

Identification of Non-Linear Differential-Algebraic Equation Models with Process Disturbances

Mohamed R.-H. Abdalmoaty, Oscar Eriksson, Robert Berez, David Broman and Håkan Hjalmarsson

Abstract—Differential-algebraic equations (DAEs) arise naturally as a result of equation-based object-oriented modeling. In many cases, these models contain unknown parameters that have to be estimated using experimental data. However, often the system is subject to unknown disturbances which, if not taken into account in the estimation, can severely affect the model's accuracy. For non-linear state-space models, particle filter methods have been developed to tackle this issue. Unfortunately, applying such methods to non-linear DAEs requires a transformation into a state-space form, which is particularly difficult to obtain for models with process disturbances. In this paper, we propose a simulation-based prediction error method that can be used for non-linear DAEs where disturbances are modeled as continuous-time stochastic processes. To the authors' best knowledge, there are no general methods successfully dealing with parameter estimation for this type of model. One of the challenges in particle filtering methods are random variations in the minimized cost function due to the nature of the algorithm. In our approach, a similar phenomenon occurs and we explicitly consider how to sample the underlying continuous process to mitigate this problem. The method is illustrated numerically on a pendulum example. The results suggest that the method is able to deliver consistent estimates.

I. INTRODUCTION

Equation-based object-oriented modeling languages [1]—such as Modelica¹, MathWorks Simscape, and VHDL-AMS—are today standard in industry when modeling and simulating complex physical dynamical systems. A key feature of these component-based languages is that they are *acausal*, meaning that the information flow between components is bidirectional. The underlying mathematical foundations of acausal models are Differential-Algebraic Equations (DAEs), which express both the dynamics and constraints between acausal components.

The differential index (or index) of a set of DAEs is the minimum number of times the DAEs have to be differentiated before one obtains a system of implicit ordinary differential equations (ODEs). The index can be seen as a distance measure between the DAEs and an underlying system of ODEs. In general, the higher the index, the more

difficult the DAEs are to solve. Many of the available numerical solvers can only handle DAEs of index 1 (see e.g., [2]), or semi-explicit DAEs of index 2. Such DAEs consists of a set of explicit ODEs coupled with algebraic constraints (see [3]).

In this paper, we consider the parameter estimation problem of a class of non-linear DAEs using experimental data. In practice, there are many sources of disturbances acting on a system, including measurement noise causing inexact measurements of the output, as well as process disturbances that can be considered as unknown inputs. The consequences of process disturbances are more difficult to predict compared to measurement noise, since the effect of the disturbance is transformed by the system dynamics. It is well known from the analysis of stochastic non-linear discrete-time systems that parameter estimation methods neglecting process disturbances in general leads to asymptotically biased estimators as well as an increase in the variance error, see e.g. [4].

A. Available methods

The parameter estimation problems of DAEs are far less treated in the literature compared to, for example, that of non-linear state-space models (see e.g., [5] and [6] for the latter). Most of the methods currently available for DAEs can only handle cases with no process disturbance (i.e., all inputs are known). Under this assumption, the methods have relatively straightforward formulations, and are usually concerned with the numerical aspects of the involved optimization problem, such as ensuring convergence to globally optimal solutions. (see e.g., [7] and [8], and the references therein).

On the other hand, cases with process disturbances are much more difficult to formulate and solve. This is mainly due to the challenges introduced by modeling disturbances as continuous-time stochastic processes. A naive treatment of DAEs with continuous-time stochastic inputs may result in variables representing physical quantities being modeled as having infinite variance, among other issues. Methods for ensuring the existence of well-defined solutions in these cases are discussed in [9], together with an approximate implementation of a particle filter for state estimation. A modified extended Kalman filter that can handle both uncertainty in the differential constraints as well as continuous-time white noise in algebraic constraints of index 1 DAEs is presented in [10], while [11] presents conditions for well-posed estimation problems for linear DAEs with process disturbances.

Although the question of whether a parameter estimation problem is well-posed in the presence of stochastic

This work was supported by the Swedish Research Council under contracts 2019-04956 and 2016-06079 (the research environment NewLEADS—New Directions in Learning Dynamical Systems), by Digital Futures, and by the Swedish Foundation for Strategic Research (FFL15-0032).

M. Abdalmoaty, R. Berez and H. Hjalmarsson are with the Division of Decision and Control Systems, EECS and Digital Futures, KTH Royal Institute of Technology, SE-100 44 Stockholm Sweden (abda, robby, hjalmarsson@kth.se).

