Constrained State Estimation Using the Unscented Kalman Filter

Rambabu Kandepu, Lars Imsland and Bjarne A. Foss

Abstract— **A simple procedure to include state inequality constraints in the Unscented Kalman Filter is proposed. With this procedure, the information of active state constraints influences the state covariance matrix, resulting in better estimates. In a numerical example, the approach outperforms the Extended Kalman Filter implemented with constraint handling via "clipping".**

I. Introduction

In the process industries one of the main goals is to make the end product at the lowest possible cost while satisfying product quality constraints. State estimation often play an important role in accomplishing this goal in process control and performance monitoring applications. There are many uncertainties to deal with in process control; model uncertainties, measurement uncertainties and uncertainties in terms of different noise sources acting on the system. In this kind of environment, representing the model state by an (approximated) probability density function (pdf) has distinct advantages. State estimation is a means to propagate the pdf of the system states over time in some optimal way. It is most common to use the Gaussian pdf to represent the model state, process and measurement noises. The Gaussian pdf can be characterized by its mean and covariance. The Kalman Filter (KF) propagates the mean and covariance of the pdf of the model state in an optimal (minimum mean square error) way in case of linear dynamic systems [5].

All practical systems possess some degree of nonlinearity. Depending on the type of process and the operating region of the process, some processes can be approximated with a linear model and the KF can be used for state estimation. In some cases the linear approximation may not be accurate enough, and state estimator designs using nonlinear process models are necessary. The most common way of applying the KF to a nonlinear system is in the form of the Extended Kalman Filter (EKF). In the EKF, the pdf is propagated through a linear approximation of the system around the operating point at each time instant. In doing so, the EKF needs the Jacobian matrices which may be difficult to obtain for higher order systems, especially in the case of time-critical applications. Further, the linear approximation of the system at a given time instant may introduce errors in the state which may lead the state to diverge over time. In other words, the linear approximation may not be appropriate for some systems. In order to overcome the drawbacks of the

R. Kandepu is with Elkem Research, N-4675 Kristiansand, Norway. `Rambabu.Kandepu@elkem.no`

L. Imsland is with Cybernetica AS, Leirfossveien 27, N-7038 Trondheim, Norway. `Lars.Imsland@cybernetica.no`

B. A. Foss is with Dept. of Engineering Cybernetics, NTNU, N-7491 Trondheim, Norway. `Bjarne.Foss@itk.ntnu.no`

EKF, other nonlinear state estimators have been developed such as the Unscented Kalman Filter (UKF) [1], the Ensemble Kalman Filter (EnKF)[4] and high order EKFs. The UKF seems to be a promising alternative for process control applications [7], [8], [12]. The UKF propagates the pdf in a simple and effective way and it is accurate up to second order in estimating mean and covariance [1].

In process control, state constraint issues are often important in connection with estimation. In this context, state constraints typically refer to the boundaries of the physical domain of estimated states (e.g., estimated concentrations should remain positive). The only approach that naturally incorporates such constraints, moving horizon estimation (MHE), has drawbacks related to online computational load since it relies on solving a nonlinear programming problem at each sample step. There are some approaches in the literature extending the EKF in this direction (e.g. [10], [11]), but as far as we are aware, there have been no such attempts for the UKF except the work reported in [13]. The aim of this paper is to demonstrate how a simple projection of the sigma points can give good constraint handling in the UKF, while applying the same projection to the EKF estimate does not give good performance.

II. UKF Algorithm

To illustrate the principle behind the UKF, consider the following example. Let $x \in \mathbb{R}^n$ be a random vector and

$$y = g(x) \tag{1}$$

be a nonlinear function, $g : \mathbb{R}^n \to \mathbb{R}^m$. The question is how the UKF compute pdf of y given the pdf of x, in other words, how to calculate the mean $\left(\overline{y}^{UKF}\right)$ and covariance $\left(P_y^{UKF}\right)$ of y, in the case of being Gaussian? Consider a set of points

$$x^{(i)}, \ i \in \{1, ..., p\}, \ p = 2n + 1$$

(similar to the random samples of a specific distribution function in Monte Carlo simulations) with each point being associated with a weight $w^{(i)}$. These sample points are termed as sigma points. Then the following steps are involved in approximating the mean and covariance: Propagate each sigma point through the nonlinear function,

