Preprints of the 21st IFAC World Congress (Virtual)
Berlin, Germany, July 12-17, 2020

MHE Based State and Parameter Estimation for Systems subjected to Non-Gaussian Disturbances

Devyani Varshney * **Sachin C. Patwardhan** * **Mani Bhushan** *
Lorenz T. Biegler **

* *Department of Chemical Engineering, Indian Institute of Technology Bombay, Mumbai, India (e-mail: sachinp@iitb.ac.in)*
** *Department of Chemical Engineering, Carnegie Mellon University, Pittsburgh, USA (e-mail:biegler@cmu.edu)*

Abstract: Moving horizon estimation (MHE) is a popular state estimation technique, particularly due to its similarity with model predictive control. The probabilistic formulation of the conventional MHE is developed under the simplifying assumption that state disturbances and measurement noise densities are Gaussian. However, many systems of interest are subjected to uncertainties that have non-Gaussian densities. In current work, we formally extend an existing probabilistic Bayesian formulation of MHE [Varshney et al., 2019] to simultaneous state and parameter estimation for systems subjected to non-Gaussian uncertainties in the state dynamics and measurement model. The efficacy of the proposed MHE has been demonstrated by conducting stochastic simulation studies on a system subjected to non-Gaussian densities. Analysis of simulation results reveals that the estimation performance of the proposed MHE formulation is superior to estimation performances of the conventional Bayesian estimators that can handle non-Gaussian densities and employ the random walk model for parameter variations.

Keywords: Moving horizon estimation, Bayesian estimation, State and parameter estimation, Non-Gaussian disturbances

1. INTRODUCTION

Nonlinear state estimation is a prerequisite for advanced process control and fault diagnosis tasks. Nonlinear state estimation techniques combine predictions from uncertain system dynamics and noisy measurements to obtain unknown states and parameters. Bayesian state estimation techniques provide a way to optimally combine the information available in the presence of such uncertainties. Sequential Bayesian techniques such as extended Kalman filter (EKF), unscented Kalman filter (UKF), etc. have become very popular due to the ease of their implementation. However, all the existing methods rely on simplifying assumptions about the probability distributions of the underlying variables to obtain a tractable optimization problem. A popular assumption in these techniques is that uncertainties are modeled as Gaussian [Patwardhan et al., 2012]. Such assumptions could be largely incorrect. This has been shown by posterior probability density function (*pdf*) of concentration for a CSTR system in Chen et al. [2004]. These approximations may be more violated when constraints are present on the variables which force their probability to be zero for certain regions [Robertson et al., 1996]. However, such assumptions, even though largely incorrect, have been applied in most of the estimation techniques for simplifications irrespective of the original problem.

Estimation problems involving non-Gaussian densities have been recently solved using sampling based methods such as particle filters (PF) [Arulampalam et al., 2002] or ensemble Kalman filter (EnKF). An alternative approach to sequential Bayesian approaches, which has become popular over the last decade, is moving horizon approach [Lopez-Negrete et al., 2011]. Moving horizon estimation (MHE) approach formulates the estimation problem as a constrained optimization problem over a moving window framework similar to the model predictive control strategy. Also, the similarity with MPC formulation makes it relatively easy to maintain when implemented in combination with MPC. Further, as MHE is posed as an optimization problem, state or disturbance constraints can be easily added to the problem. Another advantage of the approach is that it provides filtered as well as smoothed estimates of states simultaneously within the same window. However, conventional MHE approaches have either been formulated in deterministic settings as weighted least squares problem or in probabilistic settings for Gaussian disturbances only [Rao, 2000]. MHE for systems subjected to non-Gaussian densities has been recently explored by Varshney et al. [2019]. However, approaches discussed so far only handle state estimation problems and unknown parameters are not considered in the problem statement.

Many parameters/inputs used in the nonlinear model are seldom known accurately. Significant mismatch between their actual values and the values assumed in the model can severely degrade estimation performance. Towards this end, general practice for parameter estimation is to assume models about the dynamic variation of the parameters and augment the parameters as additional states [Simon, 2006]. In this approach, dynamics associated with unmeasured parameter/input are usually modeled as a random walk process. Performance of the state and parameter estimator critically depends on the tuning of this random walk model. Such augmentation based parameter estimation approaches have also been applied within the MHE framework [Kühl et al., 2011, Shen et al., 2019]. Recently, Isaksson et al. [2015] proposed a batch formulation for state and parameter estimation in a manner similar to MHE. They converted their MAP (Maximum-a-posterior) formulation to ML (maximum likelihood) formulation by the inclusion of a bias correction term. However, such bias correction term addition would not be possible for systems subjected to non-Gaussian noise. Therefore, to the best of our knowledge, no probabilistic formulation of MHE is available for state and parameter estimation for nonlinear systems subjected to non-Gaussian disturbances. In our earlier preliminary work [Varshney et al., 2019], we have introduced a probabilistic framework of MHE for state estimation of systems subjected to non-linear Gaussian disturbances. In current work, we extend the formulation of Varshney et al. [2019] to state and parameter estimation for systems subjected to non-Gaussian disturbances. To begin with, we formulate a batch probabilistic Bayesian framework, which systematically incorporates the non-Gaussian noise in the formulation without making any simplifying assumptions. The resulting *pdf* is then maximized to obtain maximum-a-posterior (MAP) estimates. For on-line tracking of slowly changing model parameters, the batch formulation is further extended to moving window formulation. The efficacy of the proposed formulation is demonstrated on a system subjected to non-Gaussian densities.

The rest of the paper is organized as follows. Section 2 describes the nonlinear system considered in current work. Proposed batch formulation and its extension to moving horizon approach for non-Gaussian disturbances are given in detail in Section 3. The simulation case studies and the results obtained are discussed in Section 4. Finally, the conclusions are detailed in Section 5.

