Cleveland State University
EngagedScholarship@CSU

MSL
ACADEMIC ENDEAVORS

Chemical & Biomedical Engineering Faculty Publications

Chemical & Biomedical Engineering Department

10-2009

Computing Arrival Cost Parameters in Moving Horizon Estimation Using Sampling Based Filters

Sridhar Ungarala
Cleveland State University, s.ungarala@csuohio.edu

Follow this and additional works at: https://engagedscholarship.csuohio.edu/encbe_facpub

Part of the Process Control and Systems Commons

How does access to this work benefit you? Let us know!

Original Citation

Ungarala, S. (2009). Computing arrival cost parameters in moving horizon estimation using sampling based filters. Journal of Process Control, 19(9), 1576-1588. doi:10.1016/j.jprocont.2009.08.002

Repository Citation

Ungarala, Sridhar, "Computing Arrival Cost Parameters in Moving Horizon Estimation Using Sampling Based Filters" (2009). *Chemical & Biomedical Engineering Faculty Publications*. 59.
https://engagedscholarship.csuohio.edu/encbe_facpub/59

Computing arrival cost parameters in moving horizon estimation using sampling based filters

Sridhar Ungarala [*]

Department of Chemical and Biomedical Engineering, 2121 Euclid Avenue, Cleveland State University, Cleveland, OH 44115, USA

1. Introduction

Consider the evolution of the state of a dynamic system modeled as

$$x_{k+1} = f(x_k) + w_k, \tag{1}$$

where $f : \mathbb{R}^n \to \mathbb{R}^n$ is generally a nonlinear vector function and $w_k \in \mathbb{R}^n$ is an independent identically distributed (i.i.d) random noise vector distributed according to the Gaussian probability density function (pdf) $p_w(w_k) = \mathcal{N}(0, Q)$. The measurements are given by

$$y_k = h(x_k) + v_k, \tag{2}$$

where $h : \mathbb{R}^n \to \mathbb{R}^p$ is a nonlinear vector function of the state and $v_k \in \mathbb{R}^p$ is i.i.d random measurement noise vector distributed according to the pdf $p_v(v_k) = \mathcal{N}(0, R)$. It is commonly assumed that the initial pdf of the state vector is known as a Gaussian pdf $p(x_0) = \mathcal{N}(\hat{x}_0, \hat{P}_0)$, conditioned on the fact that no measurements

are available yet. The problem considered in this paper is intentionally restricted to additive noise terms to simplify the discussion and relate to existing MHE literature. It is possible to widen the scope to systems that are nonlinear in noise terms, such as $f(x_k, w_k)$ and $h(x_k, v_k)$, provided f^{-1} and h^{-1} exist as vector functions of the respective noise terms.

Given the set of noisy measurements $Y = \{y_j : j = 1, \dots, k\}$, it is desired to optimally estimate the states $X = \{x_j : j = 0, \dots, k\}$. All statistical information about the discrete state trajectory is embodied in the joint pdf of the trajectory, conditioned on the measurements. If it is possible to derive the joint conditional pdf up to a constant of proportionality or approximate it in closed-form, the state trajectory may be estimated by locating the mode via the following optimization problem

$$\max_X p(X|Y) \quad \forall k. \tag{3}$$

The optimizing state trajectory $\overline{X}$ is known as the modal trajectory. This problem is also referred to as full information estimation [1,2]. As new measurements are accumulated, the size of the discrete state sequence to be determined by optimization grows. The full

information approach turns impractical very soon as the computational burden increases without bound.

In order to avoid the unbounded growth of the measurement set and the accompanying optimization problem, only a fixed window or horizon of measurements are allowed to accumulate at the beginning. After the first batch, subsequent processing is limited to the same size of the batch by appending a new measurement to the batch and discarding the earliest measurement in the batch. This trimming of the full information estimator is referred to as the moving horizon estimator (MHE) [3,4]. For a fixed horizon $m \geqslant 1$, let $X_m = \{x_j : j = k - m + 1, \ldots, k\}$ and $Y_m = \{y_j : j = k - m + 1, \ldots, k\}$, then MHE is posed as maximizing the joint conditional pdf of the states in the horizon with respect to X_m

$$\max_{X_m} p(X_m | Y_m) \quad \forall k \geqslant m. \tag{4}$$

Similar to the given initial pdf $p(x_0)$ for the full information estimator, it is assumed in MHE that the conditional pdf of the state just before the horizon $p(x_{k-m} | Y_{1:k-m})$ is available, where $Y_{1:k-m} = \{y_j : j = 1, \ldots, k - m\}$. This conditional pdf, serving as a summary of the past information not included in the horizon, is used to initialize the predictions of the state pdf inside the horizon. The first prediction is the *a priori* pdf of the state at the beginning of the horizon $p(x_{k-m+1} | Y_{1:k-m})$, from which a penalty term known as the arrival cost is formulated for the MHE objective function [5,6]. Initialization of MHE is equivalently understood as the formulation of the arrival cost term.

The arrival cost information of a horizon may be propagated to the subsequent horizon by updating the conditional pdf $p(x_{k-m} | Y_{1:k-m})$, which is independent of the optimizing $\overline{X}_m$. In general, the formulation of arrival cost is an *ad hoc* procedure and the problem formulations in Eqs. (3) and (4) are not equivalent to each other because $\overline{X}_m$ may not lie on the modal trajectory $\overline{X}$. Thus, the implementation of MHE incorporates a tradeoff. The MHE limits the optimization computational load, while requiring repeated initialization as the horizon moves forward. This entails a knowledge of the evolution of the conditional pdf $p(x_{k-m} | Y_{1:k-m})$, which is generally not possible to follow without approximations because a *complete* description of the conditional pdf is parameterized by *all* of its moments [7].

On one end of the tradeoff scale, the simplest way to approximate the conditional pdf is to assume it to be a uniform density, which means that MHE can use only the measurement information included in the most recent horizon and all the past information is ignored. In order to compensate for the lack of prior information, more information is needed in large horizons resulting in high computational cost. On the other end, if the conditional pdf is somehow determined by following its evolution in between measurements and subsequent conditional update, the most sensible choice of horizon size for MHE is $m = 1$, which is equivalent to working with the conditional pdf itself. As a compromise, one would seek to approximate the conditional pdf in terms of a finite number of parameters useful for the initialization of MHE and implement MHE in horizons small enough to allow for realtime computation. The horizon should also be large enough to compensate for the under-parameterization of the conditional pdf.

It is common in MHE implementations to assume that the conditional pdf $p(x_{k-m} | Y_{1:k-m})$ and the *a priori* pdf $p(x_{k-m+1} | Y_{1:k-m})$ of the horizon are shape invariant multivariate Gaussian pdfs [1–6]. Approximate nonlinear filters are used to propagate and update only two parameters of the pdfs, the mean and the covariance. The most frequently used methods being the extended Kalman filter (EKF) and its variants such as higher order filters [4]. For horizons larger than one, the inherent delay in the horizon allows the use of approximate nonlinear smoothers instead of filters [4,8].

The existing methods of recursively updating the arrival cost based on functional approximations of nonlinearities by truncated Taylor series can lead to unpredictable behavior. In many cases the covariance used to weight the arrival cost diverges even for higher order EKF implementations and fails to be a reliable measure of the quality of the knowledge of the state. Poorly formulated arrival cost forces large horizons, which pose hindrances for fast, realtime estimation [2]. The Gaussian assumptions also break down for many nonlinear systems because of the tendency to exhibit multiple modes in both the *a priori* and conditional densities [9–11]. The initialization of MHE with the best choice of arrival cost term is an open issue, which also leaves the computational complexity of MHE implementation as an open challenge. It has been conjectured that, arrival cost can be based on uniform density if realtime global optimization can be implemented in large horizons, because Gaussian approximation of a multimodal pdf can produce distorted estimates for all future states [2,11]. It would constitute a brute force implementation on one end of the tradeoff scale mentioned before.

Typically any nonlinear filter capable of propagating the conditional mean and covariance may be used to compute the arrival cost in MHE [3]. However, sufficient motivation exists to investigate novel methods to properly parameterize the arrival cost. Sampling based nonlinear filters are a departure from the traditional approximate filters based on functional linearization of the nonlinearity such as the EKF and its variants. It is possible to address two concerns in the existing implementation of MHE based on EKF. First, the formulation of arrival cost based on Gaussian assumption can be retained but the moments can be obtained very accurately without resorting to linearization. Second, some sampling based filters can provide numerical approximations of the *entire* density rather than just the first two moments. Additional computation may be tolerated to find non-Gaussian closed-form approximations of the densities instead of allocating resources to large horizons that are poorly initialized all the time. These two issues are explored in detail using simulation examples in this communication.

In this paper, three distinct sampling based methods are suggested for initializing or formulating the arrival cost for MHE. They include the deterministic sampling based unscented Kalman filter (UKF), the random sampling based class of nonlinear filters called particle filter (PF) and the aggregate Markov chain based cell filter (CF). The choice of these three methods is motivated by their relatively small online computational demand compared to other non-traditional filters such as the grid based approaches [9,12–14].

The UKF utilizes a minimal set of deterministically sampled support points called sigma points and associated weights to represent the state pdf, assumed Gaussian. The transformation of the sigma points through any nonlinearity yields the ensemble mean and covariance of the points accurately to the second order approximation of the nonlinearity [15]. Since linearization is avoided, the UKF is a Jacobian free Kalman filter for nonlinear systems. Significant performance gains have been demonstrated with UKF compared to EKF with about the same order of computational complexity [16–18].

The particle filter is a broad class of sequential Monte Carlo techniques that attempt to solve the recursive Bayesian inference problem without requiring simplifying assumptions about nonlinearity of the models or non-Gaussianity of the densities. A great deal of interest is generated by the utility of these simple, accurate and fast algorithms for the generally infinite dimensional nonlinear filter [2,10,19–22]. The central idea is to represent the non-Gaussian densities by a large number of samples or particles distributed accordingly and update the samples and weights conditioned on measurement information according to Bayes rule. The sample mean and covariance converge independent of the state dimen-

sion, which can provide excellent statistics to formulate the arrival cost in MHE with Gaussian assumption. The samples can also be used in kernel density estimators [23] to formulate more relevant nonparametric arrival cost terms.

Unlike the particle filter, the state pdf in the cell filter is approximated as a piecewise constant probability mass vector (pmv), which is similar to the point mass approach of the grid based methods. The main idea in CF is the development of an aggregate Markov chain for describing the temporal dynamics of the state pdf over a discretized state space comprising a finite number of cells [24–26]. The transition probability matrix is computed offline using propagation of samples from the cells. While the grid methods perform expensive online propagation of grids and point masses, the main online task in the CF is limited to a linear transformation of the pmv. However, the CF shares the disadvantage of Bellman's "curse of dimensionality" with the grid methods. Apart from accurately determining the moments, closed-form curve fits to the pmv can be used to formulate realistic arrival cost terms.