O. Eriksson, D. Broman are with the Division of Software and Computer Systems, EECS and Digital Futures, KTH Royal Institute of Technology, SE-100 44 Stockholm Sweden (oerikss, dbro@kth.se)

¹<https://www.modelica.org/>

disturbances has been explored before, we have not found any complete tractable method that can be used for *solving* the problem. For instance, guidelines for implementing the Maximum Likelihood Estimator (MLE) for linear DAEs are given in [11], but no such estimator is implemented. In [12], the potential of using a particle filter to compute an approximate MLE for non-linear DAEs is pointed out. The challenges with this approach are also discussed, with the conclusion that it is non-trivial to implement: it involves minimizing a non-smooth approximation of the likelihood function. In addition, due to the use of particle filters, the approach is limited to DAEs that can be transformed into a (semi-explicit) state-space form. In some cases, several approximations and heuristics are used to achieve this, such as removing some process disturbances from the model whenever they appear at locations that hinder the aforementioned transformation. To the best of the authors' knowledge, there are currently no general reliable methods for parameter estimation of non-linear DAEs with process disturbances.

B. Contributions

In this paper, we present a method for estimating parameters of non-linear DAEs with process disturbances modeled as continuous-time stochastic processes. Our approach can be applied to well-defined general high-index non-linear DAEs that can be reduced to a manageable form, so that off-the-shelf numerical solvers (such as [2]) can be used. Moreover, we provide details that allow for the implementation of the method in a numerically tractable way. In particular, we propose a method for modeling and simulating the continuous-time process disturbance. Under certain standard assumptions (such as ergodic data and smooth model parameterization), the proposed estimator is consistent; namely the estimator converges, asymptotically in the data size, to the true parameter. This is supported by a numerical simulation experiment.

II. PROBLEM FORMULATION

We consider DAE models on a general form, given by

$$F(\dot{x}(t), x(t), u(t), w(t); \theta) = 0 \quad (1a)$$

$$y(t) = q(x(t), u(t); \theta) + v(t), \quad (1b)$$

where, at time $t \in \mathbb{R}$, $x(t) \in \mathbb{R}^{n_x}$ denotes the unobserved state of the system, $u(t) \in \mathbb{R}^p$ denotes a known external input, and $w(t) \in \mathbb{R}^{m_w}$ denotes the process disturbance, modeled as a continuous-time stochastic process. Further assumptions on this process can be found in Section III-B. The system's output is given by $y(t) \in \mathbb{R}^m$, and measurement noise is denoted by $v(t)$. We make no assumptions on the distribution of $v(t)$ except that $\mathbb{E}[v(t)] = 0$ for all t . The model is further assumed to be smoothly parametrized by a constant unknown parameter vector $\theta \in \mathbb{R}^{d_\theta}$, to be estimated. Since $u(t)$ can be chosen by the user, it is assumed known for all values of t , while the output $y(t)$ must be sampled at discrete time instants, which we denote $\{t_k\}$. We assume that we have access to a data set

$$D_N(T) = \{(y(t_k), u(s)) : k = 1, \dots, N, \quad s \in [0, T]\}$$

for some time T , such that for each k the time instant $t_k \in [0, T]$. The objective is to estimate the parameters θ using the data set $D_N(T)$.

III. PROPOSED METHOD

A. Estimation method

The proposed method is a type of prediction error method (see [13, Ch. 7]), recently developed in [14] for the identification of stochastic non-linear discrete-time models. The basic idea relies on the use of predictors that are simpler to compute than the optimal mean-square error one-step ahead predictor, which is known to be intractable for most non-linear stochastic systems. One of the predictors proposed in [10] is the mean of the model's output. It is well known that among constant predictors the mean minimizes the mean-squared error and from this property one can derive consistency of a parameter estimator based on this predictor. In the context of dynamical systems, the mean will be a time-varying function depending on the known input signals $u(t)$, but independent of the observed output $\{y(t_k)\}$. We propose the use of an estimator equivalent to the OE-QPEM estimator in [10] adapted to our context, which means that the estimate is obtained by minimizing the cost function

$$J_N(\theta) = \frac{1}{N} \sum_{k=1}^N \|y(t_k) - \mathbb{E}[y(t_k; \theta); \theta]\|^2, \quad (2)$$

where the $\mathbb{E}[\cdot; \theta]$ denotes the expected value taken over both the measurement noise and the process disturbances. This expected value can be computed analytically only in special cases, and for rather simple models. In general, and because of non-linearities in the model through which the process disturbances pass, the mean of the model's output is usually analytically intractable. Therefore, we borrow an idea used in [15] and [16], where the expected value is estimated using Monte Carlo simulations. For this purpose, we simulate M independent realizations of the output of (1) with $v(t) = 0$, which we denote by $\hat{y}^{(m)}(\cdot; \theta)$, $m = 1, \dots, M$, and estimate the cost function (2) as

$$\hat{J}_{N,M}(\theta) = \frac{1}{N} \sum_{k=1}^N \|y(t_k) - \bar{y}_M(t_k; \theta)\|^2, \quad (3a)$$

$$\bar{y}_M(t_k; \theta) = \frac{1}{M} \sum_{m=1}^M \hat{y}^{(m)}(t_k; \theta). \quad (3b)$$