2. SYSTEM DESCRIPTION

Consider the following discrete time system:

$$\mathbf{x}_{k+1} = \mathbf{F}(\mathbf{x}_k, \mathbf{u}_k, \boldsymbol{\theta}) + \mathbf{w}_k \tag{1a}$$
$$\mathbf{y}_k = \mathbf{h}(\mathbf{x}_k) + \mathbf{v}_k \tag{1b}$$

In (1a), $\mathbf{x}_k \in \mathbb{R}^n$ represents the states of the system, $\mathbf{u}_k \in \mathbb{R}^m$ represents manipulated inputs at instant k, $\boldsymbol{\theta} \in \mathbb{R}^p$ represents unknown model parameters/ unmeasured disturbances while $\mathbf{w}_k$ represents the process noise. In (1b), $\mathbf{y}_k \in \mathbb{R}^r$ represents the measurements of a subset of states, while $\mathbf{v}_k \in \mathbb{R}^r$ represents measurement noise. Function $\mathbf{F} : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}^n$ in (1a) represents the nonlinear state dynamics while function $\mathbf{h} : \mathbb{R}^n \to \mathbb{R}^r$ in (1b) represents the nonlinear observation model. Both the functions $\mathbf{F}, \mathbf{h}$ could be non-differentiable or discontinuous. Further, in the following sections, the known quantities ($\mathbf{u}_k$) are dropped for notational simplicity in (1). The filtering problem is to find a point estimate for $\mathbf{x}_k$ and $\boldsymbol{\theta}$ governed by dynamics in (1a), using available measurements $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_k$ which are related to the states as in (1b).

3. BATCH AND MOVING HORIZON ESTIMATION

To proceed with the development of the proposed batch formulation, the following additional standard assumptions are introduced:

Assumption 1 Initial state of the system ($\mathbf{x}_0$) is a random variable with known *pdf*, $p(\mathbf{x}_0)$.
Assumption 2 $\mathbf{w}_k$ and $\mathbf{v}_k$ are independent white stochastic processes with known *pdf*, $p(\mathbf{w}_k)$ and $p(\mathbf{v}_k)$ respectively.
Assumption 3 Unknown model parameter ($\boldsymbol{\theta}$) is a random variable with known prior *pdf*, $p(\boldsymbol{\theta})$.
Assumption 4 $\mathbf{x}_0$ & $\boldsymbol{\theta}$ are independent of process disturbances $\mathbf{w}_k$ and measurement noise $\mathbf{v}_k$.

3.1 Batch Estimation Problem

Along similar lines of the formulation given by Varshney et al. [2019], we consider the problem of estimating the initial state, $\mathbf{x}_0$, $\boldsymbol{\theta}$ and the disturbance sequence $\mathbf{W}_0^{N-1} = \{\mathbf{w}_0, \mathbf{w}_1, \ldots, \mathbf{w}_{N-1}\}$ using a batch of measurements available from instant 1 to N, i.e. $\mathbf{Y}_1^N = \{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N\}$, and dynamic model (1). Towards this goal, we first construct the posterior *pdf*, $p(\mathbf{x}_0, \mathbf{W}_0^{N-1}, \boldsymbol{\theta}|\mathbf{Y}_1^N)$ by application of Bayes' rule to the joint density function $p(\mathbf{x}_0, \mathbf{W}_0^{N-1}, \boldsymbol{\theta}, \mathbf{Y}_1^N)$, which yields

$$p(\mathbf{x}_0, \mathbf{W}_0^{N-1}, \boldsymbol{\theta}|\mathbf{Y}_1^N) = \frac{p(\mathbf{Y}_1^N|\mathbf{x}_0, \mathbf{W}_0^{N-1}, \boldsymbol{\theta}) \times p(\mathbf{x}_0, \mathbf{W}_0^{N-1}|\boldsymbol{\theta}) \times p(\boldsymbol{\theta})}{p(\mathbf{Y}_1^N)} \tag{2a}$$
$$= \alpha_1 \underbrace{p(\mathbf{Y}_1^N|\mathbf{x}_0, \mathbf{W}_0^{N-1}, \boldsymbol{\theta})}_{\text{Term I}} \underbrace{p(\mathbf{x}_0, \mathbf{W}_0^{N-1}|\boldsymbol{\theta})}_{\text{Term II}} p(\boldsymbol{\theta}) \tag{2b}$$

where α_1 is a scaling constant that does not depend on the unknown parameters, states and disturbances. We can further simplify Term II in (2b). Using Assumptions 1, 2 and 3, the *pdf* in Term II can be expressed as follows:

$$p(\mathbf{x}_0, \mathbf{W}_0^{N-1}|\boldsymbol{\theta}) = p(\mathbf{x}_0) \times \prod_{j=0}^{N-1} p(\mathbf{w}_j) \tag{3}$$

Now applying Bayes' rule to Term I in (2b), we can rewrite the *pdf* as follows:

$$p(\mathbf{Y}_1^N|\mathbf{x}_0, \mathbf{W}_0^{N-1}, \boldsymbol{\theta}) = p(\mathbf{y}_N, \mathbf{Y}_1^{N-1}|\mathbf{x}_0, \mathbf{W}_0^{N-1}, \boldsymbol{\theta}) \tag{4a}$$
$$= p(\mathbf{y}_N|\mathbf{Y}_1^{N-1}, \mathbf{x}_0, \mathbf{W}_0^{N-1}, \boldsymbol{\theta}) p(\mathbf{Y}_1^{N-1}|\mathbf{x}_0, \mathbf{W}_0^{N-1}, \boldsymbol{\theta}) \tag{4b}$$

