Constrained State Estimation - A Review

Nesrine Amor, Ghulam Rasool, and Nidhal C. Bouaynaya

Abstract—The real-world applications in signal processing generally involve estimating the system state or parameters in nonlinear, non-Gaussian dynamic systems. The estimation problem may get even more challenging when there are physical constraints on the system state. This tutorial-style paper reviews the Bayesian state estimation for (non)linear state-space systems and introduces the formulation of constrained state estimation in such scenarios. Specifically, we start by providing a review of unconstrained state estimation using Kalman filters (KF) for the linear systems and their extensions for nonlinear state-space systems, including extended Kalman filters (EKF), unscented Kalman filters (UKF), and ensemble Kalman filters (EnKF). Next, we present particle filters (PFs) for nonlinear state-space systems. Finally, we review constrained state estimation using various filtering techniques and highlight the advantages and disadvantages of the different constrained state estimation approaches.

I. INTRODUCTION

The states of many dynamical systems are confined within constrained regions owing to relevant physical laws, geometric considerations, and kinematic limits, such as material balance, maximum power rating, bounds on the output of actuators and plants, and speed constraints in road networks [1], [2]. The mathematical formulation of such constraints involves defining a set of (non)linear relationships in the form of (in)equalities constraining the state estimates. One possible solution is to incorporate the constraints into the state-space model of the system. However, generally, that is not possible without a significant increase in the complexity of the model. During the state estimation process, taking constraints into account leads to accurate and physically relevant estimates. For instance, the exploitation of a known road network has been proven effective in the tracking of ground vehicles [3]. Similarly, in a maritime scenario, the knowledge of shipping lanes and sea/land distinction can improve the tracking and detection performance [4].

The state-space representation provides an extremely flexible framework for modeling discrete-time dynamical systems using two mathematical relationships, i.e., the state transition and observation. The former captures the evolution of the hidden state over time, and the latter provides the noisy measurement of a (non)linear function of the state. The descriptive power of the state-space representation comes at the expense of intractability. Generally, it is not possible to obtain an analytic solution to the state estimation problem except for a small number of cases, e.g., linear systems with Gaussian noise.

In the Bayesian framework, the inference of the hidden state given all available observations relies upon estimating the posterior density function (pdf). For the linear systems with Gaussian noise, the closed-form optimal solution is given by the Kalman filter (KF). On the other hand, for nonlinear and non-Gaussian state-space models, two fundamental techniques have been emerging, parametric and nonparametric [5]. The parametric methods include the extended Kalman filter (EKF), unscented Kalman filter (UKF), ensemble Kalman filter (EnKF), and moving horizon estimation. The nonparametric techniques are based on sequential Monte Carlo methods and include particle filters (PF) [5].

This paper provides a tutorial-style review of constrained state estimation methods for KF, EKF, UKF, EnKF, moving horizon estimation, and PF. In Section III, we start by presenting the problem statement. Section IV provides detailed description of unconstrained Bayesian state estimation techniques. Section V presents the literature available in constrained state estimation for KF, EKF, EnKF, and moving horizon estimation. In Section VI, we formulate and review of the constrained PF problem. Finally, Section VII summarizes and concludes the paper.

II. PROBLEM STATEMENT

A. System Definition

We consider a general state-space representation defined by the state transition and measurement models for discrete-time systems, given by:

\[ x_{t+1} = f_t(x_t) + w_t, \]
\[ y_t = h_t(x_t) + v_t, \]

where \( x_t \in \mathbb{R}^{n_x} \) and \( y_t \in \mathbb{R}^{n_y} \) represent the hidden state and observation at time \( t \), respectively, \( t \in \mathbb{N} \) represents the time index and \( n_x \) and \( n_y \) are state and output dimensions. In out settings, \( f_t : \mathbb{R}^{n_x} \to \mathbb{R}^{n_x} \) and \( h_t : \mathbb{R}^{n_x} \to \mathbb{R}^{n_y} \) are (non)linear mappings, \( w_t \) and \( v_t \) are zero-mean process and measurement noise with known PDFs \( p(w) \) and \( p(v) \). Both noise sequences are uncorrelated with each other and the initial condition of the state \( x_0 \) given by \( p(x_0) \).

The mapping functions \( f_t(x_t) \) and \( h_t(x_t) \) can be defined in term of PDFs, i.e.,

\[ p(x_t|x_{t-1}) = p_{w_t}(x_t - f_t(x_{t-1})), \]
\[ p(y_t|x_t) = p_{v_t}(y_t - h_t(x_t)). \]

B. Optimal State Estimation

The state estimation problem aims at finding the hidden state \( x_t \) using all available measurements up to time \( t \), \( y_{1:t} = [y_1, \ldots, y_t] \). The solution to this problem is the density of the system state conditioned on the measurements; either joint PDF \( p(x_1, \ldots, x_t|y_{1:t}) \) or the marginal PDF \( p(x_t|y_{1:t}) \) [5].
We make two assumptions here, (1) the system state follows a first-order Markov process, i.e., 
\[ p(x_t|x_{t-1}, y_{1:t}) = p(x_t|x_{t-1}) \] and 2) measurements are conditionally independent given the system state, i.e., 
\[ p(y_t|x_1, ..., x_t, y_1, ..., y_{t-1}) = p(y_t|x_t). \]

In Bayesian estimation framework, a recursion can be defined to estimate posterior PDF \( p(x_t|y_{1:t}) \) using the prior PDF \( p(x_t|y_{1:t-1}) \), transition PDF \( p(x_t|x_{t-1}) \) and the likelihood PDF \( p(y_t|x_t) \). We consider the marginal posterior, however, the same results hold true for the joint conditional PDF also. Using Bayes rule and Chapman-Kolmogorov equation, the posterior PDF can be computed recursively using the following two-step relationship [3]:

Prediction Step

\[
p(x_t|y_{1:t-1}) = \int p(x_t|y_{1:t-1}) p(x_t|y_{1:t-1}) \, dx_{t-1} \tag{5}
\]

Update Step

\[
p(x_t|y_{1:t}) = \frac{p(y_t|x_t)p(x_t)}{p(y_t)} \tag{6}
\]

For a general case with nonlinear state transition \( f_t(x_t) \) and observation functions \( h_t(x_t) \) and non-Gaussian noise sequences \( w_t \) and \( v_t \), eqs. (5) and (6) are only a conceptual solution owing to the intractable integrals. For the case of linear systems with Gaussian noise, a closed-form optimal solution in the form of KF exists [6].