In the following sections, the probabilistic interpretation [4,5] and the initialization of moving horizon estimation is laid out first. Approximation of the arrival cost term using Gaussian assumption is then discussed. The use of unscented Kalman filter [27], particle filter and cell filter to provide the two summary statistics for arrival cost is described. A more flexible formulation of arrival cost based on numerical approximation of the state pdf provided by PF and CF is considered. Then, two simulation examples from the literature are included to demonstrate the benefits of sampling based initialization of MHE, followed by conclusions.

2. Moving horizon estimation

2.1. Probabilistic interpretation

Using Bayes rule, the joint conditional density in a horizon is written up to a constant of proportionality as

$$p(X_m|Y_m) \propto p(Y_m|X_m)p(X_m|Y_{1:k-m}), \quad (5)$$

where $p(X_m|Y_{1:k-m}) = p(x_{k-m+1}, \ldots, x_k|y_1, \ldots, y_{k-m})$, i.e., the *a priori* joint pdf of the states in the horizon conditioned on the measurements before the horizon and $p(Y_m|X_m) = p(y_{k-m+1}, \ldots, y_k|x_{k-m+1}, \ldots, x_k)$ is the joint likelihood function. Assuming X_m is a first order Markovian sequence, the joint conditional density is

$$p(X_m|Y_m) = c_1 \prod_{j=k-m+1}^{k} p(y_j|x_j) \prod_{j=k-m+1}^{k-1} p(x_{j+1}|x_j)p(x_{k-m+1}|Y_{1:k-m}), \quad (6)$$

where c_1 is a constant and $p(y_j|x_j)$ is the likelihood function due to each measurement in the horizon. The evolution of the state pdf inside the horizon is described by the state transition probability density function $p(x_{j+1}|x_j)$ and initialized by the *a priori* density of the state at the beginning of the horizon $p(x_{k-m+1}|Y_{1:k-m})$.

The state transition pdf is defined by $p_w(x_{k+1} - f(x_k))$, which is given as $\mathcal{N}(0, Q)$ and the likelihood function is defined by $p_v(y_k - h(x_k))$ given as $\mathcal{N}(0, R)$. The restrictive assumption of additive noise is useful here, but it is not necessary if the models are invertible and the Jacobians exist. Thus, the joint pdf up to a constant of proportionality c_2 is,

$$p(X_m|Y_m) = c_2 \prod_{j=k-m+1}^{k} e^{\left(-\frac{1}{2}\|y_j - h(x_j)\|_{R^{-1}}^2\right)} \prod_{j=k-m+1}^{k-1} e^{\left(-\frac{1}{2}\|x_{j+1} - f(x_j)\|_{Q^{-1}}^2\right)} \\ \times p(x_{k-m+1}|Y_{1:k-m}), \quad (7)$$

where $\|\alpha\|_A^2$ is the quadratic form $\alpha^{\mathrm{T}}A\alpha$. The negative logarithm of the joint density is the MHE cost function that is minimized for state estimates,

$$\min_{X_m} \sum_{j=k-m+1}^{k} \|y_j - h(x_j)\|_{R^{-1}}^2 + \sum_{j=k-m+1}^{k-1} \|x_{j+1} - f(x_j)\|_{Q^{-1}}^2 \\ + \Gamma(x_{k-m+1}). \quad (8)$$

The terms in the two summations are collectively known as stage costs related to the uncertainties in the system model and measurements respectively. The last term $\Gamma(\cdot)$ is called the initial penalty or the arrival cost related to the uncertainty in the *a priori* information at the beginning of the horizon based on the past measurements

$$\Gamma(x_{k-m+1}) = -\ln p(x_{k-m+1}|Y_{1:k-m}), \quad (9)$$

where the *a priori* density is required in closed-form up to a constant of proportionality.

One of the original motivations for posing the state estimation problem in the optimization framework is the ability to impose physically meaningful constraints on the variables. When considering physically meaningful support for the joint conditional density, state constraints are posed $x_k \in \mathbb{X}_k$, which may not be necessarily convex

$$\mathbb{X}_k = \{x_k : x^l \leqslant g_1(x_k) \leqslant x^u\}. \quad (10)$$

Mass and energy balances and constitutive relationships can also impose multivariate algebraic equality constraints,

$$g_2(x_k) = 0. \quad (11)$$

If necessary, it is possible to consider lower and upper bounds on the noise variables by limiting to polyhedral convex sets [5].

2.2. Initialization of MHE

The stage costs are completely specified by the system model, the measurement equation and the noise densities. On the other hand, the formulation of the arrival cost term $\Gamma(x_{k-m+1})$ is typically a user-defined initialization step in the implementation of MHE. In order to formulate the arrival cost term $\Gamma(\cdot)$, the *a priori* pdf $p(x_{k-m+1}|Y_{1:k-m})$ at the beginning of the horizon is needed, which in turn must be predicted from a summary of the past information, i.e., the conditional pdf $p(x_{k-m}|Y_{1:k-m})$.

- **Prediction:** The desired prediction for initializing MHE is achieved by making use of the fact that,

$$p(x_{k-m+1}|Y_{1:k-m}) = \iint p(x_{k-m+1}|x_{k-m})p(x_{k-m}|Y_{1:k-m})\mathrm{d}x_{k-m}, \quad (12)$$

which is the Chapman–Kolmogorov equation. The state transition pdf, with c_3 a constant, is

$$p(x_{k-m+1}|x_{k-m}) = c_3 e^{\left(-\frac{1}{2}\|x_{k-m+1} - f(x_{k-m})\|_{Q^{-1}}^2\right)}. \quad (13)$$

Unfortunately, the integration is generally not feasible without simplifying assumptions in order to obtain a closed-form $p(x_{k-m+1}|Y_{1:k-m})$. After the arrival cost at $k - m + 1$ is initialized for the horizon and the estimates are optimized, the horizon is moved forward to $k - m + 2$. An update must be performed on the conditional pdf to summarize the past, which now includes $k - m + 1$.

- **Update:** The update uses Bayes rule to modify the previous prediction in light of the measurement y_{k-m+1}

$$p(x_{k-m+1}|y_{k-m+1}) = \frac{p(y_{k-m+1}|x_{k-m+1})p(x_{k-m+1}|Y_{1:k-m})}{\int p(y_{k-m+1}|x_{k-m+1})p(x_{k-m+1}|Y_{1:k-m})\mathrm{d}x_{k-m+1}}, \quad (14)$$

where the likelihood function, with c_4 a constant, is

$$p(y_{k-m+1}|x_{k-m+1}) = c_4 e^{\left(-\frac{1}{2}\|y_{k-m+1} - h(x_{k-m+1})\|_{R^{-1}}^2\right)}. \quad (15)$$

Eqs. (12) and (14) together form the conceptual predictor–corrector solution for a general Bayesian estimator that recursively determines the evolution of the conditional density. Closed-form solutions are generally possible only for linear systems and some scalar cases. It must be noted that the Bayesian estimator is also subject to all applicable constraints on the random variables, so that the support of the conditional pdf is meaningful.

3. Approximation of arrival cost I

For nonlinear systems, the *a priori* pdf $p(x_{k-m+1}|Y_{1:k-m})$ is generally infinite dimensional. Closed-form approximations may be written by making two practical assumptions, (1) the pdf is approximately represented by a fixed number of suitable parameters, for example, a few moments or weighted samples or piecewise constant approximation and (2) the evolution of the pdf is followed by approximate prediction and conditional update of these parameters.

One of the most common simplifying assumptions used in MHE implementations is that the *a priori* pdf at the beginning of the horizon is a multivariate Gaussian $p(x_{k-m+1}|Y_{1:k-m}) = \mathcal{N}(\tilde{x}_{k-m+1}, \widetilde{P}_{k-m+1})$, completely represented by the first two moments. By taking the negative logarithm of the *a priori* pdf, the arrival cost in Eq. (8) is now expressed as

$$\Gamma(x_{k-m+1}) = \|x_{k-m+1} - \tilde{x}_{k-m+1}\|^2_{\widetilde{P}^{-1}_{k-m+1}}. \tag{16}$$

Similarly, it may be assumed that the past information is also adequately summarized by the mean vector and covariance matrix of the conditional pdf, implicitly assuming that it is a multivariate Gaussian $p(x_{k-m}|Y_{1:k-m}) = \mathcal{N}(\hat{x}_{k-m}, \widehat{P}_{k-m})$. In this section, three different approaches containing different assumptions are considered to approximately implement the Bayesian estimation from Eqs. (12) and (14) and obtain the parameters of arrival cost for MHE. The traditional MHE initialization by EKF is also included for comparison with the deterministic sampling based unscented Kalman filter, the random sampling based particle filter and the aggregate Markov chain based cell filter.

3.1. Arrival cost using extended Kalman filter

All pdfs are represented by the mean vector and the covariance matrix and the dynamics of the moments are approximately realized through first order Taylor series approximations of the system nonlinear function f and the measurement nonlinear function h about chosen reference trajectories.

• **Prediction:** Given the summary of past information $\hat{x}_{k-m}$ and $\widehat{P}_{k-m}$, the *a priori* mean and covariance for arrival cost are computed via propagation through the system model linearized about the reference trajectory $\hat{x}_{k-m}$,

$$\tilde{x}_{k-m+1} = f(\hat{x}_{k-m}), \tag{17}$$

$$\widetilde{P}_{k-m+1} = F_{k-m}\widehat{P}_{k-m}F^{\mathrm{T}}_{k-m} + Q, \tag{18}$$

where the Jacobian matrix is evaluated at the previous EKF estimate

$$F_{k-m} = \left.\frac{\partial f}{\partial x}\right|_{x=\tilde{x}_{k-m}}. \tag{19}$$

• **Update:** Using the measurement equation, linearized about $\tilde{x}_{k-m+1}$, the time-varying filter gain K_{k-m+1} is computed

$$K_{k-m+1} = \widetilde{P}_{k-m+1}H^{\mathrm{T}}_{k-m+1}\left[H_{k-m+1}\widetilde{P}_{k-m+1}H^{\mathrm{T}}_{k-m+1} + R\right]^{-1}, \tag{20}$$

where the Jacobian matrix is evaluated at the predicted mean

$$H_{k-m+1} = \left.\frac{\partial h}{\partial x}\right|_{x=\tilde{x}_{k-m+1}}. \tag{21}$$

The conditional mean and covariance are updated by

$$\hat{x}_{k-m+1} = \tilde{x}_{k-m+1} + K_{k-m+1}\left(y_{k-m+1} - h(\tilde{x}_{k-m+1})\right), \tag{22}$$

$$\widehat{P}_{k-m+1} = \widetilde{P}_{k-m+1} - K_{k-m+1}H_{k-m+1}\widetilde{P}_{k-m+1}. \tag{23}$$

The recursive formulation of EKF is simple and computationally efficient. However, the EKF is prone to divergence because the estimation covariance $\widehat{P}_{k-m+1}$ is generally not a reliable measure of the goodness of the estimate $\hat{x}_{k-m+1}$. In many cases the covariance becomes too small, causing the EKF to place high confidence on the inaccurate estimate, which then continues to exacerbate the deviation of the estimates from the true states. Another disadvantage of the EKF is the necessity for the Jacobians of the nonlinear function to exist, which may be difficult to compute even when existent.

