

Document Version

Final published version

Citation (APA)

González, R. A., Cedeño, A. L., Tiels, K., & Oomen, T. A. E. (2025). Truncated Gaussian Noise Estimation in State-Space Models. In *Proceedings of the IEEE 64th Conference on Decision and Control (CDC 2025)* (pp. 482-487). (Proceedings of the IEEE Conference on Decision and Control). IEEE.
<https://doi.org/10.1109/CDC57313.2025.11312032>

Important note

To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright

In case the licence states "Dutch Copyright Act (Article 25fa)", this publication was made available Green Open Access via the TU Delft Institutional Repository pursuant to Dutch Copyright Act (Article 25fa, the Taverne amendment). This provision does not affect copyright ownership.
Unless copyright is transferred by contract or statute, it remains with the copyright holder.

Sharing and reuse

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

**Green Open Access added to [TU Delft Institutional Repository](#)
as part of the Taverne amendment.**

More information about this copyright law amendment
can be found at <https://www.openaccess.nl>.

Otherwise as indicated in the copyright section:
the publisher is the copyright holder of this work and the
author uses the Dutch legislation to make this work public.

Truncated Gaussian Noise Estimation in State-Space Models

Rodrigo A. González, Angel L. Cedeño, Koen Tiels and Tom Oomen

Abstract—Within Bayesian state estimation, considerable effort has been devoted to incorporating constraints into state estimation for process optimization, state monitoring, fault detection and control. Nonetheless, in the domain of state-space system identification, the prevalent practice entails constructing models under Gaussian noise assumptions, which can lead to inaccuracies when the noise follows bounded distributions. With the aim of generalizing the Gaussian noise assumption to potentially truncated densities, this paper introduces a method for estimating the noise parameters in a state-space model subject to truncated Gaussian noise. Our proposed data-driven approach is rooted in maximum likelihood principles combined with the Expectation-Maximization algorithm. The efficacy of the proposed approach is supported by a simulation example.

Index Terms—Noise estimation; Truncated Gaussian distribution; EM Algorithm; Nonlinear System Identification.

I. INTRODUCTION

The goal of system identification is to derive a mathematical model of a dynamical system using input-output data [1]. A critical aspect of this process involves quantifying the uncertainty associated with the estimated model, as it directly impacts analysis, prediction, and robust control performance. These uncertainties may arise from structural simplifications in the model formulation, inaccuracies in parameter estimation, and measurement noise present during system identification experiments [2], [3]. Consequently, accurately modeling noise and additive uncertainties remains a fundamental challenge, significantly influencing both model precision and robustness. Traditionally, two dominant paradigms have been employed for uncertainty modeling: (1) the probabilistic approach, which characterizes noise using Gaussian probability density functions (PDFs), and (2) the set-membership approach, a distribution-free method that assumes noise to be deterministically bounded [4]. While the probabilistic framework provides a statistical interpretation of uncertainties, it often fails to account for hard constraints on noise magnitudes. Conversely, the set-membership approach incorporates prior knowledge in the form of amplitude bounds but does not directly consider statistical insights that could enhance model accuracy.

Incorporating signal constraints, such as bounded disturbances, into the estimation process enhances the accuracy and reliability of system state estimates [5], thereby supporting better decision-making in areas such as control, system

identification, signal processing, and navigation. In system identification specifically, incorporating signal constraints strengthens the physical interpretability of solutions and facilitates the integration of prior knowledge about unknown parameters, leading to improved generalization performance.

Constraints on system noise have been incorporated into system identification through the set-membership approach [6], which assumes amplitude-bounded noise. This framework has led to the development of various distribution-free model estimators [7]–[9]. However, existing efforts have primarily focused on system estimators, with limited attention given to parametric characterizations of bounded noise in state-space representations. Conversely, state-space modeling has predominantly relied on Gaussian assumptions for state and output noise [10]–[13]. Since the Gaussian PDF has unbounded support, these approaches inherently disregard noise amplitude constraints, which may compromise the accuracy of uncertainty characterization. Other methods such as [14] explore correlation methods for noise identification, with focus on noise moment estimation in linear systems.

A promising intermediate solution bridges the set-membership and Gaussian paradigms by employing truncated Gaussian PDFs. This approach enables more realistic noise modeling while retaining the mathematical tractability and interpretability of Gaussian distributions. Applications span diverse fields: autonomous robotics, where sensor noise is bounded but non-uniform [15]; industrial control, where disturbances have operational limits [16]; and biomedical systems such as glucose monitoring, where sensor errors must be constrained [17]. Truncated statistics have also been studied in regression and classification problems [18], probability distribution fitting [19], and have been used as prior distributions for non-parametric system identification [20].

To bridge the gap between the probabilistic and set-membership approaches in noise estimation, this paper introduces a hybrid modeling strategy that integrates the strengths of both methods by employing truncated Gaussian distributions to parametrize the noise statistics in state-space models. This formulation allows for probabilistic reasoning while respecting known physical bounds on disturbances, and can be particularly important for systems where noise exhibits both a probabilistic structure and hard physical limits.

In summary, the main contributions of this work are:

- C1 We propose an iterative procedure, based on the EM algorithm [21] to compute the noise parameters of a state-space model subject to truncated Gaussian noise. In particular, our approach enables the estimation of the mean and covariance of the Gaussian prior to truncation, as well as the upper and lower truncation

This work was supported in part by the grant ANID-Fondecyt 3240181, by the ANID-Basal Project AFB240002 (AC3E), and by E2TECH, University of Santiago of Chile. R. A. González, K. Tiels and T. Oomen are with the Mechanical Engineering Department of Eindhoven University of Technology, Eindhoven, The Netherlands. A. L. Cedeño is with the University of Santiago of Chile (USACH), Faculty of Engineering, Electrical Engineering Department, Chile. T. Oomen is also with Delft Center for Systems and Control, Delft University of Technology, The Netherlands.

limits. Our method requires particle smoothing [22] to match the sample and population moments of the truncated Gaussian noise at each EM iteration.

C2 We show the effectiveness of the proposed approach through extensive Monte Carlo runs performed on a simulation example.

The paper is structured as follows. In Section II we present the problem setup, and in Section III we revisit the maximum likelihood and EM iterations. The proposed identification method under truncated Gaussian noise PDFs is derived in Section IV, while Section V discusses implementation aspects. A simulation example can be found in Section VI, and Section VII provides concluding remarks.

