Computation of Arrival Cost for Moving Horizon Estimation via Unscented Kalman Filtering

Cheryl C Qu, Juergen Hahn∗
Artie McFerrin Department of Chemical Engineering, Texas A & M University,
College Station, TX 77843-3122

Abstract

Moving horizon estimation alleviates the computational burden of solving a full information estimation problem by considering a finite horizon of the measurement data, however, it is non-trivial to determine the arrival cost. A commonly used approach for computing the arrival cost is to use a first order Taylor series approximation of the nonlinear model and then apply an extended Kalman filter. In this paper, an approach to compute the arrival cost for moving horizon estimation based on an unscented Kalman filter is proposed. The performance of such a moving horizon estimator is compared with the one based on an extended Kalman filter and illustrated in a case study.

Keywords: Moving Horizon Estimation; Arrival Cost; Unscented Kalman Filter

1 Introduction

For the last two decades, optimization-based moving horizon estimation (MHE) has found widespread use for nonlinear constrained problems [1], [2], [3]. Compared to full information estimation (FIE), MHE reduces the computational burden by considering a finite horizon of the available measurements, however, it is non-trivial to summarize the effect of the discarded data on the current states, which is the so called arrival cost. For linear unconstrained systems, the standard Kalman filter covariance update formula can be used to express the arrival cost explicitly. However, for nonlinear or constrained systems, a general analytical expression for the arrival cost is rarely available. Tenny and Rawlings [4] estimate the arrival cost by approximating the constrained, nonlinear system as an unconstrained linear time-varying system and applying linearization and an extended Kalman filter (EKF).

Unscented Kalman filters (UKF), as proposed by Julier and Uhlmann [5], avoid the linearization in the Kalman filter update formula by an unscented nonlinear transformation. By

∗Corresponding author. Tel: +1 979 845 3568; fax: +1 979 845 6446. E-mail address: hahn@tamu.edu (J. Hahn).
carefully choosing a set of sigma points, which capture the true mean and covariance of a
given distribution and then passing the means and covariances of estimated states through a
nonlinear transformation, UKF is capable of estimating the posterior means and covariances
accurately to an order higher than two. Therefore UKF can improve upon EKF performance
for nonlinear systems as has been demonstrated for some applications [5]. These advantages
of UKF over EKF form the motivation of this work where a MHE with computation of the
arrival cost via UKF is proposed and its performance is illustrated in a case study.

The paper is organized as follows: In Section 2, a brief review of MHE and UKF is
presented. The MHE with the arrival cost determined by UKF for nonlinear constrained
estimation is then proposed in Section 3. Section 4 compares the performance of the MHE
via EKF to that via UKF for nonlinear state estimation, and concluding remarks are given
in Section 5.

2 Preliminaries

2.1 Moving Horizon Estimation

From a perspective of Bayesian theory, the constrained state estimation problem can
be formulated as the solution of a full information estimation problem [6]. However, real-
time implementations of FIE are often not feasible because the size of the problem grows
unbounded with the number of points in time considered. One strategy to make the estimation
problem tractable is to bound the problem size by employing a moving horizon approach.
The optimization problem defining MHE is of the following form

\[
\min_{x_0, \{w_k\}_{k=0}^{T-1}} \phi_T(x_0, \{w_k\}) = \min_{z, \{w_k\}_{k=T-N}^{T-1}} \sum_{k=T-N}^{T-1} v_k' R^{-1} v_k + w_k' Q^{-1} w_k + \theta_{T-N}(z). \tag{1}
\]

subject to

\[
\begin{align*}
x_k &= f(x_{k-1}, u_{k-1}, w_{k-1}) \\
y_k &= h(x_k, u_k) + v_k \\
x_k &\in X, \ w_k \in W, \ v_k \in V
\end{align*}
\tag{2}
\]

where the sets \(X, W\) and \(V\) are constrained, \(N\) is the size of horizon length, \(x_k := x(k; z, \{w_j\}_{j=T-N}^{k-1})\)
denotes the solution of the system (2) at time \(k\) when the initial state is \(z\), \(\{w_j\}_{j=T-N}^{k-1}\) is the
process noise sequence from time \(T - N\) to \(k - 1\) and \(v_k := y_k - h(x_k, u_k)\). \(\theta_{T-N}(z)\) is referred
to as the arrival cost, which summarizes the effect of the data \(\{y_k\}_{k=0}^{T-N-1}\) on the state \(x_{T-N}\)
and makes it possible to transform the optimization problem into one of lower dimension. However, the best choice of the arrival cost remains an open issue for MHE.

