Data-Based Filters for Non-Gaussian Systems With Unknown Output Noise Covariance

Elham Javanfar ⬤ and Mehdi Rahmani ⬤, *Member, IEEE*

Abstract—This paper proposes linear and nonlinear filters for a non-Gaussian dynamic system with an unknown nominal covariance of the output noise. The challenge of designing a suitable filter in the presence of an unknown covariance matrix is addressed by focusing on the output data set of the system. Considering that data generated from a Gaussian distribution exhibit ellipsoidal scattering, we first propose the weighted sum of norms (SON) clustering method that prioritizes nearby points, reduces distant point influence, and lowers computational cost. Then, by introducing the weighted maximum likelihood, we propose a semi-definite program (SDP) to detect outliers and reduce their impacts on each cluster. Detecting these weights paves the way to obtain an appropriate covariance of the output noise. Next, two filtering approaches are presented: a cluster-based robust linear filter using the maximum a posterior (MAP) estimation and a cluster-based robust nonlinear filter assuming that output noise distribution stems from some Gaussian noise resources according to the ellipsoidal clusters. At last, simulation results demonstrate the effectiveness of our proposed filtering approaches.

Index Terms—Data-based filter, maximum likelihood estimation, unknown covariance, weighted maximum likelihood estimation, weighted sum-of-norms clustering.

I. INTRODUCTION

DUE to science and technology's rapid advances, a significant number of data are being generated in various engineering fields, including but not limited to satellite-based remote sensors, time-series systems, and telecommunication data [1]. This has made it imperative to analyze and process big data in contemporary engineering design, particularly in the areas of modeling, control, and estimation.

By appearing complex dynamics in different real-world systems such as robotics, aerospace, transportation, power grid, etc., we face significant uncertainties and less knowledge in our designs [2]. In this regard, traditional methods and principles in controller design, system monitoring, and performance evaluation are challenging or infeasible. Most approaches depend on accurate physical and dynamic models and complete information about design parameters. Obtaining an exact model is hard or impossible for complex systems. Different methods to model systems can be involved in four main cate-

gories: analytical, numerical, data-driven, and hybrid models [3]. In the era of machine learning, data-driven approaches offer powerful tools for identifying dynamical systems without requiring a deep understanding of the model structure [4], [5].

Data-based methods in dynamic systems have been presented primarily in control problems and rarely in state estimation. These methods play a crucial role in system identification and balancing the lack of knowledge about essential design parameters. During the last decades since the Industrial Revolution, system identification has been a critical element in most practical complex designs. Moreover, in some cases, despite access to the system model, many conventional estimation and control structures may not be applicable due to a lack of essential information. Recently, advancements in computational capabilities, iterative learning, reinforcement learning, and deep learning have given rise to new online and offline approaches to compensate for the problems mentioned above [6]−[9].

Estimation in dynamic systems is performed to obtain approximations of the system parameters using information from a model and any available measurements. Among all estimation and filtering methods, the Kalman filter is pervasive. It is a filter that uses the Bayesian rule to express the posterior probability in terms of the likelihood and the prior distributions [10]. The classical Kalman filtering theory has two main assumptions. The first is the accuracy of prior knowledge of the system model and statistical noise features, and the second is the Gaussianity of noises. In many practical systems, noises are non-Gaussian due to environmental conditions, sensor failure, manufacturing activities, etc. Heavy-tailed noises are most important among these different kinds of noises [11], [12]. Various filters have been designed to work out the filtering problem against heavy-tailed noises, e.g., see [13]−[18]. Although these filters have suitable performance, they have limitations like knowing the accurate nominal relevant covariance matrices or comparative threshold. Since non-Gaussian heavy-tailed noises increase the covariance value of a distribution [19], filters need some considerations to compensate for it. Lack of exact models and essential knowledge about the system and noise features degrade the performance of filters; therefore, data-driven filtering methods are gaining popularity. In [20], a direct data-driven filter has been designed with no mathematical model for linear time-invariant dynamic systems with bounded disturbances and noises. The authors use set membership to estimate the sets of solutions. Furthermore, using quantized measurements,

[21] proposes a data-driven filter to minimize the worst-case estimation error in the presence of bounded noises. Also, it introduces an $\mathcal{L}_2$-$\mathcal{L}_\infty$ approximately-optimal worst-case filter through linear programming technique. Recently, data-driven unknown input observer and state estimator for linear time-invariant systems are investigated in [22].

Considering all of the mentioned points, data-driven techniques have not been effectively applied in state estimation problems. Moreover, despite having information about dynamic models, conventional filtering methods are sometimes inefficient because of severe environmental noise conditions and lack of statistical features. Generally, covariance matrix estimation methods can be divided into four major groups including correlation method [23], the maximum-likelihood method [24], the covariance matching method [25] and the Bayesian method [26]. Unfortunately, not only do most of these approaches assume that noises have Gaussian distribution but also they suffer from some problems such as sensitivity to outliers, reaching non-invertible covariance matrices, etc. Therefore, there is a solid need to design an appropriate filter for dynamic systems in the presence of non-Gaussian noises with less or no information about noise covariances. For non-Gaussian systems, a finite set of higher-order moments of the state and measurement noises is obtained in [27] using the correlation measurement difference method, such that the observable matrix is full rank. Moreover, some recent works such as [28], [29] present a set of suitable filters against inaccurate covariances of the process and measurement noises for non-Gaussian dynamic systems using the variational Bayesian method. Since these filters are designed based on the probability density functions, they are more sensitive to initial values of distributions' parameters. Considering all stated problems for obtaining noise covariance, using systems output data can be an effective way to compensate for deficit knowledge. This motivates us to propose a data-based filter for time-variant non-Gaussian systems without depending on the output nominal noise and accurate process noise covariances. In the proposed approach, we assume that system output data are accessible.

Grouping data based on their likeness is an approved concept in various science fields. In statistics, however, it can be done based on two different situations where there is prior information to gain more about the group structure or not. Unavailable information necessitates unsupervised learning tools or clustering algorithms. In other words, the problem of dividing a given set of data points with high uniformity within the groups and low diversity between groups is called clustering. Clustering is ubiquitous in machine learning, pattern recognition, statistics, image processing, and biology. Some important clustering algorithms are hierarchical clustering, Gaussian mixture models (GMMs), and K-mean clustering [30]. Each of these methods bears some disadvantages. Time complexity and nonexisting mathematical objectives are primary defects of hierarchical clustering [31]. Long computation time, falling into local optimum, and deciding are the dominant shortages of Gaussian mixture models [32], [33]. Also, sensitivity to the initial condition and considerably different clustering results are the focal paucities of K-means clustering [34], [35]. These methods are generally beset by local minima, which are sometimes significantly suboptimal. Recently, sum-of-norms (SON) clustering has been introduced that ensures a unique global minimizer [36], [37] and covers all the problems mentioned earlier.

Clustering the system data plays a vital role in our proposed filtering scheme. The shape representation of a cluster is also vital in preserving the data features. Ellipsoidal clusters are common because many data observations are normally distributed [38]. We cluster the data using the sum-of-norms clustering method according to this characteristic and the above-discussed advantages. In this method, because of choosing a threshold value as a meter for Euclidian distances, each cluster may suffer from some outliers. This changes the ellipsoidal shape due to pushing the cluster center closer to the outliers. We propose a semi-definite program based on the weighted maximum likelihood estimation (MLE) to decrease their bad effects. By reducing the outlier's effect, we find the robust covariance of each cluster as the covariance of output noise for data belonging to that cluster. Finally, we propose cluster-based linear and nonlinear filters. To sum it up, the goals and contributions of the proposed approach can be listed as follows.

1) To design a data-based filter against heavy-tailed noises with unknown output and inaccurate process noise covariance.

2) To present the idea of ellipsoidal clustering to compensate for less knowledge about noises' statistical features.

3) To suggest the weighted SON clustering to improve regularization and the performance of the conventional SON.

4) To propose the weighted MLE to decrease the effect of outliers in each cluster and keep the clusters' ellipsoidal shape.

5) To present two data-based filtering approaches, including cluster-based linear and cluster-based non-linear filters.

The remainder of the paper is organized as follows. Section II formally reviews some prerequisite and briefly refers to the main problem. The clustering steps, a new SDP to reduce the effects of outliers, and the proposed cluster-based filters are discussed in Section III. The notable specifications and features of the proposed filtering approaches are presented in Section IV. The simulation results are given in Section V before concluding the paper in Section VI.

Notations: The paper uses the following standard notations. $\mathbb{R}^m$ and $\mathbb{R}^{m \times r}$ signify the m-dimensional Euclidean space and the set of all $m \times r$ real matrices, respectively. $\mathcal{N}(\cdot)$ designates the multivariate Gaussian probability density function. $\mathbf{0}$ and $\mathbf{I}$ represent zero and identity matrices with appropriate dimensions, respectively. Furthermore, $\prod$ shows the product operation. The symbol "$*$" in matrices stands for the symmetric terms. Also, $\log(\cdot)$ indicates the natural logarithm operation.