$$y^{(i)} = g(x^{(i)})$$

- the mean is approximated by the weighted average of the transformed points,

$$\overline{y}^{UKF} = \sum_{i=0}^{p} w^{(i)} y^{(i)}, \quad \overset{p}{\underset{i=0}{\Sigma}} w^{(i)} = 1$$
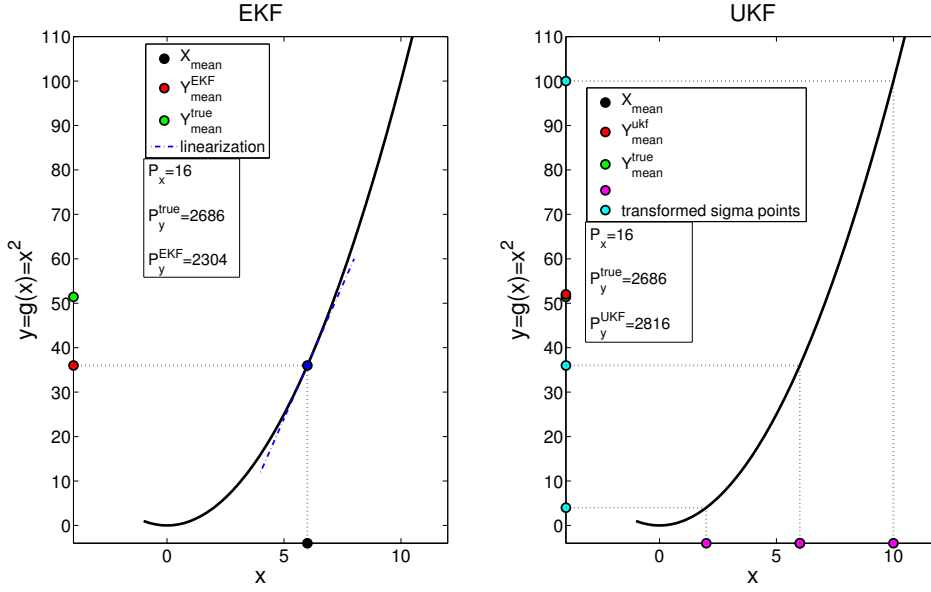
Fig. 1.   Conceptual difference between EKF and UKF

- and the covariance is computed by the weighted outer product of the transformed points,

$$P_y^{UKF} = \sum_{i=0}^{p} w^{(i)} \left(y^{(i)} - \overline{y}^{UKF} \right) \left(y^{(i)} - \overline{y}^{UKF} \right)^T.$$

Both the sigma points and the weights are computed deterministically through a set of conditions given in [1].

The conceptual difference between the principle behind the familiar EKF and the UKF is illustrated in Figure 1.

Algorithm

The UKF algorithm is presented below; for background theory, refer to [2], [1] and [3]. Let a general nonlinear system be represented by the following standard discrete time equations:

$$x_k = f(x_{k-1}, v_{k-1}, u_{k-1}) \quad (2)$$

$$y_k = h\left(x_k, n_k, u_k\right). \quad (3)$$

An augmented state at time instant k,

$$x_k^a \triangleq \begin{bmatrix} x_k \\ v_k \\ n_k \end{bmatrix} \quad (4)$$

is defined. The augmented state dimension is

$$N = n_x + n_v + n_n. \quad (5)$$

Similarly, the augmented state covariance matrix is built from the covariance matrices of x, v and n,

$$P^a \triangleq \begin{bmatrix} P_x & 0 & 0 \\ 0 & P_v & 0 \\ 0 & 0 & P_n \end{bmatrix} \quad (6)$$

where P_v and P_n are the process and observation noise covariance matrices.