Notation: All vectors and matrices are written in bold, and all vectors are column vectors, unless transposed. The set of vectors $\{\mathbf{a}_i\}_{i=1}^N$ is denoted as $\mathbf{a}_{1:N}$. If \mathbf{a} and $\mathbf{b} \in \mathbb{R}^n$ with elements a_i and b_i respectively, then the notation $\mathbf{a} \leq \mathbf{b}$ corresponds to the element-wise inequality $a_i \leq b_i$ for all $i \in \{1, \dots, n\}$. Depending on the context, we also refer to these elements as $\{\mathbf{a}\}_i$ and $\{\mathbf{b}\}_i$. The support of the function $f(\cdot)$ is denoted as $\text{supp}(f(\cdot))$, and $\mathbf{1}_{[\mathbf{a}, \mathbf{b}]}(\mathbf{x})$ is the indicator function of the hyperrectangle in \mathbb{R}^n with two vertices \mathbf{a} and \mathbf{b} , i.e., $\mathbf{1}_{[\mathbf{a}, \mathbf{b}]}(\mathbf{x}) = 1$ if $\mathbf{a} \leq \mathbf{x} \leq \mathbf{b}$, and zero otherwise.

II. SYSTEM SETUP AND PROBLEM FORMULATION

Consider the following discrete-time state-space system

$$\mathbf{x}_{t+1} = \mathbf{f}_t(\mathbf{x}_t, \mathbf{u}_t) + \mathbf{w}_t \quad (1a)$$

$$\mathbf{y}_t = \mathbf{g}_t(\mathbf{x}_t, \mathbf{u}_t) + \mathbf{v}_t, \quad (1b)$$

where \mathbf{x}_t , \mathbf{u}_t , and \mathbf{y}_t denote the state, input, and output, respectively, with dimensions $\mathbf{x}_t \in \mathbb{R}^n$, $\mathbf{u}_t \in \mathbb{R}^m$, and $\mathbf{y}_t \in \mathbb{R}^p$. The functions $\mathbf{f}_t(\cdot)$ and $\mathbf{g}_t(\cdot)$ are assumed known and map to the appropriate dimensions. These functions may be given by a white-box physical model [23], or may also be available from a previous system identification step. For simplicity, we assume the initial state \mathbf{x}_1 is known, which is reasonable when the system starts from rest or when transient data are removed. Otherwise, it can be estimated from data using an extended approach beyond the scope of this paper.

We may also write the system in (1) in the compact form

$$\boldsymbol{\xi}_t = \mathbf{h}_t(\boldsymbol{\zeta}_t) + \boldsymbol{\eta}_t,$$

with

$$\boldsymbol{\xi}_t = \begin{bmatrix} \mathbf{x}_{t+1} \\ \mathbf{y}_t \end{bmatrix}, \quad \boldsymbol{\zeta}_t = \begin{bmatrix} \mathbf{x}_t \\ \mathbf{u}_t \end{bmatrix}, \quad \mathbf{h}_t(\cdot) = \begin{bmatrix} \mathbf{f}_t(\cdot) \\ \mathbf{g}_t(\cdot) \end{bmatrix}, \quad \boldsymbol{\eta}_t = \begin{bmatrix} \mathbf{w}_t \\ \mathbf{v}_t \end{bmatrix}.$$

The process and measurement noise, which are arranged in the vector $\boldsymbol{\eta}_t$, are assumed to be truncated normal distributed and uncorrelated with \mathbf{x}_1 . The noise vector is described by

$$\boldsymbol{\eta}_t \sim \mathcal{TN}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{a}, \mathbf{b}), \quad (2)$$

see Fig. 1 for a graphical interpretation. The probability density function (PDF) of $\boldsymbol{\eta}_t$ is given by

$$p(\boldsymbol{\eta}_t) = \frac{\exp\{-\frac{1}{2}(\boldsymbol{\eta}_t - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{\eta}_t - \boldsymbol{\mu})\}}{\int_{\mathbf{a}-\boldsymbol{\mu}}^{\mathbf{b}-\boldsymbol{\mu}} \exp\{-\frac{1}{2}\mathbf{p}^\top \boldsymbol{\Sigma}^{-1}\mathbf{p}\} d\mathbf{p}} \mathbf{1}_{[\mathbf{a}, \mathbf{b}]}(\boldsymbol{\eta}_t). \quad (3)$$

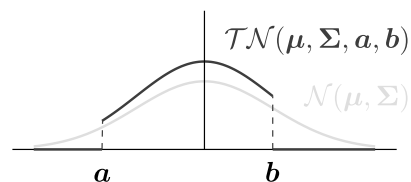


Fig. 1. Untruncated Gaussian (gray), and its truncated version (black), with lower and upper truncation bounds \mathbf{a} and \mathbf{b} , respectively.

The quantities $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ correspond to the mean and covariance of the unconstrained normal distribution, and the lower and upper bounds of the truncated normal distribution are defined as \mathbf{a} and \mathbf{b} , respectively. Note that the PDF describing $\boldsymbol{\eta}_t$ differs from that of a censored Gaussian, where the probabilities of the events $\boldsymbol{\eta}_t = \mathbf{a}$ and $\boldsymbol{\eta}_t = \mathbf{b}$ are strictly positive for finite-valued \mathbf{a} and \mathbf{b} .

The problem of interest is stated next.

Problem 2.1: Given N input-output samples $\{\mathbf{u}_t, \mathbf{y}_t\}_{t=1}^N$, and assuming that \mathbf{f}_t and \mathbf{g}_t in (1) are known or previously estimated, our goal is to obtain estimates for the parameters of the truncated normal noise distribution, including the truncation limits. The parameters of interest are arranged in a vector $\boldsymbol{\beta}$ of the form

$$\boldsymbol{\beta} = [\boldsymbol{\mu}^\top, \text{vec}\{\boldsymbol{\Sigma}\}^\top, \mathbf{a}^\top, \mathbf{b}^\top]^\top. \quad (4)$$

The next section introduces the statistical tools that are used for solving this problem.

III. MAXIMUM LIKELIHOOD AND THE EM ALGORITHM

It is widely known that the maximum likelihood (ML) method provides consistent and asymptotically efficient estimators in many practical settings [24]. Such method seeks the parameter vector that maximizes the likelihood that the output data takes the observed values. More formally, the ML estimator is obtained as

$$\hat{\boldsymbol{\beta}}_{\text{ML}} = \arg \max_{\boldsymbol{\beta}} \ell(\boldsymbol{\beta}), \quad (5)$$

where $\ell(\boldsymbol{\beta}) = \log p(\mathbf{y}_{1:N} | \boldsymbol{\beta})$. This optimization problem is challenging to solve even in the linear system and unconstrained Gaussian case, since it involves a non-convex cost function that requires robust implementations to be adopted for obtaining the global maximum [10]. One way to address this problem is to apply the Expectation-Maximization (EM) algorithm [21], which provides a sequence of estimates $\hat{\boldsymbol{\beta}}^{(k)}$, $k = 1, 2, \dots$, that monotonically increase the likelihood, converging under general conditions to a local or global maximum of $\ell(\boldsymbol{\beta})$ if properly initialized [25].