An approximate estimator is then defined as $\hat{\theta}_{N,M} = \min_{\theta} \hat{J}_{N,M}(\theta)$. This estimator is expected to possess favorable asymptotic properties. Under suitable regularity conditions, a uniform version of the law of large numbers implies that almost surely $\hat{J}_{N,M}(\theta) \rightarrow J_N(\theta)$ uniformly in θ as $M \rightarrow \infty$. If we assume that the data $\{y(t_k)\}$ has been generated using (1) with the unknown true value of the parameters $\theta = \theta_*$, the estimator $\hat{\theta}_{N,M}$ converges to θ_* as $N, M \rightarrow \infty$ because, as mentioned above, the mean is the constant predictor minimizing the mean-squared error.²

²convergence to a parameter θ_* can be similarly established for the practical case where no θ_* exists.

The benefit of this method over the MLE is that it bypasses the intractable computations of the likelihood function, i.e. the probability density function of the output given a value of the parameters θ . It only requires the ability to simulate the output of the DAEs model, and therefore is able to avoid computational difficulties related to non-linear filtering and some of the limitations of particle filters methods.

Note that, conditioned on the observed data, the cost function (2) is a deterministic quantity. On the other hand, conditioned on the observed data, the cost function $\hat{J}_{N,M}(\theta)$ is stochastic and depends on the particular realizations of $\hat{y}^{(m)}(\cdot; \theta)$. This can cause several difficulties: If the realizations of $\hat{y}^{(m)}(\cdot; \theta)$ are computed independently for different values of θ , there are no guarantees that $\hat{J}_{N,M}(\theta)$ will be a continuous function of θ . Many optimization algorithms that can be used to find the minimizers, e.g. the Levenberg-Marquardt algorithm, require the differentiability of the cost function and only find local minima. Therefore, we need to reduce the variability of the approximate cost function and increase its smoothness, to alleviate the risk of the algorithm getting stuck in a local minimum that is caused by the algorithm itself. Ideally, we would like to generate the realizations $\hat{y}^{(m)}(\cdot; \theta)$ so that $\hat{J}_{N,M}(\theta)$ satisfies the two following properties:

- 1) When $\hat{J}_{N,M}(\theta)$ is computed several times for the same value of θ , it should always return the same value.
- 2) The mapping $\theta \rightarrow \hat{J}_{N,M}(\theta)$ should be as smooth as possible, so that small changes in θ cause small changes in the value of $\hat{J}_{N,M}(\theta)$.

The second property is in theory made possible since we assume the model to be smoothly parameterized by θ . These properties guide the approach taken in our implementation, which takes measures to reduce the variability of $\hat{J}_{N,M}(\theta)$ and improve its smoothness compared to the case when $\hat{y}^{(m)}(\cdot; \theta)$ are generated independently for different θ .

B. Disturbance modeling

Several approaches can be used to model (and simulate) continuous-time stochastic disturbance. We will assume that the continuous-time stochastic process $w(t)$ is stationary with a rational spectrum. Then its second-order properties can be modeled as filtered white noise, with a model given by

$$dx_w(t) = A(\theta)x_w(t) + B(\theta)dz_c(t) \quad (4a)$$

$$w(t) = Cx_w(t), \quad (4b)$$

where $dz_c(t)$ is a process with orthogonal increments and incremental variance $\mathbb{E}[dz_c(t)dz_c^T(t)] = I$, where I is the identity matrix.

It is important to note that a solution $x(t)$ of (1) can depend on the derivatives of the input $u(t)$, as well as derivatives of the disturbance $w(t)$ (see [17] for details). This can be a problem, since the derivatives of $w(t)$ do not necessarily have finite variance. To avoid this problem, we will assume that the rational spectrum of the process disturbance (4) has a sufficiently high pole excess, as was done in [12]. A pole excess of $2p_w$ guarantees that $w(t)$ has

$p_w - 1$ derivatives (in quadratic mean) with finite variance. This is a central assumption that has to be imposed, otherwise the DAEs model might not be well-defined.