The updated EKF estimates can be adjusted to reconcile with simple linear equality constraints using methods based on Lagrange multipliers and projection methods [28,29]. More general state constraints may be imposed on the iterated form of the EKF implemented as a separate MHE in a horizon of one to formulate the arrival cost.

3.2. Arrival cost using unscented Kalman filter

The state pdf is implicitly assumed to be a Gaussian and it is represented by a fixed number of deterministically chosen samples or sigma points and the associated weights $\{\chi^{(i)}, W^{(i)}: i = 0, \ldots, 2n\}$, where n is the dimension of the state. The first two weighted moments of the sigma points exactly match the mean and covariance of the Gaussian density [16].

Given the summary of the past information $\hat{x}_{k-m}$ and $\widehat{P}_{k-m}$, $2n$ sigma points are defined symmetrically around the mean vector using the covariance matrix with the mean taken as $(2n + 1)$th sigma point,

$$\hat{\chi}^{(0)}_{k-m} = \hat{x}_{k-m}, \tag{24}$$

$$\hat{\chi}^{(i)}_{k-m} = \hat{x}_{k-m} + \sqrt{n+\lambda}\left(\sqrt{\widehat{P}_{k-m}}\right)_i, \quad i = 1, \ldots, n, \tag{25}$$

$$\hat{\chi}^{(i)}_{k-m} = \hat{x}_{k-m} - \sqrt{n+\lambda}\left(\sqrt{\widehat{P}_{k-m}}\right)_{i-n}, \quad i = n+1, \ldots, 2n, \tag{26}$$

where $\left(\sqrt{\widehat{P}_{k-m}}\right)_i$ is the ith column of the matrix square root, which is typically computed using the Cholesky decomposition. The corresponding set of weights are defined as follows [16]

$$W^{(0)}_a = \frac{\lambda}{n+\lambda}, \tag{27}$$

$$W^{(0)}_c = \frac{\lambda}{n+\lambda} + 1 - \alpha^2 + \beta, \tag{28}$$

$$W^{(i)}_a = W^{(i)}_c = \frac{1}{2(n+\lambda)}, \quad i = 1, \ldots, 2n, \tag{29}$$

where the subscript a denotes weights for computing the mean and c for computing the covariance. The constant $\lambda = \alpha^2(n+\kappa) - n$ is a scaling parameter and the positive constants α, β and κ are tuning parameters. Typical values for the constants are $\alpha = 0.5$, $\beta = 2$ and $\kappa = 3 - n$.

• **Prediction:** All the sigma points are propagated through the system model, with the system noise taken as its expected value, to obtain the transformed sigma points

$$\tilde{\chi}^{(i)}_{k-m+1} = f\left(\hat{\chi}^{(i)}_{k-m}\right), \quad i = 0, \ldots, 2n. \tag{30}$$

The *a priori* mean for arrival cost is determined by the weighted mean of the sigma points. The *a priori* covariance matrix is the

weighted covariance adjusted to account for the uncertainty introduced by the system noise

$$\tilde{x}_{k-m+1} = \sum_{i=0}^{2n} W_a^{(i)} \tilde{\chi}_{k-m+1}^{(i)\leftarrow}, \tag{31}$$

$$\tilde{P}_{k-m+1} = \sum_{i=0}^{2n} W_c^{(i)\leftarrow} \left[\tilde{\chi}_{k-m+1}^{(i)} - \tilde{x}_{k-m+1} \right] \left[\tilde{\chi}_{k-m+1}^{(i)} - \tilde{x}_{k-m+1} \right]^{\mathrm{T}} + Q. \tag{32}\leftarrow$$

- **Update:** A new set of $2n+1$ sigma points $\{\tilde{\chi}_{k-m+1}^{(i)\leftarrow} : i = 0, \ldots, 2n\}$ are defined around the *a priori* mean $\tilde{x}_{k-m+1}$ using $\tilde{P}_{k-m+1}$. Each sigma point is used as the argument in the measurement equation, with the measurement noise taken as its expected value,

$$\tilde{\gamma}_{k-m+1}^{(i)} = h\left(\tilde{\chi}_{k-m+1}^{(i)\leftarrow} \right) , \quad i = 0, \ldots, 2n. \tag{33}\leftarrow$$

The expected value of the output $\tilde{y}_{k-m+1}$ and the covariance matrix S_{k-m+1} are obtained from the sample statistics of the points $\tilde{\gamma}_{k-m+1}^{(i)\leftarrow}$

$$\tilde{y}_{k-m+1} = \sum_{i=0}^{2n} W_a^{(i)} \tilde{\gamma}_{k-m+1}^{(i)\leftarrow}, \tag{34}$$

$$S_{k-m+1} = \sum_{i=0}^{2n} W_c^{(i)\leftarrow} \left[\tilde{\gamma}_{k-m+1}^{(i)} - \tilde{y}_{k-m+1} \right] \left[\tilde{\gamma}_{k-m+1}^{(i)\leftarrow} - \tilde{y}_{k-m+1} \right]^{\mathrm{T}} + R. \tag{35}\leftarrow$$

The cross-covariance between $\tilde{\chi}^{(i)\leftarrow}$ and $\tilde{\gamma}^{(i)}$ is

$$T_{k-m+1} = \sum_{i=0}^{2n} W_c^{(i)\leftarrow} \left[\tilde{\chi}_{k-m+1}^{(i)} - \tilde{x}_{k-m+1} \right] \left[\tilde{\gamma}_{k-m+1}^{(i)} - \tilde{y}_{k-m+1} \right]^{\mathrm{T}}. \tag{36}\leftarrow$$

The time-varying filter gain is defined as

$$K_{k-m+1} = T_{k-m+1} S_{k-m+1}^{-1}. \tag{37}\leftarrow$$

Finally, the conditional mean and covariance are updated similar to the EKF update equations

$$\hat{x}_{k-m+1} = \tilde{x}_{k-m+1} + K_{k-m+1}(y_{k-m+1} - \tilde{y}_{k-m+1}), \tag{38}$$

$$\widehat{P}_{k-m+1} = \tilde{P}_{k-m+1} - K_{k-m+1} S_{k-m+1} K_{k-m+1}^{\mathrm{T}}. \tag{39}\leftarrow$$

The nonlinear functions are more effectively handled through statistical linearization instead of functional approximation using Taylor series. When the sigma points are transformed through a nonlinearity, the transformed mean and covariance are captured with accuracy to the second order Taylor series expansion of the nonlinear function [16]. The advantages of UKF over EKF include more accurate error propagation and the absence of Jacobian computations.

Inequality constraints on the states can be used to alter the definition of sigma points and adjust the weights accordingly. Each time the sigma points are redefined, the sigma points located outside the constraints are projected to the constraint boundary [18]. A similar approach to constraints handling was used by Vachhani and coworkers [30]. However, their constrained optimization approach to update the sigma points is incorrect because the user-defined sigma points are functions of the state estimate and optimality of the sigma points may not be defined independently [31]. Equality constraints are not so straightforward to handle because the sigma points are generally not subject to multivariate relationships. Constraints on the noise variables and systems that are nonlinear in the noise terms can be handled by the augmented form of the UKF described next.

3.3. Arrival cost using augmented unscented Kalman filter

In the augmented form of the unscented Kalman filter (aUKF) the augmented state vector and covariance matrix are defined as

$$\hat{\mathbf{x}}_{k-m} = \left[\hat{x}_{k-m}^{\mathrm{T}}, 0, 0 \right]^{\mathrm{T}}, \tag{40}$$

$$\widehat{\mathbf{P}}_{k-m} = \mathrm{diag}(\widehat{P}_{k-m}, Q, R). \tag{41}\leftarrow$$

Recall that the system noise vector has the same dimension n as the state vector and the measurement noise vector dimension is p. Let $\mathbf{n} = 2n + p$ be the dimension of the augmented state vector and define $2\mathbf{n} + 1$ sigma points $\widehat{\mathbf{X}}_{k-m}^{(i)\leftarrow}$ using the mean vector $\hat{\mathbf{x}}_{k-m}$ and the covariance matrix $\widehat{\mathbf{P}}_{k-m}$. The corresponding weights $W_a^{(i)\leftarrow}$ and $W_c^{(i)\leftarrow}$ are defined using $\mathbf{n}$.

- **Prediction:** The sigma points are propagated through the system model for $i = 1, \ldots, 2\mathbf{n}$

$$\widetilde{\mathbf{X}}_{k-m+1}^{(i)\leftarrow}(1:n) = f\left(\widehat{\mathbf{X}}_{k-m}^{(i)}(1:n) \right) + \widehat{\mathbf{X}}_{k-m}^{(i)}(n+1:2n), \tag{42}$$

where $\widehat{\mathbf{X}}_{k-m}^{(i)}(1:n)$ are the elements of sigma point corresponding to the state while the elements $\widehat{\mathbf{X}}_{k-m}^{(i)}(n+1:2n)$ are from the system noise. Clearly it is possible to use system models that are nonlinear in noise terms. If the noise is subject to boxed constraints, the sigma points from the infeasible region are projected to the constraint boundary before using in the model. The *a priori* mean vector and covariance matrix for arrival cost are

$$\tilde{x}_{k-m+1} = \sum_{i=0}^{2\mathbf{n}} W_a^{(i)} \widetilde{\mathbf{X}}_{k-m+1}^{(i)}(1:n), \tag{43}$$

$$\tilde{P}_{k-m+1} = \sum_{i=0}^{2\mathbf{n}} W_c^{(i)\leftarrow} \left[\widetilde{\mathbf{X}}_{k-m+1}^{(i)\leftarrow}(1:n) - \tilde{x}_{k-m+1} \right] \left[\widetilde{\mathbf{X}}_{k-m+1}^{(i)\leftarrow}(1:n) - \tilde{x}_{k-m+1} \right]^{\mathrm{T}}. \tag{44}\leftarrow$$

- **Update:** The previously transformed sigma points are then re-transformed through the measurement equation for $i = 1, \ldots, 2\mathbf{n}$

$$\widetilde{\mathbf{Y}}_{k-m+1}^{(i)\leftarrow} = h\left(\widetilde{\mathbf{X}}_{k-m}^{(i)}(1:n) \right) + \widetilde{\mathbf{X}}_{k-m}^{(i)}(2n+1:\mathbf{n}), \tag{45}$$

where $\widetilde{\mathbf{X}}_{k-m}^{(i)}(2n+1:\mathbf{n})$ are the elements corresponding to the measurement noise. Again it is possible to use measurement equations that are nonlinear in the noise and incorporate boxed constraints on noise. The expected value of the output $\tilde{y}_{k-m+1}$ is the weighted mean of $\widetilde{\mathbf{Y}}_{k-m+1}^{(i)}$ as in UKF and the covariance is

$$S_{k-m+1} = \sum_{i=0}^{2\mathbf{n}} W_c^{(i)\leftarrow} \left[\widetilde{\mathbf{Y}}_{k-m+1}^{(i)\leftarrow} - \tilde{y}_{k-m+1} \right] \left[\widetilde{\mathbf{Y}}_{k-m+1}^{(i)} - \tilde{y}_{k-m+1} \right]^{\mathrm{T}}, \tag{46}\leftarrow$$

and the cross-covariance is

$$T_{k-m+1} = \sum_{i=0}^{2\mathbf{n}} W_c^{(i)\leftarrow} \left[\widetilde{\mathbf{X}}_{k-m+1}^{(i)\leftarrow}(1:n) - \tilde{x}_{k-m+1} \right] \left[\widetilde{\mathbf{Y}}_{k-m+1}^{(i)\leftarrow} - \tilde{y}_{k-m+1} \right]^{\mathrm{T}}. \tag{47}\leftarrow$$

The remaining equations for the filter gain and the update of the conditional mean and covariance are the same as those shown for UKF.