The EM algorithm requires defining latent variables, which, if known, could ease the optimization of the likelihood function. When studying state-space systems, it is natural to take the state sequence $\mathbf{x}_{1:N+1}$ as the latent variables [11], [26]. Thus, we consider the joint PDF

$$\begin{aligned} \mathcal{L}(\boldsymbol{\beta}) &= p(\mathbf{x}_{1:N+1}, \mathbf{y}_{1:N} | \boldsymbol{\beta}) \\ &= p(\mathbf{x}_1) \prod_{t=1}^N p(\mathbf{x}_{t+1}, \mathbf{y}_t | \mathbf{x}_t, \boldsymbol{\beta}), \end{aligned} \quad (6)$$

where the Markov property of the state-space formulation has been exploited. The EM algorithm is described in Algorithm 1. In practice, the EM algorithm terminates when a maximum number of iterations has been reached, or when the relative error of the estimates is below a predefined tolerance factor.

Algorithm 1 EM algorithm for the computation of $\hat{\beta}_{\text{ML}}$

- 1: Choose an initial estimate $\hat{\beta}^{(1)}$
- 2: **for** $k = 1, 2, \dots$ **do**
- 3: **E-step:** Compute the expectation

$$\mathcal{Q}(\beta, \hat{\beta}^{(k)}) = \mathbb{E} \left\{ \log \mathcal{L}(\beta) | \mathbf{y}_{1:N}, \hat{\beta}^{(k)} \right\}, \quad (7)$$

where $\mathcal{L}(\beta)$ is given by (6).

- 4: **M-step:** Solve the optimization problem

$$\hat{\beta}^{(k+1)} = \arg \max_{\beta} \mathcal{Q}(\beta, \hat{\beta}^{(k)}). \quad (8)$$

- 5: **end for**
-

IV. PARAMETER ESTIMATION OF TRUNCATED GAUSSIAN DISTURBANCES

In this section, we propose the algorithm to compute the maximum likelihood estimator for β , the parameters associated with the truncated Gaussian noise distribution. To this end, we address the computation of the \mathcal{Q} function that aims at solving Problem 2.1, and we provide a characterization of its maximizer for each iteration of the EM algorithm.

For a known \mathbf{x}_1 , the \mathcal{Q} function is proportional to

$$\mathcal{Q}(\beta, \hat{\beta}^{(k)}) \propto \sum_{t=1}^N \mathbb{E} \left\{ \log p(\mathbf{x}_{t+1}, \mathbf{y}_t | \mathbf{x}_t, \beta) | \mathbf{y}_{1:N}, \hat{\beta}^{(k)} \right\}, \quad (9)$$

where

$$p(\mathbf{x}_{t+1}, \mathbf{y}_t | \mathbf{x}_t, \beta) = \mathbf{1}_{[a,b]}(\boldsymbol{\xi}_t - \mathbf{h}_t(\zeta_t)) \times \frac{\exp\left\{-\frac{1}{2}(\boldsymbol{\xi}_t - \mathbf{h}_t(\zeta_t) - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{\xi}_t - \mathbf{h}_t(\zeta_t) - \boldsymbol{\mu})\right\}}{\int_{a-\boldsymbol{\mu}}^{b-\boldsymbol{\mu}} \exp\left\{-\frac{1}{2}\mathbf{p}^\top \boldsymbol{\Sigma}^{-1}\mathbf{p}\right\} d\mathbf{p}}. \quad (10)$$

Lemma 4.1 reformulates the optimization problem solved in each EM iteration to estimate the parameters $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, \mathbf{a} , and \mathbf{b} .

Lemma 4.1: Define $\mathcal{A}_t := \text{supp}(p(\mathbf{x}_t, \mathbf{x}_{t+1} | \mathbf{y}_{1:N}, \hat{\beta}^{(k)}))$. The next EM iteration of β given $\hat{\beta}^{(k)}$ is given by the solution of the following optimization problem

$$\hat{\beta}^{(k+1)} = \arg \min_{\beta} \mathcal{V}^{(k)}(\beta) \quad (11a)$$

$$\text{s.t. } \mathbf{a} \leq \boldsymbol{\xi}_t - \mathbf{h}_t(\zeta_t) \leq \mathbf{b} \text{ for all } (\mathbf{x}_t, \mathbf{x}_{t+1}) \in \mathcal{A}_t, t = 1, \dots, N, \quad (11b)$$

where

$$\mathcal{V}^{(k)}(\beta) = \frac{1}{2} \text{tr} \left\{ \boldsymbol{\Sigma}^{-1} (\boldsymbol{\Phi} - \boldsymbol{\Psi} \boldsymbol{\mu}^\top - \boldsymbol{\mu} \boldsymbol{\Psi}^\top + \boldsymbol{\mu} \boldsymbol{\mu}^\top) \right\} \quad (12a)$$

$$+ \log \int_a^b \exp \left\{ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right\} d\mathbf{x}, \quad (12b)$$

and where we have defined the following quantities:

$$\boldsymbol{\Psi} = \frac{1}{N} \sum_{t=1}^N \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} (\boldsymbol{\xi}_t - \mathbf{h}_t(\zeta_t)) p(\mathbf{x}_{t+1}, \mathbf{x}_t | \mathbf{y}_{1:N}, \hat{\beta}^{(k)}) d\mathbf{x}_{t+1} d\mathbf{x}_t, \quad (13a)$$

$$\boldsymbol{\Phi} = \frac{1}{N} \sum_{t=1}^N \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} (\boldsymbol{\xi}_t - \mathbf{h}_t(\zeta_t)) (\boldsymbol{\xi}_t - \mathbf{h}_t(\zeta_t))^\top \times p(\mathbf{x}_{t+1}, \mathbf{x}_t | \mathbf{y}_{1:N}, \hat{\beta}^{(k)}) d\mathbf{x}_{t+1} d\mathbf{x}_t. \quad (13b)$$