III. UNCONSTRAINED STATE ESTIMATION

A. Kalman Filters for Linear Systems with Additive Gaussian Noise

The state transition function \( f_t(x_t) \) and the observation \( h_t(x_t) \) function are both linear functions and are represented by matrices \( F_t \in \mathbb{R}^{n_x \times n_x} \) and \( H_t \in \mathbb{R}^{n_y \times n_x} \) respectively. The noise sequences are both additive and Gaussian, and defined as \( w_t \sim \mathcal{N}(0, Q_t) \) and \( v_t \sim \mathcal{N}(0, R_t) \). The initial system state \( x_0 \) is also assumed to be known with a Gaussian distribution \( \mathcal{N}(x_0, P_0) \). The Kalman filter, a minimum mean square error (MMSE) estimator, is defined with two steps, the prediction step and the update step. We use ‘\( \hat{\cdot} \)’ to represent the estimate, ‘\( \dagger \)’ for the matrix transpose operation, and ‘\( \sim \)’ and ‘\( \dagger \)’ symbols present quantities before and after a state or observation update is applied.

Prediction step

\[
\hat{x}_{t-1} = F_{t-1} \hat{x}_{t-1} + K \left( y_t - H_t \hat{x}_t \right) \tag{8a}
\]

Update step

\[
K_t = P_{yy} H_t^T \left( H_t P_{yy} H_t^T + R_t \right) \tag{8b}
\]

\[
\hat{x}_t = \hat{x}_{t-1} + K_t \left( y_t - H_t \hat{x}_t \right) \tag{8c}
\]

\[
P_t = (I - K_t H_t) P_{yy} \tag{8d}
\]

where \( K \) represents the Kalman gain matrix. We can show that KF is the best linear unbiased estimator (BLUE) for all linear systems with additive noise.

B. Extended Kalman Filter (EKF)

The EKF assumes that the nonlinear state transition \( f_t(x_t) \) and/or observation \( h_t(x_t) \) functions can be linearized using the Taylor series [7], [8].

\[
F_t = \frac{\partial f_t(x_t)}{\partial x_t} \bigg|_{\hat{x}_{t-1}^\dagger} \\
H_t = \frac{\partial h_t(x_t)}{\partial x_t} \bigg|_{\hat{x}_{t-1}^\dagger} \tag{9}
\]

The error covariance \( P_t \) is propagated (using eqs. 7b, 7c and 7d) with the linearized functions \( F_t \) and \( H_t \) from eq. (9). The state estimate \( \hat{x}_t \) (eq. 7a) and \( \hat{y}_t \) (eq. 8a) are calculated using nonlinear functions \( f_t(x_t) \) and \( h_t(x_t) \) as given below:

\[
\hat{x}_t = f_t(\hat{x}_{t-1}^\dagger) \tag{10}
\]

\[
\hat{y}_t = h_t(\hat{x}_t) \tag{11}
\]

The computational complexity for the calculation of \( F_t \) and \( H_t \) matrices may prohibit the use of EKF in some applications [8]. Furthermore, the linearization of nonlinear systems may introduce errors in the estimation of the state, and in the worst-case, the filter may diverge especially for highly nonlinear systems [9].

C. Unscented Kalman Filter (UKF)

UKF addresses the issues raising from the approximation operations and limit the performance of EKF [10]. UKF uses a set of carefully chosen deterministic samples, called sigma points, to propagate the mean and covariance of the posterior distribution of the state. The set of sigma points capture the true mean and covariance of the posterior for the case of Gaussian distribution when propagated through the nonlinear system.

Considering the system state \( x_t \) with mean \( \hat{x}_t \) and covariance \( P_t \), we can chose sigma points \( \chi_{j,t} \in \mathbb{R}^{n_x}, j = 0, 1, \ldots, 2n \) using following relationships [9–12]:

\[
\chi_{j,t} = \hat{x}_t + \sqrt{(2n+1) \cdot P_t} \cdot \sigma_j \tag{12a}
\]

\[
\chi_{j,t} = \hat{x}_t - \sqrt{(2n+1) \cdot P_t} \cdot \sigma_j \tag{12b}
\]

\[
\sigma_j = \begin{cases} 1 & j = 0 \\ \sqrt{\frac{2}{n} + \frac{2}{n+1}} & j = 1, \ldots, n \\ \sqrt{\frac{2}{n} + \frac{2}{n+1}} & j = n+1, \ldots, 2n \end{cases} \tag{12c}
\]
\(X_{0,t} = \hat{x}_t, \) \(\)\(\)\(\)\(\)
\(X_{j,t} = \hat{x}_t + \left(\sqrt{(n + \lambda)P_t}\right)_j \) \(j = 1, \ldots, n\) \(\)\(\)\(\)
\(X_{j,t} = \hat{x}_t - \left(\sqrt{(n + \lambda)P_t}\right)_{j-n} \) \(j = n + 1, \ldots, 2n\) \(\)\(\)
\(W_{0}^{(m)} = \frac{\lambda}{n + \lambda}, \) \(\)\(\)
\(W_{0}^{(c)} = \frac{\lambda}{n + \lambda} + (1 - \alpha^2 + \beta), \) \(\)\(\)
\(W_{j}^{(m)} = W_{j}^{(c)} = \frac{1}{2(n + \lambda)}, \) \(j = 1, \ldots, 2n,\)

where \(\left(\sqrt{(n + \lambda)P_t}\right)_j\) represents the \(j\)th row of the matrix square root. \(\lambda = \alpha^2(n + \kappa)\) is a scaling parameter. \(\alpha\) determines the spread of the sigma points around the mean \(\hat{x}_t\) and is set to a small value, e.g., \(1e - 3\) \([10]\). \(\kappa\) is secondary scaling parameter and is generally set to \(0\). \(\beta\) encodes the prior knowledge about the distribution of \(x\) and is set to \(2\) for the Gaussian distribution \([10]\).