II. PRELIMINARIES AND PROBLEM FORMULATION

The concepts listed below will be used to achieve our key goals.

A. Sum-of-Norms Clustering

We are interested in dividing a set of observations, $\{X_j\}_{j=1}^N \in \mathbb{R}^d$, into different clusters such that the close points to each other are assigned to the same cluster based on the Euclidian

meter. We do not know the number of clusters, and it is unnecessary to be large. Assume that each cluster has a centroid in μ_j and is a subset of $\mathbb{R}^d$. The SON clustering problem is presented as follows [37]:

$$\min_{\mu_1 \cdots \mu_N} \sum_{j=1}^{N} \|x_j - \mu_j\|^2 + \lambda \sum_{j=2}^{N} \sum_{i<j} \|\mu_i - \mu_j\|_p \tag{1}$$

in which $p \geq 1$, and $\lambda > 0$ can be regarded as a parameter that controls the trade-off between the first term in (1) and the number of clusters. The term of $\sum_{j=1}^{N} \|x_j - \mu_j\|^2$ presents the sum-of-squares error, where μ_j is the centroid of the cluster containing x_j. If the corresponding μ's are the same, two different x's belong to the same cluster. This is the result of the second term in (1), which is a regularization term. In addition, we choose $p = 2$ in the proposed approach, but other choices are possible. After finding the center of the clusters using the optimization problem (1), the data are fitted to each cluster based on the spatial threshold.

B. Multivariate Gaussian Distribution

The Gaussian distribution, also known as the normal distribution, resembles a symmetrical bell shape. Let x be a random vector on $\mathbb{R}^p$. It has the following probability density function:

$$p(x) = \frac{1}{(2\pi)^{\frac{p}{2}}} (\det \Xi)^{\frac{-1}{2}} \exp\{-\frac{1}{2}(x-\mu)^T \Xi^{-1}(x-\mu)\} \tag{2}$$

where $\Xi \in \mathbb{R}^{p \times p}$ is the positive definite covariance matrix, and μ is the mean.

Remark 1: $(x-\mu)^T \Xi^{-1}(x-\mu)$, is a square of Mahalanobis distance. It corresponds to the actual probability of the occurrence of the observation.

C. Maximum Likelihood Estimation

Maximum likelihood estimation is a popular way of obtaining practical estimators. Cramer Rao Lower Bound (CRLB), Gaussian PDF, and unbiasedness are among MLE's asymptotic properties. Consider a set of i.i.d of data points $Y \in \mathbb{R}^{m \times n}$, containing n observations, which have a Gaussian distribution with mean μ and covariance Ξ. The likelihood function, under the normality assumption, can be written as

$$f(Y,\mu,\Xi) = \left(\frac{1}{\sqrt{(2\pi)^m |\Xi|}}\right)^n \exp\{-\frac{1}{2}(y_1-\mu)^T \Xi^{-1}$$

$$\times (y_1-\mu) - \cdots - \frac{1}{2}(y_n-\mu)^T \Xi^{-1}(y_n-\mu)\}. \tag{3}$$

When dealing with large data sets, we frequently seek statistical and mathematical models to simplify their presentations. One of the first questions we ask is whether the data can be fitted with a normal distribution. This entails estimating the normal distribution's mean and covariance. They are usually computed using the conventional MLE method by the following problem:

$$\max_{\mu,\Xi} \sum_{i=1}^{n} \log\left(f(y_i,\mu,\Xi)\right). \tag{4}$$

In this problem, the best estimates of the mean and variance

are obtained by taking the partial derivative with respect to μ and Ξ of the log-likelihood function and setting it to zero. As a result, we get

$$\mu = \sum_{i=1}^{n} \frac{1}{n} y_i$$

$$\Xi = \sum_{i=1}^{n} \frac{1}{n} y_i y_i^T - \mu \mu^T. \tag{5}$$

According to the normality condition, the log-likelihood function consists of the sum of the squared Mahalanobis distances. Consequently, if outlier data exist, the mean and covariance are pushed toward the outliers.

Remark 2: Maximizing the logarithm of the likelihood cost function is equivalent to minimizing the Mahalanobis distance. Outliers make the Mahalanobis distance large.

D. Kalman Filter

The Kalman filter is the most famous state estimator for linear Gaussian systems, but its performance is degraded in the presence of non-Gaussian noises. It can be derived from Bayesian recursive relations. In this regard, prediction and filtering steps are achieved as follows:

Prediction:

$$p(x_{k+1}|\mathbf{y^k}) = \mathcal{N}(x_{k+1} : \hat{x}'_{k+1}, P'_{k+1})$$

$$\hat{x}'_{k+1} = A_k \hat{x}_k$$

$$P'_{k+1} = A_k P_k A_k^T + Q_k \tag{6}$$

Filter:

$$p(x_k|\mathbf{y^k}) = \mathcal{N}(x_k : \hat{x}_k, P_k)$$

$$\hat{x}_k = \hat{x}'_k + K_k(y_k - C_k \hat{x}'_k)$$

$$P_k = (I - K_k C_k)P'_k(I - K_k C_k)^T + K_k R_k K_k^T \tag{7}$$

where the gain matrix is given by

$$K_k = P'_k C_k^T \left(C_k P'_k C_k^T + R_k\right)^{-1}. \tag{8}$$

E. System Model

We assume that output measured data are produced from the following linear state-space dynamic model:

$$x_{k+1} = A_k x_k + v_k$$

$$y_k = C_k x_k + w_k$$

$$\mathbf{y^k} = [y_1, y_2, \dots, y_k] \tag{9}$$

where $x_k \in \mathbb{R}^{n_x}$ and $y_k \in \mathbb{R}^{n_y}$ are, respectively, the state and the measurement signals. $A_k \in \mathbb{R}^{n_x \times n_x}$ and $C_k \in \mathbb{R}^{n_y \times n_x}$ are dynamic and output matrices, respectively. Process noise, v_k, is a non-Gaussian noise vector with zero mean and inaccurate nominal covariance, Q_k. Also, measurement noise sequence, w_k, is a non-Gaussian noise vector with zero mean and unknown nominal covariance matrix, R_k. It is remarkable that the process noise stems from internal factors while output noise comes from external sources; therefore, the process experiences less intense noise than the system's output. Moreover, all measurements up to and including time k are pre-

sented by $\mathbf{y^k}$. We assume that the initial condition x_0 and the system's noises are mutually independent, satisfying the following relation:

$$\mathbb{E}\left(\begin{bmatrix} x_0 - \bar{x}_0 \\ v_k \\ w_k \end{bmatrix} \begin{bmatrix} * \end{bmatrix}^T\right) = \begin{bmatrix} P_0 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & Q_k & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & R_k \end{bmatrix} \quad (10)$$

where $\bar{x}_0$ is the expectation of x_0.

This paper aims to design data-based filters for the output data set, y, produced by (9). It is assumed that the output noise nominal covariance is unknown. In the presence of outliers, conventional MLE estimators are affected by both good and bad observations. To compensate for this defect, after applying the proposed clustering method, we try to detect outliers in each cluster to decrease their effects on changing the shape of clusters and the covariances. Then, we present a linear filter based on the moving horizon estimation technique by restating the conventional MAP estimation problem for a measurement data set. Moreover, considering that data in each ellipsoidal cluster originated from a Gaussian distribution, we assume that there are ε_l ellipsoidal clusters with Gaussian specifications for the output noise. By doing so, we propose a novel data-based nonlinear filter.

III. MAIN RESULTS

We intend to design a suitable filter for non-Gaussian linear dynamic systems with unknown output and inaccurate process noise covariance. Regarding our idea of output measurements clustering, first, we propose the weighted SON clustering method that improves the conventional SON's performance in a large data set.

A. Weighted Sum-of-Norms Clustering

We propose the following weighted SON clustering to mitigate the influence of distances between cluster centers on the clustering performance and enhance the computational efficiency of the conventional SON clustering method (1):

$$\min_{\mu_1 \cdots \mu_N} \sum_{j=1}^{N} \|x_j - \mu_j\|^2 + \lambda \sum_{j=2}^{N} \sum_{i<j} \zeta_{ij} \|\mu_i - \mu_j\|_p \quad (11)$$

where $\zeta_{ij} \geq 0$ is a weighting parameter.