3.4. Particle filter

3.4.1. Importance sampling

Let the set $\{x^{(i)} : i = 1, \ldots, N\}$ denote random samples of the state vector distributed according to the state pdf $p(x)$. The expected value of a real function of the state $\phi(x)$ can be approximately evaluated by the sample mean of the function values,

$$\mathrm{E}[\phi(x)] = \int \phi(x) p(x) \mathrm{d}x \approx \frac{1}{N} \sum_{i=1}^{N} \phi(x^{(i)}). \tag{48}$$

This Monte Carlo sampling based approach to compute the moments of a multidimensional pdf is practically appealing because

the convergence of the sample mean to the expected value is independent of the dimension of the state vector. However, if the state pdf is not amenable to generate the samples or particles, this convenient Monte Carlo approach becomes impractical.

Let $q(x)$ be a pdf that is similar to the state pdf $p(x)$, such that they have the same support. Suppose it is possible to generate the samples $\{x^{(i)} : i = 1, \ldots, N\}$ distributed according to the pdf $q(x)$ instead, then the desired expectation can be computed as

$$E[\phi(x)] \approx \sum_{i=1}^{N} \phi(x^{(i)}) u^{(i)},$$ (49)

where the normalized weights are defined as

$$u^{(i)} = \frac{p(x^{(i)})/q(x^{(i)})}{\sum_{i=1}^{N} p(x^{(i)})/q(x^{(i)})}.$$ (50)

The pdf $q(x)$ is known as the importance or proposal density and $u^{(i)}$ are the importance weights. Sequential Importance Sampling (SIS) is a general Monte Carlo algorithm for approximately implementing the conceptual solution for recursive Bayesian estimation outlined in Eqs. (12) and (14). No assumptions are made about the shape of the conditional pdf, which is approximately represented by a finite set of random samples and associated weights. The moments necessary to draw the state estimates are then computed based on the samples and weights.

The SIS algorithm for state estimation is also known by the generic name of particle filter. Its various versions differ from each other in the choice of the importance density and the task of when and how resampling is implemented. The resampling task essentially discards the particles with very low weights in favor of enriching the ones with significant weights. The particular SIS variant called Sampling Importance Resampling (SIR), also known as the "bootstrap" filter, is a seminal and intuitively simple implementation of the particle filter [19]. Since then, many improved choices of the importance density and resampling methods have been introduced. The SIR is considered here for discussion only due to its simplicity. The reader is advised to explore alternate implementations of the particle filter as well [10,32].

3.4.2. Arrival cost using particle filter

- **Prediction:** Given the samples or particles $\{\hat{x}_{k-m}^{(i)} : i = 1, \ldots, N\}$ distributed according to the conditional density $p(x_{k-m}|Y_{1:k-m})$, it is desired to generate the samples of the *a priori* density $p(x_{k-m+1}|Y_{1:k-m})$. Such samples $\{\tilde{x}_{k-m+1}^{(i)} : i = 1, \ldots, N\}$ are generated by first sampling the noise vectors $\{w_{k-m}^{(i)} : i = 1, \ldots, N\}$ from the system noise pdf $p_w(w_{k-m}) = \mathcal{N}(0, Q)$ and then propagating the conditional samples through the system model

$$\tilde{x}_{k-m+1}^{(i)} = f(\hat{x}_{k-m}^{(i)}) + w_{k-m}^{(i)}, \quad i = 1, \ldots, N.$$ (51)

It is obvious that systems that are nonlinear in noise terms, non-Gaussian noise and constraints are readily handled.

In the SIR algorithm, the importance density is conveniently chosen as the *a priori* density $p(x_{k-m+1}|Y_{1:k-m})$ because its support includes that of the conditional density, which is more narrowly supported. Also in SIR, resampling is performed at every time instance, as a consequence of which the importance weights are fixed as $\{u_{k-m}^{(i)} = 1/N : i = 1, \ldots, N\}$. The mean and covariance for formulating the arrival cost are determined as

$$\tilde{x}_{k-m+1} = \frac{1}{N} \sum_{i=1}^{N} \tilde{x}_{k-m+1}^{(i)},$$ (52)

$$\tilde{P}_{k-m+1} = \frac{1}{N} \sum_{i=0}^{N} \left(\tilde{x}_{k-m+1}^{(i)} - \tilde{x}_{k-m+1} \right) \left(\tilde{x}_{k-m+1}^{(i)} - \tilde{x}_{k-m+1} \right)^{\mathrm{T}}.$$ (53)

- **Update:** For the particular choice of the *a priori* density as the importance density, Bayes rule yields the updated weights as,

$$u_{k-m+1}^{(i)} = \frac{p(y_{k-m+1}|\tilde{x}_{k-m+1}^{(i)})}{\sum_{i=1}^{N} p(y_{k-m+1}|\tilde{x}_{k-m+1}^{(i)})}, \quad i = 1, \ldots, N,$$ (54)

where the likelihood function is evaluated at each sample by

$$p(y_{k-m+1}|\tilde{x}_{k-m+1}^{(i)}) = c_4 e^{-\frac{1}{2}\|y_{k-m+1} - h(\tilde{x}_{k-m+1}^{(i)})\|_{R^{-1}}^2}.$$ (55)

The resampling procedure generates the conditional samples of the state $\{\hat{x}_{k-m+1}^{(i)} : i = 1, \ldots, N\}$ such that $\Pr[\hat{x}_{k-m+1}^{(i)} = \tilde{x}_{k-m+1}^{(j)}] = u_{k-m+1}^{(j)}$. The corresponding weights are reset to $\{u_{k-m+1}^{(i)} = 1/N : i = 1, \ldots, N\}$ after resampling. The updated conditional pdf is approximately represented as

$$p(x_{k-m+1}|y_{k-m+1}) \approx \frac{1}{N} \sum_{i=1}^{N} \delta(x_{k-m+1} - \hat{x}_{k-m+1}^{(i)})$$ (56)

It is not necessary to compute the conditional moments because the entire sample set is stored and the prediction step is revisited at the next time instance. The advantages of the SIR algorithm are, it is not necessary to propagate the importance weights because after resampling they are always fixed at $1/N$. The computation of the updated importance weights is simple because they are proportional to the likelihood function. In the prediction stage, the particles can be propagated in parallel via a single vectorized function call to the system model. Additional screening steps can be incorporated to impose constraints on the particles [21]. The disadvantages are, the SIR filter requires more samples because the importance density is unaware of the measurement information and the frequent resampling can lead to rapid collapse of the samples into a few values, a phenomenon known as degeneracy [19].

3.5. Cell filter

3.5.1. Aggregate Markov chain

According to the constraints imposed on the state vector, the temporal dynamics of system are confined to $\mathbb{X}$. Let all the state space outside the constraints $\overline{\mathbb{X}} = \mathbb{R}^n - \mathbb{X}$ be a single sink cell $z^{(0)}$, where the state trajectories are terminated. Let each state variable in $\mathbb{X}$ be discretized into a collection of intervals, such that N indivisible n-dimensional cells are formed in the state space. Together with the sink cell the state cell space is defined by the collection of numbered cells $\mathbb{Z} = \{z^{(i)} : i = 0, 1, \ldots, N\}$, where z corresponds to an n-tuple position identifier at the center of the cell.

Let the probability density supported over $\mathbb{X}$ be coarsely represented by a piece-wise constant function or probability mass vector (pmv) over $\mathbb{Z}$

$$\mathbf{p}(z) = [m^{(0)}, m^{(1)}, \ldots, m^{(N)}]^{\mathrm{T}}, \quad m^{(0)} = 0 \text{ and } \sum_{i=0}^{N} m^{(i)} = 1.$$ (57)

The point-to-point dynamics of the system in state space are coarsely represented as transitions in cell space between cells containing the points. The probabilistic behavior of the cell transitions can be associated with a stationary Markov chain. Transition to the sink cell are considered to be terminal. Let $p^{(ij)}$ for $i, j = 0, 1, \ldots, N$, be the probabilities of transition to cell $z^{(i)}$ from cell $z^{(j)}$, such that an $(N + 1) \times (N + 1)$ stochastic matrix $\mathbf{P}$ is defined.

By definition $p^{(0,0)} = 1$ and $p^{(0j)} = p^{(i0)} = 0$ for $i, j = 1, \ldots, N$. The other transition probabilities are determined by random simulations in the cell space via a procedure known as Generalized Cell Mapping [33]. In each cell $z^{(j)}$, a number of uniformly sampled initial conditions are chosen, $x^{(s)}, s = 1, \ldots, S$, while the same number of random numbers are sampled as $w^{(s)}$ from the process noise

density $\mathcal{N}(0, Q)$. The initial conditions and the random noise samples are transformed through the model to locate image points $x^{(r)}$

$$x^{(r)} = f(x^{(s)}) + w^{(s)}. \qquad (58)$$

The image points will be scattered among several image cells close to z^j. The probability of transition from cell $z^{(j)}$ to cell $z^{(i)}$ is

$$p^{(ij)} = \frac{\mathbf{r}}{\mathbf{s}}, \quad i, j = 1 \ldots, N, \qquad (59)$$

where $\mathbf{r}$ are the number of image points located in the image cell z^i. Sampling of N $\mathbf{s}$ initial conditions over the entire cell space can be performed in one step and the N $\mathbf{s}$ simulations can be carried out efficiently in parallel by a single vectorized function call to the system model. It is a one time computational burden to model the system as a stationary aggregate Markov chain. The evolution of state density is approximately realized through the linear transformation

$$\mathbf{p}(z_{k+1}) = \mathbf{P}\mathbf{p}(z_k), \qquad (60)$$

which is a discretized version of the Chapman–Kolmogorov equation. $\mathbf{P}$ is the $(N+1) \times (N+1)$ cell transition probability matrix.