The sigma points \(\{X_{j,t}\}_{j=0}^{2n}\) (\(n\)-dimensional vectors) are passed through the nonlinear function, e.g.,

\[Y_{j,t} = g(X_{j,t}), \quad j = 1, \ldots, 2n,\]

where \(g\) is a nonlinear function, e.g., state transition \(f\) or observation \(h\). The mean and covariance after the transformation can be calculated using:

\[\bar{y} \approx \sum_{j=0}^{2n} W_{j}^{(m)} Y_{j}, \]

\[P_y \approx \sum_{j=0}^{2n} W_{j}^{(c)} [Y_{j} - \bar{y}] [Y_{j} - \bar{y}]^T.\]

The mathematical relations described in eqs. \([12] - [20]\) are referred to as uncentered transformation (UT).

UKF uses UT (eqs. \([12] - [20]\) and KF relations (eqs. \([7]\) and \([8]\) to estimate the state mean and covariance without having to linearize the nonlinear functions \(f, an\)h\) and calculating the Jacobian matrices. However, UKF has its own limitations, e.g., it may not work well with systems that have nearly singular covariance matrices. This problem is linked to the matrix square root operation performed on the covariance matrices using Cholesky decomposition, which, in turn, can be computationally demanding also. It is also important to mention that all three filters described above, i.e., KF, EKF and UKF assume that the prior distribution function follows Gaussian distribution. The approximations obtained with at least \(2n + 1\) sampling points are accurate to the 3rd order for the additive Gaussian noise and for all types of nonlinear functions and at least to the 2nd for non-Gaussian inputs \([13]\).

D. Moving Horizon Estimation (MHE)

Moving horizon estimation (MHE) is an optimization approach that can be used to estimate the unknown system state \([14], [15], [16], [17]\). MHE employs an iterative procedure that relies on linear or nonlinear programming to find the desired solution, i.e., the system state. For maximum \textit{a posteriori} estimate, the state estimation problem can be expressed as:

\[\hat{x}_t := \arg\max_{x_t} p(x_t | Y_t).\]

For a moving horizon of size \(h \in \{0, t\}\), we can determine

\[\hat{x}_{t-h, \ldots, t} = \arg\max_{x_{t-h, \ldots, t}} p(x_{t-h, \ldots, t} | Y_t).\]

Using Markovian assumption and Bayes rules, we get

\[p(x_{t-h, \ldots, t} | Y_t) \propto \prod_{j=t-h}^{t-1} \frac{p(y_j | x_j)}{p(x_{j+1} | x_j) p(x_{t-h} | Y_{t-h-1}).\]

With Gaussian assumption and taking logarithm, we have

\[\arg\min_{x_{t-h, \ldots, t}} \sum_{j=t-h}^{t-1} \|y_j - h_j(x_j)\|_{P_j^{-1}}^2 + \|x_{j+1} - f_j(x_j)\|_{Q_j}^2 + \|x_{t-h} - \bar{x}_{t-h}\|_{\Pi_{t-h}}^2,\]

where the last term is the arrival cost and for \(t = h\), \(\Pi_{t-h} = \Pi_0\), i.e., the initial covariance of the state estimate at time \(t = 0\).

One motivation for the development of MHE was to formulate a mathematical optimization problem where constraints on the system state can be naturally incorporated into the estimation (now optimization) framework \([13]\).

E. Ensemble Kalman Filter (EnKF)

The ensemble Kalman filter (EnKF) belong to the a broader class of sequential Monte Carlo methods \([19], [20], [21]\). In the most general form, an EnKF is based on the premise that it is sufficient to estimate first two moments of the probability density functions of interest, i.e., \(p(x_t | Y_{t-1})\) and \(p(x_t | Y_t)\), for the time-update and measurement-update steps \([22]\).

EnKF is initialized with \(N\) samples (or particles) that are sampled from a given probability distribution function. At each subsequent time step, \(N\) samples are drawn from process and observation noise distribution functions and propagated through system dynamics to compute a set of transformed particles.

Given that we have \(N\) particles \(\{ \hat{x}_{i-1}^{t+1} \}_{i=1}^{N}\)

Prediction step

\[\hat{x}_{i}^{-} = f_{i-1} \hat{x}_{i-1}^{t+1} + w_i,\]

\[\hat{y}_{i}^{-} = h_{i}(\hat{x}_{i}^{t+1}) + v_i,\]

\[\bar{x}_{t}^{-} = \frac{1}{N} \sum_{i=1}^{N} \hat{x}_{i}^{-},\]

\[\bar{y}_{t}^{-} = \frac{1}{N} \sum_{i=1}^{N} \hat{y}_{i}^{-},\]

\[P_{xy} = \frac{1}{N-1} \sum_{i=1}^{N} [\hat{x}_{i}^{-} - \bar{x}_{t}^{-}] [\hat{y}_{i}^{-} - \bar{y}_{t}^{-}]^T,\]

\[P_{y} = \frac{1}{N-1} \sum_{i=1}^{N} [\hat{y}_{i}^{-} - \bar{y}_{t}^{-}] [\hat{y}_{i}^{-} - \bar{y}_{t}^{-}]^T.\]
Update step

\[ K_t = P_{xy} P_y^{-1} \]
\[ \hat{x}_{t\mid t}^{-} = \hat{x}_{t\mid t-1}^{-} + K_t [y_t - (H_t \hat{x}_{t\mid t-1}^{-}) + v_t] \]
\[ \hat{x}_{t\mid t}^+ = \frac{1}{N} \sum_{i=1}^{N} \hat{x}_{t\mid t}^{i+} \]
\[ P_t = \frac{1}{N-1} \sum_{i=1}^{N} \left[ \hat{x}_{t\mid t}^{i+} - \hat{x}_{t\mid t}^{-} \right] \left[ \hat{x}_{t\mid t}^{i+} - \hat{x}_{t\mid t}^{-} \right]^T \]

The estimate accuracy of the EnKF depends on the number of samples \( N \). The difference between EnKF and particle filters is that EnKF makes the assumption that all probability density functions are Gaussian and can be represented only by the mean and covariance. EnKF can be used for systems with non-Gaussian probability distribution functions. However, as only first two moments are being used in the prediction and the update steps, the estimation results may not be accurate for systems that do not follow the Gaussian assumption.