For simulating (1), the disturbance model (4) can be exactly discretized at the time instants $\{\tau_k\}$, $k = 1, \dots, N_w$ (see e.g., [18]). This results in the discrete-time disturbance model

$$x_w^{(m)}(\tau_{k+1}; \theta) = A_d(\Delta_k; \theta)x_w^{(m)}(\tau_k; \theta) + B_d(\Delta_k; \theta)z_m(\tau_k) \quad (5a)$$

$$w^{(m)}(\tau_k; \theta) = Cx_w^{(m)}(\tau_k; \theta) \quad (5b)$$

with $A_d(\Delta_k; \theta) = e^{A(\theta)\Delta_k}$ and $B_d(\Delta_k; \theta)$ chosen as the Cholesky factor of the matrix $D_d(\Delta; \theta)$ given by

$$\int_0^{\Delta_k} e^{A(\theta)(\Delta_k-s)} B(\theta) B(\theta)^T \left(e^{A(\theta)(\Delta_k-s)} \right)^T ds.$$

In this paper, we will assume that $w(t)$ is a Gaussian process³, and therefore the innovation process $\{z_m(\tau_k)\}_k$ is a Gaussian discrete-time white noise process with zero mean and variance $\mathbb{E}[z_m(\tau_k)z_m^T(\tau_k)] = I$. The potentially time-varying sampling interval is given by $\Delta_k = \tau_{k+1} - \tau_k$. Using the results from [19], the matrices A_d and B_d can be readily computed as follows: First compute

$$\hat{C} = \begin{bmatrix} -A(\theta) & B(\theta)B^T(\theta) \\ 0 & A^T(\theta) \end{bmatrix},$$

where 0 represents a zero matrix of appropriate dimensions, and then compute the matrix exponential

$$e^{\hat{C}\Delta} = \begin{bmatrix} A_d^{-1}(\Delta; \theta) & A_d^{-1}(\Delta; \theta)D_d(\Delta; \theta) \\ 0 & A_d^T(\Delta; \theta) \end{bmatrix}.$$

The matrix $A_d(\Delta; \theta)$ can be obtained from the lower diagonal block of $e^{\hat{C}\Delta}$, and then left-multiplied with the upper right block to obtain $D_d(\Delta; \theta)$.

IV. IMPLEMENTATION DETAILS

In this paper, we choose the DAE solver SUNDIALS IDA [2] which is based on DASPK [20], a DASSL [21] implementation. These are variable-step, general purpose index 1 (and semi-explicit index 2) DAE solvers, able to handle stiff problems, and usually the default solvers in equation-based object-oriented modeling languages⁴. In this section, we describe how to obtain samples of the process disturbance at the time instants the solver needs, and how to generate the disturbance realization in a way that reduces the variability of the cost function (3a).

A. Sampling process disturbance

If the solver used for simulating (1) uses fixed step sizes, the disturbance model can be discretized to obtain (5) for a single fixed value of $\Delta_k = \Delta$, equal to the solver's step size. Then, using the discrete-time model, one can obtain samples of the disturbance at desired time instants. However,

³this assumption is not limiting, since a large class of processes can be modeled as transformations of Gaussian ones. Such transformations can be part of the model (see for instance the model in V-A)

⁴e.g. OpenModelica and Dymola. Simscape defaults to ode23t or daessc for DAEs, both variable step solvers, depending on version.

in adaptive step solvers the step sizes are not known a priori. This means that we do not know the time instants at which the solver will request samples of the process disturbance, and we therefore cannot simply form a model on the form (5) with fixed Δ . The solver also does not necessarily request samples in chronological order, which removes the option of constructing the discrete-time model as requests from the solver arrive.

A way to solve this problem is to sample the process disturbance a priori and uniformly at N_w time instants $\{\tau_k\}_{k=1}^{N_w}$, and then compute values between the sample instants using some form of interpolation. For example, polynomial interpolation or spline interpolation can be used to obtain values of the disturbance $w^{(m)}(t; \theta)$ between sampling instants, i.e. when $\tau_k < t < \tau_{k+1}$ for some k . The simplest method would be to use linear interpolation between the two values of $w^{(m)}(t; \theta)$ closest in time to the requested time instants. This is the interpolation method that we tested in the numerical experiments in Section V. Interpolating is an approximative method and will result in a realization of the process disturbance that does not have the same exact spectrum as the process disturbance modeled by (4). However, the approximation error can be reduced by sampling the disturbance in a dense grid with large N_w .

B. Fixing noise realizations

It is important to note that, if the samples $\{w^{(m)}(\tau_k; \theta)\}$ are generated anew for every value of the parameters θ , it is possible that the cost function (3a) will not vary smoothly in θ , and it will not give the same value if computed twice for the same value of the parameters. In order to solve this issue, and reduce the variability of the cost function, we use an approach based on common random numbers (see [22]). This is achieved in the dynamical setting considered here by using the same M realizations of the process disturbance every time the cost function is computed. However, since the disturbance model could potentially be parameterized by θ , we instead fix M realizations of the innovation process $\{z_m(\tau_k)\}$ that generates the process disturbance. Then one can use (5) to generate the realizations of the process disturbance for any value of θ . If the model is smoothly parametrized, the cost function should also vary smoothly with θ .

