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

Nested Sampling Approach to Set-membership Estimation ⋆

Radoslav Paulen * **Lucian Gomoescu** ** **Benoît Chachuat** **

* *Faculty of Chemical and Food Technology, Slovak University of Technology in Bratislava, Bratislava, Slovakia*
** *Centre for Process Systems Engineering, Department of Chemical Engineering, Imperial College London, United Kingdom*

Abstract: This paper is concerned with set-membership estimation in nonlinear dynamic systems. The problem entails characterizing the set of all possible parameter values such that given predicted outputs match their corresponding measurements within prescribed error bounds. Most existing methods to tackle this problem rely on outer-approximation techniques, which perform poorly when the parameter host set is large due to the curse of dimensionality. An adaptation of nested sampling—a Monte Carlo technique introduced to compute Bayesian evidence—is presented herein. The nested sampling algorithm leverages efficient strategies from Bayesian statistics for generating an inner-approximation of the desired parameter set. Several case studies are presented to demonstrate the approach.

Keywords: Monte Carlo sampling, Nested sampling, Set-membership estimation, Bounded-error identification.

1. INTRODUCTION

Mathematical modelling has become an integral part of process design and control methodologies as well as operations optimization. A typical model development procedure is divided into two main phases, namely specification of the model structure followed by parameter estimation. The latter phase, often referred to as model fitting, proceeds by determining parameter values for which the model predictions closely match the available process measurements.

Most commonly, parameter estimation is posed as an optimization problem that minimizes the gap between the measurements and the model predictions, for instance in the least-squares sense. Several factors can impair a successful and reliable estimation procedure in practice. First of all, structural model mismatch is inherent to the modeling exercise, and it is illusive to look for the 'true' parameter values in this context. And even in the absence of structural mismatch, fitting a set of experimental data perfectly generally proves impossible due to various sources of uncertainty. A measurement's accuracy is always tied to the resolution of the corresponding apparatus and measured data are furthermore corrupted with noise.

Of the available approaches that account for uncertainty in parameter estimation, the focus in this paper is on *set-membership* estimation (SME) (Walter, 1990). This approach entails the determination of *all* parameter values—referred to as the feasible parameter set subsequently—that are consistent with the measurements under given uncertainty scenarios. Specifically, we consider the case where the uncertainty enters the estimation problem in the form of bounded measurement errors. An inherent advantage of this approach over more traditional parameter estimation is that no (consistent) solution to the problem will be lost, and this can help detect problems arising due to lack of identifiability. Moreover, the estimation process does not rely on a particular statistical description of the uncertainty, as is typically the case when applying maximum likelihood or Bayesian estimation techniques.

Most existing approaches to tackling SME for general nonlinear dynamic systems aim to outer-approximate the feasible parameter set. Set-theoretic methods construct an enclosure of the feasible parameter set in the form of a subpaving, which relies on the ability to compute tight bounds on the set of responses of the dynamic system (Raissi et al., 2004; Kieffer and Walter, 2011; Streif et al., 2012; Hast et al., 2015; Paulen et al., 2016). By contrast, optimization-based methods compute a simple enclosure of the feasible parameter set—usually a box—by solving a collection of dynamic optimization subproblems to global optimality (Gottu Mukkula and Paulen, 2017; Walz et al., 2018). Methods have also been developed that compute an inner-approximation of the feasible parameter set (Streif et al., 2013). However, all of these approaches are afflicted by the curse of dimensionality and often perform poorly when the parameter host set is large.

The focus in this paper is on sampling techniques to inner-approximate the feasible parameter set, which has received little attention in the literature so far. Previous work includes the use of gridding techniques (**?**) and Markov Chain Monte Carlos (MCMC) methods (Bai et al., 2015). Herein, we present a novel algorithm based on nested sampling that leverages efficient strategies from Bayesian statistics for generating a dense sample in the feasible parameter set. We demonstrate this approach with several case studies.