Further, with given set of $\boldsymbol{\theta}$, $\mathbf{W}_0^{N-1}$ and $\mathbf{x}_0$, recursive use of state dynamics (1a) yields:

$$\mathbf{x}_1 = \mathbf{F}(\mathbf{x}_0, \boldsymbol{\theta}) + \mathbf{w}_0, \ldots, \mathbf{x}_N = \mathbf{F}(\mathbf{x}_{N-1}, \boldsymbol{\theta}) + \mathbf{w}_{N-1} \tag{5}$$

Eq. (5) allow us to write first density function in (4b) as follows:

$$p(\mathbf{y}_N|\mathbf{Y}_1^{N-1}, \mathbf{x}_0, \mathbf{W}_0^{N-1}, \boldsymbol{\theta}) = p(\mathbf{y}_N|\mathbf{x}_N, \mathbf{X}_1^{N-1}, \mathbf{Y}_1^{N-1}) \tag{6}$$

where, $\mathbf{X}_1^{N-1} = \{\mathbf{x}_1, \mathbf{x}_1, \ldots, \mathbf{x}_{N-1}\}$. Using arguments similar to (6), the second density function in (4b) can be obtained as given below:

$$p(\mathbf{Y}_1^{N-1}|\mathbf{x}_0, \mathbf{W}_0^{N-1}, \boldsymbol{\theta}) = p(\mathbf{Y}_1^{N-1}|\mathbf{x}_N, \mathbf{X}_1^{N-1}) \tag{7}$$

According to (1b), $\mathbf{y}_k$ depends on $\mathbf{x}_k$ alone and is independent of past or future system variables. Therefore, (6) and (7) can be written as follows:

$$p(\mathbf{y}_N|\mathbf{x}_N, \mathbf{Y}_1^{N-1}, \mathbf{X}_1^{N-1}) = p(\mathbf{y}_N|\mathbf{x}_N) \tag{8}$$
$$p(\mathbf{Y}_1^{N-1}|\mathbf{x}_N, \mathbf{X}_1^{N-1}) = p(\mathbf{Y}_1^{N-1}|\mathbf{X}_1^{N-1}) \tag{9}$$

Using arguments similar to (8) for $p(\mathbf{Y}_1^{N-1}|\mathbf{X}_1^{N-1})$, (9) can be further simplified as follows:

$$p(\mathbf{Y}_1^{N-1}|\mathbf{X}_1^{N-1}) = p(\mathbf{y}_{N-1}|\mathbf{x}_{N-1})\dots p(\mathbf{y}_1|\mathbf{x}_1) \tag{10}$$

According to (1b), $\mathbf{y}_k = \mathbf{h}(\mathbf{x}_k) + \mathbf{v}_k$, it follows that, $\mathbf{y}_k$ depends only on $\mathbf{v}_k$ for given $\mathbf{x}_k$. Thus, the final expression for the *pdf* (10) requires the computation of $p(\mathbf{y}_k|\mathbf{x}_k) = p\big(\mathbf{h}(\mathbf{x}_k) + \mathbf{v}_k|\mathbf{x}_k\big)$, which can be computed by using $\mathbf{v}_k \sim p(\mathbf{v}_k)$ as follows:

$$p(\mathbf{y}_k|\mathbf{x}_k) = p\Big(\mathbf{h}(\mathbf{x}_k) + \mathbf{v}_k|\mathbf{x}_k\Big) = p_{\mathbf{v}_k}(\mathbf{y}_k - \mathbf{h}(\mathbf{x}_k)) \tag{11}$$

Using (11), (10) and (4b), (4a) can be rewritten as follows:

$$p(\mathbf{Y}_1^N|\mathbf{x}_0, \mathbf{W}_0^{N-1}, \boldsymbol{\theta}) = p(\mathbf{Y}_1^N|\mathbf{X}_1^N) = \prod_{j=1}^{N} p_{\mathbf{v}_j}(\mathbf{y}_j - \mathbf{h}(\mathbf{x}_j)) \tag{12}$$

Further, rearranging the terms in (2b) and substituting (3) and (12) in (2b), we can re-write (2b) as follows:

$$\begin{aligned} &p(\mathbf{x}_0, \mathbf{W}_0^{N-1}, \boldsymbol{\theta}|\mathbf{Y}_1^N) \\ &= \alpha_1 p(\mathbf{x}_0) \times p(\boldsymbol{\theta}) \times \prod_{j=0}^{N-1} p(\mathbf{w}_j) \times \prod_{j=1}^{N} p_{\mathbf{v}_j}(\mathbf{y}_j - \mathbf{h}(\mathbf{x}_j)) \end{aligned} \tag{13}$$

Since the conditional density $(p(\mathbf{x}_0, \mathbf{W}_0^{N-1}, \boldsymbol{\theta}|\mathbf{Y}_1^N))$ in Bayesian estimation is called the posterior density, the estimate that maximizes this density is called the *maximum a posteriori* (MAP) estimate. Constructing MAP estimates of $(\mathbf{x}_0, \mathbf{W}_0^{N-1}, \boldsymbol{\theta})$ requires maximizing $p(\mathbf{x}_0, \mathbf{W}_0^{N-1}, \boldsymbol{\theta}|\mathbf{Y}_1^N)$ w.r.t. $(\mathbf{x}_0, \mathbf{W}_0^{N-1}, \boldsymbol{\theta})$. Hence, the following maximization problem is solved w.r.t $\mathbf{x}_0, \boldsymbol{\theta}$ and $\mathbf{W}_0^{N-1}$ to obtain the estimates:

$$\begin{aligned} &\max_{\mathbf{x}_0, \mathbf{W}_0^{N-1}, \boldsymbol{\theta}} p(\mathbf{x}_0, \mathbf{W}_0^{N-1}, \boldsymbol{\theta}|\mathbf{Y}_1^N) \\ &= p(\mathbf{x}_0) \times p(\boldsymbol{\theta}) \times \prod_{j=0}^{N-1} p(\mathbf{w}_j) \times \prod_{j=1}^{N} p_{\mathbf{v}_j}(\mathbf{y}_j - \mathbf{h}(\mathbf{x}_j)) \\ &\text{s.t.} \quad \mathbf{x}_j = \mathbf{F}(\mathbf{x}_{j-1}, \boldsymbol{\theta}) + \mathbf{w}_{j-1} \quad \forall j \in \{1, \dots, N\} \end{aligned} \tag{14}$$