- *Initialization at $k = 0$:*

$$\widehat{x}_0 = E[x_0], \quad P_{x_0} = E\left[(x_0 - \widehat{x}_0)(x_0 - \widehat{x}_0)^T\right]$$

$$\widehat{x}_0^a = E[x^a] = E[\widehat{x}_0 \quad 0 \quad 0]^T$$

$$P_0^a = E\left[(x_0^a - \widehat{x}_0^a)(x_0^a - \widehat{x}_0^a)^T\right] = \begin{bmatrix} P_x & 0 & 0 \\ 0 & P_v & 0 \\ 0 & 0 & P_n \end{bmatrix}$$

- *For $k = 1, 2, ...\infty$:*

  1) Calculate $2N + 1$ sigma-points based on the present state covariance:

  $$\mathbf{X}_{i,k-1}^a \triangleq \begin{cases} \widehat{x}_{k-1}^a, & i = 0 \\ \widehat{x}_{k-1}^a + \gamma \mathbf{S}_i, & i = 1, ..., N \\ \widehat{x}_{k-1}^a - \gamma \mathbf{S}_i, & i = N+1, ..., 2N \end{cases} \quad (7)$$

  where $\mathbf{S}_i$ is the ith column of the matrix

  $$S = \sqrt{P_{k-1}^a}.$$

  In (7) γ is a scaling parameter [2],

  $$\gamma = \sqrt{N + \lambda}, \quad \lambda = \alpha^2(N + \kappa) - N$$

  where α and κ are tuning parameters. We must choose $\kappa \geq 0$ to guarantee the semi-positive definiteness of the covariance matrix, a good default choice is $\kappa = 0$. The parameter α, $0 \leq \alpha \leq 1$, controls the size of the sigma-point distribution and it should ideally be a small number [2].

  The ith sigma point (augmented) is the ith column of the sigma point matrix,

  $$\mathbf{X}_{i,k-1}^a = \begin{bmatrix} \mathbf{X}_{i,k-1}^x \\ \mathbf{X}_{i,k-1}^v \\ \mathbf{X}_{i,k-1}^n \end{bmatrix}$$

where the superscripts x, v and n refer to a partition conformal to the dimensions of the state, process noise and measurement noise respectively.

2) Time-update equations:
Transform the sigma points through the state-update function,

$$\mathbf{X}_{i,k|k-1}^x = f\left(\mathbf{X}_{i,k-1}^x, \mathbf{X}_{i,k-1}^v, u_{k-1}\right), \quad (8)$$
$$i = 0, 1, ..., 2N$$

Calculate the apriori state estimate and apriori covariance,

$$\widehat{x_k^-} = \sum_{i=0}^{2N} \left(w_m^{(i)} \mathbf{X}_{i,k|k-1}^x\right), \quad (9)$$

$$P_{x_k}^- = \sum_{i=0}^{2N} w_c^{(i)} \left(\mathbf{X}_{i,k|k-1}^x - \widehat{x_k^-}\right)\left(\mathbf{X}_{i,k|k-1}^x - \widehat{x_k^-}\right)^T \quad (10)$$

The weights $w_m^{(i)}$ and $w_c^{(i)}$ are defined as,

$$w_m^{(0)} = \frac{\lambda}{N+\lambda}, \quad i=0,$$
$$w_c^{(0)} = \frac{\lambda}{N+\lambda} + (1 - \alpha^2 + \beta), \quad i=0,$$
$$w_m^{(i)} = w_c^{(i)} = \frac{1}{2(N+\lambda)}, \quad i=1,...,2N$$

where β is a non-negative weighting parameter introduced to affect the weighting of the zeroth sigma-point for the calculation of the covariance. This parameter (β) can be used to incorporate knowledge of the higher order moments of the distribution. For a Gaussian prior the optimal choice is $\beta = 2$ [2].