3.5.2. Arrival cost using cell filter

- **Prediction:** Given the conditional pmv $\mathbf{p}(z_{k-m}|Y_{1:k-m})$ as a piecewise constant approximation of the conditional pdf $p(x_{k-m}|Y_{1:k-m})$, the *a priori* pmv is computed by the transformation

$$\mathbf{p}(z_{k-m+1}|Y_{1:k-m}) = \mathbf{P}\mathbf{p}(z_{k-m}|Y_{1:k-m}). \qquad (61)$$

The mean and covariance for arrival cost are obtained as

$$\tilde{x}_{k-m+1} = \sum_{i=0}^{N} m_{k-m+1|k-m}^{(i)} z^{(i)}, \qquad (62)$$

$$\widetilde{P}_{k-m+1} = \sum_{i=0}^{N} m_{k-m+1|k-m}^{(i)} (z^{(i)} - \tilde{x}_{k-m+1})(z^{(i)} - \tilde{x}_{k-m+1})^{\mathrm{T}}, \qquad (63)$$

where $z^{(i)}$ indicates the coordinates of the center of the cell.

- **Update:** The likelihood function $p(y_{k-m+1}|x_{k-m+1})$ is approximated by the piecewise constant likelihood mass vector (lmv) as

$$\mathbf{l}(y_{k-m+1}|z_{k-m+1}) = c_4 e^{-\frac{1}{2}\|y_{k-m+1} - h(z^{(i)})\|_{R^{-1}}^2}, \quad i = 1, \ldots, N, \qquad (64)$$

with $\mathbf{l}(y_{k-m+1}|z_{k-m+1}^{(0)}) = 0$. The *a posteriori* conditional pmv is updated using Bayes rule

$$\mathbf{p}(z_{k-m+1}|Y_{1:k-m+1}) = \frac{\mathbf{l}(y_{k-m+1}|z_{k-m+1}) \odot \mathbf{p}(z_{k-m+1}|Y_{1:k-m})}{\sum_{i=0}^{N} \mathbf{l}(y_{k-m+1}|z_{k-m+1}^{(i)}) \mathbf{p}(z_{k-m+1}^{(i)}|Y_{1:k-m})}, \qquad (65)$$

where $\odot$ is the Haddamard product.

It is not necessary to compute the moments of the conditional pmv because the entire pmv is used at the next prediction step for assigning the arrival cost. Constraints on the states are easily handled by limiting the cells to constraint surfaces and noise constraints are imposed during cell mapping [25]. The obvious disadvantage of the cell filter approach is the explosion of the cell numbers with dimension. Although the online computational cost is mostly limited to the linear transformation of the pmv, limitations on memory and storage make the CF unsuitable for unconstrained and high dimensional systems.

4. Approximation of arrival cost II

The approximation of the *a priori* density $p(x_{k-m+1}|Y_{1:k-m})$ outlined at the beginning of Section 3 may be considered as *parametric* estimation of the density. It is presumed that prior knowledge of the system enabled the user to specify a Gaussian structure to the pdf. One may, of course, choose this structure simply as a matter of convenience and familiarity to handle it. Emphasis is then placed on finding good estimates of the mean vector $\tilde{x}_{k-m+1}$ and the covariance matrix $\widetilde{P}_{k-m+1}$ via prediction by various nonlinear filters. The UKF propagation of these parameters through nonlinearity are accurate to the second order Taylor series approximation, which is a clear improvement over EKF. However, since the conditional update in UKF is performed only on two moments, the conditional pdf may not be adequately described. On the other hand, the estimates of $\tilde{x}_{k-m+1}$ and $\widetilde{P}_{k-m+1}$ by PF and CF are inspired by the diversity of the particles and the spread of probability mass points respectively. Hence, the estimates can be expected to be closer to the true mean and covariance of the non-Gaussian *a priori* pdf. Furthermore, the conditional update of PF and CF reveals approximations of the entire pdf rather than just two moments. The results of PF and CF point to the possibility of more realistic, admittedly more ambitious due to increased computational demands, attempts to approximate the arrival cost term.

It is more likely for the *a priori* pdf to be non-Gaussian than the conditional density because the measurement information typically narrows the support and suppresses unlikely modes, especially with linear measurement equations. In this section, consider that $p(x_{k-m+1}|Y_{1:k-m})$ exhibits sufficient non-Gaussian characteristics such as lack of symmetry and multiple modes, and consequently the first two moments make inadequate representation of the pdf even when they are accurately estimated by UKF, PF or CF. By relaxing the rigid specification of a Gaussian structure to $p(x_{k-m+1}|Y_{1:k-m})$, one can consider broader *nonparametric* estimation of the entire density, therefore formulating nonparametric arrival cost. The EKF and UKF are clearly not suitable for the task.

Consider the samples available from the particle filter after prediction $\{\tilde{x}_{k-m+1}^{(i)} : i = 1, \ldots, N\}$. The most commonly utilized nonparametric density estimation technique is the kernel density estimator [23]. The basic form of the estimator applied to the *a priori* density is

$$p(x_{k-m+1}|Y_{1:k-m}) \approx \frac{1}{Nb^N} \sum_{i=1}^{N} K\left(\frac{x - \tilde{x}_{k-m+1}^{(i)}}{b}\right), \qquad (66)$$

where b is called the bandwidth or smoothing parameter and $K(x)$ is the kernel function. Typically the kernel function is chosen as a symmetric unimodal pdf such as the multivariate Gaussian density $\mathcal{N}(0, 1)$, i.e,

$$K(x) = \frac{1}{(2\pi)^{n/2}} e^{(-\frac{1}{2}x^{\mathrm{T}}x)}. \qquad (67)$$

The kernel density estimate is a summation series of bell curves centered at the samples. The width of the curves is adjusted with the bandwidth parameter b. If b is too small, then spuriously sharp peaks will arise in the density estimate at the samples and if b is too large, legitimately sharp features such as multiple modes will get smoothed over. The proper choice of b is important to capture the non-Gaussian features of the underlying density. An optimum bandwidth b is frequently chosen by searching for the one that minimizes the integrated square error of the kernel estimate [23].

The typical size of the sample set employed in PF in most applications is not inordinately large because the motivation for PF is to compute the conditional mean efficiently. It is known that the moments converge independently of the dimension of the state according to the law of large numbers, allowing for reasonably sized sample sets. On the contrary, the motivation for kernel density estimation can be entirely different. Silverman [23] showed that the necessary size of the sample set grows exponentially with dimension if the intent is to ensure that the density at any point is obtained within a small relative mean error. Nearly a million

samples are required in ten dimensions even for estimating a smooth unimodal pdf using Gaussian kernels. It is not clear if measures such as point accuracy or reproduction of long tails in the density are the appropriate measures for the purpose of nonparametric arrival cost. If important features in the shape of the density such as peaks and asymmetry are identified, that alone is a significant improvement over the Gaussian assumption. The cautious user of kernel density estimator will weigh the computational cost of the kernel estimator to keep the MHE horizon small against the computational cost incurred in larger horizons based on simple two parameter arrival cost. Detailed studies of the computational tradeoff are necessary.

The direct formula in Eq. (66) for kernel estimate is computationally inefficient and rarely used as such. For high dimensional and large sample sets, calls to the exponential function of the Gaussian kernel can easily get out of hand while searching for optimum bandwidth. The kernel density estimation is a convolution of the samples with the kernel, hence in practice, it is performed using the Fourier transform via the far more efficient FFT. The Fourier transform methods also allow for extremely fast search for the best possible bandwidth. Interested reader should consider the optimal choices of b, several adaptive methods, alternate kernel functions and other density estimators in the monograph by Silverman [23].

The prediction from the cell filter produces a piece-wise constant approximation of $p(x_{k-m+1}|Y_{1:k-m})$ as the pmv $\mathbf{p}(z_{k-m+1}|y_{k-m})$. A closed-form approximation of the pdf is necessary to formulate the nonparametric arrival cost. One way is to generate samples distributed according to the discrete density and then use the kernel density estimation approach. Another way is to find a curve or surface fit of the discrete probability mass points as a continuous function over the state space. The additional computational demand can quickly turn prohibitive for surface fitting over high dimensional state spaces. It must be noted though that for low dimensional nonlinear systems where the arrival cost is more realistically formulated by nonparametric means, the performance gains can justify the additional complexity and computational cost of kernel estimator or surface fit of the discrete density as illustrated in a simulation example in the next section.

5. Simulation examples

Two simulation examples are presented in this section to demonstrate the use of sampling based nonlinear filters to initialize MHE. The first example is a two dimensional nonlinear system with constrained noise where the arrival cost is formulated parametrically with bivariate Gaussians. The second example is a benchmark univariate nonlinear system with a time-varying parameter for which no assumption about the shape of the state pdfs is used and the arrival cost is formulated nonparametrically. For performance comparison among various methods, the estimation accuracy is measured by the mean squared error (MSE),

$$\text{MSE} = \frac{1}{Kn} \sum_{k=1}^{K} \sum_{i=1}^{n} (x_k(i) - \bar{x}_k(i))^2, \qquad (68)$$

where K is the number of measurements and n is the length of the state vector. When appropriate, plots of a typical sample path are used in discussion, but the MSE shown in tables are based on averages of 100 realizations implemented in MATLAB on 3.2 GHz Intel Xeon processor running 64-bit Linux.

5.1. Example 1

Consider the following two dimensional discrete-time nonlinear system [6],

$$x_{k+1}(1) = 0.99x_k(1) + 0.2x_k(2), \qquad (69)$$

$$x_{k+1}(2) = -0.1x_k(1) + \frac{ax_k(2)}{b + x_k^2(2)} + w_k, \qquad (70)$$

where $w_k = |\xi_k|$ with $\xi_k \sim \mathcal{N}(0,1)$ and the constants are $a = 0.5$ and $b = 1$. A single linear measurement is available as

$$y_k = x_k(1) - 3x_k(2) + v_k, \qquad (71)$$

where $v_k \sim \mathcal{N}(0, 0.1^2)$. The initial state of the system is $x_0 = [1,0]^T$.

Process noise in the implementations of EKF and UKF is assumed as zero mean with unit variance, hence the non-negativity constraint is ignored. During the aUKF prediction step the sigma points of process noise are constrained by moving the negative value to the constraint boundary at zero. In PF and CF the system is simulated with random input as absolute values of samples with zero mean and unit variance. For all the filters, the state estimate is taken as the mean of the conditional pdf, although it is possible for PF and CF to provide the mode.

The particle filter is implemented with a 1000 samples. For the cell filter, state space containing $x(1) \in [0,8]$ and $x(2) \in [-1,3]$ is arbitrarily chosen to define 100 × 50 cells. The bounds are artificial because there are no constraints on the state variables. The transition probability matrix is computed with 500 samples per cell, i.e., 2.5 million initial conditions are mapped forward once using a single vectorized function call to the system model. The computation time for the transition probability matrix $\mathbf{P}$ is about 15 s.

Moving horizon estimation is implemented in horizons of $m = 3, 6$ and 12, with the weighting parameters $Q = 1$ and $R = 0.1^2$. The initial state pdf is assumed to be a bivariate Gaussian, and the a priori mean $\tilde{x}_{k-m+1}$ and covariance $\tilde{P}_{k-m+1}$ provided by the nonlinear filters are used to formulate the arrival cost. It is a reasonable assumption for this example because the pdfs display only one peak and some asymmetry due to the one sided process noise.