Various methods, including EKF, UKF, MHE, and EnKF may not accurately estimate systems state when the underlying system dynamics are highly nonlinear and/or the noise in the system does not follow Gaussian distribution. Particle filters (PFs) are able to handle nonlinear systems with non-Gaussian noise. PFs are flexible and simple simulation-based numerical approaches used for estimating the system state in a sequential manner.

**F. Particle Filters (PFs)**

PFs solve the optimal estimation problem in nonlinear and non-Gaussian dynamic systems by incorporating sequential Monte Carlo sampling within the Bayesian filtering framework \([24, 25, 26, 27, 28]\). The goal is to estimate the posterior density for the state using Bayesian recursion. PFs approximate the posterior PDF using a group of samples (also called particles) \( x_t^i \) and their associated weights \( w_t^i \geq 0 \) as:

\[ \hat{p}(x_t | Y_t) = \sum_{i=1}^{N} w_t^i \delta(x_t - x_t^i), \]

(27)

where \( \delta(.) \) represents the Dirac delta function and \( N \) is the number of particles.

The conditional mean estimate of the state is given by the weighted mean of the particles as follows:

\[ \hat{x}_t = E[x_t | Y_t] \approx \sum_{i=1}^{N} w_t^i x_t^i. \]

(28)

The particles are required to be sampled from the true posterior PDF, which is not available. Therefore, another PDF, usually referred to as the importance distribution or the proposal distribution \( p(x_t | x_{t-1}) \), which is generally easy to sample from, is defined \([29]\).

The importance weight of every particle is given by:

\[ \tilde{w}_t^i = \frac{w_t^i}{\sum_{j=1}^{N} w_t(j)}, \]

(29)

\[ w_t^i = \frac{w_t^i}{\sum_{j=1}^{N} w_t(j)} \]

(30)

It has been shown that PFs converge asymptotically, as \( N \to \infty \), towards the optimal filter in the mean square error sense \([24]\).

The selection of correct proposal distribution \( q(x_t | x_{t-1}) \) is an important step in using PFs. Some studies have proposed to use either EKF \([30, 31, 24]\) or UKF \([32]\) to generate the importance PDF. At each step, an EKF or UKF is run for each particle to generate the mean and the covariance of the proposal PDF. Later, the particles are drawn from the newly found PDFs. The obvious advantage is that EKF and UKF take into account the most recent measurement while estimating the mean and the covariance. Further discussion on the proposal PDFs can be found in \([24, 33, 34, 28]\).

Despite the selection of appropriate proposal densities, the sequential importance sampling algorithm may degenerate and never converge. The normalized weights of all but one particle degenerate to zero, referred to as sample impoverishment. In order to avoid such a degeneracy problem, re-sampling is usually performed. Re-sampling will eliminate particles with low weight and multiply samples with high weights. Re-sampling algorithms are discussed in \([35, 36]\). The algorithm 1 details the steps for particle filtering.

**Algorithm 1 Particle filter**

Generate \( x_0^{(j)} \sim q_0(x_0) \), then calculate \( w_0^j = p(y_0 | x_0^j) \) and normalize the weights.

for \( t = 1, 2, \cdots, T \) (where \( T \) : time length) do

for \( i = 1, 2, \cdots, N \) (where \( N \) is the number of particles) do

Generate new samples from an accessible proposal distribution \( x_t^{(i)} \sim q_t(x_t) \).

Calculate the weights \( w_t^i \) of \( x_t^{(i)} \); then, normalize the weights.

end for

Re-sample to obtain equally weighted particles \( \{x_t^{(i)} \mid \frac{1}{N}\}_{i=1}^{N} \).

Compute the weighted mean \( \hat{x}_t = \sum_{i=1}^{N} w_t^i x_t^{(i)} \).

end for
IV. CONstrained STATE ESTIMATION

Many engineering applications, such as vision-based systems, chemical processes, target tracking, biomedical systems, navigation and robotics, can be modeled using state-space framework. In most of these dynamical system, the system state may be subject to constraints that arise from physical laws, natural phenomena, or model restrictions [6], [57], [38]. There may not be an easy and viable way to incorporate these constraints directly into the state-space model or the estimation framework [19]–[41].

We consider a set of constraints $C_t$, including linear and nonlinear, defined as:

$$a_t \leq \phi_t(\hat{x}_t) \leq b_t,$$

where $\phi_t$ represents the constraint function at time $t$. In the case when $\phi$ is an identity transformation, eq. (31) reduces to a simple interval constraint on the mean of the state. The constraints $C_t$ can be hard or soft, where the estimation algorithms required to satisfy soft constraints approximately [42].

A. Linear Systems Subject to Constraints

1) Linear Constraints: For linear Gaussian systems, linear constraints can be incorporated by directly within the KF framework as presented in eqs. 7 and 8. The widely used methods are based on model reduction methodology [43], [44], [45], pseudo-measurements, also called perfect measurement [46], [47], [48], [49], state estimate projection [47], [44], [50], and gain projection [47], [44].

2) Nonlinear Constraints: For nonlinear constraints, a closed-form solution may not be possible, even for linear systems. The adopted methods rely on linear approximation of the nonlinear constraints using Taylor series expansion [51], [50] or on direct numerical optimization of the nonlinear problem [43], [14]. Iterative application of the constraint linearization operation at each measurement time is used to get closer to the constraint satisfaction with each iteration [52]. Furthermore, PDF truncation methods have also been proposed that truncate Gaussian PDFs estimated by KF at the constraint bounds [8], [53].