Using a constant weight for each point in the objective function of clustering can lead to suboptimal results because distant points with low similarity values would have the same impact as nearby points with high similarity values. To address this issue, we introduce the concept that the affinity weight, ζ_{ij}, quantifies the similarity between x_j and x_i, and assigns higher weights to nearby points and lower weights to distant ones. This approach can also reduce the impact of noise and outliers and consequently, result in more accurate and robust clustering. This requires that the weight ζ_{ij} be obtained from a similarity function based on a statistical similarity measure criterion. The similarity function is typically a strictly monotonically decreasing continuous function in the $[0 \, \infty)$ range with a positive second-order derivative.

Remark 3: The similarity measure criterion provides freedom in selecting the similarity function. One of the most widely recognized functions is the Gaussian kernel.

We know that the data in the same cluster have similar statistical features. Hence, to obtain output noise covariance, first, we decide to cover a set of n points $\mathbf{y^k}$, by some ellipses $(\epsilon_1, \epsilon_2, \ldots, \varepsilon_\ell)$ using the weighted SON clustering. Now, at the first step, Algorithm 1 is presented for clustering purpose.

Algorithm 1 Clustering Steps

Required:
Data point $\{X_j\}_{j=1}^{N} \in \mathbb{R}^d$ and similarity function f.
Steps:
1) Set parameter $\lambda > 0$ in (11).
2) Run optimization (11).
3) Find mean and covariance of each cluster based on (5).
4) Detect unsuitable clusters and re-cluster them.
5) Find new mean and covariance of each cluster by (5).

Remark 4: To have more efficiency in Algorithm 1, a re-clustering process is welcome for unsuitable clusters. In this regard, clusters with a small number of data are removed and combined with the other clusters, and clusters with large estimated covariances are divided into some smaller clusters. By doing so, a balance is made between the number of clusters and their number of data.

B. Detecting and Reducing the Effect of Outliers

The presented clustering algorithm results in some ellipsoidal clusters for output data of the system. According to the fact that Gaussian data have ellipsoidal scattering, we will encourage the use of the filters developed for Gaussian systems for each cluster. Note that the obtained clusters may suffer from some outlier data. Outliers can change the covariance of that cluster and create remarkable bias in the conventional MLE. For this reason, we are supposed to make a difference between outliers and good data to improve the performance of the MLE. In this regard, we need to minimize the sum of the smallest squared Mahalanobis distances while outliers with larger squared values are excluded. Under such a circumstance, we introduce a variable ω_i in the range of $[0 \, 1]$, so that if the squared Mahalanobis distance is larger, the corresponding ω_i is smaller. This approach is useful in state estimation to detect and decrease the effect of outliers. According to this concept, we consider the following weighted MLE:

$$\max_{\mu, \Xi} \sum_{i=1}^{n} \omega_i \log\left(f(y_i, \mu, \Xi)\right)$$

$$\sum_{i=1}^{n} w_i = 1. \quad (12)$$

We can rewrite the cost function in (12) by replacing the likelihood function (3) and using the features of the logarithm function as follows:

$$\mathcal{L}(\mu, \Xi) = \sum_{i=1}^{n} -\frac{\omega_i}{2} \{\log(2\pi)^m + \log\det\Xi + \|y_i - \mu\|_{\Xi^{-1}}^2\}$$

$$\sum_{i=1}^{n} w_i = 1. \quad (13)$$

One crucial point is that the above log-likelihood function is not concave or convex.

If we fix the parameter ω in the optimization problem (12), we can easily obtain the mean and covariance variables by the following lemma.

Lemma 1: Consider the weighted MLE problem (12), the mean and covariance, μ and Ξ, are obtained by

$$\mu_w = \sum_{i=1}^{n} w_i y_i$$

$$\Xi_w = \sum_{i=1}^{n} w_i y_i y_i^T - \mu_w \mu_w^T. \tag{14}$$

Proof: By taking partial derivations of (12) with respect to μ and Ξ and setting them tos zero, it is straightforward to obtain (14). ∎

Due to a lack of knowledge of good and bad observations and the problem's non-convexity, we cannot determine the appropriate weights and utilize the weighted MLE. Iterative algorithms, like iteratively reweighted least squares [39], can solve this problem by updating the weights in inverse proportion to Mahalanobis distances. However, slow convergence and trap into local optimal solutions are some of their weaknesses. In the following, with the help of a suitable objective function and introducing probability criterion on all observations, we will present an optimization problem to set the outliers with smaller weights for solving the above-mentioned issues. To this end, considering the following relations:

$$d_i = (y_i - \mu_\chi)^T \Xi_\chi^{-1} (y_i - \mu_\chi)$$

$$\text{prob}(y_i) = (2\pi)^{-\frac{n_y}{2}} (\det \Xi_\chi)^{-\frac{1}{2}} \exp(\frac{-d_i}{2}) \tag{15}$$

and using the fact that outliers have low probability and large Mahalanobis distances, we encourage using probability of occurrence as weights to decrease the effect of outliers in our proposed method. We introduce a positive weight vector $\chi = (\chi_1, \ldots, \chi_n)$ with $\sum_{i=1}^{n} \chi_i = 1$. Therefore, inspiring Remark 2, the mean and covariances of the weighted MLE (14), and the squared Mahalanobis distances (15), we present the following optimization problem to obtain optimal values of χ and d:

$$\min_{\chi, \mu_\chi, \Xi_\chi, d} L(\chi, d)$$

$$\text{s.t.} \quad \mu_\chi = \sum_{i=1}^{n} \chi_i y_i$$

$$\Xi_\chi = \sum_{i=1}^{n} \chi_i y_i y_i^T - \mu_\chi \mu_\chi^T$$

$$d_i = (y_i - \mu_\chi)^T \Xi_\chi^{-1} (y_i - \mu_\chi)$$

$$\sum_{i=1}^{n} \chi_i = 1$$

$$\chi \geq 0. \tag{16}$$

In (16), the loss function must be monotonically non-decreasing function of χ_i and d_i. This function should be cho-

sen such that χ_i is inversely proportional to the corresponding Mahalanobis distance, d_i. Doing so allows all outlier observations with large d_i to receive small weight χ_i. On the other hand, we ought to choose a suitable cost function such that the sum of all Mahalanobis distances is small. To satisfy all of these conditions, in the optimization problem (16), we propose the cost function as

$$L(\chi, d) = \sum_{i=1}^{n} \chi_i^2 (d_i + 1). \tag{17}$$

In what follows, we present a theorem to obtain the best values of weights, χ_i, which have an essential role in outliers detection and decreasing their effects.

Theorem 1: The optimization problem (16) with the cost function (17) can be reformulated as the following SDP problem:

$$\min_{\chi, \delta} \sum_{i=1}^{n} \delta_i$$

$$\text{s.t.} \begin{pmatrix} 1 & \chi_i & \sum_{i=1}^{n} \chi_i y_i^T \\ \chi_i & \delta_i & \chi_i y_i^T \\ * & * & \sum_{i=1}^{n} \chi_i y_i y_i^T \end{pmatrix} \geq 0, \forall i$$

$$\sum_{i=1}^{n} \chi_i = 1$$

$$\chi \geq 0. \tag{18}$$

Proof: For the cost function (17), assume an upper bound, δ_i, such that $\chi_i^2 (d_i + 1) \leq \delta_i$. Thus, minimization of this cost can be formulated as follows:

$$\min_{\chi, \delta} \sum_{i=1}^{n} \delta_i$$

$$\text{s.t.} \quad \chi_i^2 (d_i + 1) \leq \delta_i, \forall i$$

$$\sum_{i=1}^{n} \chi_i = 1$$

$$\chi \geq 0. \tag{19}$$

The first constraint can be reformulated as

$$\delta_i - \chi_i^2 \geq \chi_i^2 (y_i - \mu_\chi)^T \Xi_\chi^{-1} (y_i - \mu_\chi)$$

$$\Rightarrow \begin{pmatrix} \delta_i - \chi_i^2 & \chi_i(y_i - \mu_\chi)^T \\ * & \Xi_\chi \end{pmatrix} \geq 0. \tag{20}$$

Now, by substituting the weighted mean and covariance from the constraints of (16), the above equation can be simplified as

$$\begin{pmatrix} \delta_i - \chi_i^2 & \chi_i y_i^T - \chi_i \mu_\chi^T \\ \chi_i y_i - \chi_i \mu_\chi & \sum_{i=1}^{n} \chi_i y_i y_i^T - \mu_\chi \mu_\chi^T \end{pmatrix} \geq 0$$

$$\Rightarrow \begin{pmatrix} \delta_i & \chi_i y_i^T \\ * & \sum_{i=1}^{n} \chi_i y_i y_i^T \end{pmatrix} - \begin{pmatrix} \chi_i \\ \sum_{i=1}^{n} \chi_i y_i \end{pmatrix}$$

$$\times \begin{pmatrix} \chi_i & \sum_{i=1}^{n} \chi_i y_i^T \end{pmatrix} \geq 0.$$

Finally, applying the Schur complement equivalence, the SDP problem in (18) is obtained. ∎

Remark 5: Since outliers are destructive in different domains and applications and cause some limitations, the proposed SDP problem in (18) can detect outliers in each Gaussian data set without needing extra designs.