Case 1: Good prior. Information about the initial condition is assumed to be a Gaussian pdf with mean vector $\hat{x}_0 = [1,0]^T$, the same as the true initial condition, and covariance matrix $\hat{P}_0 = I_n$, the identity matrix. The predicted mean by each filter is shown as dotted lines in Figs. 1 and 2, with the error bars indicating the standard deviation. The dashed lines are the conditionally updated mean estimates by the filters and the solid line is the true state.

Ideally when the predicted mean is close to the true state, the variance should be small and vice versa. It is reasonable to require the spread of three standard deviations to straddle the true state. The EKF predicted mean is unsatisfactory because for $x(1)$ the
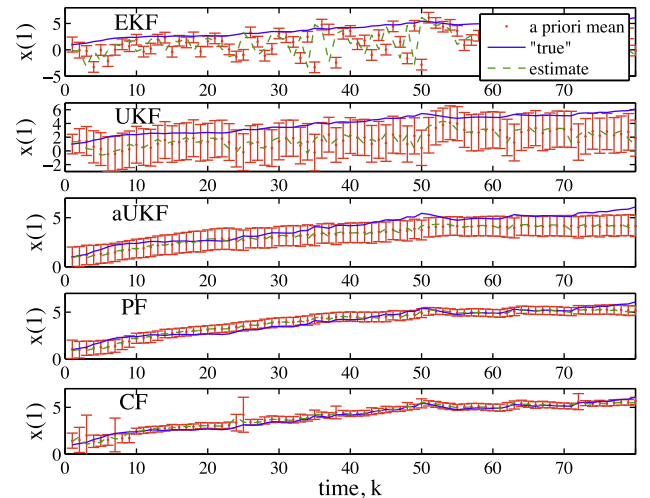

Fig. 1. Case 1: Good prior. Arrival cost parameters by filters for the state $x(1)$. Dotted lines with error bars indicate the predicted mean and standard deviation.
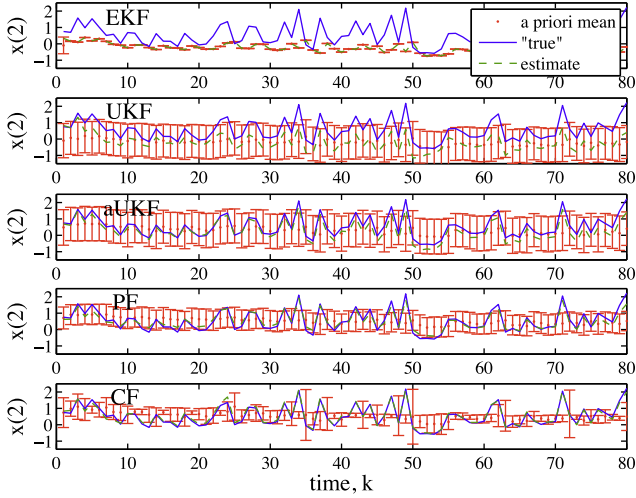
Fig. 2. Case 1: Good prior. Arrival cost parameters by filters for the state $x(2)$. Dotted lines with error bars indicate the predicted mean and standard deviation.

mean is biased and erratic, while for $x(2)$ it remained mostly close to zero. The arrival cost is significantly deficient because the small predicted variance placed high confidence in the mean while assigning insignificant probability near the true state.

The UKF predicted mean is biased but followed the true trend to some extent and the larger variance shows lowered confidence. It is an indication that the moments are propagated more accurately through nonlinearity by the unscented transform compared to Taylor series approximation in EKF. The use of the non-negativity constraint on the system noise helped the aUKF predicted mean to move closer to the true state, while the variance is also reduced. The predicted means by both PF and CF are slightly improved compared to aUKF, but the variances are considerably smaller. The moments computed by PF and CF are without the Gaussian assumption, whereas accuracy to second order approximation by the unscented transformation is guaranteed only for a Gaussian density. It is evident that all three sampling based methods provided vastly superior formulation of the arrival cost than EKF. The computational cost for processing one set of measurements by MHE is 6.6 CPU sec with EKF, 6.6 CPU sec with UKF, 7.6 CPU sec with the augmented form of UKF, 7.2 CPU sec with PF and 7.2 CPU sec with CF. In this example, while the code is not optimized for computation, it is evident that the sampling based filters add only a small amount of computational load to the MHE optimization load.

The average MSE results of all the estimators are summarized in Table 1, where the standard deviation of MSE is shown in parenthesis. MHE initialization method is indicated in the column headings and the first row is the MSE of the filter estimates. Although

Table 1
Case 1: Good prior. Average mean squared error of 100 realizations with the standard deviation shown in parenthesis.

Initialization method	Average MSE				
	EKF	UKF	aUKF	PF	CF
Filter, mean estimate	4.9 (0.7)	3.2 (0.5)	0.67 (0.33)	0.15 (0.11)	0.06 (0.05)
MHE, $m = 3$	2.0 (0.3)	1.2 (0.3)	0.55 (0.21)	0.13 (0.10)	0.08 (0.05)
MHE, $m = 6$	1.1 (0.2)	0.27 (0.10)	0.25 (0.10)	0.11 (0.07)	0.06 (0.04)
MHE, $m = 12$	0.61 (0.13)	0.07 (0.05)	0.07 (0.05)	0.06 (0.04)	0.04 (0.03)

the error comparison is not relevant between the filters and MHE because the former are mean estimates and the latter is mode estimate, they can be close in this case due to the unimodal conditional pdfs. Initialization by EKF required increasing horizons with rapidly increasing computational demand. The results with sampling based methods indicate that much smaller horizons can be used effectively. In this example, these filters by themselves provided fairly accurate estimates, i.e., the conditional pdfs are well approximated. Operating close to one end of the tradeoff scale, no dramatic improvement in performance is seen by MHE. However, a case is made for using MHE in the next two simulation case studies.

Case 2: Poor prior. The initial condition is assumed to be $\hat{x}_0 = [4, 3]^T$, away from the true initial condition $x_0 = [1, 0]^T$. The covariance matrix is chosen as $\widehat{P}_0 = 0.5^2 I_n$, which results in little probability near the true state. Estimates of the arrival cost parameters are displayed in Figs. 3 and 4. The unscented filter predictions are slow to converge but the uncertainty spread is acceptable. The CF recovered from the poor prior information with the fastest convergence. The PF predictions diverged considerably and the small variance placed insignificant probability near the true state result-
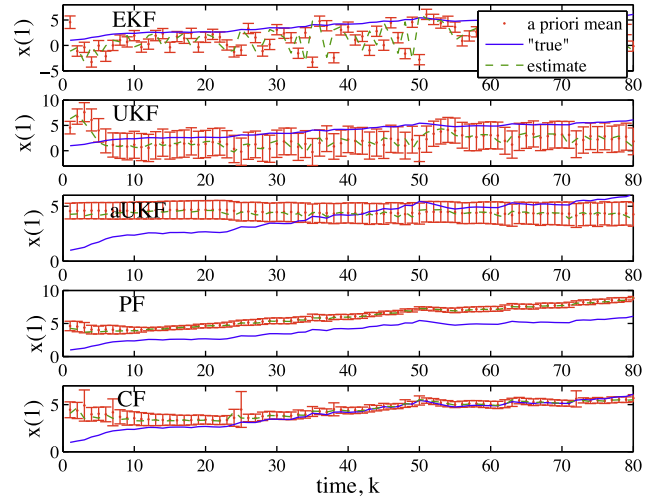

Fig. 3. Case 2: Poor prior. Arrival cost parameters by filters for the state $x(1)$. Dotted lines with error bars indicate the predicted mean and standard deviation.
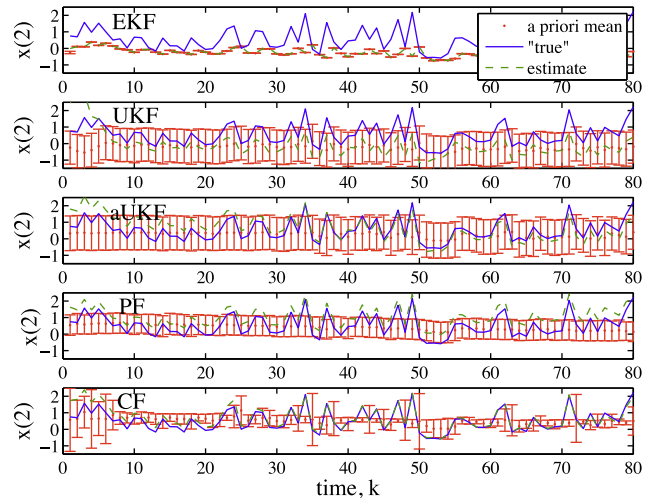

Fig. 4. Case 2: Poor prior. Arrival cost parameters by filters for the state $x(2)$. Dotted lines with error bars indicate the predicted mean and standard deviation.

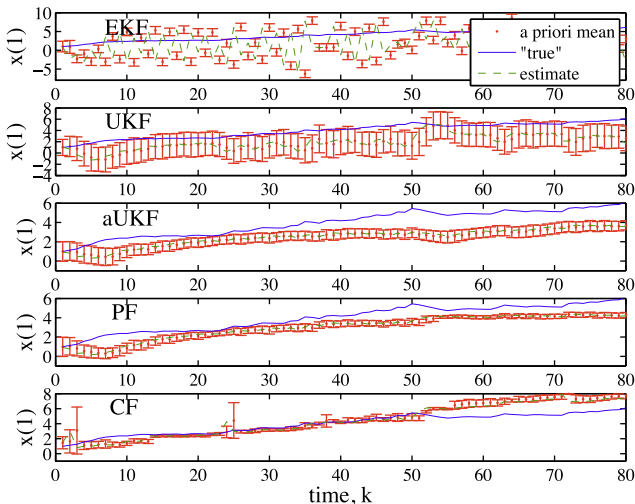| Initialization method	average MSE				
	EKF	UKF	aUKF	PF	CF
Filter, mean estimate	5.0 (0.7)	3.8 (0.5)	1.01 (0.22)	15.1 (12.4)	0.37 (0.10)
MHE, $m = 3$	2.1 (0.4)	1.4 (0.3)	0.65 (0.18)	15.2 (14.2)	0.23 (0.08)
MHE, $m = 6$	1.1 (0.2)	0.28 (0.10)	0.27 (0.10)	14.4 (13.2)	0.10 (0.04)
MHE, $m = 12$	0.62 (0.13)	0.07 (0.05)	0.07 (0.05)	14.3 (14.2)	0.04 (0.03)

ing in poor arrival cost. Implementations of SIR particle filters often struggle to converge when the prior information is poor [20].