V. NUMERICAL EXPERIMENT

In this section, we will demonstrate the performance of our proposed method using a numerical simulation experiment. In Section V-A we perform parameter estimation on a pendulum, modeled using DAEs. In Section V-B we demonstrate the effects of using fixed white noise realizations, as described in Section IV-B, for computations of the cost function. For the implementation we used Julia⁵ and the DiffEq package [23] with the SUNDIALS IDA solver.

⁵<https://julialang.org/>

A. Parameter estimation

We consider a model of a pendulum subject to a drag and expressed in Cartesian coordinates. This model is of interest due to its high index (index 3): a naive index reduction leads to a numerically unstable solution. The model is defined as the following DAEs:

$$m\ddot{x}_1(t) = x_3(t)x_1(t) - k|\dot{x}_1(t)|\dot{x}_1(t) + u(t) + w^2(t) \quad (6a)$$

$$m\ddot{x}_2(t) = x_3(t)x_2(t) - k|\dot{x}_2(t)|\dot{x}_2(t) - mg \quad (6b)$$

$$L^2 = x_1^2(t) + x_2^2(t) \quad (6c)$$

where all signals are scalars. The variables $x_1(t)$ and $x_2(t)$ denote the position of the pendulum-arm endpoint in the plane and $x_3(t)$ the tension per unit length in the pendulum-arm. The signal $u(t)$ is a known input function, $w(t)$ is an unknown process disturbance, and $v(t)$ a measurement noise as described in Section II. It is sufficient for this demonstration to model these as scalar signals which simplifies the numerical treatment. With a vector-valued disturbances, the correlation of the components would additionally have to be taken into account. The system output

$$y(t) = \arctan\left(\frac{x_1(t)}{-x_2(t)}\right) + v(t) \quad (7)$$

is the angle ϕ between the pendulum-arm and the negative vertical axis. We assume $\phi_0 = 0$ and that the pendulum is initially at rest. The parameters m , g , k , and L denote the mass, the gravitational force, the drag coefficient, and the length of the pendulum-arm, respectively. We assume that m , g , and k are known and attempt to estimate the pendulum-arm length, so here $\theta = L$.

Equations (6) describe a DAE model with differential index 3 which is not suitable for direct numerical solving [24]. We therefore transform (6) into a stabilized, index 1, first-order formulation (see [25]) as

$$\dot{x}_1(t) = x_4(t) - 2\dot{x}_6(t)x_1(t) \quad (8a)$$

$$\dot{x}_2(t) = x_5(t) - 2\dot{x}_6(t)x_2(t) \quad (8b)$$

$$m\dot{x}_4(t) = \tilde{\dot{x}}_3(t)x_1(t) - k|x_4(t)|x_4(t) + u(t) + w^2(t) \quad (8c)$$

$$m\dot{x}_5(t) = \tilde{\dot{x}}_3(t)x_2(t) - k|x_5(t)|x_5(t) - mg \quad (8d)$$

$$L^2 = x_1^2(t) + x_2^2(t) \quad (8e)$$

$$0 = x_4(t)x_1(t) + x_5(t)x_2(t) \quad (8f)$$

where $x_4(t)$ and $x_5(t)$ are velocities, $x_6(t) = 0$ is a dummy variable introduced to maintain a consistent number of variables and equations, and $\tilde{\dot{x}}_3(t) = x_3(t)$. Equation (8f) is the first derivative of (8e) with respect to time and with velocities substituted for positions.

We fix the parameters $m = 0.3$ kg, $g = 9.81$ m/s², $k = 0.05$ kg m⁻¹, and choose the parameter $\theta_0 = L = 6.25$ m for the true system. To simulate the process disturbance $w(t)$ and fixing the input $u(t)$ we use the model (5) with

$$A = \begin{bmatrix} 0 & 1 \\ -4^2 & -0.8 \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 0 \end{bmatrix}, \quad (9)$$