3) Measurement-update equations:
Transform the sigma points through the measurement update function,

$$\mathbf{Y}_{i,k|k-1} = h\left(\mathbf{X}_{i,k|k-1}^x, \mathbf{X}_{k-1}^n, u_k\right), \quad (11)$$
$$i = 0, 1, ..., 2N$$

and the mean and covariance of the measurement vector is calculated,

$$\widehat{y_k^-} = \sum_{i=0}^{2N} w_m^{(i)} \mathbf{Y}_{i,k|k-1},$$

$$P_{\overline{y}_k} = \sum_{i=0}^{2N} w_c^{(i)} \left(\mathbf{Y}_{i,k|k-1} - \widehat{y_k^-}\right)\left(\mathbf{Y}_{i,k|k-1} - \widehat{y_k^-}\right)^T.$$

The cross covariance is calculated according to

$$P_{x_k y_k} = \sum_{i=0}^{2N} w_c^{(i)} \left(\mathbf{X}_{i,k|k-1}^x - \widehat{x_k^-}\right)\left(\mathbf{Y}_{i,k|k-1} - \widehat{y_k^-}\right)^T$$

The Kalman gain is given by,

$$K_k = P_{x_k y_k} P_{\overline{y}_k}^{-1}$$

and the UKF estimate and its covariance are computed from the standard Kalman update equations,

$$\widehat{x}_k = \widehat{x_k^-} + K_k\left(y_k - \widehat{y_k^-}\right), \quad (12)$$
$$P_{x_k} = P_{x_k}^- - K_k P_{\overline{y}_k} K_k^T. \quad (13)$$

III. STATE ESTIMATION WITH CONSTRAINTS

Constraints on states to be estimated are important model information that is often not used in state estimation. Typically, such constraints are due to physical limitations on the states; for instance, estimated concentrations should remain positive. In Kalman filter theory, there is no general way of incorporating these constraints into the estimation problem. However, the constraints can be incorporated in the KF by projecting the unconstrained KF estimates onto the boundary of the feasible region at each time step [10], [11]. An other way of nonlinear state estimation with constraints is Moving Horizon Estimation (MHE), in which the constraints can be included in the estimation problem in a natural way [9]. In MHE, the state trajectory is computed taking state constraints into account at the expense of solving a nonlinear programming problem at each time step. The numerical optimization at each time step may be a challenge in time-critical applications. In this section, a new and simple method is introduced to handle state constraints in the UKF and it is compared to the standard way of constraint handling in the EKF, known as 'clipping' [6].
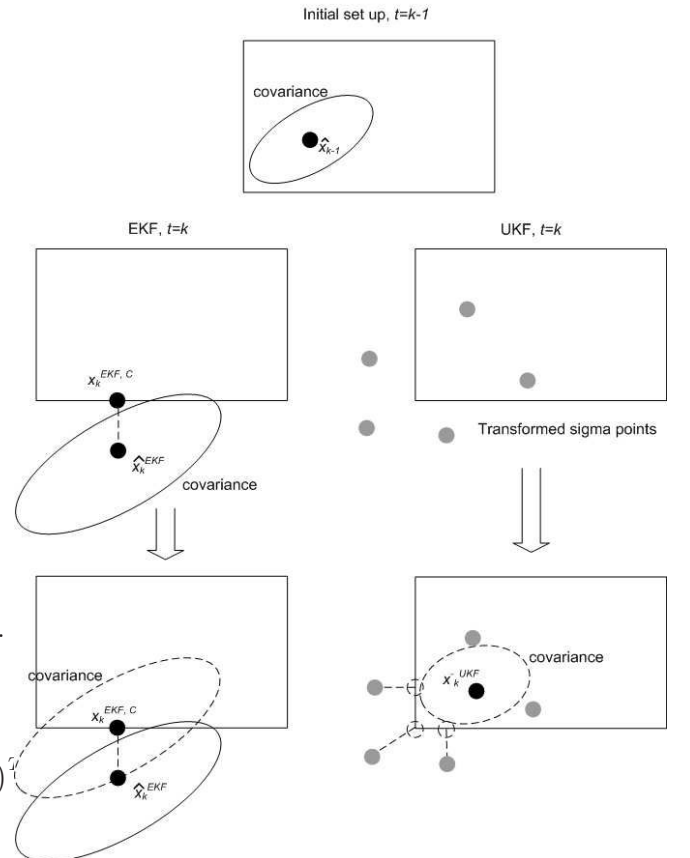