Proof. After applying the conditional expectation with respect to $\mathbf{y}_{1:N}$ and $\hat{\beta}^{(k)}$ to the log joint density in (10), we find that the maximizer of $\mathcal{Q}(\beta, \hat{\beta}^{(k)})$ in (9) must satisfy

$$\mathbb{E} \left\{ \log \mathbf{1}_{[a,b]}(\boldsymbol{\xi}_t - \mathbf{h}_t(\zeta_t)) | \mathbf{y}_{1:N}, \hat{\beta}^{(k)} \right\} \geq C$$

for some constant $C \in \mathbb{R}$ (otherwise, \mathcal{Q} is ill-defined). Thus, we must have $\mathbf{a} \leq \boldsymbol{\xi}_t - \mathbf{h}_t(\zeta_t) \leq \mathbf{b}$ for all $(\mathbf{x}_t, \mathbf{x}_{t+1}) \in \text{supp}(p(\mathbf{x}_t, \mathbf{x}_{t+1} | \mathbf{y}_{1:N}, \hat{\beta}^{(k)}))$. Under this constraint, we have

$$\begin{aligned} & \mathbb{E} \left\{ \log \mathbf{1}_{[a,b]}(\boldsymbol{\xi}_t - \mathbf{h}_t(\zeta_t)) | \mathbf{y}_{1:N}, \hat{\beta}^{(k)} \right\} \\ &= \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} \log \mathbf{1}_{[a,b]}(\boldsymbol{\xi}_t - \mathbf{h}_t(\zeta_t)) p(\mathbf{x}_t, \mathbf{x}_{t+1} | \mathbf{y}_{1:N}, \hat{\beta}^{(k)}) d\mathbf{x}_{t+1} d\mathbf{x}_t \\ &= 0. \end{aligned}$$

On the other hand, after exploiting properties of the trace and the linearity of the expectation operator,

$$\begin{aligned} & \sum_{t=1}^N \mathbb{E} \left\{ (\boldsymbol{\xi}_t - \mathbf{h}_t(\zeta_t) - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{\xi}_t - \mathbf{h}_t(\zeta_t) - \boldsymbol{\mu}) | \mathbf{y}_{1:N}, \hat{\beta}^{(k)} \right\} = \\ & \text{tr} \left\{ \boldsymbol{\Sigma}^{-1} \sum_{t=1}^N \mathbb{E} \left\{ (\boldsymbol{\xi}_t - \mathbf{h}_t(\zeta_t) - \boldsymbol{\mu})(\boldsymbol{\xi}_t - \mathbf{h}_t(\zeta_t) - \boldsymbol{\mu})^\top | \mathbf{y}_{1:N}, \hat{\beta}^{(k)} \right\} \right\}, \end{aligned}$$

which, after expanding the expectation, leads to (12a) with matrices given by (13). Since the integral term in the denominator in (10) is deterministic and does not depend on t , we immediately obtain (12b). \square

The constraints associated with the optimization problem in Lemma 4.1 describe the parameter space where all state trajectories described by $\hat{\beta}^{(k)}$ are feasible. The condition stated in (11b) is fulfilled when $\beta = \hat{\beta}^{(k)}$, indicating that $\mathcal{Q}(\beta, \hat{\beta}^{(k)})$ is well defined in a region that contains $\hat{\beta}^{(k)}$. As a result, there exists a parameter vector $\hat{\beta}^{(k+1)}$ that satisfies these constraints and minimizes $\mathcal{V}^{(k)}(\beta)$.

Remark 4.1: The optimization problem in Lemma 4.1 generalizes the unconstrained Gaussian case for linear time-invariant (LTI) systems in Lemma 3.1 of [10]. Indeed, if $a_i \rightarrow -\infty$ and $b_i \rightarrow \infty$, the constraints in (11b) are always satisfied and the log-int-exp term in (12b) is replaced by a log-det expression by observing that, for any $\mathbf{S} \succ \mathbf{0}$,

$$\int_{\mathbb{R}^n} \exp \left\{ -\frac{1}{2} \mathbf{x}^\top \mathbf{S}^{-1} \mathbf{x} \right\} d\mathbf{x} = \sqrt{\det(2\pi\mathbf{S})}. \quad (14)$$

Replacing accordingly and setting $\boldsymbol{\mu} = \mathbf{0}$ and $\mathbf{h}_t(\boldsymbol{\eta}_t) = \boldsymbol{\Gamma}\boldsymbol{\eta}_t$, i.e., the LTI case, we reach Eq. (22) of [10].

Next, we characterize the solution of the optimization problem in (11). The set of nonlinear equations that the parameters in β must satisfy at the optimal point are derived

next, which constitutes Contribution C1 of this paper. To ease the notation, we have excluded the iteration indices on the parameters $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, \mathbf{a} , and \mathbf{b} .

Theorem 4.2: The global optimum of $\mathcal{Q}(\boldsymbol{\beta}, \hat{\boldsymbol{\beta}}^{(k)})$ in (9) with respect to $\boldsymbol{\beta}$ in (4), for fixed values of $\hat{\boldsymbol{\beta}}^{(k)}$ and $\mathbf{y}_{1:N}$, is given by $\hat{\boldsymbol{\beta}}^{(k+1)} = [\hat{\boldsymbol{\mu}}^\top, \text{vec}\{\hat{\boldsymbol{\Sigma}}\}^\top, \hat{\mathbf{a}}^\top, \hat{\mathbf{b}}^\top]^\top$, where the parameters solve the following set of nonlinear equations:

$$\begin{cases} \mathcal{M}_1(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}, \hat{\mathbf{a}}, \hat{\mathbf{b}}) = \boldsymbol{\Psi}, & (15a) \\ \mathcal{M}_2(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}, \hat{\mathbf{a}}, \hat{\mathbf{b}}) = \boldsymbol{\Phi}, & (15b) \\ \{\hat{\mathbf{a}}\}_i = \inf_{(\mathbf{x}_t, \mathbf{x}_{t+1}) \in \mathcal{A}_t, t=1, \dots, N} \{\boldsymbol{\xi}_t - \mathbf{h}_t(\boldsymbol{\zeta}_t)\}_i, & (15c) \\ \{\hat{\mathbf{b}}\}_i = \sup_{(\mathbf{x}_t, \mathbf{x}_{t+1}) \in \mathcal{A}_t, t=1, \dots, N} \{\boldsymbol{\xi}_t - \mathbf{h}_t(\boldsymbol{\zeta}_t)\}_i, & (15d) \end{cases}$$

where $\mathcal{M}_i(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{a}, \mathbf{b})$ represents the i th moment of a Gaussian with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$ when truncated to a hyperrectangle with vertices \mathbf{a} and \mathbf{b} , and where $\boldsymbol{\Psi}$ and $\boldsymbol{\Phi}$ are given by (13a) and (13b), respectively.

