Adaptive M-estimation for Robust Cubature Kalman Filtering

Changliang Zhang†, Ruirui Zhi†, Tiancheng Li‡ and Juan M. Corchado‡

†Northwestern Polytechnical University, China ‡University of Salamanca, Spain

Introduction

As a l_1/l_2 norms-based estimation method, Huber's M-estimation [1] has provided an efficient method to deal with measurement outliers for robust filtering, which has been applied to the cubature Kalman filter (CKF), namely M-estimation based robust CKF (HCKF) and its square-root version (HSCKF). In particular, the CKF/SCKF is suitable for high dimensional systems and enjoys better numerical stability, computational efficiency and accuracy[2-4].

Huber's M-estimation is based on generalized maximum likelihood estimation that can handle clutter in both processes and measurements. The robust M-estimation methods however cannot adjust the noise covariance adaptively when it does not match with the truth. To further handle abnormal measurement noise, an adaptive method is proposed in this paper to adjust the measurement noise covariance used in the Huber's M-estimation approach **based on the difference between actual and theoretical/predicted innovation covariance,** to gain adaptivity in addition to the robustness. The approach is applied within the HCKF and AHCKF, leading to adaptive HCKF (AHCKF) and adaptive HSCKF (AHSCKF). Simulation results on a typical target tracking model have demonstrated their advantages over existing approaches in terms of estimate accuracy, outlier-robustness and reliability.

Huber's M-estimation

As usual, the estimation model can be described as a discrete-time state-space model.

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

where k is the time, $x_k \in R^n$ is the state, $y_k \in R^{n_y}$ is the measurement, $f(\cdot)$ and $h(\cdot)$ are the state transition and measurement functions, w_k and v_k are the zero-mean Gaussian process noise and zero-mean Gaussian measurement noise with covariance matrixes Q and R respectively.

Given the cost function $J(x) = \rho\left(\Sigma^{-1/2}(x-\mu)\right)$ (comparing estimate x to truth u), more specifically, supposing the predicted state $\hat{x}_{k|k-1}$ and the measurement matrix H_k, the measurement function can be reformulated in matrix as a linear regression model

$$\begin{bmatrix} y_k - h(x_{k|k-1}) + H_k \hat{x}_{k|k-1} \\ \hat{x}_{k|k-1} \end{bmatrix} = \begin{bmatrix} H_k \\ I \end{bmatrix} x_k + \begin{bmatrix} v_k \\ -\delta x_k \end{bmatrix} \quad \Rightarrow \quad z_k = M_k x_k + u_k$$

where H_k is the measurement matrix, δx_k is state prediction error with covariance $P_{k|k-1}$.

Huber's M-estimation minimizes the cost function: $J(x_k) = \sum_{i=1}^{n_y+n} \rho(u_i)$

where u_i refers to the ith component of the residual vector $u = M_k x_k - z_K$, and the function $\rho(\cdot)$ is known as the "score function", which can be defined as

$$\rho(u_i) = \begin{cases} \frac{1}{2}u_i^2, & |u_i| \le c \\ c|u_i| - \frac{1}{2}c^2, & |u_i| > c \end{cases}$$

where c is a tuning threshold.

The state estimate error covariance matrix can be iteratively computed from $P_{k|k} = (M_k^T W M_k)^{-1}$. Usually, the number of iterations is only one.

Adaptive M-estimation based CKF/SCKF

If something occur "unexpectedly", e.g. (estimated) noise covariance differs significantly from the truth, the filter will deteriorate or even fail. For this, to **render the M-estimation immunized to abnormal measurements**, we propose to online compare the innovation covariance between the theoretical expectation (prediction) and the truth for abnormal detection.

Given that the received measurements are normal, the true innovation covariance Σ_k equals approximately to the calculated one $P_{yy,k|k-1}$. (the calculation is given in the paper)

Two strategies to compute the true innovation covariance:

- Sliding window strategy to compute Σ_k is based on the former m>1 innovation estimator.
- Use the current innovation vector only for computing the innovation covariance.

When an abnormal measurement occurs, we can get an equation about the variant scalar factor a_k.

$$a_k = \left(tr(\Sigma_k) - tr(P_{y,k|k-1})\right) / tr(R_k)$$

where $tr(\cdot)$ denotes the trace of the related matrix.