For unconstrained, linear systems, the arrival cost can be expressed explicitly since the MHE optimization simplifies to the Kalman filter and its covariance update formula can be used [7]. Subject to the initial condition $\Pi_0$ and assuming the matrix $\Pi_{T-N}$ is invertible, the arrival cost can then be expressed as

$$\theta_{T-N}(z) = (z - \hat{x}_{T-N})' \Pi_{T-N}^{-1} (z - \hat{x}_{T-N}) + \phi_{T-N}^*.$$  

(3)

where $\hat{x}_{T-N}$ denotes the optimal estimate at time $T-N$ given all of the measurements $y_k$ from time 0 to $T-N-1$, $\phi_{T-N}^*$ represents the optimal cost at time $T-N$ and $\Pi_{T-N}$ is computed from the Kalman filter covariance update

$$\Pi_T = A \Pi_{T-1} A^T + GQG^T - A \Pi_{T-1} C^T (C \Pi_{T-1} C^T + HH^T)^{-1} C \Pi_{T-1} A^T.$$  

(4)

The solution to the problem described by equations (1) and (3) is the unique optimal pair $(z^*, \{\hat{w}^*_k\}_{k=T-N})$ and it can be integrated to yield the optimal state estimates $\{\hat{x}^*_k\}_{k=T-N+1}$, where $\hat{x}^*_k := x(k; z^*, \{\hat{w}^*_j\}_{j=T-N})$ denotes the optimal estimate of the system at time $k$ when the initial state is $z^*$ and the estimated process noise sequence is $\{\hat{w}^*_j\}_{j=T-N}$.

For constrained, linear systems, general analytical expressions for the arrival cost are not available. One reasonable strategy is to approximate the arrival cost by the one for the unconstrained problem. The approximation is exact when the inequality constraints are inactive. For nonlinear systems, Tenny and Rawlings estimate the arrival cost by approximating a constrained, nonlinear system as an unconstrained, linear time-varying system [4]. In their work the model functions $f(.)$ and $h(.)$ in Eq.(2) are supposed to be sufficiently smooth so that a first-order Taylor series approximation of the model can then be applied, i.e. $A_k := \frac{\partial f}{\partial x}|_{\hat{x}_{k-1}}$, $C_k := \frac{\partial h}{\partial x}|_{\hat{x}_{k-1}}$, $G_k := \frac{\partial f}{\partial w}|_{w_k}$ and $H_k := \frac{\partial h}{\partial v}|_{v_k}$ can be obtained. The arrival cost $\theta_{T-N}(z)$ in Eq.(3) can be computed by solving the matrix Riccati Eq.(4) subject to the initial condition $\Pi_0$.

This is called the MHE problem with an arrival cost computed by EKF.

The horizon length $N$ is a tuning parameter for MHE. As a general rule, the larger the horizon length, the more accurate the estimation results will be, however, this comes at the expense of an increase of the computational burden. A practical rule of thumb is that the length of the horizon should not be less than the number of the system states. Rao and Rawlings recommend to choose the horizon length as twice the order of the system [7].
2.2 Unscented Kalman Filter

An unscented Kalman filter is the application of the unscented transformation to recursive estimation. In the unscented transformation procedure, a set of weighed sigma points are deterministically chosen from the statistics of the transformation so that certain properties of these points (e.g., a given mean and covariance) match those of the prior distribution. These sigma points are propagated through a nonlinear mapping and then weighted means and covariances are computed. One way to find a set of sigma points that have the same first two moments and all higher odd-ordered central moments as the given distribution is given by the following:

1. Augment the system state vector to an $n^a = n + q + r$ dimensional vector $x^a = [x^T \ w^T \ v^T]^T$

to obtain its augmented means and covariances,

$$
\hat{x}^a_{k-1|k-1} = \begin{pmatrix}
\hat{x}_{k-1|k-1} \\
0^{q \times 1} \\
0^{r \times 1}
\end{pmatrix}
$$

$$
P^a_{k-1|k-1} = \begin{pmatrix}
P_{k-1|k-1} & 0^{n \times q} & 0^{n \times r} \\
0^{q \times n} & Q_{k-1} & P^{wv}_{k-1} \\
0^{r \times n} & P^{vw}_{k-1} & R_{k-1}
\end{pmatrix}
$$

(5)

where $n$ is the dimension of the original state vector, $q$ and $r$ are the dimensions of the system and measurement noise vectors and $P^{vw}_{k-1}$ and $P^{wv}_{k-1}$ are the correlations between the system and measurement noise. For ease of computation, $P^{vw}_{k-1}$ and $P^{wv}_{k-1}$ are usually set to zero.

2. Generate a set of $2n^a + 1$ symmetric sigma points

$$
\chi^a_{k-1} = \hat{X}^a_{k-1|k-1} + \begin{pmatrix}
0 \\
\sqrt{(n^a + \kappa)P^a_{k-1|k-1}} \\
-\sqrt{(n^a + \kappa)P^a_{k-1|k-1}}
\end{pmatrix}
$$

(6)

where $\hat{X}^a_{k-1|k-1}$ is the expanded $n^a \times (2n^a + 1)$ matrix with $\hat{x}^a_{k-1|k-1}$ as each column. $\kappa \in \mathbb{R}$ is a parameter which could be any positive or negative number with the exception of $\kappa = -n_a$. $\kappa$ can be used to fine tune the higher order moments of the approximation. The more higher order moments are taken into account, the less the overall prediction error will be. For a Gaussian distribution $x(k)$, a useful heuristic is to select $n^a + \kappa = 3$ [8].

After a set of sigma points is selected, each of them is propagated through the nonlinear model functions $f(.)$ and $h(.)$. Weighted means and covariances are then computed from the transformed set of points. Finally the Kalman filter gain is calculated from the covariances and the predicted states are updated based on the available measurements. This procedure results in the following equations defining an unscented Kalman filter:
Prediction equations:

\[
\chi_k^x = f(\chi_{k-1}^x, \chi_{k-1}^w, u_{k-1})
\]

\[
\hat{x}_{k|k-1} = \sum_{i=1}^{2n^a+1} W_i \chi_{i,k}^x
\]

\[
\gamma_k = h(\chi_k^x, \chi_{k-1}^v, u_k)
\]

\[
\hat{y}_k = \sum_{i=1}^{2n^a+1} W_i \gamma_{i,k}
\]  

\[(7)\]

Update equations:

\[
P_{k|k-1} = \sum_{i=1}^{2n^a+1} W_i [\chi_{i,k}^x - \hat{x}_{k|k-1}][\chi_{i,k}^x - \hat{x}_{k|k-1}]^T
\]

\[
P_{y,k} = \sum_{i=1}^{2n^a+1} W_i [\gamma_{i,k} - \hat{y}_k][\gamma_{i,k} - \hat{y}_k]^T
\]

\[
P_{xy,k} = \sum_{i=1}^{2n^a+1} W_i [\chi_{i,k}^x - \hat{x}_{k|k-1}][\gamma_{i,k} - \hat{y}_k]^T
\]

\[
K_k = P_{xy,k} P_{y,k}^{-1}
\]

\[
P_{k|k} = P_{k|k-1} - K_k P_{y,k} K_k^T
\]

\[
\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k (y_k - \hat{y}_k)
\]  

\[(10)\]

where \( \chi^a = [(\chi^x)^T_{1 \times n} (\chi^w)^T_{1 \times q} (\chi^v)^T_{1 \times r}]^T \) and \( W_i \) are weights calculated by

\[
W_i = \begin{cases} 
\frac{n}{2(n_a+n_r)}, & \text{if } i=1; \\
\frac{1}{2(n_a+n_r)}, & \text{otherwise.}
\end{cases}
\]  

\[(14)\]

3 Computing Arrival Cost of MHE by UKF

As mentioned in Section 2.1, algebraic expressions for the arrival cost, which account for data not included in the estimation window, are not available for the majority of systems. Therefore an approximation of the arrival cost is required to implement an estimator.