⋆ Corresponding author: radoslav.paulen@stuba.sk

2. PROBLEM STATEMENT

Consider a process model in the form parametric ODEs:

$$\dot{\boldsymbol{x}}(t;\boldsymbol{p}) = \boldsymbol{f}(\boldsymbol{x}(t;\boldsymbol{p}),\boldsymbol{p}), \quad \boldsymbol{x}(t_0;\boldsymbol{p}) = \boldsymbol{h}(\boldsymbol{p}), \tag{1a}$$

$$\hat{\boldsymbol{y}}(t;\boldsymbol{p}) = \boldsymbol{g}(\boldsymbol{x}(t;\boldsymbol{p}),\boldsymbol{p}), \tag{1b}$$

where $\boldsymbol{x}$ denotes the n_x-dimensional vector of process states; $\boldsymbol{p}$, the n_p-dimensional vector of process (a priori unknown) parameters; and $\hat{\boldsymbol{y}}$, the n_y-dimensional vector of model outputs (predictions).

Given a set of output measurements $\boldsymbol{y}_\mathrm{m}$ at N time points $t_1,\dots,t_N$, *classical* estimation seeks for *one* particular instance $\boldsymbol{p}_\mathrm{e} \in \mathbb{R}^{n_p}$ for which the (possibly weighted) residuals between these measurements and the corresponding model outputs $\hat{\boldsymbol{y}}$ are minimized. In contrast, *set-membership* (bounded-error) estimation accounts for the fact that the actual process outputs, $\boldsymbol{y}_\mathrm{p}$, are only known within some bounded measurement errors $\boldsymbol{E} := [-\bar{\boldsymbol{e}},\bar{\boldsymbol{e}}]$:

$$\boldsymbol{y}_\mathrm{p}(t_i) \in \boldsymbol{y}_\mathrm{m}(t_i) + [-\bar{\boldsymbol{e}},\bar{\boldsymbol{e}}] := \boldsymbol{Y}_i. \tag{2}$$

Here, the objective is to characterize the set $\mathcal{P}_\mathrm{e}$ of *all* values of $\boldsymbol{p}$ such that $\hat{\boldsymbol{y}}(t_i;\boldsymbol{p}) \in \boldsymbol{Y}_i$ for every $i = 1,\dots,N$:

$$\mathcal{P}_\mathrm{e} := \left\{ \boldsymbol{p} \in \mathcal{P}_0 \;\middle|\; \begin{array}{l} \exists \boldsymbol{x}, \hat{\boldsymbol{y}} \text{ such that:} \\ \dot{\boldsymbol{x}}(t;\boldsymbol{p}) = \boldsymbol{f}(\boldsymbol{x}(t;\boldsymbol{p}),\boldsymbol{p}), \\ \boldsymbol{x}(t_0;\boldsymbol{p}) = \boldsymbol{h}(\boldsymbol{p}) \\ \hat{\boldsymbol{y}}(t_i;\boldsymbol{p}) = \boldsymbol{g}(\boldsymbol{x}(t_i;\boldsymbol{p}),\boldsymbol{p}) \\ \boldsymbol{e}(t_i;\boldsymbol{p}) := \hat{\boldsymbol{y}}(t_i;\boldsymbol{p}) - \boldsymbol{y}_\mathrm{m}(t_i) \in [-\bar{\boldsymbol{e}},\bar{\boldsymbol{e}}] \\ \forall t \in [t_0,t_N], \forall i \in \{1,\dots,N\} \end{array} \right\}. \tag{3}$$

Notice that the above formulation encompasses estimation problems in algebraic equations too, e.g., when the initial-value problem (1a) is replaced by a static model or the output function in (1b) does not depend on the differential states.

3. NESTED SAMPLING FOR BAYESIAN INFERENCE

In 2004, Skilling (2004) proposed a new Monte Carlo method called *Nested Sampling* for approximating the expected value of a positive-definite function F of a random variable $\boldsymbol{p} \sim f$:

$$\mathbb{E}[F(\cdot) \mid f(\boldsymbol{p})] = \int_{\boldsymbol{p}} F(\boldsymbol{p})\, f(\boldsymbol{p})\, \mathrm{d}\boldsymbol{p}. \tag{4}$$

The particular focus in Skilling's paper was on computing the Bayesian evidence Z, namely the expected value of a likelihood function (the function F) given a prior distribution (the probability density function f of $\boldsymbol{p}$) for given model parameters (the random variables $\boldsymbol{p}$).