The adaptive factor a_k will be close to 1 when and only when the measurement is normal. This provides a measure to detect the measurement abnormality: once abnormal measurements occur, a larger factor a_k shall be (calculated and) used for adjusting.

In general, given $y_k \sim N(\hat{y}_{k|k-1}, P_{yy,k|k-1})$, $e_k \sim N(0, P_{yy,k|k-1})$ we have

$$\bar{e}_k = V_k^T(y_k - \hat{y}_{k|k-1}) = V_k^T e_k$$
$$\Lambda_k = V_k^T P_{yy,k|k-1} V_k$$

where the column vector of V_k is standardized feature vector, satisfying $V_k^T = V_k^{-1}$ and $P_{yy,k|k-1} = V_k \Lambda_k V_k^T$, so we have $\bar{e}_k \sim N(0, \Lambda_k)$ which can be used for normality test as it only holds for the normal measurement [5], which treats e_k as a whole unit without distinguishing its dimensions.

In contrast, we propose to treat each dimension separately for finer (also more sensitive) test. This is easy to implement and feasible because Λ_k is a diagonal matrix for which the elements of vector $\bar{e}_k$ are independent with each other (we assume conditional independence between dimensions).

The measurement data is identified to be abnormal when and only when at least one dimension does not meet the normality test. The adaptive factor a_k is then given as follows

$$a_k = \begin{cases} \dfrac{tr(\Sigma_k) - tr(P_{y,k|k-1})}{tr(R_k)}, & \text{any } i: |\bar{e}_{i,k}| > N_{i,\alpha_i} \\ 1, & \text{otherwise} \end{cases}$$

The probability of abnormal $P\{|\bar{e}_{i,k}| > N_{i,\alpha_i}\} = 1 - \alpha_i, 0 < 1 - \alpha_i < 1$, a smaller test threshold N_{i,α_i} implies more sensitive detection, depending on the user's preference. We recommend $N_{i,\alpha_i} \in [\Lambda_{i,k}, 5\Lambda_{i,k}]$, in which $N_{i,\alpha_i} = 3\Lambda_{i,k}$ corresponds to a confidence coefficient 99.74%.

Once abnormal is confirmed, the gained adaptive factor a_k will be used to replace the measurement noise covariance R by $a_k R$ in the KF, such as the HCKF or HSCKF. This will enable the filter to detect abnormal measurement and to apply adjusting by a_k when abnormal is detected. This leads to the so-called adaptive HCKF (AHCKF) and adaptive HSCKF (AHCKF) respectively in our paper.

However, when the innovation sequence is large, adaptive and robust methods turn out to be two opposite strategies [6]. Therefore, our approach is exposed to the risk of over-reaction to the sensor data, We will further consider this in the future work. One potential solution is to fit sensor data over time [7]

Simulation

To verify the effectiveness of the proposed adaptive M-estimation method in both AHCKF and AHSCKF, we use the same target tracking model as given in [2] with the same parameters. The true target is initialized with $N(x_0, P_{0|0})$, where $x_0 = [1000\text{m} \;\; 300\text{ms}^{-1} \;\; 1000\text{m} \;\; 0\text{ms}^{-1} \;\; -3°\text{s}^{-1}]$, $P_{0|0} = \text{diag}[100\text{m}^2 \;\; 10\text{m}^2\text{s}^{-2} \;\; 100\text{m}^2 \;\; 10\text{m}^2\text{s}^{-2} \;\; 100\text{mrad}^2\text{s}^{-2}]$.

The measurement data generated by the normal measurement model are denoted as '***data*0**' namely the healthy data. On the other hand, morbid conditions is also simulated based on ***data*0** morbid measurement data by uding a deviations $[75\text{ m} \;\; 0.2\text{ rad}]^T$ on the measurement noise w_k when $20 < k < 60$ and $mod(k, 8) = 0$, where $mod(a, b)$ gives the remainder of a/b, giving ***data1***. More types of abnormal measurement situations are simulated and analysed in the paper.