Until now, we have presented a method to cluster the data set and an optimization problem to compute the weights in the MLE for reducing the effect of outliers. Note that we can obtain the covariance of each cluster using two methods, the conventional MLE and the weighted MLE. The second method is robust against outliers thanks to obtaining the covariance according to reducing their impact. Now, we are ready to introduce our proposed filtering approaches.

C. Cluster-Based Linear Filter

Since the proposed approach incorporates the new weighted SON clustering and the new weighted MLE, we deal with a data set belonging to Gaussian distributions with different statistical features in each cluster. Given this circumstance, in this section, instead of using the existing recursive filters, such as the Kalman filter, to estimate or predict the states according to the available information at each instant, we intend to estimate the entire states at once using the whole data set.

Theorem 2: The state estimations for the output data set, $y_{1:N}$, belonging to the system (9), can be obtained as

$$\hat{\mathcal{X}} = -(\tilde{O}^T \tilde{O})^{-1} \tilde{O}^T \tilde{\mathcal{F}} \tilde{\mathcal{Y}} \tag{21}$$

where

$$\tilde{\mathcal{Y}} = \begin{pmatrix} \hat{x}_0 \\ y_0 \\ \vdots \\ y_N \end{pmatrix}, \quad \mathcal{X} = \begin{pmatrix} x_0 \\ \vdots \\ x_N \end{pmatrix}, \quad \tilde{\mathcal{F}} = \begin{pmatrix} P_0^{-\frac{1}{2}} & 0 \\ \hline 0 & 0 \\ 0 & \mathcal{R}_y \end{pmatrix}$$

$$\tilde{O} = \begin{pmatrix} P_0^{-\frac{1}{2}} & 0 \\ \hline Q_x \\ \hline \mathcal{R}_x \end{pmatrix} \mathcal{R}_y = \begin{pmatrix} R_0^{-\frac{1}{2}} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & R_N^{-\frac{1}{2}} \end{pmatrix}$$

$$Q_x = \begin{pmatrix} -Q_0^{-\frac{1}{2}} A_0 & Q_0^{-\frac{1}{2}} & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & -Q_{N-1}^{-\frac{1}{2}} A_{N-1} & Q_{N-1}^{-\frac{1}{2}} \end{pmatrix}$$

$$\mathcal{R}_x = \begin{pmatrix} -R_0^{-\frac{1}{2}} C_0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & -R_N^{-\frac{1}{2}} C_N \end{pmatrix}.$$

Proof: Based on the Bayesian method and using the fact that process and measurement noises are independent, we have

$$p(x_{0:N}) = \prod_{i=0}^{N-1} p(x_{i+1}|x_i) p(x_0)$$

$$p(y_{0:N}|x_{0:N}) = \prod_{i=0}^{N} p(y_i|x_i).$$

Thus, the posterior is given by

$$\hat{\mathcal{X}} = \arg\min_{x_{0:N}} \|x_0 - \hat{x}_0\|_{P_0^{-1}}^2 + \sum_{i=0}^{N-1} \|x_{i+1} - A_i x_i\|_{Q_i^{-1}}^2$$

$$+ \sum_{i=0}^{N} \|y_i - C_i x_i\|_{R_i^{-1}}^2. \tag{22}$$

According to the fact that $\|x\|_{w^{-1}} = (w^{\frac{-1}{2}} x)^T (w^{\frac{-1}{2}} x)$, the cost function in (22) can be rewritten as

$$\{P_0^{-\frac{1}{2}}(x_0 - \hat{x}_0)\}^T \{P_0^{-\frac{1}{2}}(x_0 - \hat{x}_0)\} + \{Q_0^{-\frac{1}{2}}(x_1 - A_0 x_0)\}^T$$

$$\times \{Q_0^{-\frac{1}{2}}(x_1 - A_0 x_0)\} + \cdots + \{Q_{N-1}^{-\frac{1}{2}}(x_N - A_{N-1} x_{N-1})\}^T$$

$$\times \{Q_{N-1}^{-\frac{1}{2}}(x_N - A_{N-1} x_{N-1})\} + \{R_0^{-\frac{1}{2}}(y_0 - C_0 x_0)\}^T$$

$$\times \{R_0^{-\frac{1}{2}}(y_0 - C_0 x_0)\} + \cdots + \{R_N^{-\frac{1}{2}}(y_N - C_N x_N)\}^T$$

$$\times \{R_N^{-\frac{1}{2}}(y_N - C_N x_N)\}. \tag{23}$$

By rearranging the terms in (23) and using the defined matrices in (21), the optimization problem (22) can be written as

$$\hat{\mathcal{X}} = \arg\min_{\mathcal{X}} (\tilde{\mathcal{F}} \tilde{\mathcal{Y}} + \tilde{O} \mathcal{X})^T (\tilde{\mathcal{F}} \tilde{\mathcal{Y}} + \tilde{O} \mathcal{X}). \tag{24}$$

Consequently, taking partial derivation with respect to $\mathcal{X}$, the estimation, $\hat{\mathcal{X}}$, in (21) is obtained. ∎

Since (21) involves the inverse of the matrix $(\tilde{O}^T \tilde{O})$, this matrix grows in size as the number of data or dimension increases. To solve this unfavorable issue, we propose the following optimization problem to obtain the moving horizon estimation (MHE) with a window length of N_w:

$$\min_{x_{t-N_w:t}} \|x_{t-N_w} - \hat{x}_{t-N_w}\|_{P_{t-N_w}^{-1}}^2 + \sum_{i=t-N_w}^{t-1} \|x_{i+1} - A_i x_i\|_{Q_i^{-1}}^2$$

$$+ \sum_{i=t-N_w}^{t} \|y_i - C_i x_i\|_{R_i^{-1}}^2. \tag{25}$$

Consequently, the steps of the proposed cluster-based robust moving horizon estimation are presented in Algorithm 2.

Remark 6: In Algorithm 2, Theorem 1 determines the best weights and detects outliers; subsequently, the elimination of outliers when computing the robust covariance matrices of each cluster results in a filter that is robust against outliers. Therefore, we call it a cluster-based robust MHE. We have a

simple (non-robust) cluster-based MHE if we omit step 2 and determine the covariance matrix by the conventional MLE.

Algorithm 2 Cluster-Based Robust Moving Horizon Estimation

Required:
Data Point $\{X_j\}_{j=1}^N \in \mathbb{R}^d$ and window length N_w.
Steps:
1) Run Algorithm 1.
2) Detect outliers of each cluster using Theorem 1.
3) Calculate robust covariance matrix of each cluster.
4) Determine cluster of each y_k.
5) Obtain the state estimation from (25).

Remark 7: Equation (21) is the closed-form solution of the optimization problem (25) in a complete interval $N_w = N$.

D. Cluster-Based Non-Linear Filter

As we discussed, each ellipsoidal set's data originates from a Gaussian distribution in the SON clustering. Since we cluster the output data with ε_l independent ellipsoidal set and considering this assumption that the intensity of the process noise is less than output noise, we can assume that there exist ε_l Gaussian resources for the output noise distribution. Considering all clusters, the total occurrence probability of the output noise is equal to the sum of the occurrence probability of each Gaussian resource. This is another idea to design a new filter. We show that this filter would be non-linear. To this end, we introduce the following probability density function for each noise resource:

$$p(w_{ik}) = \beta_{ki}\mathcal{N}(w_k : 0, R_{ki}),\ i = 1 : \varepsilon_l$$

$$\sum_{i=1}^{\varepsilon_l}\beta_{ki} = 1,\quad \beta_{ki} = \frac{n_i}{n} \tag{26}$$

where i is the number of clusters, n_i is the number of data in each cluster, and n is the number of data in a data set. Due to the data clustering, it is impossible to determine which noise distributions ($i = 1 : \varepsilon_l$) occur at each specific time. In this condition, to solve the estimation problem using the Bayesian method, it is necessary to consider ε_l possibilities and the potential influence of each noise distribution on the measurements by incorporating separate prior states ($\acute{x}_{01} : \acute{x}_{0\varepsilon_l}$) corresponding to each noise distribution. To this end, we obtain the conditional PDF of $p(x_k|\mathbf{y^k})$. Let us start with $k = 0$ and determine $p(x_0|\mathbf{y^0})$ using the measurement y_0; then, we can generalize it to other times. First, the probability density function $p(y_0|x_{0i})$ can easily be obtained as

$$p(y_0|x_{0i}) = \beta_{ki}\mathcal{N}(y_0 : C_0 x_{0i}, R_{0i}). \tag{27}$$

Using above explanations and substituting (27) into the Bayesian rule result in

$$p(x_0|\mathbf{y^0}) = \frac{\displaystyle\sum_{i=1}^{\varepsilon_l}\mathcal{N}(x_0 : \acute{x}_{0i}, \acute{P}_{0i})\beta_{ki}\mathcal{N}(y_0 : C_0 x_{0i}, R_{0i})}{\displaystyle\sum_{i=1}^{\varepsilon_l} p_i(y_0|\mathbf{y^{-1}})}. \tag{28}$$