B. Nonlinear Systems Subject to Constraints

Linear constraints can be added directly in various estimation frameworks, i.e., EKF, UKF, or EnKF. The nonlinear constraints can linearized and then incorporated in the filter.

1) Constrained State Estimation Using EKF: The system model (both dynamics and observation) and constraints can be linearized and used within the EKF framework for the constrained state estimation [54]. In some cases, the convergence properties of EKF can be improved by applying the filter iteratively for enforcing the constraint [55], [56].

For the equality constraints, the measurement-argumentation approach can be used [57], [58]. The measurement model can be augmented using the equality constraints and then EKF can be used for the linearized observation model. Obviously, measurement-argumentation approach is limited to equality constraints only.

Smoothing constrained kalman filter (SCKF) can be used with the EKF [52]. The SCKF constraints the system iteratively and the final constraint solution may not be accurate. This approach is only valid for equality constraints.

A modified extended Kalman filter has been introduced by Prakash et al. to handle the constraint imposed on the state estimation for a non-linear stochastic dynamic system [56]. Hence, Prakash et al. proposed two schemes to modify the prior and posterior distributions based on generating samples from truncated multivariate normal distribution.

Recursive Nonlinear Dynamic Data Reconciliation (RNDDR) suitably combine advantages of both EKF and Nonlinear Dynamic Data Reconciliation (NDDR). The NDDR is a nonlinear optimization based strategy to estimate system parameters and states. Due to the optimization-based formulation of the problem, constraints on states and unknown parameters can be added in a natural way. However, the application of the NDDR for online estimation of states and parameters can be computationally prohibitive due to at each time-step, a nonlinear constrained optimization problem is solved.

Recently, Zixiao et al. introduced a constrained dual extended Kalman filter algorithm that works in an alternating manner [61]. This algorithm is based on a parameter estimation and state prediction technique. The inequality constraints have been incorporated using optimization procedure. This proposed algorithm is more difficult to implement. However, it has the advantage of better convergence potential and algorithm stability.

In [62], the inequality constraints have been incorporated into the EKF using a gradient projection method. The performance of the proposed technique has been tested using synthetic model based on Gaussian functions.

2) Constrained State Estimation with UKF: Many approaches have been proposed to incorporate the constraints within UKF. During the update step, sigma points can be projected onto the constrained interval using sigma point projection approach [63], [64]. This approach perform projection after generation of sigma points and then after passing sigma points through system dynamics.

Optimization in measurement-update has been proposed to reformulate the measurement-update step to integrate constraints within the standard UKF [65]. This technique proposes to solve a quadratic or nonlinear optimization problem at each step of the algorithm.

Unscented Recursive Nonlinear Dynamic Data Reconciliation (URNDDR) is an extension of RNDDR where the EKF is replaced with the UKF [66], [67], [68]. All types of constraints are taken into consideration by solving the constrained optimization problem for each sigma point. The algorithm may be computationally expensive because the optimization problem is solved for all sigma points. Recently Kadu, Mandela et al. introduced improvements to the URNDDR algorithm to address general constraints in the generation process of sigma points and computational issues [69].

Julier and LaViola proposed a two-step approach for tackling nonlinear equality constraints. Using the UKF approach, all the selected sigma points are projected onto the constrained
surface individually in the first step, while in the second step the final estimate by the filter is again projected onto the constrained surface [70]. The authors presented a details discussion on the need for two-step projection for nonlinear constraints both for samples and their statistics (i.e., moments, which are expected value and covariance).

Teixeira et al. extended the UKF for different types constraints. These approaches include equality constrained UKF, projected UKF, measurement-augmented UKF, constrained UKF, constrained-interval UKF, interval UKF, sigma point UKF, truncated UKF, truncated interval UKF, project-ed-interval UKF [48], [71], [72], [73], [74]. PDF truncation approach applied by Teixeira et al. is applicable for linear interval constraints only [8].

Straka et al. proposed a truncated unscented kalman filter (TUKF) algorithm to solve the non-linear and non gaussian system with constraints on the state estimation [78]. The main idea of TUKF is to use the PDF truncation approach.

Recently, a truncated randomized unscented Kalman filter (TRUKF) was presented in [76]. The TRUKF is based on the randomized unscented Kalman filter algorithm (RUKF) and a pdf truncation technique. The main idea of TRUKF scheme is to introduce a truncation step within RUKF that was proposed for the unconstrained estimation problem. However, the application of TRUKF on synthetic data results in computational costs compared to UKF and TUKF (truncated UKF) [76].

Alireza et al. proposed a novel approach called the constrained iterated unscented Kalman filter (CIUKF) [77]. This approach combines the advantage of iterations and use of constraints to provide an accurate bounded dynamic state estimation. The constraints in the UKF are incorporated by projection the sigma points that are outside the feasible region to the boundary of this region to obtain constrained sigma points.

Calabrese et al. introduced an approach to integrate the constraints within the UKF based on two mainly approaches [78]. In the first approach, all sigma points that violate the constrained region are moved onto the feasible region during the prediction step. In the second approach, all transformed sigma points that violate the constrained region are projected to constraints boundary only when the updated state estimate exceeds the boundary in the correction step.

Variants of the algorithms seek to improve the performance and computational issues of the original UKF-based method under additional constraints [76], [79], [80].

3) Constrained state estimation using EnKF: A constrained state estimation approach can also use the EnKF [81], [82], where an initial ensemble of samples are drawn from a truncated normal distribution. For each iteration after the prediction step, a transformation is applied to project the violating samples onto the boundary.

Parakash et al. presented constrained state estimation using the EnKF [82], [83]. The authors propose to generate group of initial samples from a truncated normal distribution. Later, for each iteration after prediction step, a transformation is applied to project the violating samples on the boundary.

A Constrained Dual Ensemble Kalman Filter (dual C-EnKF) has emerged recently in [82]. The dual C-EnKF algorithm combines the C-EnKF algorithm proposed by Prakash in [81] for incorporating constraints within EnKF and the dual EnKF algorithm proposed in [84] to reduce the number of particles.