The main idea behind the presented MHE algorithm is to make use of the advantages that UKF offers over EKF for approximating the arrival cost. A set of weighted sigma points \( \chi_k \) are selected for computing the arrival cost based on two considerations: Firstly the means and covariances of the set of sigma points need to match those of the prior distribution at time \( k = T - N \) in the absence of active bounds. Secondly the distribution of the selected sigma points should be within the feasible region if bounds are active. If \( s_{k,i} = \pm(\sqrt{P_{k|k}}), i = 1, \ldots, n \) are considered as the directions along which the sigma points are selected and \( r_{k,i} \) as the step
sizes along these directions, then the step sizes for the selected sigma points in Eq.(6) for a general UKF are all equal to $\sqrt{n_a + \kappa}$. In other words all the sigma points chosen are located symmetrically around the current estimate. This selection of sigma points can be applied to the MHE via UKF when the constraints are inactive. However, such selection is not adequate when constraints are active. To better approximate the covariance and then the arrival cost in the presence of active constraints, the set of sigma points is chosen in each direction with a step size of

$$r_{k,i} = \min(\sqrt{n_a + \kappa}, (x_{U,i} - \hat{x}_{k|k,i}^a)/s_{k,i}, (x_{L,i} - \hat{x}_{k|k,i}^a)/s_{k,i}).$$  \hspace{1cm} (15)$$

where $x_{U,i}$ and $x_{L,i}$ are the upper and lower bounds in the direction $s_{k,i}$. The central point $\hat{x}_{k|k,b}^a$ is the same as in the general UKF formulation. The rest of selected sigma points may be asymmetrically around the central point due to the presence of active constraints. These sigma points are identical to those for conventional UKF for an unconstrained problem or if the constraints are not active. However, under the condition that the current estimate is close to the bounds, the selection process takes the constraints into account so that none of the selected sigma points violate the constraints on the state variables.

Considering that weights of all sigma points sum up to unity and that these are the same as those for UKF in Eq.(14) in the absence of active bounds, weights $W_i$ for each sigma point can be calculated as follows:

$$W_i = \begin{cases} \frac{\kappa}{2(n_a + \kappa)}, & \text{if } i=1; \\ ar_i + b, & \text{otherwise.} \end{cases}$$  \hspace{1cm} (16)$$

where

$$a = \frac{2\kappa - 1}{2(n_a + \kappa)(S_r - (2n_a + 1)(\sqrt{n_a + \kappa}))}$$

$$b = \frac{1}{2(n_a + \kappa)} - \frac{2\kappa - 1}{2\sqrt{n_a + \kappa}(S_r - (2n_a + 1)(\sqrt{n_a + \kappa}))}$$

$$S_r = \sum_{i=1}^{2n_a} r_i.$$  \hspace{1cm} (17)$$

The procedure to compute the weights with constraints is similar to the work presented by Vachhani et al. [9], where a complete mathematical derivation of the weight equations albeit for different purposes can be found.

After a set of sigma points is obtained, each of these sigma points is instantiated through the nonlinear model functions $f(.)$ and $h(.)$ to obtain the transformed sets $\chi_k^x$ and $\gamma_k$.

$$\chi_k^x = f(\chi_{k-1}^x, \chi_{k-1}^w, u_{k-1})$$

$$\gamma_k = h(\chi_k^x, \chi_{k-1}^w, u_k).$$  \hspace{1cm} (18)$$
The weighted means of system states and measurements and weighted covariances of process and observation noise are then computed from the transformed sets. It should be noted that the weighted predicted estimates  $\hat{x}_{k|k-1}$ may not satisfy the constraints.