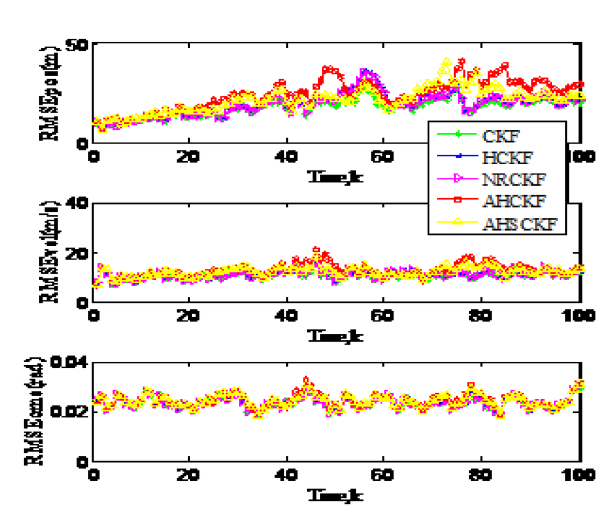

Fig.1 RMSE of position, velocity and turn rate when using '***data*0**'

In Fig.1, on '***data*0**', the performances of all CKFs are almost the same, indicating that robust filtering design is not necessary in this case while it will not cause side effects (except slightly additional computation).

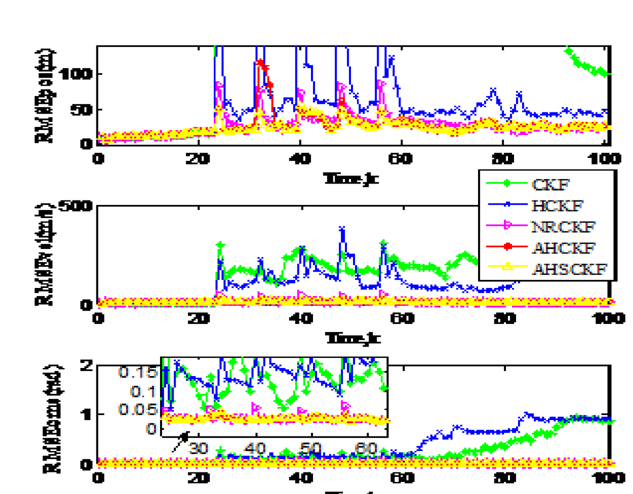

Fig.2 RMSE of position, velocity and turn rate when using '*data*1'

In Fig. 2 on '***data1***', the estimation errors given by HCKF and AHCKF have significant "peaks" (indicating very worse accuracy as compared to the proposed adaptive filter) when abnormal measurements occur. In these cases, HCKF does not work well, because the score function rescales the measurement covariance weakly. In contrast, AHCKF benefits much from the use of the adaptive factor a_k to adapt according to abnormal noises.

Conclusion

An adaptive M-estimation method has been proposed to adjust the measurement noise covariance for robust recursive estimation to accommodate abnormal measurement noise, which is quite easy to realize. It has been implemented in two versions of robust CKFs: HCKF and HSCKF. Simulations on a typical target tracking model demonstrate the effectiveness of the proposed method.

References

[1] P.J. Huber, "Robust estimation of a location parameter," The Annals of Mathematical Statistics, vol. 35, pp.73-101, 1964.

[2] I. Arasaratnam, S. Haykin, "Cubature kalman filters," IEEE Trans. Automatic Control, vol. 54, pp. 1254-1269, 2009.

[3] Y. Huang, L.H. Wu, and F. Sun. Robust Cubature Kalman filter based on Huber M estimator. Control and Decision, vol. 29, no.3, p. 572-576, 2014.

[4] K. Li, B. Hu, L. B. Chang, Y. Li, Robust square-root cubature Kalman filter based on Huber's M-estimation methodology, Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering June, vol. 229 no. 7 1236-1245, 2015.

[5] C. Hajiye, H.E. Soken. A novel adaptive unscented Kalman filter for pico satellite attitude estimation[C] //PHYSCON 2011, September 5-8, 2011, León, Spain. 2011: 1-7.

[6] G.B. Chang, "Kalman filter with both adaptivity and robustness," Journal of Process Control, vol. 24, pp. 81-87, 2014.

[7] T. Li, J. Prieto, J.M. Corchado, *Fitting for Smoothing: A Methodology for Continuous-time Target Track Estimation*, 2016 International Conference on Indoor Positioning and Indoor Navigation (IPIN), 4-7 October 2016, Madrid, Spain