Raghu et al. introduced two algorithms to incorporate the constraints into the EnKF [85]. The first algorithm uses the projection-based method. The second algorithm relies on the use of a technique for soft constrained covariance localization. Simulations results showed that the second proposed algorithm provide better estimation of the unknown states compared to the first proposed algorithm.

4) Constrained state estimation using MHE: It is evident from formulation of the MHE [24] that constraints can be incorporated into its framework in a natural way [14], [15], [16], [86], [87]. However, there are multiple issues with MHE framework, i.e., 1) the computational effort especially for nonlinear optimization problem (constraints and/or objective functions) [15], [87], [88], [89]; 2) calculation of the arrival cost [90]; 3) Gaussian assumption for densities that results in simplification of the relations, i.e., from Eq. [23] to [24]; and 4) selection of optimal horizon (h) size to balance performance and computational load [90].

Recently, Garcia Tirado et al. introduced an approach for constrained estimation problem depends on the MHE and the game theoretical approach to the $H_{\infty}$ filtering with constraint handling [91]. The theoretical of the proposed approach with constraints mainly based on a modified Lyapunov theory for optimization-based systems.

It is important to emphasize that almost all methods that we present for the constrained state estimation have an underlying assumption of linearity or Gaussianity, an unrealistic presumption in most real-world applications. Moreover, the presence of constraints, such as bounds, on the states implies that the conditional state densities are non-Gaussian. Furthermore, the UKF and EnKF-based methodologies systematically constrain all their sigma points and ensemble samples with no mathematical ground or justification.

V. CONSTRAINED STATE ESTIMATION USING PARTICLE FILTERS

Particle filters are widely used for latent state estimation/tracking in dynamic systems where systems dynamics or observation models are nonlinear, or the system/observation noise are non-additive or follow non-Gaussian distributions [29]. The technique of PFs is based on powerful sampling that is aimed to find an optimal estimate by exploiting a set of random weighted samples called the particles. These particles are used to approximate the posterior density of the state and later find the statistics of interest [29]. Due to the complex nature of computations in PFs, it is not straightforward to incorporate constrains on the latent state. Systematic efforts to incorporate constraints imposed on the unknown state in PFs are limited and heuristic in nature.

A. Acceptance/Rejection Approach

An acceptance-rejection approach was proposed for nonlinear inequality constraints [92], [93]. This approach focused on retaining particles that fell within the constrained interval
and rejecting all violating constraint region. However, their approach does not make any assumption on the distributions and can guarantee the validity of particles, and thus retains the general properties of the particle filter. Besides, in certain cases, the number of particles may reduce which may further lead to a decrease in estimation accuracy and computationally efficiency.

In such cases, most of the particle may violate the constraint and the algorithm may fail. Also, unconstrained sampling from followed by verification against constraints (especially nonlinear) may be computationally more demanding than sampling directly from the constrained region [90, 92, 93].

B. Optimization/Projection Based Approach

Shao et al. presented a two stages approach to deal with constraints [94]. In the first stage, a set of particles were drawn randomly without considering state constraints. In the second stage, all particles which did not satisfy constraints, were projected into the feasible region using an optimization formulation. However, by applying optimization method to impose the particle to be within the constraining interval, the obtained particles are no longer considered as representative samples of the posterior distribution of the state. At every sampling instant, solving multiple optimization problems may make the algorithm computationally expensive.

Xiong et al. proposed an adaptive constrained particle filter (ACPF) algorithm that uses all particles to accurately estimate the state [95]. The approach is based on sample size testing that allows calculation of the number of particles needed to obtain a desired state estimate. Xiong et al. combined the sample size test with the generic particle filter to deal with constraints and to address the particle number problem in PF. The simulation results showed that, in terms of root mean square error (RMSE) performance, convergence, and running time, the ACPF approach exhibits a better estimation compared to the constrained PF proposed in [94] and unconstrained PF.

Li et al. used a series of constrained optimization to incorporate the inequality constraints into the auxiliary particle filter by modifying the priori pdf [96]. The auxiliary particle filter consists of resampling and sampling steps at each time step. In the resampling steps, it selected particles with a lower likelihood and/or far from the feasible region. It then performed a series of constrained optimization to transform the center of transition distribution into a feasible region.

Recently, Hongwei et al. introduced a constrained multiple model particle filtering (CMMPPF) method to solve the constrained high dimensional state estimation problems [97]. The proposed approach divided the problem of target tracking into two sub-problems: i) motion model estimation and model-conditioned state filtering as stated in the Rao–Blackwellised theorem; ii) The hidden state estimation is formulated using the multiple switching dynamic models in a jump Markov system framework. In order to incorporate the constraints within the proposed approach, a modified sequential importance resampling (MSIR) method based on a series of optimizations is used to generate model particles that can be constrained in the feasible region.

C. Constrained Importance Distributions

The particles can be drawn from a proposal distribution having support on the constrained region only. There are different approaches to incorporating constraints in the PF using the proposal distribution presented in literature. Density truncation can be performed analytically in case of multivariate Gaussian distribution [58, 98]. Perfect Monte Carlo simulations can be used to estimate first two moments of the truncated proposal distribution. Samples are drawn from a distribution of interest and all constraint-violating samples are rejected. Leftover samples are used to estimate mean and covariance of the truncated PDF.

The importance distribution can be constrained also, i.e., in the sampling step, particles are drawn from an importance distribution which has its support on the constrained region only [99, 100]. Constrained-EKF, constrained-UKF and or constrained-EnKF are used at each iteration to generate constrained proposal distributions. Specifically, a constrained-filter (EKF, UKF or EnKF) is used for each particle and an estimate of mean and covariance of constrained distribution is found. Particles are drawn from the newly found proposal distribution. An analytical solution for PDF truncation is also proposed [99, 100].