Nested sampling estimates the Bayesian evidence by transforming the multi-dimensional integral (4) over the prior density into the following one-dimensional integral:

$$\mathbb{E}[F(\cdot) \mid f(\boldsymbol{p})] = \int_0^1 F(P)\, \mathrm{d}P, \tag{5}$$

where the variable $P(\lambda)$ corresponds to the integral over the subset contained within the iso-contour $F(\boldsymbol{p}) = \lambda$:

$$P(\lambda) = \int_{\{\boldsymbol{p}:F(\boldsymbol{p})>\lambda\}} f(\boldsymbol{p})\, \mathrm{d}\boldsymbol{p}. \tag{6}$$

Using this reformulation the problem reduces to finding an ordered set of values $\{(P_i,\lambda_i) : i = 1,\dots,M\}$ in order to approximate the one-dimensional integral in (5) as:

Algorithm 1 Nested sampling (N_live, N_prop, F, f)

1: Create a set $\mathcal{L} \leftarrow \{\boldsymbol{p}_j \sim f(\boldsymbol{p}) \,|\, j \in \{1,\dots,N_\mathrm{live}\}\}$
2: Initialize $N_\mathrm{nest} \leftarrow 0$ and $\mathcal{D} \leftarrow \emptyset$
3: Set $F^\star \leftarrow \min_{\boldsymbol{p}\in\mathcal{L}} F(\boldsymbol{p})$ and $\boldsymbol{p}^\star \leftarrow \arg\min_{\boldsymbol{p}\in\mathcal{L}} F(\boldsymbol{p})$
4: **repeat**
5: Create a set $\mathcal{R} \leftarrow \{\bar{\boldsymbol{p}}_k \sim f(\boldsymbol{p}) \,|\, k \in \{1,\dots,N_\mathrm{prop}\}\}$
6: **for all** $\bar{\boldsymbol{p}}_k \in \mathcal{R}$ **do**
7: **if** $F(\bar{\boldsymbol{p}}_k) \geq F^\star$ **then**
8: Update $\mathcal{L} \leftarrow \mathcal{L} \cup \{\bar{\boldsymbol{p}}_k\} \setminus \{\boldsymbol{p}^\star\}$
9: Update $F^\star$ and $\boldsymbol{p}^\star$ (as in line 3)
10: Get the mass $P_\mathrm{nest} \leftarrow \exp(-N_\mathrm{nest}/N_\mathrm{live})$
11: Update $\mathcal{D} \leftarrow \mathcal{D} \cup \{N_\mathrm{nest}, P_\mathrm{nest}, \boldsymbol{p}^\star, F^\star\}$
12: Increment $N_\mathrm{nest} \leftarrow N_\mathrm{nest} + 1$
13: **end if**
14: **end for**
15: **until** Stop Criterion
16: **return** live points $\mathcal{L}$ and dead points $\mathcal{D}$

$$\mathbb{E}[F(\cdot) \mid f(\boldsymbol{p})] \approx \sum_{i=1}^{M} \lambda_i\, (P_{i-1} - P_i) =: \hat{Z}. \tag{7}$$

Although $F(P)$ is unknown, we can turn to Monte Carlo methods for the probabilistic association of prior volumes P_i with iso-contours, $\lambda_i = F(P_i)$ based on (6).

A pseudo-code for the nested sampling algorithm is presented in Algorithm 1. The initialization proceeds by sampling N_live points from the distribution f (line 1), called *live points*, and sorting them by F value to identify the lowest value $F^\star$ and corresponding point $\boldsymbol{p}^\star$ (line 3).

Each iteration starts by creating a set $\mathcal{R}$ of N_prop replacement proposals (line 5). In the Bayesian estimation context, this step requires sampling from the prior subject to a likelihood constraint, which can prove challenging for non-uniform prior distributions. This step is also critical to the overall efficiency of nested sampling and to avoid biasing the estimation towards low or high likelihood regions. We also note that the largest computational effort is typically incurred by the evaluation of the replacement proposals and executing these evaluations in parallel can therefore lead to considerable speed-up.

Each replacement proposal $\bar{\boldsymbol{p}}_k$ is evaluated and replaces $\boldsymbol{p}^\star$ in the live point set $\mathcal{L}$ if its F value is larger than $F^\star$. Upon repeating this process the live samples move through nested shells of increasing F value and the enclosed probability mass in each nest is progressively reduced.