The denominator in (28) involves the integration of the sum

of likelihoods and priors for each noise distribution. This ensures that the total probability of all states of x_0 given y_0 equals one. In this vein, the density function (28) can be written as (29).

$$p(x_0|\mathbf{y^0}) = \frac{\displaystyle\sum_{i=1}^{\varepsilon_l}\mathcal{N}(x_0 : \acute{x}_{0i}, \acute{P}_{0i})\beta_{ki}\mathcal{N}(y_0 : C_0 x_{0i}, R_{0i})}{\displaystyle\int (\sum_{i=1}^{\varepsilon_l}\mathcal{N}(x_0 : \acute{x}_{0i}, \acute{P}_{0i})\beta_{ki}\mathcal{N}(y_0 : C_0 x_{0i}, R_{0i}))dx_0}. \tag{29}$$

Considering the Bayesian rule for obtaining the conventional Kalman filter and relations in (6) and (7), it holds that $\mathcal{N}(x_0 : \hat{x}'_0, P'_0)\,\mathcal{N}(y_0 : C_0 x_0, R_0) = \mathcal{N}(x_0 : \hat{x}_0, P_0)\,\mathcal{N}(y_0 : C_0\hat{x}_0, C_0\acute{P}_0 C_0^T + R_0)$. To achieve our proposed filter structure and using this relation among density functions, the density function $p(x_0|\mathbf{y^0})$ is given by (30).

$$p(x_0|\mathbf{y^0})$$
$$= \frac{\displaystyle\sum_{i=1}^{\varepsilon_l}\beta_{0i}\mathcal{N}(x_0 : \hat{x}_{0i}, P_{0i})\mathcal{N}(y_0 : C_0\hat{x}'_{0i}, C_0\acute{P}_{0i}C_0^T + R_{0i})}{\displaystyle\sum_{i=1}^{\varepsilon_l}\beta_{0i}\mathcal{N}(y_0 : C_0\hat{x}'_{0i}, C_0\acute{P}_{0i}C_0^T + R_{0i})},$$
$$= \sum_{i=1}^{\varepsilon_l}\alpha_{0i}\mathcal{N}(x_0 : \hat{x}_{0i}, P_{0i}). \tag{30}$$

For simplicity, we assume

$$\zeta_{0i} = \mathcal{N}(y_0 : C_0\hat{x}'_{0i}, C_0 P'_{0i}C_0^T + R_{0i}). \tag{31}$$

Using the Bayesian rule to inspire the strides of obtaining the Kalman filter, and after setting out the results, the proposed filter structure consists of the prediction and filtering steps as follows:

Prediction:

$$\hat{x}'_{(k+1)i} = A_k\hat{x}_{ik}$$
$$P'_{(k+1)i} = A_k P_{ki}A_k^T + Q_k$$
$$\alpha'_{(k+1)i} = \alpha_{ki}$$
$$\acute{\varepsilon}_{l(k+1)} = \varepsilon_{lk} \tag{32}$$

Filter:

$$\hat{x}_{ki} = \hat{x}'_{ki} + K_{ki}(y_k - C_k\hat{x}'_{ki})$$
$$P_{ki} = (I - K_{ki}C_k)P'_{ki}(I - K_{ki}C_k)^T + K_{ki}R_{ki}K_{ik}^T$$
$$\zeta_{ki} = \mathcal{N}(y_k : C_k\hat{x}'_{ki}, C_k P'_{ki}C_k^T + R_{ki})$$
$$\alpha_{ki} = \frac{\beta_{ki}\zeta_{ki}}{\sum_{i=1}^{\varepsilon_l}\beta_{ki}\zeta_{ki}}$$
$$\varepsilon_{lk} = \acute{\varepsilon}_{lk}. \tag{33}$$

Moreover, the filter gain can be obtained as

$$K_{ik} = P'_{ki}C_k^T\left(C_k P'_{ki}C_k^T + R_{ki}\right)^{-1}. \tag{34}$$

Now, after computing the estimation of each cluster, to obtain the unique estimation for the state, $\hat{x}_k$, we have

$$\hat{x}_k = \mathbb{E}[x_k|\mathbf{y^k}] = \int x_k \sum_{i=1}^{\varepsilon_l} \alpha_{ki} \mathcal{N}(x_k : \hat{x}_{ki}, P_{ki}) dx_k$$

$$= \sum_{i=1}^{\varepsilon_l} \alpha_{ki} \hat{x}_{ki}. \tag{35}$$

Subsequently, $\hat{x}'_{k+1}$ can be achieved as

$$\hat{x}'_{k+1} = \mathbb{E}[x_{k+1}|\mathbf{y^k}]$$

$$= \int x_{k+1} \sum_{i=1}^{\hat{\varepsilon}_{l(k+1)}} \alpha_{(k+1)i} \mathcal{N}(x_{k+1} : \hat{x}'_{(k+1)i}, P'_{(k+1)i}) dx_{k+1}$$

$$= \sum_{i=1}^{\hat{\varepsilon}_{l(k+1)}} \alpha_{(k+1)i} \hat{x}'_{(k+1)i}. \tag{36}$$

Moreover, using (35) and (36), covariances of estimation error, P_k and P'_{k+1}, are given by

$$P_k = \mathbb{E}[(x_k - \hat{x}_k)(x_k - \hat{x}_k)^T | \mathbf{y^k}] = \int (x_k - \hat{x}_k)(x_k - \hat{x}_k)^T$$

$$\sum_{i=1}^{\varepsilon_l} \alpha_{ki} \mathcal{N}(x_k : \hat{x}_{ki}, P_{ki}) dx_k$$

$$= \int (x_k - \sum_{i=1}^{\varepsilon_l} \alpha_{ki} \hat{x}_{ki})(x_k - \sum_{i=1}^{\varepsilon_l} \alpha_{ki} \hat{x}_{ki})^T \sum_{i=1}^{\varepsilon_l} \alpha_{ki}$$

$$\mathcal{N}(x_k : \hat{x}_{ki}, P_{ki}) dx_k$$

$$= \sum_{i=1}^{\varepsilon_l} \alpha_{ki} [\int x_k x_k^T \mathcal{N}(x_k : \hat{x}_{ki}, P_{ki}) dx_k - \sum_{i=1}^{\varepsilon_l} \alpha_{ki} \hat{x}_{ki}$$

$$\int x_k^T \mathcal{N}(x_k :, \hat{x}_{ki}, P_{ki}) dx_k$$

$$- \int x_k \mathcal{N}(x_k : \hat{x}_{ki}, P_{ki}) dx_k (\sum_{i=1}^{\varepsilon_l} \alpha_{ki} \hat{x}_{ki})^T + \hat{x}_k \hat{x}_k^T]$$

$$= \sum_{i=1}^{\varepsilon_l} \alpha_{ki} (P_{ki} + \hat{x}_{ki} \hat{x}_{ki}^T - \hat{x}_k \hat{x}_{ki}^T - \hat{x}_{ki} x_k^T + \hat{x}_k \hat{x}_k^T)$$

$$= \sum_{i=1}^{\varepsilon_l} \alpha_{ki} [P_{ki} + (\hat{x}_k - \hat{x}_{ki})(\hat{x}_k - \hat{x}_{ki})^T] \tag{37}$$

and

$$P'_{k+1} = \mathbb{E}[(x_{k+1} - \hat{x}'_{k+1})(x_{k+1} - \hat{x}'_{k+1})^T | \mathbf{y^k}]$$

$$= \sum_{i=1}^{\hat{\varepsilon}_{l(k+1)}} \alpha_{(k+1)i} [P'_{(k+1)i} + (\hat{x}'_{k+1} - \hat{x}'_{(k+1)i})$$

$$\times (\hat{x}'_{k+1} - \hat{x}'_{(k+1)i})^T]. \tag{38}$$

According to the main filter's relation (35), this filtering approach consists of a bank of ε_l Kalman filters. Although the prediction and filtering relations in (31), (32) and (33) are linear similar to the Kalman filter, because of using the nonlinear parameters ζ_{ki} and α_{ki} in (33) for obtaining state estimation $\hat{x}_k$ in (35), the proposed filter in this part is nonlinear.

As the last step, Algorithm 3 is presented for the proposed cluster-based robust nonlinear filter. We have a simple clus-

ter-based nonlinear filter by removing the second step in Algorithm 3 and using the conventional MLE.