Straka et al. [75] proposed a truncated unscented particle filter to incorporate the constraints in the unscented particle filter. Proposal density is generated using the UKF and before sampling particles from it, a truncation procedure, in accordance with constraints on the system, is performed to form a truncated proposal density. The truncated proposal density is formed using perfect Monte Carlo or importance-sampling. The truncation procedure produces mean and variance of the proposal density, which is assumed to be Gaussian.

On the other hand, Yu et al. proposed a truncated unscented particle filter to handle nonlinear constraints [101]. Their technique combines both PF based accept/reject approach and the unscented Kalman filter. Authors start by drawing the importance distribution for sampling new particles using the unscented Kalman filter and then applied the truncated probability estimation using accept/reject approach. However, this technique also used the accept/reject approach and thus suffers from same limitations as mentioned earlier.

Similarly, Pirard et al. proposed two approaches to solve the problem of target tracking in the presence of constraints [102]. The first approach was based on the extension of the Rao-Blackwellized PF (RBPF) to handle hard constraints. The technique of RBPF or marginalized PF can be used only when the state can be divided into two parts such as a linear part and a non-linear part, and where the constraints only depend on the non-linear part. This approach used the accept/rejection technique that may lead to a reduction the estimation accuracy. The second approach was based on the proposal distributions adapted to the constraints. This approach drew a proposal distribution which guaranteed that particles were drawn from constrained regions. This could be quite ineffective when a substantial part of the drawn particles is located in outside constrained regions.

Ungarala et al. proposed constrained Bayesian state es-
D. Challenges in Constrained Particle Filtering

The particle filtering represents the state-of-the-art for estimation in nonlinear/non-Gaussian dynamical systems; however, incorporation of constraints on the hidden state (e.g., non-negativity) is a challenging problem.

1) On The Convergence of Constrained Particle Filtering: By constraining every particle in the PF, the PoDeT approach will always result in an estimation error of the posterior density unless the density has a bounded support. In particular, if the unconstrained distribution naturally satisfies the constraints and has unbounded support, the PoDeT will fail in rendering the unconstrained density. An evaluation of the PoDeT estimation error is discussed in (108), where we derived the optimality bounds of the PoDeT approach. We have shown that the estimation error is bounded by the area of the state posterior density that does not include the constraining interval. Specifically, if the density is well localized, i.e., most of the unconstrained posterior density were inside the constraining interval, then the PoDeT error will be upper bounded. On another hand, if the density is not well-localization, i.e., the constrained interval was not contained most of the posterior distribution, then the PoDeT error will be bounded from below. In particular, if the constraining interval covers a small area of the density, then the PoDeT error will be large (108).

2) Mean Density Truncation: Based on the PoDeT estimation error results, the mean density truncation (MDT) method has been recently proposed in (109). MDT deals with errors of the PoDeT by proposing a new strategy to satisfy the constraint on the conditional mean estimate rather than the posterior density itself (109). The authors generated $N$ unconstrained particles from the importance distribution as in the standard PF. If these $N$ weighted particles satisfied the constraints on the mean, the constrained state estimate is considered optimal. If not, an $(N+1)^{st}$ particle was drawn from a high probability region to enforce constraint on the weighted mean. In case if the $(N+1)^{st}$ particle was not sufficient to guarantee the constraints on the conditional mean, authors generated additional particles iteratively one by one (2, then 3, then 4 ..., etc.) until the constraints were satisfied. One drawback of this approach was evident when the proposal distribution was poorly chosen. In that case, it may take a large number of particles to move the mean of the conditional distribution into the feasible region. Consequently, the iterative technique of generating particles one by one may make the algorithm inefficient and computationally expensive.

Later, authors introduced systematic and inductive procedure to ensure that the constraints were satisfied with the generation of $N$ particles, referred to as the Iterative Mean Density Truncation (IMeDeT) (107). It is important to notice that, this perturbation approach remains very different from the PoDeT approach where the original constraint is imposed on every particle, while the $N$-particle IMeDeT enforces the desirable constraint only on the conditional mean estimate. However, the only drawback of IMeDeT is that inductively choosing particles $j = 1, \ldots, N$ to satisfy constraints on the conditional mean at every time step may lead to computation-
ally expensive.

Recently, we addressed the limitations of MDT and IMeDeT by proposing a new strategy based on perturbing the unconstrained density with only one particle [110]. We referred to this technique as Mean Density Truncation (MeDeT). In MeDeT, we choose one particle in a special way to satisfy the constraints on the mean and construct a sequence of densities satisfying constraints [110]. We start by generating $N$ unconstrained particles from the importance sampling distribution. If the conditional mean estimate using the $N$-order approximation satisfies the constraints, we retain these particles. Otherwise, after the resampling step, we remove a particle with the lowest weight and located closest to the boundary of the feasible region and replace it with another particle that is drawn from the high probability region. The process of remove/add one particle can be viewed as the "minimal perturbation" of the unconstrained density with using only one particle.

Keeping in view the way constrained are implemented during filtering, we can classify constraints as soft constraints and hard constraints. In the techniques of MDT, IMeDeT and MeDeT, the state satisfies the constraint on the conditional mean, so the constraints are considered soft constraints. On the other hand, for PoDeT, the constraints are enforced on each particle and are considered hard constraints.

We hope that this paper encourages further research in the development of more algorithms that constrain the state estimate rather than the density itself.

VI. CONCLUSION

This paper reviews the advances approaches to incorporate the constraints within state estimation. We provided a critical review of constrained Bayesian state estimation algorithms using Kalman filter for linear systems and extended Kalman filter, unscented Kalman filter, ensemble Kalman filter, moving horizon estimation and particles filters for nonlinear systems.

ACKNOWLEDGMENT

This work was supported by the National Science Foundation under Grants NSF CCF-1527822 and NSF ACI-1429467.