Table 2 is a summary of the estimation performance. Even when large horizons are used, initialization by PF failed due to the divergence of the predicted mean and very small variance for $x(1)$. The MHE results indicate that estimation in horizons converged far more quickly, whereas the filters are slower to converge. Horizons initialized by CF are about half the size of the horizons initialized by UKF and aUKF for similar MSE values.
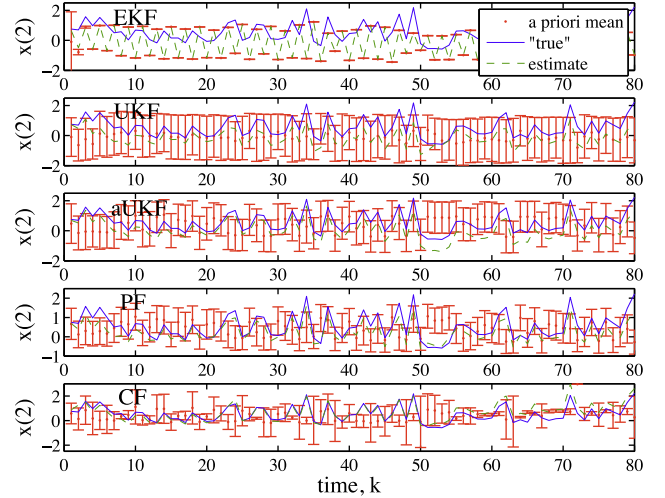
Case 3: Plant-model mismatch. The filters and MHE are provided with a modified model containing the parameters $a = -1$ and $b = 0.5$. This model's predictions for $x(2)$ have increased mean and variance compared to the plant signal. See Figs. 5 and 6 for the predicted mean and standard deviation and Table 3 for the MSE summary. Plant-model mismatch is tolerated well by estimation in large horizons, whereas the filters showed poor performance. The CF typically yields the best approximation of the state pdf, however the high fidelity to the model is detrimental in this case. Initialization by CF of large horizons did not improve the MSE appreciably because the predicted variance is often too small for $x(1)$ with little probability near the true state, which weighed down the MHE.

5.2. Example 2

The second illustrative example is a more challenging benchmark problem in nonlinear estimation research [10,19,24]. The dy-

Fig. 5. Case 3: Plant-model mismatch. Arrival cost parameters by filters for the state $x(1)$. Dotted lines with error bars indicate the predicted mean and standard deviation.

Fig. 6. Case 3: Plant-model mismatch. Arrival cost parameters by filters for the state $x(2)$. Dotted lines with error bars indicate the predicted mean and standard deviation.

Initialization method	MSE				
	EKF	UKF	aUKF	PF	CF
Filter, mean estimate	8.0 (0.9)	2.8 (0.4)	1.9 (0.6)	1.37 (0.62)	0.42 (0.29)
MHE, $m = 3$	2.5 (0.3)	1.1 (0.2)	0.80 (0.28)	0.83 (0.46)	0.52 (0.34)
MHE, $m = 6$	1.7 (0.2)	0.39 (0.12)	0.32 (0.12)	0.35 (0.19)	0.41 (0.40)
MHE, $m = 12$	1.2 (0.2)	0.14 (0.07)	0.14 (0.07)	0.14 (0.08)	0.23 (0.13)

namic model and measurement equation of a univariate time-varying nonlinear system are

$$x_{k+1} = \frac{x_k}{2} + \frac{25x_k}{1 + x_k^2} + 8\cos(1.2k) + w_k, \tag{72}$$

$$y_k = \frac{x_k^2}{20} + v_k, \tag{73}$$

where $w_k \sim \mathcal{N}(0, 10)$ and $v_k \sim \mathcal{N}(0, 1)$. Both the a priori and conditional densities of this system tend to be bimodal most of the time. The nonlinear measurement equation leads to a symmetric bimodal likelihood function for $y_k > 0$, which is the case more frequently. It is gross simplification to assume that the mean and variance can adequately describe the shape of the state pdf at any given time. The two parameter form of arrival cost is not advisable for initializing MHE for this system even in large horizons. Therefore, EKF and UKF are not suitable for MHE initialization.

The SIR particle filter and the cell filter are initialized with $\hat{x}_0 = 1$, the same as the true initial condition and $\hat{P}_0 = 1$. One thousand samples are used for the particle filter. The cell space is defined by 300 cells uniformly split over $x \in [-30, 30]$. The transition probability matrix is computed using 400 samples per cell in about 0.04 s. $\mathbf{P}$ is recomputed at every sampling time because the system contains a time-varying parameter.

The a priori pdf of the initial state of a horizon is expressed in the form of a polynomial of arbitrary degree J,

$$p(x_{k-m+1}|Y_{1:k-m}) \approx \psi(x_{k-m+1}) = \sum_{j=0}^{J} a_j x_{k-m+1}^j, \tag{74}$$

with the requirement that $\psi(x_{k-m+1}) \geq 0$. The value of this polynomial at the center of each cell $z \in \mathbb{Z}$ is available from the cell filter prediction as the discrete density or pmv $\mathbf{p}(z_{k-m+1}|y_{k-m})$. The polynomial coefficients $\{a_j\}$ are determined by least squares polynomial fit using MATLAB `polyfit` function. The best fit is automatically chosen as the one with the smallest error among the fits for $J = 1, \ldots, 50$. Such a large number may not be necessary for fitting bimodal functions, but polynomials exhibit poor point accuracy where the function is flat towards the ends of the defining interval. In addition to the least squares criterion, the fit is also subjected to non-negativity condition over the chosen support of x for the pdf

The predicted samples from the particle filter $\{\tilde{x}_{k-m+1}^{(i)-}\}$ are used for kernel density estimation by MATLAB `ksdensity` function, which is fast because it uses a bandwidth parameter in terms of the number of samples, not necessarily an optimal choice. The result is a discrete density sampled over the range of the particles. Hence, a polynomial fit is again used to approximate the *a priori* density in closed-form from the discrete density representing the samples.

The two filters are first used to estimate the conditional mean and also the conditional mode by maximizing the conditional pmv in the cell filter and the discrete kernel density estimate in PF. On the average, the mean estimates are closer to the true state than the mode estimates [24] as indicated by the average MSE in the bar chart in Fig. 7. MHE in a horizon of $m = 1$ is implemented by initializing with the polynomial fit for $p(x_{k-m+1}|Y_{1:k-m})$ provided by the cell filter. Not surprisingly, the MSE in Fig. 7 indicates that the MHE mode estimates are very close to the cell filter mode estimates, verifying that MHE in $m = 1$ is the conditional density when accurately initialized. The same result is obtained by initializing MHE with the polynomial fit provided by the particle filter.

The pitfalls of mode estimates from the filters are highlighted in the sample path in Fig. 8, where arrows mark several occasions when the mode estimates are of the opposite sign implying that the wrong mode is chosen. Closer inspection of the densities in one occurrence, at $k = 60$ shown in Fig. 9, reveals the difficulty of choice. The top panels show that (a) the *a priori* density has a stronger mode to the left of zero, (b) the symmetric bimodal likelihood function can narrow the support under the predicted modes but cannot alter the relative strengths of the peaks and (c) the conditional density has a strong mode to the left of zero while the true state is to the right. The measurement information is not enough to significantly improve the mode prediction. Therefore, the error of
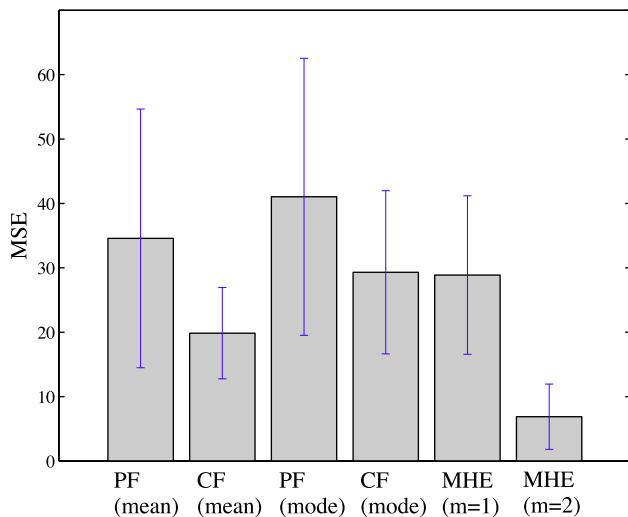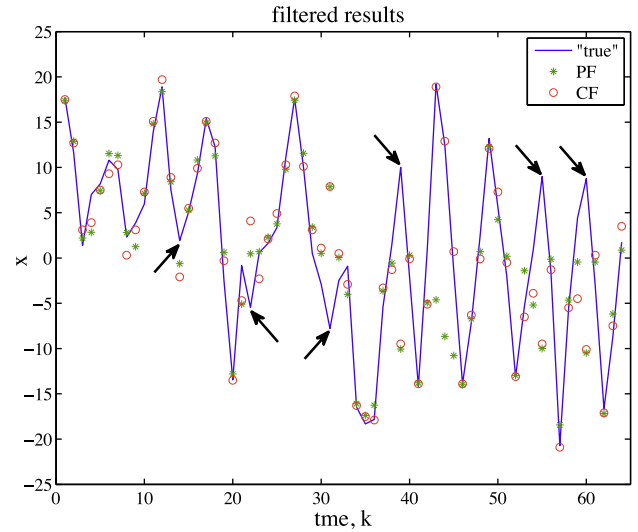

Fig. 8. Arrows mark the states for which the mode estimates by filters are of the opposite sign.

the mean estimate will be smaller. The bottom panels show the situation at $k = 61$, where the likelihood function suppressed multiple modes and the conditional density is unimodal resulting in a good estimate. Note that unlike traditional filters, the task of choosing a point estimate from the conditional pdf has no influence on the behavior of density based filters such as PF and CF.