Proof. We begin by deriving the truncation bound updates (15c) and (15d). Since the optimal value $\hat{\boldsymbol{\beta}}^{(k+1)}$ must satisfy the constraints $\hat{\mathbf{a}} \leq \boldsymbol{\xi}_t - \mathbf{h}_t(\boldsymbol{\zeta}_t) \leq \hat{\mathbf{b}}$ for all $(\mathbf{x}_t, \mathbf{x}_{t+1}) \in \text{supp}(p(\mathbf{x}_t, \mathbf{x}_{t+1} | \mathbf{y}_{1:N}, \hat{\boldsymbol{\beta}}^{(k)}))$, the following inequalities must hold for $i = 1, \dots, n$:

$$\begin{cases} \{\hat{\mathbf{a}}\}_i \leq \inf_{(\mathbf{x}_t, \mathbf{x}_{t+1}) \in \mathcal{A}_t, t=1, \dots, N} \{\boldsymbol{\xi}_t - \mathbf{h}_t(\boldsymbol{\zeta}_t)\}_i, & (16) \\ \{\hat{\mathbf{b}}\}_i \geq \sup_{(\mathbf{x}_t, \mathbf{x}_{t+1}) \in \mathcal{A}_t, t=1, \dots, N} \{\boldsymbol{\xi}_t - \mathbf{h}_t(\boldsymbol{\zeta}_t)\}_i. \end{cases}$$

We prove equality (15c); the same logic is applied to prove the result for $\hat{\mathbf{b}}$. To argue by contradiction, consider the global optimum $\hat{\boldsymbol{\beta}}^{(k+1)}$ and assume that there exists some $i^* \in \{1, \dots, n\}$ such that $\{\hat{\mathbf{a}}\}_{i^*}$ is strictly less than the right-hand side of (16). Let $\tilde{\boldsymbol{\beta}}^{(k+1)}$ (with lower truncation limit $\tilde{\mathbf{a}}$) be the parameter vector that is equal to $\hat{\boldsymbol{\beta}}^{(k+1)}$ in all entries except $\{\hat{\mathbf{a}}\}_{i^*}$, at which it is equal to the right-hand side of (16). This parameter vector also satisfies the constraints of the optimization problem in (11b). Since any Gaussian PDF has support on \mathbb{R}^n , we have the strict inequality

$$\int_{\tilde{\mathbf{a}} - \boldsymbol{\mu}}^{\hat{\mathbf{b}} - \boldsymbol{\mu}} \exp\left\{-\frac{1}{2}\mathbf{x}^\top \boldsymbol{\Sigma}^{-1}\mathbf{x}\right\} d\mathbf{x} > \int_{\hat{\mathbf{a}} - \boldsymbol{\mu}}^{\hat{\mathbf{b}} - \boldsymbol{\mu}} \exp\left\{-\frac{1}{2}\mathbf{x}^\top \boldsymbol{\Sigma}^{-1}\mathbf{x}\right\} d\mathbf{x},$$

which leads to $\mathcal{V}^{(k)}(\hat{\boldsymbol{\beta}}^{(k+1)}) > \mathcal{V}^{(k)}(\tilde{\boldsymbol{\beta}}^{(k+1)})$, i.e., $\hat{\boldsymbol{\beta}}^{(k+1)}$ is not the global optimum. This proves (15c).

Fixing $\mathbf{a} = \hat{\mathbf{a}}$ and $\mathbf{b} = \hat{\mathbf{b}}$, the optimizer is characterized by the first-order optimality conditions of the unconstrained problem in (11). The following gradients are derived using standard matrix differentiation rules (see, e.g., [27]):

$$\begin{aligned} \frac{\partial \mathcal{V}^{(k)}(\boldsymbol{\beta})}{\partial \boldsymbol{\mu}} &= \boldsymbol{\Sigma}^{-1} \left[\boldsymbol{\Psi} - \boldsymbol{\mu} - \frac{\int_{\hat{\mathbf{a}} - \boldsymbol{\mu}}^{\hat{\mathbf{b}} - \boldsymbol{\mu}} \mathbf{x} \exp\{-\frac{1}{2}\mathbf{x}^\top \boldsymbol{\Sigma}^{-1}\mathbf{x}\} d\mathbf{x}}{\int_{\hat{\mathbf{a}} - \boldsymbol{\mu}}^{\hat{\mathbf{b}} - \boldsymbol{\mu}} \exp\{-\frac{1}{2}\mathbf{x}^\top \boldsymbol{\Sigma}^{-1}\mathbf{x}\} d\mathbf{x}} \right], \\ \frac{\partial \mathcal{V}^{(k)}(\boldsymbol{\beta})}{\partial \boldsymbol{\Sigma}} &= -\frac{1}{2} \boldsymbol{\Sigma}^{-1} (\boldsymbol{\Phi} - \boldsymbol{\Psi} \boldsymbol{\mu}^\top - \boldsymbol{\mu} \boldsymbol{\Psi}^\top + \boldsymbol{\mu} \boldsymbol{\mu}^\top) \boldsymbol{\Sigma}^{-1} \\ &\quad + \frac{1}{2} \boldsymbol{\Sigma}^{-1} \frac{\int_{\hat{\mathbf{a}} - \boldsymbol{\mu}}^{\hat{\mathbf{b}} - \boldsymbol{\mu}} \mathbf{x} \mathbf{x}^\top \exp\{-\frac{1}{2}\mathbf{x}^\top \boldsymbol{\Sigma}^{-1}\mathbf{x}\} d\mathbf{x}}{\int_{\hat{\mathbf{a}} - \boldsymbol{\mu}}^{\hat{\mathbf{b}} - \boldsymbol{\mu}} \exp\{-\frac{1}{2}\mathbf{x}^\top \boldsymbol{\Sigma}^{-1}\mathbf{x}\} d\mathbf{x}} \boldsymbol{\Sigma}^{-1}. \end{aligned} \quad (17)$$

Note that the integral expressions above can be written in terms of the moments of a truncated Gaussian distribution as follows:

$$\begin{aligned} \frac{\int_{\hat{\mathbf{a}} - \boldsymbol{\mu}}^{\hat{\mathbf{b}} - \boldsymbol{\mu}} \mathbf{x} \exp\{-\frac{1}{2}\mathbf{x}^\top \boldsymbol{\Sigma}^{-1}\mathbf{x}\} d\mathbf{x}}{\int_{\hat{\mathbf{a}} - \boldsymbol{\mu}}^{\hat{\mathbf{b}} - \boldsymbol{\mu}} \exp\{-\frac{1}{2}\mathbf{x}^\top \boldsymbol{\Sigma}^{-1}\mathbf{x}\} d\mathbf{x}} &= \mathcal{M}_1 - \boldsymbol{\mu}, \\ \frac{\int_{\hat{\mathbf{a}} - \boldsymbol{\mu}}^{\hat{\mathbf{b}} - \boldsymbol{\mu}} \mathbf{x} \mathbf{x}^\top \exp\{-\frac{1}{2}\mathbf{x}^\top \boldsymbol{\Sigma}^{-1}\mathbf{x}\} d\mathbf{x}}{\int_{\hat{\mathbf{a}} - \boldsymbol{\mu}}^{\hat{\mathbf{b}} - \boldsymbol{\mu}} \exp\{-\frac{1}{2}\mathbf{x}^\top \boldsymbol{\Sigma}^{-1}\mathbf{x}\} d\mathbf{x}} &= \mathcal{M}_2 - \mathcal{M}_1 \boldsymbol{\mu}^\top - \boldsymbol{\mu} \mathcal{M}_1^\top + \boldsymbol{\mu} \boldsymbol{\mu}^\top, \end{aligned}$$