\[
\hat{x}_{k|k-1} = \sum_{i=1}^{2n^a+1} W_i \hat{x}_{i,k}
\]

\[
\hat{y}_k = \sum_{i=1}^{2n^a+1} W_i \gamma_{i,k}
\]

\[
P_{k|k-1} = \sum_{i=1}^{2n^a+1} W_i [\chi_{i,k} - \hat{x}_{k|k-1}] [\chi_{i,k} - \hat{x}_{k|k-1}]^T
\]

\[
P_{y,k} = \sum_{i=1}^{2n^a+1} W_i [\gamma_{i,k} - \hat{y}_k] [\gamma_{i,k} - \hat{y}_k]^T
\]

\[
P_{xy,k} = \sum_{i=1}^{2n^a+1} W_i [\chi_{i,k} - \hat{x}_{k|k-1}] [\gamma_{i,k} - \hat{y}_k]^T
\]

Finally the matrix $P_k$ is calculated from the filter gain and is used to compute the arrival cost in Eq.(20). The MHE problem described in the Eq.(1) and (3) is solved with the approximated arrival cost to obtain the updated estimates  $\hat{x}_{k|k}$ as the solutions:

\[
K_k = P_{xy,k} P_{y,k}^{-1}
\]

\[
P_{k|k} = P_{k|k-1} - K_k P_{y,k} K_k^T
\]

\[
\theta_k(z) = (z - \hat{x}_k)' P_k^{-1}(z - \hat{x}_k) + \phi_k^*
\]

The proposed approach to approximate the arrival cost for MHE does not require linearization of the system and measurement functions. As in any MHE filter, the matrices $Q$ and $R$ can be chosen to take uncertainty in the model and measurement noise into account and the size of the estimation horizon serves as an additional tuning parameter.

4 Case Study

To illustrate the performance of MHE based on UKF (uMHE) compared against the one based on EKF (eMHE), both algorithms have been applied to a variety of models and a large number of scenarios such as different operating conditions, different tuning parameters $Q$ and $R$, and different process and measurement noise. Due to space constraints only one representative case study is shown in this section. Monte Carlo simulations with 50 sample points have been conducted for each procedure so as not to bias results to one set of data. The performance is evaluated by the overall mean-squared error (MSE). MSE is first averaged
over all simulations for each time point and then over time to to take the behavior over the entire time horizon into account.

The model of this case study is a nonisothermal continuous stirred tank reactor with coolant jacket dynamics, where the following exothermic irreversible reaction between sodium thiosulfate and hydrogen peroxide is taking place:

\[
2\text{Na}_2\text{S}_2\text{O}_3 + 4\text{H}_2\text{O}_2 \rightarrow \text{Na}_2\text{S}_3\text{O}_6 + \text{Na}_2\text{SO}_4 + 4\text{H}_2\text{O}
\] (21)

The capital letters A and B are used to denote the chemical compounds \(\text{Na}_2\text{S}_2\text{O}_3\) and \(\text{H}_2\text{O}_2\) in the following. The reaction kinetic law is reported in the literature to be [10]:

\[
-r_A = k_0 e^{-E/RT} C_A C_B
\]

where \(k_0\) is the pre-exponential factor, \(E\) is the activation energy, \(R\) is the gas constant, \(T\) is the temperature, and \(C_A\) and \(C_B\) are the concentrations of species A and B, respectively. A stoichiometric proportion of species A and B in the feed stream is assumed which results in \(C_B(t) = 2C_A(t)\). A mole balance for species A and energy balances for the reactor and the cooling jacket result in the following nonlinear process model:

\[
\frac{dC_A}{dt} = \frac{F}{V}(C_{Ain} - C_A) - 2k(T)C_A^2
\]

\[
\frac{dT}{dt} = \frac{F}{V}(T_{in} - T) + \frac{2}{\rho c_p}(\Delta H)R\frac{k(T)C_A^2}{2} - \frac{UA}{V\rho c_p}(T - T_j)
\]

\[
\frac{dT_j}{dt} = \frac{F_w}{V}(T_{jin} - T_j) + \frac{UA}{V_w\rho c_{pw}}(T - T_j)
\]

where \(F\) is the feed flow rate, \(V\) is the volume of the reactor, \(C_{Ain}\) is the inlet feed concentration, \(T_{in}\) is the inlet feed temperature, \(F_w\) is the feed flow rate of the cooling jacket, \(V_w\) is the volume of the cooling jacket, \(T_{jin}\) is the inlet coolant temperature, \(c_p\) is the heat capacity of the reacting mixture, \(c_{pw}\) is the heat capacity of the coolant, \(\rho\) is the density of the reaction mixture, \(\rho_w\) is the density of the coolant, \(U\) is the overall heat-transfer coefficient, and \(A\) is the area over which the heat is transferred. The process parameter values are given in the work by Rajaraman et al. [11].