Since the filter mode estimates are frequently inaccurate for this system, the question arises if MHE is suitable because MHE in $m = 1$ yields similar results. The situation looks more encouraging for MHE implemented in $m = 2$. Consider the horizon containing $k = 60$ and 61, the top panels in Fig. 10 display (a) a polynomial fit of the *a priori* density $p(x_{60}|y_{59})$ of the horizon and (b) the transition probability density $p(x_{61}|x_{60})$ formed like an S-shaped hill and the bottom panels display (c) the symmetric bimodal bivariate likelihood function and (c) the joint conditional density with a single mode located on the right side of zero for x_{60} and around zero for x_{61}. The benefits of drawing modes estimates from such joint densities are clearly evident in the sample path shown in Fig. 11, where most of the erroneous mode estimates from MHE in $m = 1$
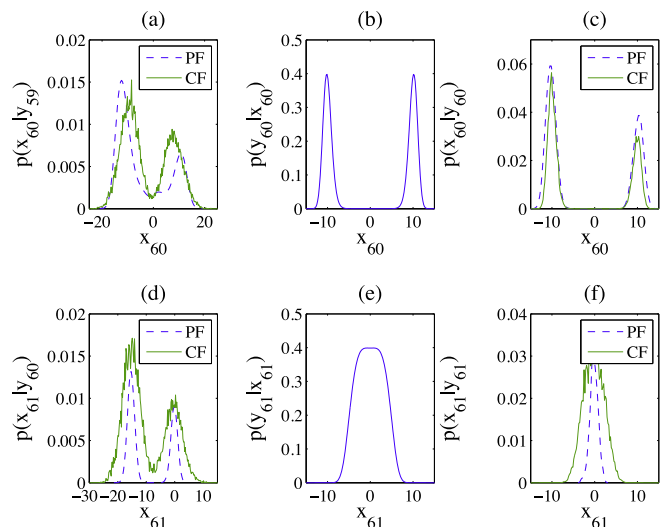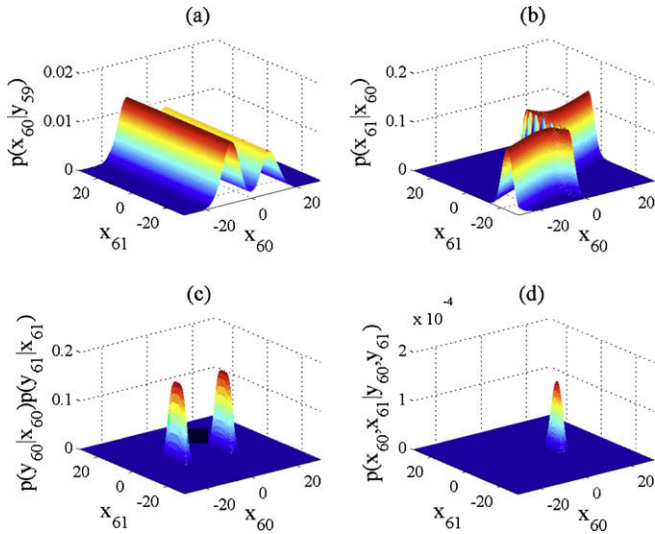

Fig. 7. Average mean squared error and standard deviation for the mean and mode estimates by filters and MHE mode estimates in horizons $m = 1$ and 2.

Fig. 9. At $k = 60$, (a) density estimate from PF and pmv from CF of *a priori* density, (b) likelihood function and (c) density estimate and pmv of conditional density. At $k = 61$, (d) *a priori* density (e) likelihood and (f) conditional density.

Fig. 10. In a horizon containing $k = 60$ and 61 (a) polynomial fit of *a priori* density $p(x_{60}|y_{59})$, (b) transition probability density $p(x_{61}|x_{60})$, (c) product of likelihood functions $p(y_{60}|x_{60})p(y_{61}|x_{61})$ and (d) joint conditional density $p(x_{60}, x_{61}|y_{60}, y_{61})$.
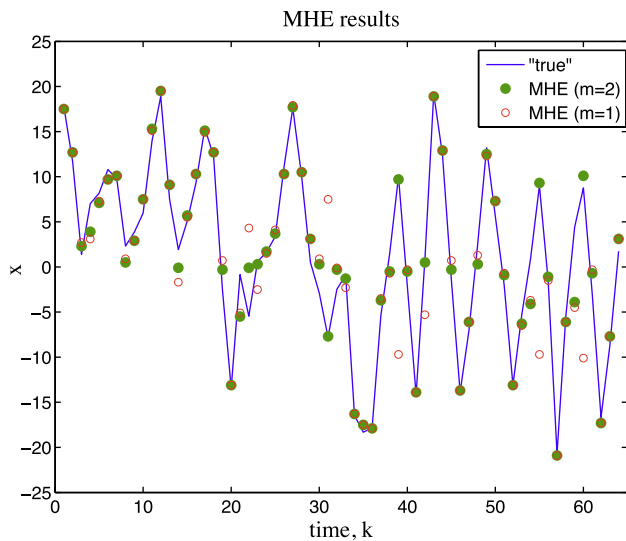
Fig. 11. Mode estimates using MHE in horizons $m = 1$ and 2.

are corrected by MHE in $m = 2$. This situation only illustrated the ease of drawing point estimates from the joint density, the conditioning of the joint density still refers to the filtering problem and not the smoothing problem. The average MSE and its variance reduced dramatically for MHE in $m = 2$ compared to the filters in the bar chart of Fig. 7. To the best of this author's knowledge such performance improvement has not been reported previously for this system using any nonlinear filter.

6. Conclusions

Initialization of moving horizon estimation by formulating the arrival cost accurately is a critical step for realtime implementation of MHE in small horizons. Traditionally, the EKF has been used to recursively update the arrival cost parameters. The drawbacks of EKF continue to keep the arrival cost as an open issue in MHE research. In this paper it is suggested that recent advances in sampling based or density based nonlinear filters can effectively

address some of the issues. It is shown that the arrival cost parameters can be very accurately computed and updated by sampling based methods without using functional linearization. This paper discussed the unscented Kalman filter, the particle filter and the cell filter as alternatives to the EKF approach to initialize MHE. The ease of implementation of these filters is also a strong motivation not to use EKF for arrival cost. The arrival cost can be more realistically formulated by avoiding the Gaussian assumption if nonparametric methods are used to find closed-form approximation of the *a priori* density. Kernel density estimation and curve fitting are discussed in this paper. The simulation examples clearly showed the superior performance of the sampling based methods over EKF. More detailed studies are needed to weigh the computational demands of nonparametric arrival cost against simply increasing the horizon length when poorly initialized.

References

[1] C.V. Rao, J.B. Rawlings, J.H. Lee, Constrained linear state estimation – a moving horizon approach, Automatica 37 (2001) 1619–1628.
[2] J.B. Rawlings, B.R. Bakshi, Particle filtering and moving horizon estimation, Comput. Chem. Eng. 30 (2006) 1529–1541.
[3] D.G. Robertson, J.H. Lee, A least squares formulation for state estimation, J. Process Control 5 (4) (1995) 291–299.
[4] D.G. Robertson, J.H. Lee, J.B. Rawlings, A moving horizon-based approach to least squares estimation, AIChE J. 42 (8) (1996) 2209.
[5] C.V. Rao, J.B. Rawlings, Constrained process monitoring: moving-horizon approach, AIChE J. 48 (1) (2002) 97–109.
[6] C.V. Rao, J.B. Rawlings, D.Q. Mayne, Constrained state estimation for nonlinear discrete-time systems: stability and moving horizon approximations, IEEE Trans. Autom. Control 48 (2) (2003) 246–258.
[7] A.H. Jazwinski, Stochastic Processes and Filtering Theory, Academic Press, New York, 1970.
[8] M.J. Tenny, J.B. Rawlings, S.J. Wright, Closed-loop behavior of nonlinear model predictive control, AIChE J. 50 (9) (2004) 2142–2154.
[9] P. Terwiesch, M. Agarwal, A discretized nonlinear state estimator for batch processes, Comput. Chem. Eng. 19 (2) (1995) 155–169.
[10] M.S. Arulampalam, S. Maskell, N. Gordon, T. Clapp, A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking, IEEE Trans. Signal Process. 50 (2) (2002) 174–188.
[11] E.L. Haseltine, J.B. Rawlings, Critical evaluation of extended Kalman filtering and moving horizon estimation, Ind. Eng. Chem. Res. 44 (2005) 2451–2460.
[12] R.S. Bucy, K.D. Senne, Digital synthesis of for nonlinear filters, Automatica 7 (1971) 287–298.
[13] G. Kitagawa, Non-Gaussian state-space modeling of nonstationary time series, J. Am. Stat. Assoc. 82 (400) (1987) 1032–1063.
[14] M. Simandl, J. Kralovec, T. Soderstrom, Advanced point-mass method for nonlinear state estimation, Automatica 42 (2006) 1133–1145.
[15] S. Julier, J. Uhlmann, H. Durrant-Whyte, A new method for the nonlinear transformation of means and covariances in filters and estimators, IEEE Trans. Automat. Control 45 (2000) 477–482.
[16] S. Julier, J. Uhlmann, Unscented filtering and nonlinear estimation, Proc. IEEE 92 (2004) 401–422.
[17] A. Romanenko, J.A.A.M. Castro, The unscented filter as an alternative to the ekf for nonlinear state estimation: a simulation case study, Comput. Chem. Eng. 28 (2004) 347–355.
[18] R. Kandepu, B. Foss, L. Imsland, Applying the unscented Kalman filter for nonlinear state estimation, J. Process Control 18 (2008) 753–768.
[19] N. Gordon, D.J. Salmond, A.F.M. Smith, Novel approach to nonlinear/non-Gaussian Bayesian state estimation, IEE Proc. F 140 (2) (1993) 107–113.
[20] W.S. Chen, B.R. Bakshi, P.K. Goel, S. Ungarala, Bayesian estimation via sequential monte carlo sampling: unconstrained nonlinear dynamic systems, Ind. Eng. Chem. Res 43 (2004) 4012–4025.
[21] L. Lang, W.S. Chen, B.R. Bakshi, P.K. Goel, S. Ungarala, Bayesian estimation via sequential monte carlo sampling – constrained dynamic systems, Automatica 43 (9) (2007) 1615–1622.
[22] T. Chen, J. Morris, E. Martin, Dynamic data rectification using particle filters, Comput. Chem. Eng. 32 (2008) 451–462.
[23] B.W. Silverman, Density Estimation for Statistics and Data Analysis, Chapman and Hall, New York, NY, 1986.
[24] S. Ungarala, Z.Z. Chen, K. Li, Bayesian state estimation of nonlinear systems using approximate aggregate Markov chains, Ind. Eng. Chem. Res. 45 (2006) 4208–4221.

[25] S. Ungarala, K. Li, Z. Chen, Constrained Bayesian state estimation using a cell filter, Ind. Eng. Chem. Res. 47 (19) (2008) 7312–7322.

[26] S. Ungarala, K. Li, The use of a cell filter for state estimation in closed-loop NMPC of low dimensional systems, J. Process Control (2008), doi:10.1016/j.jprocont.2008.04.020.

[27] C.C. Qu, J. Hahn, Computation of arrival cost for moving horizon estimation via unscented Kalman filtering, J. Process Control 19 (2) (2009) 358–363.

[28] S. Ungarala, E. Dolence, K. Li, Constrained extended Kalman filter for nonlinear state estimation, in: Proceedings of the IFAC DYCOPS-8, Cancun, Mexico, 2007, pp. 63–68.

[29] D. Simon, Optimal State Estimation, Wiley Interscience, Hoboken, NJ, 2006.

[30] P. Vachhani, S. Narasimhan, R. Rengaswamy, Robust and reliable estimation via unscented recursive nonlinear dynamic data reconciliation, J. Process Control 16 (2006) 1075–1086.

[31] S. Ungarala, Comments on Robust and reliable estimation via unscented recursive nonlinear dynamic data reconciliation, J. Process Control (2008), doi:10.1016/j.jprocont.2008.08.001.

[32] B. Ristic, S. Arulampalem, N. Gordon, Beyond the Kalman Filter: Particle Filters for Tracking Applications, Artech House, Boston, 2004.

[33] C.S. Hsu, Cell-to-Cell Mapping: A Method of Global Analysis for Nonlinear Systems, Applied Mathematical Sciences, Vol. 64, Springer-Verlag, New York, 1987.