Thus, solving (14) yields the MAP estimates $\Big(\mathbf{x}_0, \mathbf{W}_0^{N-1}, \boldsymbol{\theta}\Big)_{MAP} \equiv \{\widehat{\mathbf{x}}_{0|N}, \widehat{\mathbf{w}}_{0|N}, \widehat{\mathbf{w}}_{1|N}, \dots, \widehat{\mathbf{w}}_{N-1|N}, \widehat{\boldsymbol{\theta}}\}$ where $\widehat{\mathbf{w}}_{j|N}$ represents smoothed estimate of $\mathbf{w}_j$. The smoothed estimates with estimated parameter, $\widehat{\boldsymbol{\theta}}$, together with recursive use of the state dynamics (1a) are used to obtain the state trajectory MAP estimates as follows:

$$\begin{aligned} \hat{\mathbf{x}}_{\mathbf{1|N}} &= \mathbf{F}(\hat{\mathbf{x}}_{0|N}, \widehat{\boldsymbol{\theta}}) + \hat{\mathbf{w}}_{0|N}, \dots, \\ \hat{\mathbf{x}}_{\mathbf{N|N}} &= \mathbf{F}(\hat{\mathbf{x}}_{N-1|N}, \widehat{\boldsymbol{\theta}}) + \hat{\mathbf{w}}_{N-1|N} \end{aligned} \tag{15}$$

Remark 1. Since the batch estimation problem for state and parameter estimation is formulated in the probabilistic framework, it facilitates dealing with systems subjected to state and measurement noise with non-Gaussian densities. When these densities are Gaussian, the optimization (14) reduces to the weighted least squares problem.

3.2 Moving Horizon Formulation

When model parameters are time invariant, then the estimated model parameters can be used to formulate a suitable state estimator for on-line monitoring and control. However, when model parameters are changing slowly with time, it becomes necessary to track the time varying model parameters simultaneously to maintain the performance of the state estimator. In this section, we modify the batch formulation to a moving horizon formulation that can track slowly changing model parameters while simultaneously estimating the states.

Consider the batch formulation translated in time to a window $[k-q, k]$ where k denotes the current time instant. Under assumption that parameter vector, $\boldsymbol{\theta}$, remains constant over the window $[k-q, k]$, the batch optimization problem, can be reformulated as follows:

$$\begin{aligned} &\max_{\mathbf{x}_{k-q}, \mathbf{W}_{k-q}^{k-1}, \boldsymbol{\theta}} p(\mathbf{x}_{k-q}, \mathbf{W}_{k-q}^{k-1}, \boldsymbol{\theta}|\mathbf{Y}_{k-q+1}^{k}) \\ &= p(\mathbf{x}_{k-q}) \times p(\boldsymbol{\theta}) \times \prod_{j=k-q}^{k-1} p(\mathbf{w}_j) \times \prod_{j=k-q+1}^{k} p_{\mathbf{v}_j}(\mathbf{y}_j - \mathbf{h}(\mathbf{x}_j)) \\ &\text{s.t.} \quad \mathbf{x}_j = \mathbf{F}(\mathbf{x}_{j-1}, \boldsymbol{\theta}) + \mathbf{w}_{j-1} \quad \text{for } j \in \{k-q+1, \dots, k\} \end{aligned} \tag{16}$$

Solving (16) leads to the MAP estimates of states, parameters and disturbances: $\Big(\mathbf{x}_{k-q}, \mathbf{W}_{k-q}^{k-1}, \boldsymbol{\theta}\Big)_{MAP} \equiv \{\widehat{\mathbf{x}}_{k-q|k}, \widehat{\mathbf{w}}_{k-q|k}, \widehat{\mathbf{w}}_{k-q+1|k}, \dots, \widehat{\mathbf{w}}_{k-1|k}, \widehat{\widehat{\boldsymbol{\theta}}}_{[k-q,k]}\}$ where, $\widehat{\mathbf{w}}_{j|k}$ represents smoothed estimate of $\mathbf{w}_j$. With recursive use of model (1) together with these optimal estimates, we can estimate the state trajectory $(\widehat{\mathbf{x}}_{k-q+1|k}, \dots, \widehat{\mathbf{x}}_{k|k})$. In the context of on-line state and parameter estimation, major difficulty is in constructing the prior densities, $p(\mathbf{x}_{k-q})$ and $p(\boldsymbol{\theta})$. This issue has been further discussed in the following section.

Remark 2. It is important to note that, unlike conventional augmentation approaches for parameter estimation, proposed MHE formulation does not assume any model associated with the unknown parameter. Therefore, tuning of the corresponding covariance for random walk model, which is not a trivial exercise, is not required in the current formulation. However, current MHE formulation requires a prior probability density function for the unknown parameter, which is assumed to be uniformly distributed, in the current work as discussed in the next section.