The probability mass contained within each nest is a random variable distributed as $P_\mathrm{nest} = t_\mathrm{nest} P_\mathrm{nest-1}$, where t_nest follows the distribution for the largest of N_live samples drawn uniformly from the interval $[0,1]$ (Feroz et al., 2009). It follows that the probability mass after N_nest successful replacements can be estimated as (line 10)

$$P_\mathrm{nest} \approx \exp(-N_\mathrm{nest}/N_\mathrm{live}). \tag{8}$$

Each discarded point $\boldsymbol{p}^*$ is referred to as a *dead point* and stored into the set $\mathcal{D}$ with its F value and probability mass.

The iterations are usually interrupted when the expected contribution from the current set of live points is less than a user-defined tolerance. For instance, this expected remaining contribution can be estimated as $\Delta Z_\mathrm{nest} = F^* P_\mathrm{nest}$, with $F^* := \max_{\boldsymbol{p}\in\mathcal{L}} F(\boldsymbol{p})$. A typical stopping

criterion in Bayesian estimation is when the remaining contribution is below 10%.

Distinctive features of nested sampling over other Monte Carlo techniques can be summarized as follows:

- The sampling proceeds by gradually moving toward higher likelihood regions and the probability mass enclosed in the nest shrinks exponentially—see (8);
- The tail of the posterior distribution is sampled sufficiently to avoid leaving out regions that contribute to the Bayesian evidence;
- The likelihood function evaluations can be easily parallelized;
- The algorithm can be applied to multimodal likelihood functions insofar as a global-search strategy is used to generate the replacement proposals.

4. NESTED SAMPLING FOR SET-MEMBERSHIP ESTIMATION

A straightfoward application of nested sampling to the set-membership (NS-SME) context entails using an indicator function to define the prior parameter distribution and the log-likelihood function (Bai et al., 2015):

$$f(\boldsymbol{p}) := \log(I_f(\boldsymbol{p})), \text{ with } I_f(\boldsymbol{p}) := \begin{cases} 1, & \text{if } \boldsymbol{p} \in \mathcal{P}_0, \\ 0, & \text{otherwise,} \end{cases} \quad (9)$$

and

$$F(\boldsymbol{p}) := \log(I_F(\boldsymbol{p})), \text{ with } I_F(\boldsymbol{p}) := \begin{cases} 1, & \text{if } \boldsymbol{p} \in \mathcal{P}_\mathrm{e}, \\ 0, & \text{otherwise.} \end{cases} \quad (10)$$

In this formulation, however, the progress of Algorithm 1 is hindered by the difficulty of generating a proposal that belongs to the feasible parameter set $\mathcal{P}_\mathrm{e}$. For problems whereby the initial parameter set $\mathcal{P}_\mathrm{e}$ is a small subset of $\mathcal{P}_0$, the likelihood of the iso-contour F^* might be stuck at $-\infty$ during most of the iterations. In response to this, we introduce the following likelihood function:

$$I_F(\boldsymbol{p}) := \begin{cases} 1, & \text{if } \boldsymbol{p} \in \mathcal{P}_\mathrm{e}, \\ \prod_{i=1}^{N} \mathrm{e}^{-\frac{1}{2}\boldsymbol{e}^\intercal(t_i;\boldsymbol{p})\boldsymbol{Q}\boldsymbol{e}(t_i;\boldsymbol{p})}, & \text{otherwise,} \end{cases} \quad (11)$$

with the diagonal matrix $\boldsymbol{Q} := [\mathrm{diag}\left(\frac{1}{3}\bar{\boldsymbol{e}}\right)]^{-2}$ and $\bar{\boldsymbol{e}}$ as in (2). In essence, (11) adds to $I_F(\boldsymbol{p})$ a Gaussian tail with $\sigma_i = \bar{e_i}/3$ for the i^{th} measurement while keeping $F(\cdot)$ steep near the boundary of $\mathcal{P}_\mathrm{e}$, which does not impair the estimation quality (see Sec. 4.1 for further discussion).

The NS-SME approach is non-intrusive since it relies on the result of model simulations at given parameter values only. This allows for SME to be applied to black-box models in process flowsheeting or CFD modeling, in principle. Furthermore, NS-SME does not assume any particular shape or connectedness of the set $\mathcal{P}_\mathrm{e}$. In practice, one can exploit the live point set $\mathcal{L}$ returned by Algorithm 1 to train a data classifier, possibly combined with a softmax function to estimate the probability of a given point to belong to $\mathcal{P}_\mathrm{e}$ (**?**).