Fig. 2.   Illustration of estimation with state constraints

Assume that the state constraints are represented by box constraints,

$$x_L \leq x \leq x_H.$$

We will illustrate the method for $x \in \mathbb{R}^2$. In case of a second order system, the feasible region by the box constraints can be represented by a rectangle as in Figure 2. The figure shows the illustration of the steps of constraint handling in case of the EKF and UKF from one time step to the next. At $t = k - 1$, the true state (x_k), its estimate $(\widehat{x}_k)$ and state covariance are selected as shown in the figure. At $t = k$, the unconstrained EKF estimate $(\widehat{x}_k^{EKF})$ is outside the feasible region and is projected to the boundary of the feasible region to get the constrained EKF estimate $(x_k^{EKF,C})$ as shown in the figure. While projecting the EKF estimate, the covariance of the EKF estimate is not changed and thus the constraints have no effect on the covariance. Hence, the covariance does not include any constraint information. This way of the handling constraints in the EKF is termed as 'clipping' in literature [6].

The constraints information can be incorporated in the UKF algorithm in a simple way during the time-update step. After the propagation of the sigma points from (8), the (unconstrained) transformed sigma points which are outside the feasible region can be projected onto the boundary of the feasible region and continue the further steps. In Figure 2, at $t = k$ three sigma points which are outside the feasible region are projected onto the boundary (lower right plot in the figure). The mean and covariance with the constrained sigma points now represents the apriori UKF estimate $(x_k^{UKF^-})$ and covariance, and they are further updated in the measurement-update step. The advantage here is that the new apriori covariance includes information on the constraints, which should make the UKF estimate more efficient (accurate) compared to the EKF estimate. An example (reversible reaction) is considered in the next section to illustrate the state estimation with constraints with the proposed method.

Extension of the proposed method to higher dimension is straightforward. Alternative linear constraints. e.g.,

$$Cx \leq d$$

are easily included by projecting the sigma points violating the inequality normally onto the boundary of the feasible region. It is observed that the new covariance obtained at a time step in the general case will be lower in size compared to the covariance obtained without projection. If the estimate after the measurement-update (refer to (12)) is outside the feasible region, the same projection technique can be applied. For more detailed description, refer [14].

The proposed algorithm is outlined below:

Algorithm (outline)

- *For* $k = 1, 2, ...\infty$:
  1) Calculate $2N+1$ sigma points based on the present state covariance according to (7) and project the sigma points which are outside the feasible region

to the boundary to obtain the constrained sigma points,

$$\mathbf{X}_{i,k-1}^{x,C} = P(\mathbf{X}_{i,k-1}^x) \; i = 0, 1, ..., 2N$$

where P refers to the projections.

2) Time-update equations:
   Transform the sigma points through the state-update function,

$$\begin{aligned} \mathbf{X}_{i,k|k-1}^x &= f\left(\mathbf{X}_{i,k-1}^{x,C}, \mathbf{X}_{i,k-1}^v, u_{k-1}\right), \\ i &= 0, 1, ..., 2N. \end{aligned}$$

Again apply the constrains on the transformed sigma points to obtain the constrained transformed sigma points,

$$\mathbf{X}_{i,k|k-1}^{x,C} = P(\mathbf{X}_{i,k|k-1}^x) \; i = 0, 1, ..., 2N.$$

Calculate the apriori state estimate and apriori covariance as given in (9) and (10) using the constrained transformed sigma points $\mathbf{X}_{i,k|k-1}^{x,C}$.

3) Measurement-update equations:
   Transform the constrained sigma points through the measurement-update function as in (11) and obtain the UKF estimate by following the same steps given in Section 2. If UKF estimate violates the constraints, the same projection technique can be used.

A. A reversible reaction example

We will here consider an example to illustrate the constraint handling capability of the UKF compared to that of the EKF. Consider the gas-phase, reversible reaction,

$$2A \xrightarrow{\overline{k}} B, \;\; \overline{k} = 0.16,$$
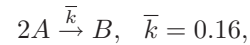
with stoichiometric matrix

$$v = \begin{bmatrix} -2 & 1 \end{bmatrix}$$

and reaction rate

$$r = \overline{k}C_A^2.$$

The state and measurement vectors are defined as

$$x = \begin{bmatrix} C_A \\ C_B \end{bmatrix}, \;\; y = \begin{bmatrix} 1 & 1 \end{bmatrix} x$$