REFERENCES

[1] N. Amor, G. Rasool, N. Bouaynaya, and R. Shterenberg, “Constrained particle filtering for movement identification in forearm prosthesis,” Signal Processing, vol. 161, pp. 25–35, 2019.
[2] D. J. Albers, P.-A. Blancquart, M. E. Levine, E. E. Seylabi, and A. Stuart, “Ensemble kalman methods with constraints,” Inverse Problems, vol. 35, no. 9, p. 095007, 2019.
[3] C. Yang, M. Bakich, and E. Blasch, “Nonlinear constrained tracking of targets on roads,” in Proceedings of the International Conference on Information Fusion, Philadelphia, PA, July 2005, pp. 235 — 242.
[4] G. Battistello and M. Ulmke, “Exploitation of a-priori information for tracking maritime intermittent data sources,” in Proceedings of the International Conference on Information Fusion, Chicago, IL, July 2011, pp. 1–8.
[5] S. Thrun, “Probabilistic robotics,” Communications of the ACM, vol. 45, no. 3, pp. 52–57, 2002.
[6] D. Simon, Optimal State Estimation: Kalman, H∞, and Nonlinear Approaches. Wiley, July 2006, p. 552.
[7] A. H. Jazwinski, Stochastic Processes and Filtering Theory. Academic Press, 1970.
[8] D. Simon, Optimal state estimation: Kalman, $H_{\infty}$ and nonlinear approaches. Wiley-Interscience, 2006.
[9] S. J. Julier and J. J. LaViola, “Unscented filtering and nonlinear estimation,” Proceedings of the IEEE, vol. 92, no. 3, pp. 401–422, 2004.
[10] E. A. Wan and R. Van Der Merwe, “The unscented kalman filter for nonlinear estimation,” in Proceedings of the IEEE 2000 Adaptive Sys-
tems for Signal Processing, Communications, and Control Symposium (Cat. No. 00EX373). Ieee, 2000, pp. 153–158.
[11] S. J. Julier and J. K. Uhlmann, “New extension of the kalman filter to nonlinear systems,” in Signal Processing, Sensor Fusion, and Target Recognition VI, vol. 3068, Orlando, FL, April 1997, pp. 182–193.
[12] S. J. Julier, J. Uhlmann, and H. Durrant-Whyte, “A new method for the nonlinear transformation of means and covariances in filters and estimators,” IEEE Transactions on Automatic Control, vol. 45, no. 3, pp. 477–482, 2000.
[13] G. A. Terejanu, “Unscented kalman filter tutorial,” Workshop on Large-Scale Quantification of Uncertainty, Sandia National Laboratories, pp. 1–6, 2009.
[14] C. V. Rao, J. B. Rawlings, and J. H. Lee, “Constrained linear state estimation using a moving horizon approach,” Automatica, vol. 37, no. 10, pp. 1619–1628, Oct. 2001.
[15] C. V. Rao and J. B. Rawlings, “Constrained process monitoring: Moving-horizon approach,” AIChE Journal, vol. 48, no. 1, pp. 97–109, 2002.
[16] C. Rao, J. Rawlings, and D. Mayne, “Constrained state estimation for nonlinear discrete-time systems: stability and moving horizon approximations,” IEEE Transactions on Automatic Control, vol. 48, no. 2, pp. 246–258, 2003.
[17] X. Shao, B. Huang, and J. M. Lee, “Constrained bayesian state estimation - a comparative study and a new particle filter based approach,” Journal of Process Control, vol. 20, no. 2, pp. 143–157, Feb. 2010.
[18] S. C. Patwardhan, S. Narasimhan, P. Jagadeesan, B. Gopaluni, and S. L. Shah, “Nonlinear bayesian state estimation: A review of recent developments,” Control Engineering Practice, vol. 20, no. 10, pp. 933–953, Oct. 2012.
[19] S. Gillijns, O. Mendoza, J. Chandrasekar, B. L. R. De Moor, D. Bernstein, and A. Ridley, “What is the ensemble kalman filter and how well does it work?” in American Control Conference, Minneapolis, MN, June 2006, pp. 4448–4453.
[20] G. Evensen, Data Assimilation: The Ensemble Kalman Filter. Springer, Aug. 2009.
[21] G. Burgers, P. van Leeuwen, and G. Evensen, “Analysis scheme in the ensemble kalman filter,” Monthly Weather Review, vol. 126, no. 6, pp. 1719–1724, Jun. 1998.
[22] J. Prakash, S. C. Patwardhan, and S. L. Shah, “Constrained nonlinear state estimation using ensemble kalman filters,” Industrial & Engineering Chemistry Research, vol. 49, no. 5, pp. 2242–2253, Mar. 2010.
[23] F. Le Gland, V. Monbet, and V. Tran, “Large sample asymptotics for the ensemble Kalman filter,” in The Oxford Handbook of Nonlinear Filtering, B. R. Dan Crisan, Ed. Oxford University Press, 2011, pp. 598–631.
[24] A. Doucet, S. Godsill, and C. Andrieu, “On sequential monte carlo sampling methods for bayesian filtering,” Statistics and Computing, vol. 10, no. 3, pp. 197–208, Jul. 2000.
[25] A. Doucet and A. M. Johansen, “A tutorial on particle filtering and smoothing: Fifteen years later,” Handbook of Nonlinear Filtering, vol. 12, pp. 656–704, 2009.
[26] N. Gordon, D. Salmond, and A. Smith, “Novel approach to nonlinear/non-gaussian bayesian state estimation,” IEE Proceedings F Radar and Signal Processing, vol. 140, no. 2, pp. 107–113, 1993.
[27] A. Doucet, N. De Freitas, and N. Gordon, Sequential Monte Carlo methods in practice. Springer, 2001.
[28] R. Van Der Merwe, “Sigma-point kalman filters for probabilistic infer-
ence in dynamic state-space models,” Ph.D., University of Stellenbosch, 2004.
[29] A. Doucet and A. M. Johansen, Handbook of Nonlinear Filtering, 2009, vol. 12, ch. A tutorial on particle filtering and smoothing: Fifteen years later, pp. 656 – 704.
[30] P. K. Findiesen, “Moving horizon state estimation of discrete time systems,” Master’s Thesis, University of Wisconsin-Madison, 1997.
[31] A. Doucet, “On sequential simulation-based methods for bayesian filtering,” Department of Engineering, University of Cambridge UK, Tech. Rep., 1998.