3.3 Computation of Arrival Cost

We can choose to construct the prior densities, or arrival costs, as follows:

$$p(\mathbf{x}_{k-q}) = p(\mathbf{x}_{k-q}|\mathbf{Y}_1^{k-q}, \boldsymbol{\theta}) \quad \text{and} \quad p(\boldsymbol{\theta}) = p(\boldsymbol{\theta}|\mathbf{Y}_1^{k-q})$$

However, the *pdf* associated with arrival cost $p(\mathbf{x}_{k-q}|\mathbf{Y}_1^{k-q}, \boldsymbol{\theta})$ is, in general, complex and difficult to compute when the model is nonlinear and the state and measurement densities are non-Gaussian. In literature, various approaches have been used to obtain the arrival cost [Lopez-Negrete et al., 2011, Robertson et al., 1996]. When state and/or measurement noise densities are non-Gaussian, it is possible to construct approximations to densities as Gaussian mixture models. However, in this work, following Lopez-Negrete et al. [2011], it is proposed to construct approximations to these densities as follows:

$$\begin{aligned} p(\mathbf{x}_{k-q}|\mathbf{Y}_1^{k-q}, \boldsymbol{\theta}) &\approx \mathcal{N}(\hat{\mathbf{x}}_{k-q|k-q}, \mathbf{P}_{k-q|k-q}) \\ p(\boldsymbol{\theta}|\mathbf{Y}_1^{k-q}) &\approx \mathcal{U}_{[\mathbf{a},\mathbf{b}]} \end{aligned}$$

where, $\mathcal{U}_{[\mathbf{a},\mathbf{b}]}$ is Uniform probability distribution function over $[\mathbf{a}, \mathbf{b}]$. The choice of $p(\boldsymbol{\theta}|\mathbf{Y}_1^{k-q})$ for prior density of $\boldsymbol{\theta}$ translates to inclusion of bound constraints on $\boldsymbol{\theta}$ in the proposed MHE formulation (16). The estimates $\hat{\mathbf{x}}_{k-q|k-q}$ and corresponding covariance $\mathbf{P}_{k-q|k-q}$ can be computed by implementing a filter that can systematically deal with non-Gaussian state disturbances, such as conventional

EnKF or PF [Lopez-Negrete et al., 2011] or a moving window based EnKF [Valluru et al., 2017], in parallel with MHE. While EnKF & PF can systematically capture the non-Gaussianity, their use can be computationally expensive due to large number of particles. To alleviate this issue, we also propose a modified EKF which can deal with non-Gaussian noise scenarios. Proposed modified EKF can deal with non-Gaussian densities for case studies discussed in current work and has been briefly described in Section 3.4. Since the densities are not Gaussian, we propose to use EnKF for computing the arrival cost. We additionally propose to use modified version of EKF approach to approximate the arrival cost term for states (Section 3.4). Further, as the parameters are changing, the unknown parameters estimated by proposed MHE approach are used to update the models used in both EnKF and modified EKF. The proposed MHE with EnKF and modified EKF approach for arrival cost of states computation are termed as MHE-EnKF and MHE-mEKF, respectively.

3.4 Modified Extended Kalman filter

EKF is generally applicable for Gaussian disturbances [Patwardhan et al., 2012]. However, in current work, we have modified the general EKF approach to accommodate non-Gaussian disturbances in the algorithm. In our modification, mean and covariance of the original non-Gaussian density is used to obtain a Gaussian approximation, which is then used in predicted and update steps of conventional EKF. The case studies presented in Section 4, involves truncated Gaussian process noise and non-Gaussian measurement noises. These are approximated as Gaussian densities as follows:

Case I: Standard normal distribution of a normal variable with mean zero and variance one, when truncated at zero from below i.e. variable is ≥ 0, has the following modified central moments (mean and variance) [Barr and Sherrill, 1999]:

$$\tilde{\mu} = \sqrt{\frac{2}{\pi}}, \quad \tilde{Q} = \left(1 - \frac{2}{\pi}\right) \tag{18}$$

Case II: Central moments (mean and variance) for Gaussian mixture model given by (24) is as follows [Söderström, 2002]:

$$\tilde{\mu} = P_1 \times \mu_1 + (1 - P_1) \times \mu_2, \tag{19}$$

$$\tilde{R} = P_1 \times (\sigma_1^2 + \mu_1^2) + (1 - P_1) \times (\sigma_2^2 + \mu_2^2) \\ -(P_1 \times \mu_1 + (1 - P_1) \times \mu_2)^2 \tag{20}$$

Above re-computed mean and variances for state and measurement noises are used to approximate the non-Gaussian densities as Gaussian distribution which are then further used during the implementation of conventional EKF. It is to be noted that the proposed modification facilitates use of EKF for specific densities considered in Case I and Case II and may not be good approximation for other non-Gaussian densities.

4. CASE STUDY

To demonstrate the efficacy of the proposed approach, a two state system subjected to non-Gaussian disturbances has been considered. This system has been adapted from Prakash et al. [2010] to evaluate performance of proposed MHE formulation for a system subjected to non-Gaussian densities. Following are the two cases studied:

Case I: Truncated Gaussian process noise with Gaussian measurement noise

Case II: Truncated Gaussian process noise with non-Gaussian measurement noise

We compare the results of proposed MHE approach with conventionally used augmentation techniques in modified EKF and EnKF (termed as mEKF$_{aug}$ and EnKF$_{aug}$), by augmenting the states with the parameter to obtain an augmented vector. Generally $\boldsymbol{\theta}$ is modeled as random walk model, i.e., $\boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_k + \mathbf{w}_{\boldsymbol{\theta},k}$ where $\mathbf{w}_{\boldsymbol{\theta},k}$ is an artificial noise term. In the case studies, the covariance matrix ($\mathbf{Q}_P$) corresponding to $\mathbf{w}_{\boldsymbol{\theta},k}$ has been tuned to get the best estimate of $\boldsymbol{\theta}$. All the computations were performed in Matlab version R2015b running on a computer with 16GB RAM and Intel-core i7 processor. The optimization problem has been solved using the 'CasADi' package in MATLAB environment. To obtain meaningful comparisons between the filtering approaches, multiple (N_r) simulation runs were performed with the length of each simulation run being N time instants. The simulation runs differed in the realizations of process and measurement noises obtained at different time instants. We compare the performance of proposed MHE variants and mEKF$_{aug}$ or EnKF$_{aug}$ using average root mean squared error (ARMSE) as performance index defined as follows:

$$\text{ARMSE}(i) = \frac{1}{N_r} \sum_{j=1}^{N_r} \left(\sqrt{\frac{1}{N} \sum_{k=1}^{N} (x^j_{(k,i)} - \hat{x}^j_{(k|k,i)})^2} \right) \tag{21}$$

where ARMSE(i) denote ARMSE for the ith state. Consider the following nonlinear dynamic system (adapted from Prakash et al. [2010]) for the two case studies being demonstrated in current section:

$$x_{1,k+1} = \theta_k \, x_{1,k} + 0.2 x_{2,k} \tag{22a}$$

$$x_{2,k+1} = -0.1 x_{1,k} + \frac{0.5 x_{2,k}}{1 + [x_{2,k}]^2} + w_k \tag{22b}$$

$$y_k = x_{1,k} - 3x_{2,k} + v_k \tag{22c}$$

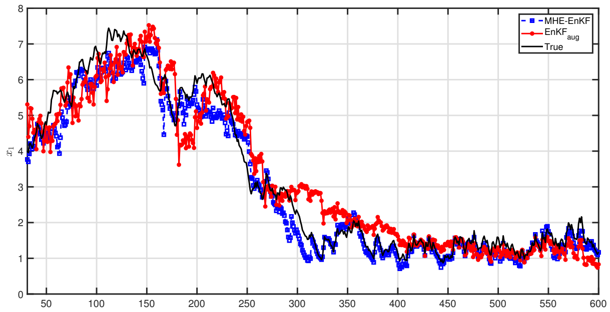
4.1 Case I

In this case study we consider the system described in (22). We further assume that $\{v(k)\}$ is a sequence of independent, zero mean, normally distributed random variables with variance 0.01 while $w(k)$ follows truncated normal distribution obtained by truncating normal distribution with unit variance at zero. Initial filter conditions for EnKF and mEKF are given as: $\mathbf{x}_{0|0} = \begin{bmatrix} 1 & 0 \end{bmatrix}^T$, $\mathbf{P}_{0|0} = \text{diag} \begin{bmatrix} 1 & 1 \end{bmatrix}$. Other implementation parameters are given in Table 1. As discussed earlier, the arrival cost in MHE is approximated as Gaussian with mean as estimates of states obtained from EnKF and modified EKF and the appropriate covariance. After substitution of the *pdf*, and taking *log* on both sides, the proposed MHE (16) is then finally converted into following minimization problem:

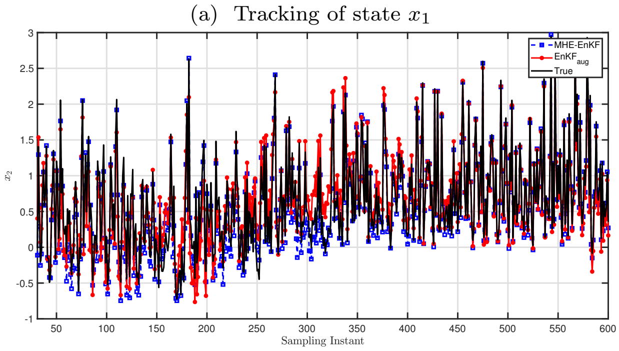
$$\min_{\substack{\mathbf{x}_{k-q}, \\ \mathbf{W}_{k-q}^{k-1}, \theta}} \left\{ ||\mathbf{x}_{k-q} - \hat{\mathbf{x}}_{k-q|k-q}||^2_{\mathbf{P}^{-1}_{k-q|k-q}} + \sum_{j=k-q}^{k-1} w_j^2 + \sum_{j=k-q+1}^{k} \frac{v_j^2}{0.01} \right\}$$
$$\text{s.t.} \quad \text{Eqs. (22)}, \quad -w_{j-1} \leq 0 \\ 0 \leq \theta \leq 2, \quad \forall j \in \{k-q+1, \ldots, k\} \tag{23}$$

Table 1. Implementation Parameters

Parameter	Value
Number of Particles for EnKF (N_s)	150
Tuned Covariance for parameter for EnKF$_{aug}$, mEKF$_{aug}$ ($\mathbf{Q}_p$)	0.99×10^{-5}
Horizon Length for MHE (q)	30
N, N_r	600, 50



(a) Tracking of state x_1



(b) Tracking of state x_2

Fig. 1. Case I: Estimates of states obtained from MHE-EnKF and MHE-mEKF for a typical simulation run

It should be noted that constraint $-w_j \leq 0$ along with w_j^2 in (23) corresponds to use of truncated Gaussian density for w_j.

Table 2 presents the ARMSE values for the state estimates. ARMSE values obtained from MHE-mEKF and MHE-EnKF are comparable, thus estimates obtained from both MHE variants are similar irrespective of the prior density obtained from different filters. Also, it can be seen that ARMSE obtained from both MHE variants are significantly lower than ARMSE values obtained from EnKF$_{aug}$ and mEKF$_{aug}$ alone. Further, to compare proposed MHE approach with augmented approach, tracking of proposed MHE-EnKF and EnKF$_{aug}$ for both states has been shown in Fig. 1. It should be noted that MHE estimates are computed after $j > 30$, i.e., when the time index is equal to or higher than the window length.