where C_j denotes the concentration of species j. It is assumed that the ideal gas law holds and that the reaction occurs in a well-mixed isothermal batch reactor. Then, from first principles, the model for this system is

$$\dot{x} = f(x) = v^T r.$$

The system is discretized with sampling interval of 0.1. The UKF and EKF are used for state estimation, with the following setup as used in [6]:

$$\begin{aligned} x_0 &= \begin{bmatrix} 3 & 1 \end{bmatrix}^T, \widehat{x}_0 = \begin{bmatrix} 0.1 & 4.5 \end{bmatrix}^T, \\ P_{x_0} &= \begin{bmatrix} 36 & 0 \\ 0 & 36 \end{bmatrix}, \;\; R_v = 10^{-6}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \\ R_n &= 10^{-2}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \end{aligned}$$

The estimation result for the unconstrained case is shown in Figure 3. The result shows that the dynamic performance of the UKF estimates is better compared to that of the EKF. The EKF performance is very similar to the reported results in [6].
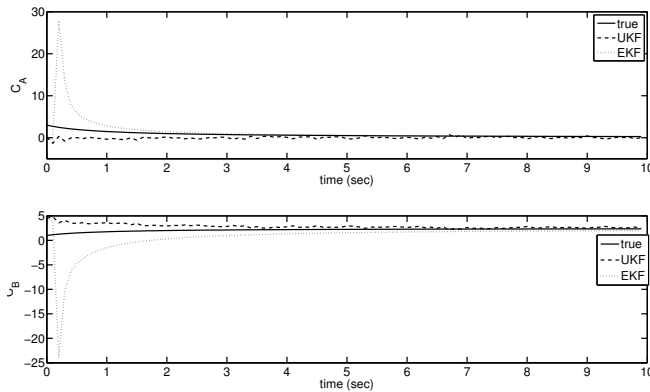

Fig. 3. Comparison of estimated states: no constraint handling

However, during the dynamic response, both the UKF and EKF estimates become negative, (meaning negative concentrations) which is not possible physically. State constraints are incorporated according to the proposed method in Section 3 for the UKF and standard "clipping" is used for the EKF. The results are shown in Figure 4. From Figure 4 the UKF estimates converge to the true states without violating the constraints. Because of the clipping in the EKF, C_A estimate of EKF did not converge to the true state and the estimate of C_B takes much longer time to converge to the true state.
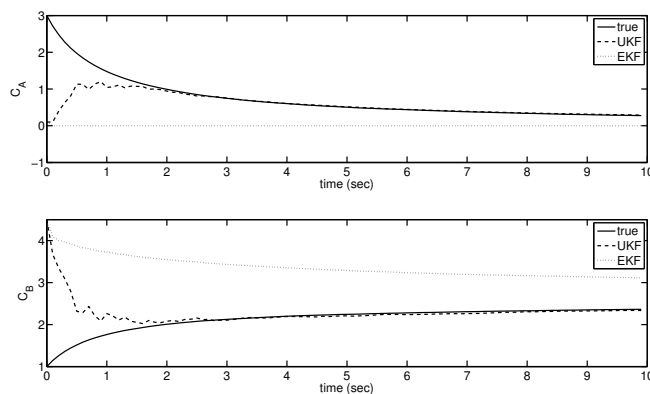

Fig. 4. Comparison of state estimates: state constraint handling

Figure 5 shows the phase portraits of the unconstrained and constrained UKF estimates for the first 4 sec. The figure also includes the corresponding covariances plotted at t=0, 1 and 3 seconds. From Figure 5, it is clear that it takes longer time for the unconstrained estimate to converge as the corresponding covariances do not include the constraint information. The constrained UKF estimate converges faster as the covariances decrease faster, which include the constraint information. The results from this example confirm that the proposed constraint handling method is promising.
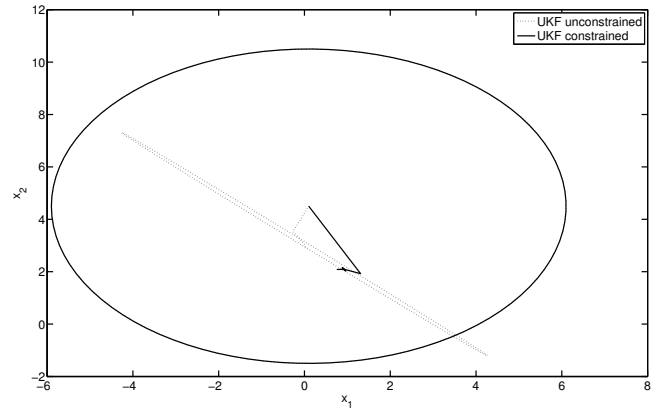