$x_0 = [0 \ 0]^T$, and $\Delta_k = \Delta = 0.01$ s. We construct three sets of realizations of (5) given (9) with sufficiently large N_w . One set of size $E = 1000$, one set of size $M = 500$, and one set of size 1. For all sets, we approximate inter-sample values using linear interpolation. The first set belongs to the true system, the second set belongs to the model, and the last realization forms the input $u(t)$, which is thus a realization of a stationary process with the same spectrum as $w(t)$. The signal forming the process disturbance is scaled with 0.2 and the input signal is scaled with 0.6. This choice of scales gives a balance between the input signal and the process noise so that the effect of the process noise is visible in the parameter estimation problem. We choose the variance of the measurement noise σ as 0.002, and the output $y(t)$ is uniformly sampled every $T_s = 0.1$ s. We form the E datasets $\{D_N^{(1)}, \dots, D_N^{(E)}\}$ by simulating the system at the true parameter θ_o . For each dataset $D_N^{(e)}$, we then minimize (3a) to find an estimate $\hat{\theta}_{N,M}^{(e)}$ on a grid in the interval $[3.85, 10.25]$ with a stepsize 0.08. We assume a transient of 500 time-steps which are omitted when computing (3a). The output function $y(t)$ is computed by the DAE solver whose relative and absolute tolerances are set to 10^{-6} and 10^{-9} , respectively. For comparison we include an output error method ignoring the processes disturbance, i.e. (8) and (7) with $w(t) = 0$ and $v(t) = 0$. We summarize the results for different choices of the number of time-steps N in Figure 1. These results show that the proposed method outperforms the output error method starting from $N = 5000$. The median and mean of the proposed method converges on θ_o while the median and mean of the output error method lies at 7.21 and 7.20, respectively, at $N = 50000$. This is due to bias introduced by neglecting the process disturbance in the output error model. In contrast, the proposed method has a bias of $|\frac{1}{E} \sum_{e=1}^E \hat{\theta}_{N,M}^{(e)} - \theta_o| = 0.0343$ at $N = 50000$.

B. Influence of fixing noise realizations

Here we demonstrate how the proposed method of fixing the noise realizations $\{z_m(\tau_k)\}$ improves smoothness of the cost function. This approach is compared to a method where, for every new value of θ , the values $\{x_w^{(m)}(\tau_k; \theta)\}$ are obtained by sampling them anew from their distribution. The cost function is computed for the same model as in Section V-A. To exaggerate the results, 20 realizations were used ($M = 20$). For larger M , because of the averaging of the different realizations, even when the a priori samples of the process disturbance are sampled anew for every value of θ the cost function will become increasingly smooth. However, as one can see in Figure 2, the proposed method clearly results in a smooth cost function even when only a few realizations are used. We expect the effect of this approach to be even more significant if e.g. a parametrized disturbance model is used.

VI. DISCUSSION

As seen from the results in the previous section, for the model used, both the bias and the variance of the proposed method approach zero as N increases. This is in contrast

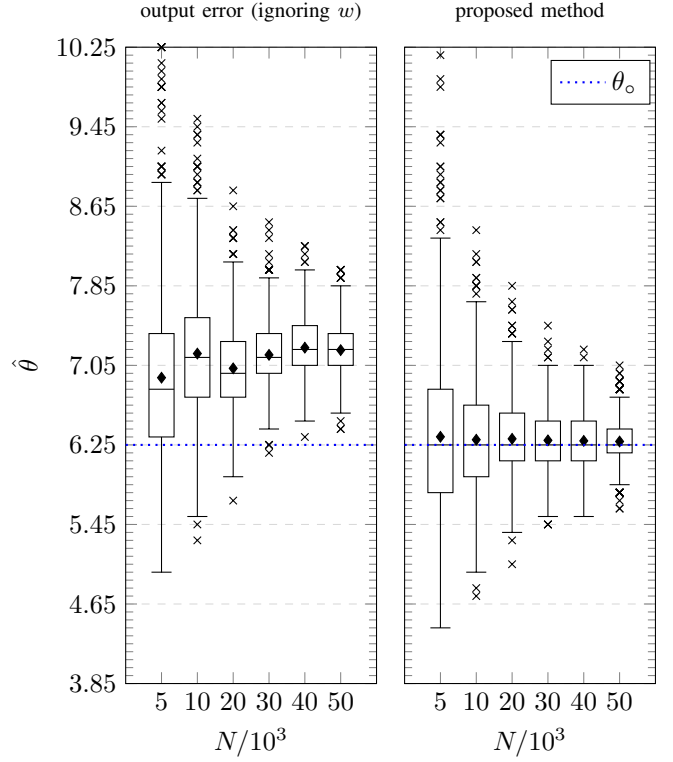


Fig. 1. Comparison between an output error method (which ignores $w(t)$) and the proposed method. The boxplots summarize the statistics of $\hat{\theta}_{N,M}^{(e)}$ over $E = 1000$ datasets. The number of time-steps N is indicated on the x-axis and the transient is assumed to be 500 time-steps. For the proposed method, $M = 500$. The dotted line indicates the true parameter $\theta_o = 6.25$. The horizontal line and the diamond in each box denotes the median and the mean, respectively.