Fig. 2 presents the tracking of the unknown parameter obtained from MHE-EnKF and EnKF$_{aug}$. It can be seen that proposed MHE-EnKF tracks better than EnKF$_{aug}$. Similar results can be observed from the ARMSE values of the parameter in Table 2. It can be seen that ARMSE values of MHE-EnKF and MHE-mEKF are significantly smaller than EnKF$_{aug}$ or mEKF$_{aug}$.

4.2 Case II

In this section, in addition to truncated Gaussian process noise, non-Gaussian measurement noise is also considered

Table 2. Case I: ARMSE values

Filter	x_1	x_2	θ
MHE-mEKF	0.3864	0.1320	0.0108
MHE-EnKF	0.3883	0.1326	0.0108
EnKF$_{aug}$	0.5451	0.1840	0.0123
mEKF$_{aug}$	0.5318	0.1796	0.0118

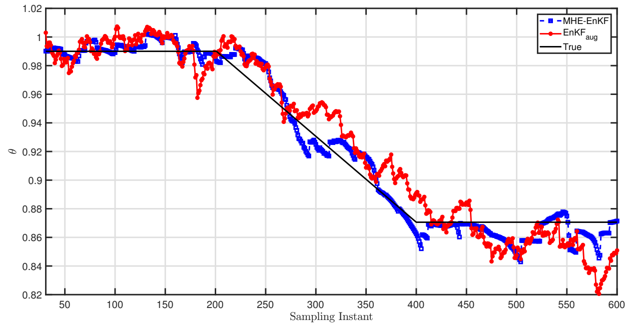


Fig. 2. Case I: Estimates of parameters obtained from MHE-EnKF and MHE-mEKF for a typical simulation run

in the case study. The *pdf* of the non-Gaussian measurement noise is as follows:

$$p(v_k) = \frac{1}{(2\pi)^{0.5}} \left\{ \frac{P_1}{\sigma_1} \exp\left[\frac{-(v_k - \mu_1)^2}{2\sigma_1^2}\right] + \frac{P_2}{\sigma_2} \exp\left[\frac{-(v_k-\mu_2)^2}{2\sigma_2^2}\right] \right\} \quad (24)$$

where, $P_1 = 0.6, P_2 = 0.4$, $\mu_1 = 0.5$, $\sigma_1 = 0.1$, $\mu_2 = -0.5$ and $\sigma_2 = 0.1$. Initial filter conditions for EnKF and mEKF are given as $\mathbf{x}_{0|0} = \begin{bmatrix}0 & 0\end{bmatrix}^T$, $\mathbf{P}_{0|0} = \text{diag}\begin{bmatrix}1 & 1\end{bmatrix}$. All the rest of the implementation parameters for both the filters are same as discussed in Section 4.1. The final optimization problem, after substitution of (24) in (16) and taking *log* on both sides, is as follows:

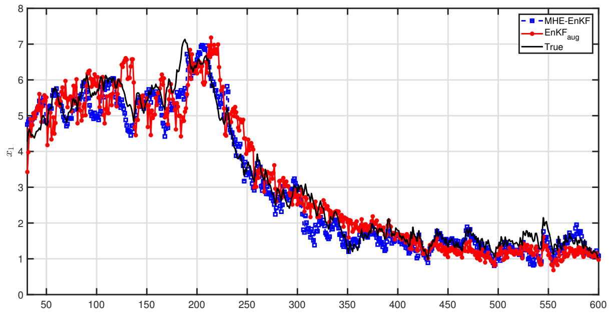
$$\min_{\mathbf{x}_{k-q}, \mathbf{W}_{k-q}^{k-1}, \theta} \Bigg\{ ||\mathbf{x}_{k-q} - \hat{\mathbf{x}}_{k-q|k-q}||^2_{\mathbf{P}^{-1}_{k-q|k-q}} + \sum_{j=k-q}^{k-1} w_j^2 - 2\sum_{j=k-q+1}^{k} log\left[\frac{P_1}{(2\pi)^{1/2}\sigma_1}\exp\left[\frac{-0.5(v_j-\mu_1)^2}{\sigma_1^2}\right] + \frac{P_2}{(2\pi)^{1/2}\sigma_2}\exp\left[\frac{-0.5(v_j-\mu_2)^2}{\sigma_2^2}\right]\right]\Bigg\}$$

$$\text{s.t. Eq. (22)}, \; -w_{j-1} \leq 0, \; 0 \leq \theta \leq 2, \; \forall j \in \{k-q+1, \ldots, k\} \quad (25)$$

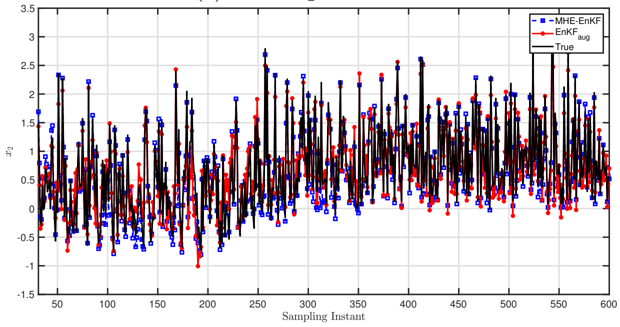
Table 3 presents the ARMSE values for MHE-EKF, MHE-EnKF, mEKF$_{aug}$ and EnKF$_{aug}$. It can be seen that the ARMSE values of both the proposed MHE variants are similar. Also, ARMSE values of both MHE variants are smaller than EnKF$_{aug}$ and mEKF$_{aug}$. Further, Figure 3 shows the tracking of both states by proposed MHE-EnKF and EnKF$_{aug}$. Fig. 4 presents the tracking of the unknown parameter and it can be seen that MHE-EnKF tracks better than EnKF$_{aug}$. Similar results can be observed from the ARMSE values of the parameter in Table 3.