Fig. 5. Phase portraits of the UKF unconstrained and constrained estimates with covariances

IV. CONCLUSIONS

In this paper a constraint handling method is proposed for the Unscented Kalman Filter (UKF). The proposed method is much simpler to implement than a Moving Horizon Estimator. The computational load for the UKF is comparable to the Extended Kalman Filter for the typical case where the Jacobians are computed numerically, and the proposed constraint handling adds minimally to this.

A natural next step is to analyse the properties of the proposed algorithm more in detail.

REFERENCES

[1] S. Julier and J. K. Uhlmann, "Unscented filtering and nonlinear estimation", *Proceedings of the IEEE*, vol. 92, 2004, pp.401-422.
[2] R. Van der Merwe, "Sigma-Point Kalman Filters for probability inference in dynamic state-space models" *PhD thesis*, 2004, Oregon Health and Science University.
[3] B. Huang and Q. Wang, "Overview of emerging Bayesian approach to nonlinear system identification". *Round tables on Non-linear Model Identification, International Workshop on Solving Industrial Control and Optimization Problems*, Cramado, Brazil, April 6-7, 2006.
[4] G. Evensen, "Data Assimilation, The Ensemble Kalman Filter", *Springer, Berlin*, 2007.
[5] A. H. Jazwinski, "Stochastic Processes and Filtering Theory", *Mathematics in Science and Engineering, vol. 64, Academic Press, New York and London*, 1970.
[6] E. L. Haseltine and J. B. Rawlings, "A Critical Evaluation of Extended Kalman Filtering and Moving Horizon Estimation", *Technical report, Texas-Wisconsin Modeling and Control Consortium (TWMCC)*, 2002-2003.
[7] A. Romanenko and J. A. A. M. Castro, "The unscented filter as an alternative to the EKF for nonlinear state estimation: a simulation case study", *Computer and Chemical Engineering*, vol. 28, 2004, pp. 347-355.
[8] B. Akin, U. Orguner and Aydin Ersak, "State Estimation of Induction Motor Using Unscented Kalman Filter", *IEEE Transactions on Control Applications*, Vol. 2, 2003, pp. 915-919.
[9] C. V. Rao, J. B. Rawlings and D. Q. Mayne, "Constrained state estimation for nonlinear discrete-time systems: stability and moving horizon approximations", *IEEE Transactions on Automatic Control*, Vol. 48, 2003, pp. 246-258.

[10] D. Simon and T. L. Chia, "Kalman Filtering with State Equality Constraints", *IEEE Transactions on Aerospace and Electronic Systems,* Vol. 38, 2002, pp. 128-136.

[11] S. Ungarala, E. Dolence and K. Li, " Constrained Extended Kalman Filter for Nonlinear State Estimation", *In proceedings of 8th International IFAC Symposium on Dynamics and Control Process Systems, Cancun, Mexico,* June 2007, vol. 2, pp. 63-68.

[12] W. Li and H. Leung, "Simultaneous Registration and Fusion of Multiple Dissimilar Sensors for Cooperative Driving", *IEEE Transactions on Intelligent Transportation Systems*, Vol. 5, 2004, pp. 84-98.

[13] P. Vachhani, S. Narasimhan and R. Rengaswamy, "Robust and reliable estimation via unscented recursive nonlinear dynamic data reconciliation", *Journal of process control,* Vol. 16, 2006, pp. 1075-1086.

[14] R. Kandepu, B. Foss and L. Imsland, "Applying the unscented kalman filter for nonlinear state estimation, *Journal of process control*, Article in press, 2008.