where \mathcal{M}_1 and \mathcal{M}_2 are the shorthand notation for $\mathcal{M}_1(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \hat{\mathbf{a}}, \hat{\mathbf{b}})$ and $\mathcal{M}_2(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \hat{\mathbf{a}}, \hat{\mathbf{b}})$, respectively. Setting the gradient with respect to $\boldsymbol{\mu}$ to zero, we directly obtain (15a). When this result is inserted in (17), with the gradient with respect to $\boldsymbol{\Sigma}$ being set to zero, we reach (15b). This concludes the proof. \square

Theorem 4.2 describes the iterations of the proposed EM algorithm. For a fixed iteration $\hat{\boldsymbol{\beta}}^{(k)}$, Eqs. (15a) and (15b) can be interpreted as method of moments equations [28], while (15c) and (15d) describe the smallest support for $\boldsymbol{\eta}_t$ that can explain the dataset under the assumption that the noise model is described exactly by $\hat{\boldsymbol{\beta}}^{(k)}$.

V. IMPLEMENTATION ASPECTS

Section IV outlined the theoretical foundations of the proposed identification method; here, important implementation aspects of the approach are discussed.

A. Initialization

A good initial estimate $\hat{\boldsymbol{\beta}}_1$ can significantly reduce the iterations required for EM algorithm to converge, and may avoid local optima. In many cases, an adequate initialization can be computed by solving the set of equations of Theorem 4.2 for the untruncated Gaussian case, i.e., $a_i \rightarrow -\infty$ and $b_i \rightarrow \infty$ for $i = 1, \dots, n + p$. This yields $\hat{\mathcal{M}}_1 = \mathbf{0}$ and $\hat{\mathcal{M}}_2 = \hat{\boldsymbol{\Sigma}} - \hat{\boldsymbol{\mu}} \hat{\boldsymbol{\mu}}^\top$, leading to the initial estimate

$$\hat{\boldsymbol{\mu}}_1 = \boldsymbol{\Psi}, \quad \hat{\boldsymbol{\Sigma}}_1 = \boldsymbol{\Phi} - \boldsymbol{\Psi} \boldsymbol{\Psi}^\top.$$

Thus, initializing the proposed EM approach with estimates from the Kalman smoothing method in [10] can lead to a fast convergence of the iterations in Theorem 4.2. This strategy is preferred when the Gaussian distribution being truncated has most of its probability density concentrated within the hyperrectangle $[\mathbf{a}, \mathbf{b}]$.

B. Computation of sample and population moments

Equations (15a) and (15b) require the computation of sample and population moments. For the sample moments $\boldsymbol{\Psi}$ and $\boldsymbol{\Phi}$, a cornerstone of the proposed method is the smoothing procedure that must be performed. Sequential Monte Carlo methods [22], also known as particle filtering and smoothing, provide a framework for estimating the state of a dynamic system based on noisy observations. The key idea is to represent the posterior distribution of the system state using a set of particles and weights, that is

$$p(\tilde{\mathbf{x}}_t | \mathbf{y}_{1:N}) \approx \sum_{i=1}^M w_t^{(i)} \delta(\tilde{\mathbf{x}}_t - \tilde{\mathbf{x}}_t^{(i)}),$$

where $\delta(\cdot)$ denotes the Dirac delta function, $\tilde{\mathbf{x}}_t := [\mathbf{x}_{t+1}^\top, \mathbf{x}_t^\top]^\top$ is the augmented state vector, $w_{t|N}^{(i)}$ and $\tilde{\mathbf{x}}_t^{(i)}$ are the i th weight and particle that represent the smoothing distribution $p(\tilde{\mathbf{x}}_t | \mathbf{y}_{1:N})$, and M is the number of particles. Using the particle-based representation of the posterior PDF, any conditional expectation depending on the augmented state can be computed using

$$\mathbb{E} \{ \mathbf{f}(\tilde{\mathbf{x}}_t) | \mathbf{y}_{1:N} \} \approx \sum_{i=1}^M w_{t|N}^{(i)} \mathbf{f}(\tilde{\mathbf{x}}_t^{(i)}), \quad (18)$$

where $\mathbf{f}(\tilde{\mathbf{x}}_t)$ is a function of the augmented state $\tilde{\mathbf{x}}_t$. In our setting, Ψ and Φ are computed using $\mathbf{f}(\tilde{\mathbf{x}}_t) = \xi_t - \mathbf{h}_t(\zeta_t)$ and $\mathbf{f}(\tilde{\mathbf{x}}_t) = (\xi_t - \mathbf{h}_t(\zeta_t))(\xi_t - \mathbf{h}_t(\zeta_t))^\top$, respectively.

With regards to the population moments, Theorem 4.2 requires the computation of the first and second moments of a truncated Gaussian distribution. For any dimension, these can be computed using Monte Carlo integration and specialized tools for sampling high-dimensional truncated Gaussians [29]. For the univariate case, certain explicit formulas can be obtained; non-recursive expressions (in terms of the Gamma function) can be found in [30]. For simplicity, we adopt the moment-generating function approach of [31].

For $k \in \mathbb{N}_0$, consider the computation of

$$\mathcal{M}_k(\mu, \sigma^2, a, b) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_a^b x^k \exp \left\{ -\frac{(x-\mu)^2}{2\sigma^2} \right\} dx. \quad (19)$$

We introduce the moment generating function

$$M(t) := \frac{1}{\sqrt{2\pi\sigma^2}} \int_a^b \exp \{ tx \} \exp \left\{ -\frac{(x-\mu)^2}{2\sigma^2} \right\} dx.$$

