 3: Particles spreading and branching off

II. Improved handling of nonlinearity

As a result of the improved handling of nonlinearity in IEKF, it is more robust than the EKF for tracking. Consider table 4, where the filters are started with an initial pose of $(0.8, 0.8, \frac{\pi}{8})$ instead of the ground-truth (0,0,0). The IEKF converges whereas the EKF fails to locate. How far off the initial position can be for the IEKF to properly locate depends on the problem at hand but the region will be bigger for the IEKF.

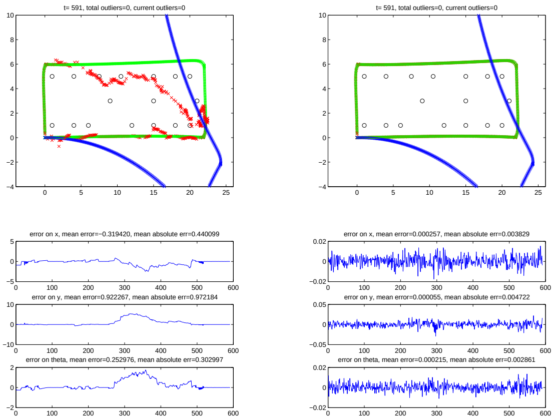

Table 4: EKF vs IEKF

Filter choice is about trade-off, consider the situation in table 5, where the EKF performs reasonably and the IEKF shows an improvement at the cost of a higher computational time. The EKF performs reasonably also for mildly non-linear functions.

Unfortunately, we could not get sensible comparisons for the UKF when more than one landmark was observed at the same time step. Therefore we could not compare the UKF with the other filters in the way we wanted. We could however show that when there was only one feature visible, the UKF has a substantially lower error than the EKF. The black crosses indicate that one more feature is actively measured. These events coincide with the error growing and eventually the UKF is far off. See table 7 where the upper plot is the UKF tracking.

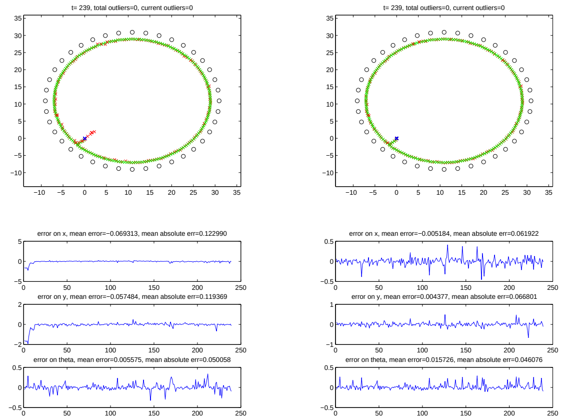

Table 5: IEKF better than EKF at cost of computational time

Zooming in on a time step where only one landmark was observed, we compared the estimated moments of the EKF, IEKF, PF and UKF, see table 6. IEKF and EKF are similar, but not equal, both in terms of μ and Σ and are overly optimistic about the estimation. The PF and UKF are also similar, but not equal. They reflect both the true position and shape of the uncertainty better than the IEKF and EKF. The optimistic EKF estimate can be a problem. It can prevent the mean from moving to a better estimate.

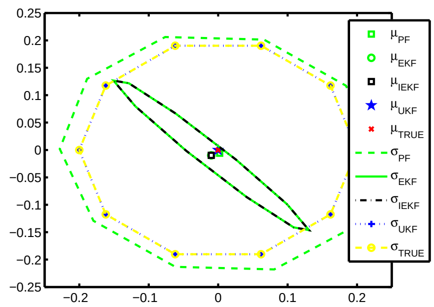

Table 6: Comparison of μ and Σ

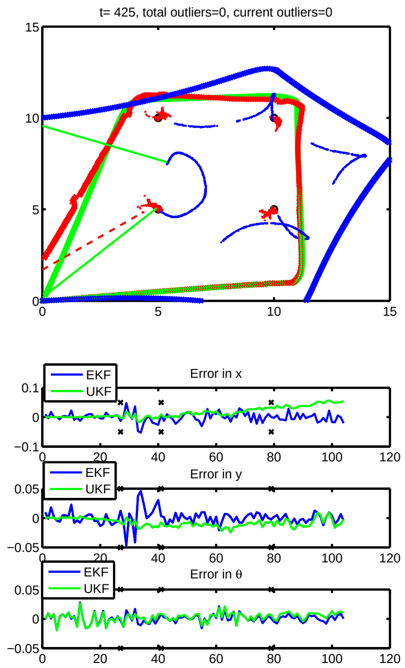

Table 7: UKF vs EKF

V. Conclusions

- It has been verified that the GMM has the property of keeping hypotheses where the PF potentially would lose them.
- It has been verified that IEKF is more accurate than the EKF and that the IEKF is more robust to use in localization.
- We have given some evidence that the UKF is more accurate than the EKF when only one feature is observed.
- The improved handling of non-linearities is useful if measurement data is very accurate.

We conclude by giving a summary of the filters in table 8.

Further work

It would have been interesting to get the UKF working properly for multiple features and to then carry out more comparisons between EKF, UKF and IEKF. One way of achieving this could be to follow along the work of [1].

References

[1] Colin McManus and Timothy D. Barfoot *A Serial Approach to Handling High-Dimensional Measurements in the Sigma-Point Kalman Filter* Proceedings of Robotics: Science and Systems,2011,Los Angeles, CA, USA,June

[2] Eric A. Wan and Rudolph van der Merwe *The Unscented Kalman Filter for Nonlinear Estimation* In Proceedings of Symposium 2000 on Adaptive Systems for Signal Processing, Communications, and Control (AS-SPCC) , pages 153 – 158, Lake Louise, Alberta, Canada, Oct. 1 – 4 2000

[3] John Folkesson *Lecture notes, KTH-EL2320 Applied Estimation*

[4] S. Thrun, W. Burgard, D. Fox *Probabilistic Robotics* (Intelligent Robotics and Autonomous Agents). The MIT Press, 2005.

Table 8: Summary of findings

Filter	Properties	
	Advantages	Disadvantages
EKF	Works well for mildly nonlinear systems	May diverge for highly nonlinear systems
GMM	Keeps valid hypotheses over time	Same as for EKF
IEKF	Increased accuracy for nonlinear measurements	May have high complexity
UKF	Captures μ and Σ better for nonlinear systems	Extend to handle multiple landmarks
PF	Asymptotically correct estimation	Very high complexity