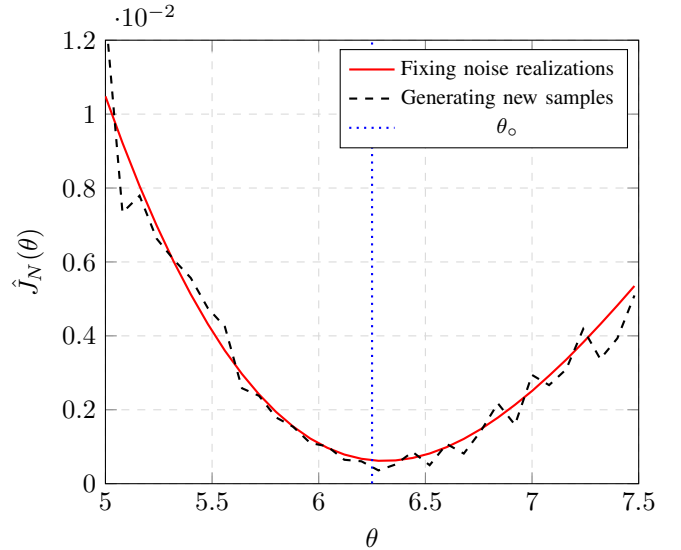


Fig. 2. Comparison of the cost function when using the proposed disturbance sampling method conditioning on pre-generated trajectory and when using the method where a disturbance realization is sampled anew for every new value of the parameters. The system was simulated for 10 s with $M = 20$ and $N_w = 100$. The true value of the parameter θ_o is marked by the dotted vertical line.

to the alternative method where the process disturbance is ignored. All realizations of $w(t)$ were approximated using linear interpolation, as described in Section IV. Recall that the target realizations are smooth, and notice that the error caused by this approximation will be relatively small whenever the sampling time Δ is sufficiently small. Thus, the obtained results would not be drastically different if exact sampling were used.

On the other hand, in cases where the process disturbance is multi-dimensional with unknown parameters, the performance of such a simple interpolation method may not be acceptable: If each component of the process disturbance is interpolated independently of the others, the interpolated realization might not capture existing dependence between the components of the disturbance. In such a situation, an exact method based on conditional sampling should be applied. Namely, the model (4) can be used to find the conditional distribution of $w^{(m)}(s; \theta)$, at an arbitrary value $s \in [0, T]$, given an a priori uniformly sampled realization $\{w^{(m)}(\tau_k; \theta)\}_k$. Such an exact sampling method has several computational challenges. For example, to retain the correct spectrum of the realization, every new sample has to be stored and used for future conditioning. In principle, the innovations used to generate the conditional samples can be a priori generated and fixed. However, this method would require a significant memory capacity, depending on how large the data set is, and how many samples are requested by the solver. A potential solution here is to use pseudo-random number generators with fixed seeds, instead of a priori generating and storing the samples. These challenges and ideas are currently on our future research agenda.

VII. CONCLUSIONS

In this contribution, we proposed a simulation-based prediction error method for estimation parameters of nonlinear DAE models with process disturbance. The method is applicable to general high-index non-linear DAEs that can be reduced to a form suitable for commonly used off-the-shelf numerical solvers (such as SUNDIALS). The process disturbances are modeled as a continuous-time stochastic process with rational spectrum. The method does not require the solution of any non-linear filtering problems, and therefore we are able to avoid some of the computational difficulties and limitations of particle filter methods. We provided an implementation where, parameter-independent, a priori generated realizations of uniformly sampled standard Gaussian process are used to approximate the used cost function. This lessens the effects due to random variations caused by the stochasticity of the algorithm itself, resulting in a smoother behavior that is less prone to issues with local minima. Under certain standard regularity conditions, the obtained estimators are expected to be consistent, which is supported by a numerical simulation experiment.