Algorithm 3 Cluster-Based Robust Non-Linear Filter

Required:
Data Point $\{X_j\}_{j=1}^N \in \mathbb{R}^d$ and ε_l.
Steps:
1) Run Algorithm 1.
2) Detect outliers of each cluster using Theorem 1.
3) Calculate robust covariance matrix of each cluster.
for $k \leftarrow 1$ to N **do**
for $\ell \leftarrow 1$ to ε_l **do**
5) Compute the parameters in the prediction and filtering steps from (32) and (33).
end for
6) Obtain the state estimation by (35).
end for

IV. DISCUSSION ON THE RESULTS

In the development of filters, a bias in the estimate can arise due to heavy-tailed noises. To compensate for this bias, the effects of outliers should be reduced. Based on the above concepts, our proposed cluster-based robust filters have a better performance than the proposed cluster-based filters, due to decreasing the filter gain (34) and the matrix $\tilde{\mathcal{F}}$ in (21) which consequently improves the filter performance. However, various factors can affect the performance of moving horizon estimators such as window length, initial state, covariance values in the beginning of window, time-variant or time-invariant characteristics of the systems' dynamic, etc. In light of the points mentioned above, we cannot explicitly claim which of the proposed filters is the best.

The impact of removing outliers (from a cluster or data set) on statistical covariance can vary depending on the distribution of the remaining data. The covariance decreases significantly when the outliers are large and far from the other data. These outliers can lead to significant variability and distortion in the covariance estimate. On the other hand, outliers consistent with the general trend of the data have different impacts on covariance if removed. It should be noted that the number of data in each cluster can also affect this phenomenon.

The number of measurement data in this process depends on the system complexity, dispersion, and constraints. Although clustering is possible for small number of data, the accuracy of covariance estimation and consequently, the precision of filtering decreases. Besides, highly dispersed data set are unreliable for clustering, while excessive data leads to a single ellipsoidal cluster due to the central limit theorem. Therefore, we suggest to check the data scatter and collect the measurement data such that the dispersion and empty positions are minimized in the whole data set.

In the proposed methods, different parameters appear in the optimization problem, filtering, and clustering. The parameters in the first and second groups are computed precisely. However, in the weighted SON clustering problem, the weights ζ_{ij} are obtained using the similarity functions. In Remark 3, the Gaussian kernel has been introduced as one of the most famous functions. Also, p is another parameter and

determines which norm is used in the clustering. We know that different norms emphasize different aspects of the data, leading to variations in the clustering outcome. The proposed approach optimally uses the Euclidean norm ($p = 2$) to capture the overall distance between data points without any bias or focusing on specific characteristic of data.

The proposed filtering algorithms are offline; therefore, they do not depend on the time properties of dynamic systems, i.e., our proposed data-based filters can be applied to both time-variant and time-invariant systems.

V. SIMULATION STUDY

In this section, we want to study the performance and effectiveness of the proposed filters, including cluster-based MHE (C-B MHE), cluster-based robust MHE (C-B RMHE), cluster-based non-linear filter (C-B NF), and cluster-based robust non-linear filter (C-B RNF). For this purpose, we exert these filters on the time-varying and practical system subject to heavy-tailed noises. Since no completely related filters exist in the literature, we compare our proposed filters with those designed for non-Gaussian systems with known parameters.

A. Example 1: Three-Tank System

We consider a three-tank system with the schematic shown in Fig. 1. The dynamics of this system can be described as follows:

$$\frac{dh_1}{dt} = \frac{1}{A_b}(Q_1 - \alpha_{13}A_c\sqrt{2g(h_1 - h_3)} - q_{L1})$$

$$\frac{dh_2}{dt} = \frac{1}{A_b}(Q_2 + \alpha_{32}A_c\sqrt{2g(h_3 - h_2)} - \alpha_{2O}A_c\sqrt{2gh_2} - q_{L2})$$

$$\frac{dh_3}{dt} = \frac{1}{A_b}(\alpha_{13}A_c\sqrt{2g(h_1 - h_3)} - \alpha_{32}A_c\sqrt{2g(h_3 - h_2)} - q_{L3})$$

(39)

where Q_1 and Q_2 are the flow rates of pumps 1 and 2, q_{Li} is the leakage flow rate of the tank i ($i = 1, 2, 3$), h_i is the level of the tank i, A_c denotes the cross-sectional area of the connecting pipe, and $q_{mn}(m \neq n)$ is the flow rate from tank m to tank n. Parameters of the system and their numerical values are tabulated in Table I.
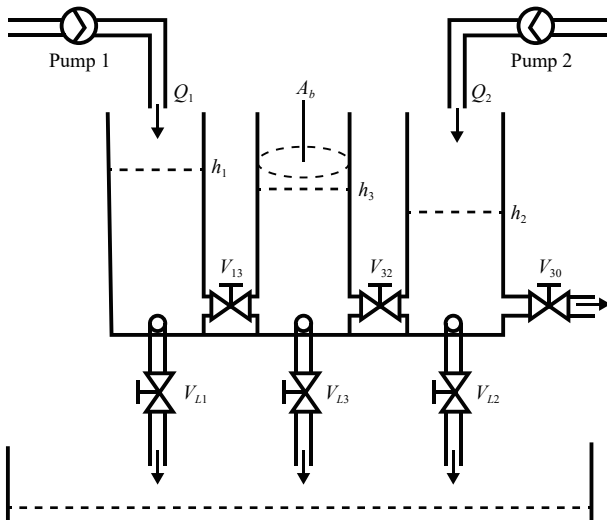

Fig. 1.　Three-tank system.

TABLE I
THREE-TANK SYSTEM PARAMETERS

Descriptions and parameters	Numerical value
Cross-sectional area of the tank (A_b)	0.0154 m^2
Cross-sectional area of the pipes (A_c)	5×10^{-5} m^2
Valve opening position (α_{ij})	0.84

Using the system parameters in Table I with operating points ($Q_1 = 5.5 \times 10^{-5}, Q_2 = 3.4 \times 10^{-5}$ m^3/s), and ($h_1 = 0.4$, $h_2 = 0.23$, $h_3 = 0.31$ m), and after linearization and discretization of the model (39), we reach the state space model of the system with the following matrices:

$$A = \begin{bmatrix} 0.9799 & 2.04 \times 10^{-4} & 0.0199 \\ 2.04 \times 10^{-4} & 0.9676 & 0.0198 \\ 0.0199 & 0.0198 & 0.9602 \end{bmatrix}$$

$$B = \begin{bmatrix} 64.279 & 0.0044 \\ 0.0044 & 63.783 \\ 0.6515 & 0.6506 \end{bmatrix}$$

$$C = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{40}$$

Process and measurement noises are non-Gaussian with the following distributions:

$$v_{i,k} = 0.9\mathcal{N}(0, Q) + 0.1\mathcal{N}(0, 10Q), \quad i = 1, 2, 3$$

$$w_{j,k} = 0.9\mathcal{N}(0, R) + 0.1\mathcal{N}(0, 1000R), \quad j = 1, 2, 3 \tag{41}$$

where $Q = R = 0.035$.