For $t = 0$ (i.e., $k = 0$ in (19)), it is known that $\mathcal{M}_0(\mu, \sigma^2, a, b) = M(0) = [\text{erf}(\bar{b}) - \text{erf}(\bar{a})] / 2$, where $\bar{b} = (b - \mu) / \sqrt{2\sigma^2}$ and $\bar{a} = (a - \mu) / \sqrt{2\sigma^2}$, and

$$\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp \{ -z^2 \} dz.$$

Moreover, $\mathcal{M}_k(\mu, \sigma^2, a, b) = M^{(k)}(0)$. By completing the square in the exponent of the integrand of $M(t)$,

$$\begin{aligned} M(t) &= \frac{\varrho(t)}{\sqrt{2\pi\sigma^2}} \int_a^b \exp \left\{ -\frac{(x - [\mu + t\sigma^2])^2}{2\sigma^2} \right\} dx \\ &= \frac{\varrho(t)}{2} \left[\text{erf} \left(\bar{b} - \frac{t\sigma}{\sqrt{2}} \right) - \text{erf} \left(\bar{a} - \frac{t\sigma}{\sqrt{2}} \right) \right], \end{aligned}$$

where $\varrho(t) = \exp \{ (t^2\sigma^2/2) + \mu t \}$. Since $\frac{d}{dt} \text{erf}(\bar{\mu} + ct) = \frac{2c}{\sqrt{\pi}} \exp \{ -(\bar{\mu} + ct)^2 \}$, we see that $M^{(k)}(0)$ can be explicitly computed in terms of the exponential and erf functions. In particular, this leads to the explicit computation of the first and second moments

$$\begin{aligned} \mathcal{M}_1(\mu, \sigma^2, a, b) &= \mu \mathcal{M}_0(\mu, \sigma^2, a, b) \\ &\quad - \sigma \left(\exp \{ -\bar{b}^2 \} - \exp \{ -\bar{a}^2 \} \right) / \sqrt{2\pi}, \\ \mathcal{M}_2(\mu, \sigma^2, a, b) &= (\mu^2 + \sigma^2) \mathcal{M}_0(\mu, \sigma^2, a, b) \\ &\quad - \sigma \left[(\mu + b) \exp \{ -\bar{b}^2 \} - (\mu + a) \exp \{ -\bar{a}^2 \} \right] / \sqrt{2\pi}. \end{aligned}$$

C. Computational methods for solving Eqs. in (15)

Once the smoothing step has been performed, upper and lower bounds for (15c) and (15d) can respectively be obtained by evaluating $\xi_t - \mathbf{h}_t(\zeta_t)$ at the particle values $\tilde{\mathbf{x}}_t^{(i)}$, and computing the maximum and minimum of each entry. Caution must be exercised, as such computations scale with the number of particles and data points.

Once $\hat{\mathbf{a}}$ and $\hat{\mathbf{b}}$ are determined, or if the truncation limits are known, the remaining task is to solve the system in (15a)-(15b). Following the approach in [19] for fitting truncated Gaussian mixtures, we employ the fixed-point iterations:

$$\begin{aligned} \hat{\mu}_{j+1} &= \Psi - \mathcal{M}_1(\mathbf{0}, \hat{\Sigma}_j, \hat{\mathbf{a}} - \hat{\mu}_j, \hat{\mathbf{b}} - \hat{\mu}_j), \\ \hat{\Sigma}_{j+1} &= \Phi + \hat{\Sigma}_j - \mathcal{M}_2(\mathbf{0}, \hat{\Sigma}_j, \hat{\mathbf{a}} - \hat{\mu}_{j+1}, \hat{\mathbf{b}} - \hat{\mu}_{j+1}), \end{aligned}$$

where $\hat{\mu}_j$ and $\hat{\Sigma}_j$ represent the j th iterates of the mean and covariance within the k th EM iteration.

VI. SIMULATION EXAMPLE

In this section, we illustrate the proposed approach via a simulation example. We study the gain in estimation accuracy when estimating truncated Gaussian noise PDFs instead of standard Gaussian ones on the system

$$\begin{aligned} x_{t+1} &= f(x_t, u_t) + w_t, \\ y_t &= g(x_t, u_t) + v_t, \end{aligned}$$

where $f(x_t, u_t) = 0.9x_t + 2u_t$, $g(x_t, u_t) = 1.6x_t + 1.2u_t$, the noises $w_t \sim \mathcal{TN}(-0.3, 1, -1.5, 2.5)$ and $v_t \sim \mathcal{N}(-0.1, 0.5)$ are uncorrelated. For simplicity we assume that the truncation bounds related to w_t are known, and we use this information to estimate the mean and covariance of the Gaussian being truncated to generate w_t (i.e., -0.3 and 1 respectively), as well as the mean and covariance of v_t .

A Monte Carlo simulation is performed with 100 runs, each with $N = 5000$ samples. Two estimators are tested: the EM algorithm that uses the Kalman smoother for computing the noise statistics (KS-EM), which is based on the work of [10], and the proposed method (TG-EM). A total of 40 EM iterations is performed for each method, and they are initialized at a uniformly-distributed random value centered at the true parameters with 10% relative error. Each smoothed PDF is approximated using 500 particles.

Boxplots containing the statistical performance of the KS-EM and TG-EM methods for each parameter of interest are presented in Figure 2. The proposed method shows clear improvements in estimation accuracy over the Kalman-based approach in 3 of the 4 parameters. Interestingly, the median of the KS-EM estimates of μ_w and Σ_w are approximately equal to the mean and variance of w_t , which are given by -0.09 and 0.66 , respectively. This is not the case for the proposed method, since the parameters μ_w and Σ_w describe the mean and covariance of the Gaussian distribution prior to truncation, and not the distribution measures of the truncated PDF itself. Only the proposed approach can adequately incorporate the truncation bounds into the estimation, leading to a more accurate characterization of the noise in the state-space model. The KS-EM estimates of μ_v are more accurate

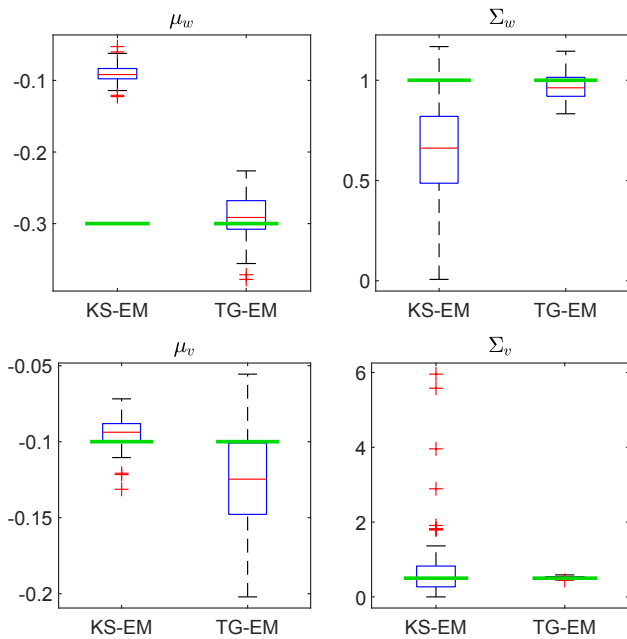


Fig. 2. Boxplots of each estimated mean and variance of the state and output noise. The green line in each boxplot represents the true value of the parameter being estimated.

than those of the TG-EM method, however, the KS-EM estimates of Σ_v show more spread and contain more outliers.