REFERENCES

- [1] D. Broman, "Meta-Languages and Semantics for Equation- Based Modeling and Simulation," Ph.D. dissertation, Department of Computer and Information Science, Linköping University, Sweden, 2010.
- [2] A. C. Hindmarsh, P. N. Brown, K. E. Grant, S. L. Lee, R. Serban, D. E. Shumaker, and C. S. Woodward, "SUNDIALS: Suite of nonlinear and differential/algebraic equation solvers," *ACM Transactions on Mathematical Software (TOMS)*, vol. 31, no. 3, pp. 363–396, 2005.
- [3] K. E. Brenan, *Numerical solution of initial-value problems in differential-algebraic equations*, ser. Classics in applied mathematics ; 14. SIAM, 1996.
- [4] A. Hagenblad, L. Ljung, and A. Wills, "Maximum likelihood identification of wiener models," *Automatica*, vol. 44, no. 11, pp. 2697–2705, 2008.
- [5] T. B. Schön, F. Lindsten, J. Dahlin, J. Wågberg, C. A. Naesseth, A. Svensson, and L. Dai, "Sequential Monte Carlo methods for system identification," *IFAC-PapersOnLine*, vol. 48, no. 28, pp. 775 – 786, 2015.
- [6] J. Schoukens and L. Ljung, "Nonlinear system identification: A user-oriented road map," *IEEE Control Systems Magazine*, vol. 39, no. 6, pp. 28–99, Dec 2019.
- [7] W. R. Esposito and C. A. Floudas, "Global Optimization for the Parameter Estimation of Differential-Algebraic Systems," *Industrial & Engineering Chemistry Research*, vol. 39, no. 5, pp. 1291–1310, 2000.
- [8] H. G. Bock, E. Kostina, and J. P. Schlöder, "Numerical Methods for Parameter Estimation in Nonlinear Differential Algebraic Equations," *GAMM-Mitteilungen*, vol. 30, no. 2, pp. 376–408, 2007.
- [9] M. Gerdin and J. Sjöberg, "Nonlinear stochastic differential-algebraic equations with application to particle filtering," in *Proceedings of the 45th IEEE Conference on Decision and Control*, 2006, pp. 6630–6635.
- [10] P. Mobed, S. Munusamy, D. Bhattacharyya, and R. Rengaswamy, "State and parameter estimation in distributed constrained systems. 1. Extended Kalman filtering of a special class of differential-algebraic equation systems," *Industrial and Engineering Chemistry Research*, vol. 56, no. 1, pp. 206–215, 2017.
- [11] M. Gerdin, T. B. Schön, T. Glad, F. Gustafsson, and L. Ljung, "On parameter and state estimation for linear differential-algebraic equations," *Automatica*, vol. 43, no. 3, pp. 416–425, 2007.
- [12] M. Gerdin, "Identification and estimation for models described by differential-algebraic equations," Doctoral Thesis, Linköpings universitet, Sweden, 2006.
- [13] L. Ljung, *System Identification: Theory for the User*, 2nd ed. Prentice Hall, 1999.
- [14] M. R.-H. Abdalmoaty and H. Hjalmarsson, "Linear prediction error methods for stochastic nonlinear models," *Automatica*, vol. 105, pp. 49–63, 2019.
- [15] M. R.-H. Abdalmoaty and H. Hjalmarsson, "Simulated Pseudo Maximum Likelihood Identification of Nonlinear Models," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 14 058–14 063, 2017.
- [16] M. R.-H. Abdalmoaty, "Identification of stochastic nonlinear dynamical models using estimating functions," Ph.D. dissertation, KTH Royal Institute of Technology, 2019.
- [17] P. Kunkel, *Differential-algebraic equations : analysis and numerical solution*, ser. EMS Textbooks in Mathematics. European Mathematical Society, 2006.
- [18] K. J. Åström, *Introduction to stochastic control theory*, ser. Mathematics in science and engineering ; v. 70. Academic Press, 1970.
- [19] C. Van Loan, "Computing integrals involving the matrix exponential," *IEEE Transactions on Automatic Control*, vol. 23, no. 3, pp. 395–404, 1978.
- [20] P. N. Brown, A. C. Hindmarsh, and L. R. Petzold, "Using krylov methods in the solution of large-scale differential-algebraic systems," *SIAM J. Sci. Comput.*, vol. 15, no. 6, p. 1467–1488, Nov. 1994.
- [21] L. R. Petzold, "Description of DASSL: a differential/algebraic system solver," Sandia National Labs., Livermore, CA (USA), Tech. Rep. SAND-82-8637; CONF-820810-21, Sep. 1982.
- [22] P. Glasserman and D. D. Yao, "Some guidelines and guarantees for common random numbers," *Management Science*, vol. 38, no. 6, pp. 884–908, 1992.
- [23] C. Rackauckas and Q. Nie, "DifferentialEquations.jl—a performant and feature-rich ecosystem for solving differential equations in Julia," *Journal of Open Research Software*, vol. 5, no. 1, 2017.
- [24] K. E. Brenan, S. L. Campbell, and L. R. Petzold, *Numerical Solution of Initial-Value Problems in Differential-Algebraic Equations*, ser. Classics in Applied Mathematics. SIAM, 1995.
- [25] C. Gear, B. Leimkuhler, and G. Gupta, "Automatic integration of euler-lagrange equations with constraints," *Journal of Computational and Applied Mathematics*, vol. 12-13, pp. 77–90, 1985.