Table 3. Case II: ARMSE values

Filter	x_1	x_2	θ
MHE-mEKF	0.3963	0.1435	0.0116
MHE-EnKF	0.3967	0.1422	0.0116
EnKF$_{aug}$	0.5409	0.2333	0.0124
mEKF$_{aug}$	0.5304	0.2306	0.0125



(a) Tracking of state x_1



(b) Tracking of state x_2

Fig. 3. Case II: Estimates of states obtained from MHE-EnKF and MHE-mEKF for a typical simulation run

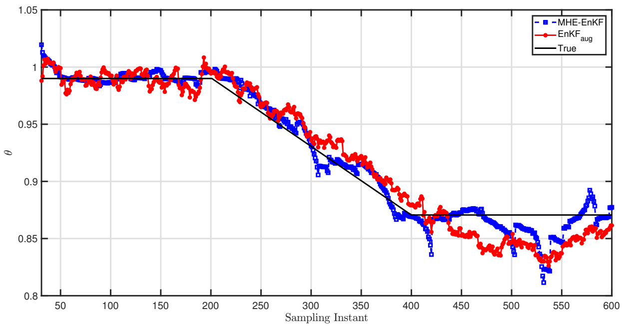


Fig. 4. Case II: Estimates of parameters obtained from MHE-EnKF and MHE-mEKF for a typical simulation run

5. CONCLUSION

In this work, we have extended our preliminary work [Varshney et al., 2019] to a probabilistic MHE formulation for simultaneous state and parameter estimation for nonlinear systems subjected to non-Gaussian disturbances. Main advantage of proposed approach is that unlike various sequential Bayesian approaches (such as mEKF$_{aug}$, EnKF$_{aug}$) which employ a random walk model for the drifting parameters that requires tuning of the covariance matrix, the proposed MHE approach does not require such models. Further, a modified version of EKF is proposed that works with Gaussian approximations of certain type of non-Gaussian densities. Efficacy of the proposed MHE formulation has been demonstrated on a two state system subjected to disturbances following truncated Gaussian and Gaussian mixture model. The arrival cost in the proposed MHE formulation is computed by using either EnKF or modified EKF. The results of stochastic simulation studies demonstrated that, proposed MHE formulation performs better than the conventional formulations that are based on the random walk model for the parameter variations. In future, the current work will be illustrated on more different types of non-Gaussian densities and will be extended to systems where the unknown parameters are present in the measurement equations as well, along with correlated disturbances.

REFERENCES

Arulampalam, M.S., Maskell, S., Gordon, N., and Clapp, T. (2002). A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking. *IEEE Transactions on Signal Processing*, 50, 174–188.

Barr, D. and Sherrill, E. (1999). Mean and variance of truncated normal distributions. *The American Statistician*, 53, 357–361.

Chen, W., Bakshi, B., Goel, P., and Ungarala, S. (2004). Bayesian estimation of unconstrained nonlinear dynamic systems. *IFAC-PapersOnLine*, 37, 263–268.

Isaksson, A.J., Sjöberg, J., Törnqvist, D., Ljung, L., and Kok, M. (2015). Using horizon estimation and nonlinear optimization for grey-box identification. *Journal of Process Control*, 30, 69–79.

Kühl, P., Diehl, M., Kraus, T., Schlöder, J., and Bock, H. (2011). A real-time algorithm for moving horizon state and parameter estimation. *Computers and Chemical Engineering*, 35, 71–83.

Lopez-Negrete, R., Patwardhan, S.C., and Biegler, L.T. (2011). Constrained particle filter approach to approximate the arrival cost in moving horizon estimation. *Journal of Process Control*, 21(6), 909–919.

Patwardhan, S., Narasimhan, S., Prakash, J., Gopaluni, R., and Shah, S. (2012). Nonlinear Bayesian state estimation : A review of recent developments. *Control Engineering Practice*, 20, 933–953.

Prakash, J., Patwardhan, S., and Shah, S. (2010). Constrained nonlinear state estimation using ensemble kalman filters. *Ind. Eng. Chem. Res.*, 49, 2242–2253.

Rao, C.V. (2000). *Moving Horizon Strategies for the Constrained Monitoring and Control of Nonlinear Discrete-Time Systems.* Ph.D. thesis, University OF Wisconsin - Madison.

Robertson, D., Lee, J., and Rawlings, J. (1996). A moving horizon-based approach for least-squares estimation. *AIChE Journal*, 42(8), 2209–2222.

Shen, J.N., Shen, J.J., He, Y.J., and Ma, Z.F. (2019). Accurate state of charge estimation with model mismatch for Li-Ion batteries: A joint moving horizon estimation approach. *IEEE Transactions on Power Electronics*, 34, 4329–4342.

Simon, D. (2006). *Optimal State Estimation: Kalman, H_∞ and Nonlinear Approaches.* Wiley-Interscience, New York, NY, USA.

Söderström, T. (2002). *Discrete-Time Stochastic Systems: Estimation and Control.* Springer-Verlag, Berlin, Heidelberg, 2nd edition.

Valluru, J., Lakhmani, P., Patwardhan, S., and Biegler, L. (2017). Development of moving window state and parameter estimators under maximum likelihood and bayesian frameworks. *Journal of Process Control*, 60, 48–67.

Varshney, D., Patwardhan, S., Bhushan, M., and Biegler, L. (2019). Batch and moving horizon estimation for systems subjected to non-additive stochastic disturbances. *IFAC-PapersOnLine*, 52, 16–21.