VII. CONCLUSIONS

In this paper, we proposed a method for computing the parameters of a truncated Gaussian noise model that affects the state and output equations of a state-space system. This approach offers flexibility in noise model estimation, as combining Gaussianity with bounded support assumptions typically provides a more accurate characterization of the noise. By considering the state vector as a hidden variable in the EM algorithm, we derived explicit expressions that the parameters must satisfy at each iteration. This method generalizes the approach in [10] for fixed system parameters, extending it from untruncated to truncated Gaussian distributions, with the option to set the truncation bounds to infinity, thus recovering the untruncated case. The computation of the EM iterations requires particle smoothing techniques, as well as the estimation of moments of truncated Gaussian PDFs. Future work includes accelerating the estimation procedure, and extending it to joint system and noise identification.

REFERENCES

- [1] T. Söderström and P. Stoica, *System Identification*. Prentice-Hall, 1989.
- [2] P. S. Maybeck, *Stochastic Models, Estimation, and Control - Vol 1*. Academic Press, 1982.
- [3] L. Ljung, *System Identification: Theory for the User*, 2nd Edition. Prentice-Hall, 1999.
- [4] M. Milanese, J. Norton, H. Piet-Lahanier, and É. Walter, *Bounding Approaches to System Identification*. Springer, 2013.
- [5] T. Alamo, J. M. Bravo, and E. F. Camacho, "Guaranteed state estimation by zonotopes," *Automatica*, vol. 41, no. 6, pp. 1035–1043, 2005.
- [6] E. Fogel and Y.-F. Huang, "On the value of information in system identification—bounded noise case," *Automatica*, vol. 18, no. 2, pp. 229–238, 1982.

- [7] H. Hjalmarsson and L. Ljung, "A unifying view of disturbances in identification," *IFAC Proceedings Volumes*, vol. 27, no. 8, pp. 383–388, 1994.
- [8] H. Hjalmarsson, "Optimally robust system identification of systems subject to amplitude-bounded stochastic disturbances," *IEEE Transactions on Automatic Control*, vol. 43, no. 7, pp. 947–953, 1998.
- [9] H. Akçay and N. At, "Membership set identification with periodic inputs and orthonormal regressors," *Signal Processing*, vol. 86, no. 12, pp. 3778–3786, 2006.
- [10] S. Gibson and B. Ninness, "Robust maximum-likelihood estimation of multivariable dynamic systems," *Automatica*, vol. 41, no. 10, pp. 1667–1682, 2005.
- [11] J. I. Yuz, J. Alfaro, J. C. Agüero, and G. C. Goodwin, "Identification of continuous-time state-space models from non-uniform fast-sampled data," *IET Control Theory and Appl.*, vol. 5, no. 7, pp. 842–855, 2011.
- [12] A. Wills, T. Schön, L. Ljung, and B. Ninness, "Identification of Hammerstein–Wiener models," *Automatica*, vol. 49, no. 1, pp. 70–81, 2013.
- [13] A. L. Cedeño, R. A. González, R. Carvajal, and J. C. Agüero, "Identification of Wiener state-space models utilizing Gaussian sum smoothing," *Automatica*, vol. 166, Article 111707, 2024.
- [14] O. Kost, J. Duník, and O. Straka, "Measurement difference method: A universal tool for noise identification," *IEEE Transactions on Automatic Control*, vol. 68, no. 3, pp. 1792–1799, 2022.
- [15] H. Durrant-Whyte and T. Bailey, "Simultaneous Localization and Mapping: Part I," *IEEE Robotics & Automation Magazine*, vol. 13, no. 2, pp. 99–110, 2006.
- [16] S. Bittanti and M. Campi, "Bounded error identification of time-varying parameters by RLS techniques," *IEEE Transactions on Automatic Control*, vol. 39, no. 5, pp. 1106–1110, 1994.
- [17] R. Hovorka, V. Canonico, L. J. Chassin, U. Haueter, M. Massi-Benedetti, M. O. Federici, T. R. Pieber, H. C. Schaller, L. Schaupp, and T. Vering, "Nonlinear model predictive control of glucose concentration in subjects with type 1 diabetes," *Physiological measurement*, vol. 25, no. 4, p. 905, 2004.
- [18] A. Ilyas, E. Zampetakis, and C. Daskalakis, "A theoretical and practical framework for regression and classification from truncated samples," in *International Conference on Artificial Intelligence and Statistics*, pp. 4463–4473, PMLR, 2020.
- [19] G. Lee and C. Scott, "EM algorithms for multivariate gaussian mixture models with truncated and censored data," *Computational Statistics & Data Analysis*, vol. 56, no. 9, pp. 2816–2829, 2012.
- [20] M. Zheng and Y. Ohta, "Bayesian positive system identification: Truncated Gaussian prior and hyperparameter estimation," *Systems & Control Letters*, vol. 148, p. 104857, 2021.
- [21] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society: Series B (Meth.)*, vol. 39, no. 1, pp. 1–22, 1977.
- [22] 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, 2000.
- [23] L. Ljung and T. Glad, *Modeling of Dynamic Systems*. Prentice-Hall, 1994.
- [24] E. L. Lehmann and G. Casella, *Theory of Point Estimation*, 2nd Edition. Springer, 1998.
- [25] C. F. J. Wu, "On the convergence properties of the EM algorithm," *The Annals of Statistics*, pp. 95–103, 1983.
- [26] R. A. González, A. L. Cedeño, M. Coronel, J. C. Agüero, and C. R. Rojas, "An EM algorithm for Lebesgue-sampled state-space continuous-time system identification," *IFAC-PapersOnLine*, vol. 56, no. 2, pp. 4204–4209, 2023.
- [27] J. R. Magnus and H. Neudecker, *Matrix Differential Calculus with Applications in Statistics and Econometrics*. John Wiley & Sons, 2019.
- [28] W. K. Newey and D. McFadden, "Large sample estimation and hypothesis testing," *Handbook of Econometrics*, vol. 4, pp. 2111–2245, 1994.
- [29] Z. I. Botev, "The normal law under linear restrictions: simulation and estimation via minimax tilting," *Journal of the Royal Statistical Society: Series B (Stat. Meth.)*, vol. 79, no. 1, pp. 125–148, 2017.
- [30] H. Ogasawara, "Unified and non-recursive formulas for moments of the normal distribution with stripe truncation," *Communications in Statistics-Theory and Methods*, vol. 51, no. 19, pp. 6834–6862, 2022.
- [31] G. M. Tallis, "The moment generating function of the truncated multi-normal distribution," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 23, no. 1, pp. 223–229, 1961.