The initial conditions and filter parameters are as follows:

\[
\hat{x}_0 = \begin{bmatrix} 0.018 & 382 & 371.3 \end{bmatrix}^T, \quad \hat{P}_0 = \text{diag}\{10^{-7}, 2.5, 2.5\},
\]

\[
Q_0 = \text{diag}\{10^{-8}, 0.25, 0.25\}, \quad R_0 = 0.25.
\] (23)

The dimension of the augmented state vector is 7 and the set of sigma points is composed of 15 elements. The additional tuning parameter of the UKF, \(\kappa\), is set to \(-4\) for the case study according to the heuristic mentioned in Section 2.2. For fairness of comparison, \(\kappa\) is
Table 1: MSE Average over 50 Monte Carlo Simulations for Varying Measurement Noise Levels and Horizon Lengths

<table>
<thead>
<tr>
<th>MSE</th>
<th>N=3 eMHE</th>
<th>N=4 eMHE</th>
<th>N=6 eMHE</th>
<th>N=10 eMHE</th>
<th>N=3 uMHE</th>
<th>N=4 uMHE</th>
<th>N=6 uMHE</th>
<th>N=10 uMHE</th>
</tr>
</thead>
<tbody>
<tr>
<td>R = 25</td>
<td>8.45</td>
<td>5.64</td>
<td>5.56</td>
<td>4.94</td>
<td>5.14</td>
<td>4.76</td>
<td>2.66</td>
<td>2.52</td>
</tr>
<tr>
<td>R = 0.25</td>
<td>2.69</td>
<td>1.45</td>
<td>1.43</td>
<td>1.21</td>
<td>0.89</td>
<td>0.84</td>
<td>0.87</td>
<td>0.70</td>
</tr>
<tr>
<td>R = 0.01</td>
<td>1.01</td>
<td>0.86</td>
<td>0.80</td>
<td>0.78</td>
<td>0.59</td>
<td>0.56</td>
<td>0.48</td>
<td>0.45</td>
</tr>
</tbody>
</table>

not further adjusted to fine tune the higher order moments of the approximation. A non-negative constraint is enforced on the concentration $C_A$ for both MHE formulations.

Based on the overall mean-squared error, the performance of each MHE is evaluated for horizon lengths $N=3$, 4, 6, and 10. Further simulations have also been carried out by varying measurement noise parameters for a fixed horizon length. Table 1 provides a summary of the results. It can be seen that uMHE performs better than eMHE for all the investigated horizon lengths and measurement noise levels. The MSEs for both uMHE and eMHE are decreasing with increasing lengths of the horizons, i.e., the performance of both uMHE and eMHE improves as more data are included in a horizon. If $N$ is chosen to be large, the arrival cost could be accurately computed with either approach. Therefore the advantages of uMHE over eMHE decrease for large $N$.

5 Conclusion

This paper presented a MHE formulation where the arrival cost is computed by UKF. The unscented transformation and a set of selected sigma points are used to compute the covariances and then the arrival cost. The selection procedure for the sigma points is the same as the one used for unscented Kalman filtering if the constraints are inactive, however, a modification is used that satisfies the state variable constraints when the constraints are active. Linearization of the model is not required for the presented approach.

While only a representative case study was included in this paper, the presented method performed slightly better than the commonly used eMHE for all investigated cases. Therefore, the method can be a promising alternative for approximating the arrival cost for MHE.
Acknowledgement

The authors gratefully acknowledge financial support from the Mary Kay O’Connor Process Safety Center and appreciate valuable feedback given by Dr. Sam Mannan.

References