The performance of the proposed filters is compared to the maximum correntropy Kalman filter (MCKF) [13] as a famous filter in the presence of heavy-tailed noises, the conventional MHE with window length $N_w = 10$, and the conventional Kalman filter. Since the dimension of the matrix $\tilde{O}$ in (21) is extensive, to have more analysis, we introduce cluster-based and cluster-based robust Kalman filters using the conventional Kalman filter with estimated covariance matrix using MLE and weighted MLE, respectively. Fig. 2 shows different filters' state estimation and MSE. Our proposed data-based filters perform better than the Kalman filter, the MCKF, and the conventional MHE designed with known and true parameters. The main reason for this deficiency is that the Kalman filter is a minimum mean square filter and is sensitive to non-Gaussian noises; therefore, its performance degrades against these noises. Also, as a prior filter against non-Gaussian noises, the maximum correntropy Kalman filter cannot guarantee a compelling performance against these noises. Moreover, although the conventional MHE uses data in a window to obtain estimations and has a weak performance against non-Gaussian noises, it outperforms the conventional Kalman filter and MCKF. This fact brings about the proposed cluster-based MHE and the proposed cluster-based robust MHE has a better response than the cluster-based Kalman filter and cluster-based robust Kalman filter. In contrast, data clustering, outlier detection, and data-based filter-
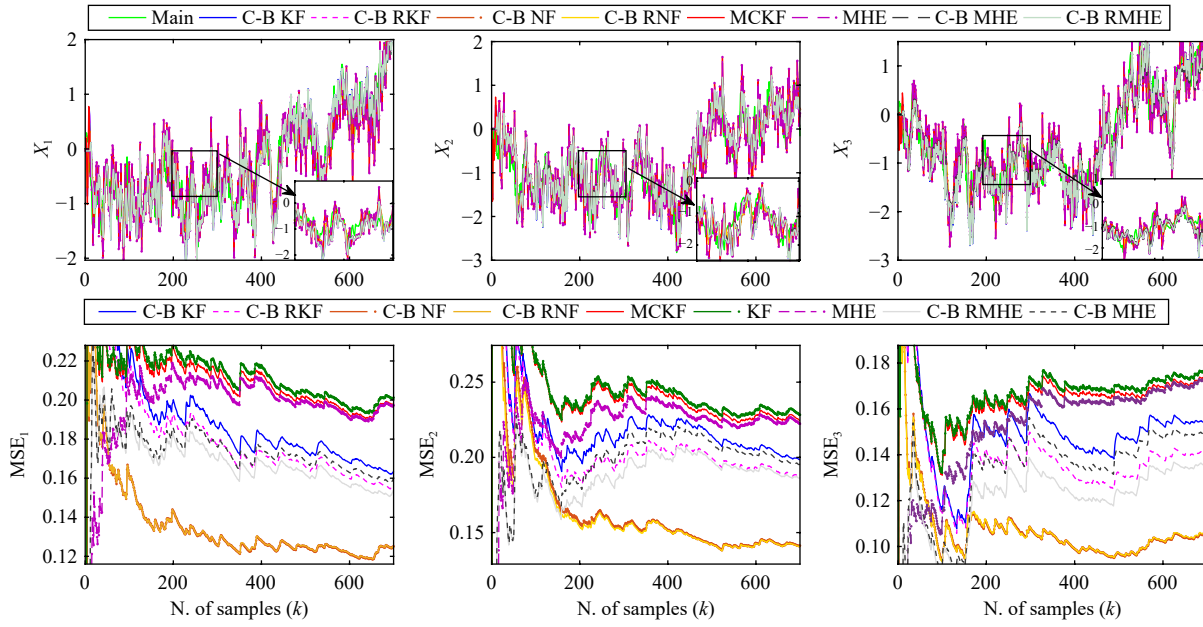
Fig. 2. State estimations and MSEs in Example 1 (C-B, R, and N are abbreviation of cluster-based, robust, and non-linear).
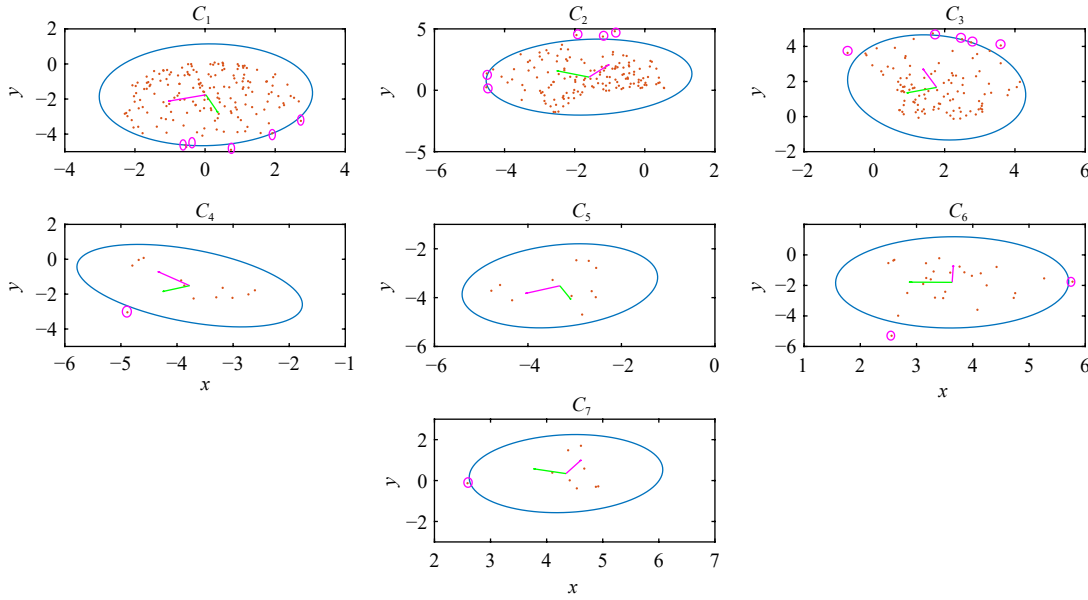

Fig. 3. Error ellipse (Iso-Contour) of the each cluster with outliers in Example 2.

ing equip the proposed approaches to perform well against non-Gaussian noises.

B. Example 2: Time-Variant System

To show the usefulness of our proposed filters for time-variant systems, we assume that output data set of the following non-Gaussian time-variant system (taken from [40] with some modification) are available:

$$x_{k+1} = \begin{bmatrix} -0.9 & 0 & -0.3+0.2\sin(0.3k) \\ 0 & 0.7 & 0.2 \\ -0.5 & 0 & 0.7+0.1\cos(0.1k) \end{bmatrix} x_k + v_k$$

$$y_k = \begin{bmatrix} 0.5+0.3\sin(0.1) & 1 & 0 \\ 0.5 & 1+0.1\cos(0.3) & 0.1 \end{bmatrix} x_k + w_k. \quad (42)$$

Process and measurement noises are non-Gaussian with the

following distributions:

$$v_{i,k} = 0.9\mathcal{N}(0,Q)+0.1\mathcal{N}(0,10Q), \quad i=1,2,3$$
$$w_{j,k} = 0.9\mathcal{N}(0,R)+0.1\mathcal{N}(0,700R), \quad j=1,2 \quad (43)$$

where $Q = R = 0.15$.

We have six ellipsoidal clusters after running Algorithm 1. Fig. 3 shows the iso-contour of these six ellipsoidal clusters along with the outliers of each cluster. Clusters' shapes without outliers emphasize the proposed optimization's proper performance. It is worth noting that if the axes of ellipsoidal are parallel to coordinate axes, the obtained covariances are diagonal. Moreover, the results of mean square error of states for the proposed data-based filters are compared in Table II. Likewise, our proposed filters have better performance than the others.

TABLE II
MEAN SQUARE ERROR OF DIFFERENT FILTERS FOR NOISES IN (43)

Filter	MSE of x_1	MSE of x_2	MSE of x_3
C-B RMHE (21)	**0.1147**	**0.0581**	**0.0889**
C-B MHE (21)	0.1177	0.0637	0.9022
C-B RMHE (25)	0.2003	0.0760	0.1367
C-B MHE (25)	0.2024	0.0832	0.1385
C-B RNF	0.1826	0.07891	0.1047
C-B NF	0.1901	0.08517	0.1066

Based on the results in Table II, the robust moving horizon estimation with a whole window performs better than other methods. This phenomenon is because MHE with a large window length in slow varying dynamic systems can capture more relevant information for estimation, and consequently, has a more reliable performance.

VI. CONCLUSION

In this paper, we aimed to develop data-based filters for linear dynamic systems against non-Gaussian heavy-tailed noises under the unknown output noise covariance condition. Inspiring that Gaussian data sets have ellipsoidal scattering, we clustered the output data set of a non-Gaussian dynamic system using the proposed weighted SON clustering method. The proposed approach can improve the performance of the SON clustering method by focusing on closely spaced points and reducing the computational cost. Outliers in each cluster can change the clusters' ellipsoidal shape and affect the filter's performance. To address this issue, we proposed an SDP problem based on the weighted maximum likelihood to detect outliers in each cluster and obtain a robust covariance matrix. We then developed four different filters using two approaches. In the first method, we presented the cluster-based MHE. Provided that output noise covariance is computed by the conventional MLE and the proposed SDP problem, we have a simple cluster-based filter and a robust cluster-based filter, respectively. In the second approach, given ε_l clusters, we assumed that there are ε_l Gaussian resources for the output noise distribution with their specific statistician features. This idea led us to extract a non-linear filter structure, presenting a cluster-based non-linear filter and a cluster-based robust non-linear filter, depending on how the covariance matrix is computed. Finally, we verified the performance of our proposed filters through simulation results on a practical system and a time-variant system.

REFERENCES

[1] J. Han and M. Kamber, *Data Mining: Concepts and Techniques*. Morgan Kaufmann, 2000.

[2] R. M. Murray, K. J. Astrom, S. P. Boyd, R. W. Brockett, and G. Stein, "Future directions in control in an information-rich world," *IEEE Control Systems Magazine*, vol. 23, no. 2, pp. 20–33, 2003.

[3] O. Nelles, *Nonlinear System Identification: From Classical Approaches to Neural Networks, Fuzzy Models, and Gaussian Processes*. Springer Nature, 2020.

[4] S. A. Ayankoso and M. K. Habib, "Data-driven modeling of a two-link flexible manipulator (TLFM)," in *Proc. IEEE 21st Int. Conf. Research and Education in Mechatronics*, 2020, pp. 1–6.

[5] P. M. Van den Hof, "System identification-data-driven modelling of dynamic systems," *Eindhoven University of Technology*, 2012.

[6] A. Xi, T. W. Mudiyanselage, D. Tao, and C. Chen, "Balance control of a biped robot on a rotating platform based on efficient reinforcement learning," *IEEE/CAA J. Autom. Sinica*, vol. 6, no. 4, pp. 938–951, 2019.

[7] P. D. Ngo, M. Tejedor, and F. Godtliebsen, "Data-driven robust control using reinforcement learning," *Applied Sciences*, vol. 12, no. 4, p. 2262, 2022.

[8] X. Wang, J. Sun, G. Wang, F. Allgöwer, and J. Chen, "Data-driven control of distributed event-triggered network systems," *IEEE/CAA J. Autom. Sinica*, vol. 10, no. 2, pp. 351–364, 2023.

[9] Y. Yang, Z. Ding, R. Wang, H. Modares, and D. C. Wunsch, "Datadriven human-robot interaction without velocity measurement using offpolicy reinforcement learning," *IEEE/CAA J. Autom. Sinica*, vol. 9, no. 1, pp. 47–63, 2021.

[10] Z. Chen, *et al.*, "Bayesian filtering: From Kalman filters to particle filters, beyond," *Statistics*, vol. 182, no. 1, pp. 1–69, 2003.

[11] T. Ortmaier, M. Groger, D. H. Boehm, V. Falk, and G. Hirzinger, "Motion estimation in beating heart surgery," *IEEE Trans. Biomedical Engineering*, vol. 52, no. 10, pp. 1729–1740, 2005.

[12] R. K. Pearson, "Outliers in process modeling and identification," *IEEE Trans. Control Systems Technology*, vol. 10, no. 1, pp. 55–63, 2002.

[13] B. Chen, X. Liu, H. Zhao, and J. C. Principe, "Maximum correntropy Kalman filter," *Automatica*, vol. 76, pp. 70–77, 2017.

[14] E. Javanfar and M. Rahmani, "Optimal $\mathcal{L}_p$ filtering for discrete-time non-Gaussian dynamic systems," *IET Control Theory & Applications*, vol. 15, no. 16, pp. 2111–2127, 2021.

[15] Y. Huang, Y. Zhang, Y. Zhao, P. Shi, and J. A. Chambers, "A novel outlier-robust Kalman filtering framework based on statistical similarity measure," *IEEE Trans. Automatic Control*, vol. 66, no. 6, pp. 2677–2692, 2020.

[16] T. Zhang, S. Zhao, X. Luan, and F. Liu, "Bayesian inference for statespace models with student-t mixture distributions," *IEEE Trans. Cybernetics*, 2022.

[17] B. Shen, X. Wang, and L. Zou, "Maximum correntropy Kalman filtering for non-Gaussian systems with state saturations and stochastic nonlinearities," *IEEE/CAA J. Autom. Sinica*, vol. 10, no. 5, pp. 1223–1233, 2023.

[18] L. Zou, Z. Wang, H. Geng, and X. Liu, "Set-membership filtering subject to impulsive measurement outliers: A recursive algorithm," *IEEE/CAA J. Autom. Sinica*, vol. 8, no. 2, pp. 377–388, 2021.

[19] E. Javanfar, M. Rahmani, and B. Moaveni, "Measurement-outlier robust Kalman filter for discrete-time dynamic systems," *ISA Transactions*, vol. 134, pp. 256–267, 2023.

[20] M. Milanese, F. Ruiz, and M. Taragna, "Direct data-driven filter design for uncertain LTI systems with bounded noise," *Automatica*, vol. 46, no. 11, pp. 1773–1784, 2010.

[21] Y. Xia, L. Dai, W. Xie, and Y. Gao, "Data-driven filter design for linear systems with quantized measurements," *IFAC-PapersOnLine*, vol. 48, no. 28, pp. 697–702, 2015.

[22] M. S. Turan and G. Ferrari-Trecate, "Data-driven unknown-input observers and state estimation," *IEEE Control Systems Letters*, vol. 6, pp. 1424–1429, 2021.

[23] J. Duník, O. Straka, and M. Šimandl, "On autocovariance least-squares method for noise covariance matrices estimation," *IEEE Trans. Automatic Control*, vol. 62, no. 2, pp. 967–972, 2016.

[24] T. B. Schön, A. Wills, and B. Ninness, "System identification of nonlinear state-space models," *Automatica*, vol. 47, no. 1, pp. 39–49, 2011.

[25] P. Matisko and V. Havlena, "Noise covariances estimation for Kalman filter tuning," *IFAC Proceedings Volumes*, vol. 43, no. 10, pp. 31–36, 2010.

[26] C. Magnant, A. Giremus, E. Grivel, L. Ratton, and B. Joseph, "Bayesian non-parametric methods for dynamic state-noise covariance matrix estimation: Application to target tracking," *Signal Processing*, vol. 127, pp. 135–150, 2016.

[27] O. Kost, J. Duník, and O. Straka, "Noise moment and parameter estimation of state-space model," *IFAC-PapersOnLine*, vol. 51, no. 15, pp. 891–896, 2018.

[28] Y. Huang, Y. Zhang, P. Shi, Z. Wu, J. Qian, and J. A. Chambers, "Robust Kalman filters based on Gaussian scale mixture distributions with application to target tracking," *IEEE Trans. Systems*, *Man*, *Cybernetics: Systems*, vol. 49, no. 10, pp. 2082–2096, 2017.

[29] H. Zhu, G. Zhang, Y. Li, and H. Leung, "A novel robust Kalman filter with unknown non-stationary heavy-tailed noise," *Automatica*, vol. 127, p. 109511, 2021.

[30] R. Xu and D. Wunsch, "Survey of clustering algorithms," *IEEE Trans. Neural Networks*, vol. 16, no. 3, pp. 645–678, 2005.

[31] P. Wittek, *Quantum Machine Learning: What Quantum Computing Means to Data Mining*. Academic Press, 2014.

[32] G. J. McLachlan and T. Krishnan, *The EM Algorithm and Extensions*. John Wiley & Sons, 2007.

[33] G. J. McLachlan and D. Peel, *Finite Mixture Models*. John Wiley & Sons, 2004.

[34] S. S. Khan and A. Ahmad, "Cluster center initialization algorithm for K-means clustering," *Pattern Recognition Letters*, vol. 25, no. 11, pp. 1293–1302, 2004.

[35] D. Arthur and S. Vassilvitskii, "K-Means++: The advantages of careful seeding," in *Proc. Symp Discrete Algorithms*, 2007, pp. 1027–1035.

[36] F. Lindsten, H. Ohlsson, and L. Ljung, "Clustering using sum-of-norms regularization: With application to particle filter output computation," in *Proc. IEEE Statistical Signal Processing Workshop*, 2011, pp. 201–204.

[37] F. Lindsten, H. Ohlsson, and L. Ljung, "Just relax and come clustering! A convexification of K-means clustering," Linkping University, The Institute of Technology, Tech. Rep., 2011.

[38] K. Villaverde, O. Kosheleva, and M. Ceberio, "Why ellipsoid constraints, ellipsoid clusters, and Riemannian space-time: Dvoretzkys Theorem revisited," in *Constraint Programming and Decision Making*.

Springer Int. Publishing, 2014, pp. 203–207.

[39] P. J. Green, "Iteratively reweighted least squares for maximum likelihood estimation, some robust and resistant alternatives," *Journal of the Royal Statistical Society: Series B* (*Methodological*), vol. 46, no. 2, pp. 149–170, 1984.

[40] M. Abolhasani and M. Rahmani, "Robust Kalman filtering for discretetime time-varying systems with stochastic and norm-bounded uncertainties," *Journal of Dynamic Systems*, *Measurement*, *Control*, vol. 140, no. 3, p. 030901, 2018.

Elham Javanfar received the Ph.D. degree in electrical engineering from Imam Khomeini Int. University, Iran, in 2022. She is currently working as a Research Assistant at the same university. Her primary research interests include estimation and filter design. She is also interested in the design of robust and nonlinear controllers, reinforcement learning, and adaptive dynamic programming.

Mehdi Rahmani (Member, IEEE) received the B.Sc., M.Sc., and Ph.D. degrees in electrical engineering from Sharif University of Technology, Iran. In 2013, he joined the Electrical Engineering Department of Imam Khomeini Int. University (IKIU), Iran, where he was the Head of the Control Engineering Department from 2016–2019. His research interests are estimation: filtering and prediction, data-driven control and estimation, nonlinear systems, distributed systems, optimization, and control theory.