4.1 Implementation

An implementation of Algorithm 1 in a Python package called DEUS (standing for *DEsign under Uncertainty using Sampling methods*) is available from: `https://github.com/omega-icl/DEUS`. Input files for all the case studies can also be retrieved from this link. At each iteration, the replacement candidates are generated by sampling in ellipsoids enclosing the current live points, which are generated using the X-Means clustering algorithm (Pelleg and Moore, 2000). The tuning parameters in DEUS include: (i) the number of live points, N_live; (ii) the number of replacement candidates at each iteration, N_prop; (iii) the initial enlargement factor of the ellipsoid; and (iv) the shrinking rate of the enlargement factor at each iteration.

DEUS is used to solve all of the numerical case studies presented below, using as default parameters an initial enlargement factor of 30% and a shrinking rate of 0.2 for the ellipsoids (Feroz et al., 2009). The algorithm is interrupted once the expected contribution from the current set of live points is below 10%. This choice is motivated by the likelihood function (11) as we expect only a negligible contribution from outside of $\mathcal{P}_\mathrm{e}$ region and a low ratio between the size of $\mathcal{P}_e$ and of $\mathcal{P}_0$. Should these assumptions fail to hold, alternative stopping criteria may be used, such as interrupting the algorithm as soon as the number of dead and live points within $\mathcal{P}_\mathrm{e}$ is greater than a chosen threshold value M.

Upon termination, the set of dead and live points with $F^* = 0$ yields an inner-approximation of P_e. A suitable value for N_live depends on the accuracy with which a user wants to characterize the set $\mathcal{P}_\mathrm{e}$. We choose N_prop as one half of the number of live points here as a heuristic. This choice can be viewed as the same trade-off as exploration against exploitation in a stochastic-search algorithm. A high value of N_prop shifts the trade-off towards exploration by generating more points that are exploited within the ellipsoidal cluster(s) at each iteration of the algorithm.

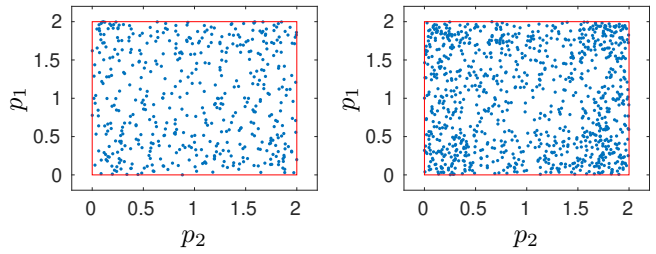
5. CASE STUDIES

This section presents several case studies that illustrate different aspects of the proposed NS-SME approach. All the results are obtained on a workstation with Intel Core i7-8565U processor at 1.80 GHz×8 with 16 GB RAM and running 64-bit Linux.

5.1 Sampling of a Simple Box

For the purpose of studying the performance of Algorithm 1 for SME, we consider the benchmark problem of characterizing an n_p-dimensional box $\mathcal{P}_\mathrm{e} := [0, 2]^{n_p}$ by sampling from an a priori box $\mathcal{P}_0 := [-10^m, 10^m]^{n_p}$ with $m \in \{1, \ldots, 5\}$ and $n_p \in \{2, \ldots, 10\}$. This problem can be cast as a parameter estimation problem, whereby the measurement function is $\boldsymbol{y}(t_i, \boldsymbol{p}) = \boldsymbol{p}$, the measurement error bound is $\bar{\boldsymbol{e}} = 1$, the true (but unknown) parameter values are $\boldsymbol{p} = (1, \ldots, 1)^\intercal$, and the actual measurements are noise-free.

We run Algorithm 1 for each value of n_p and m with $N_\mathrm{live} = 150$ and 300. The resulting inner-approximations of the set $\mathcal{P}_\mathrm{e}$ in Fig. 1 are for $n_p = 10$ and $m = 5$. Only the 2D projections onto the subspace (p_1, p_2) are represented since the other 2D projections look qualitatively alike. The effect of the number of live points on the density (accuracy) of the obtained inner-approximation is clear. The resulting



(a) $N_{\text{live}} := 150$ (b) $N_{\text{live}} := 300$

Fig. 1. Inner-approximation (blue dots) of a 10D box (in red) projected on the subspace (p_1, p_2).

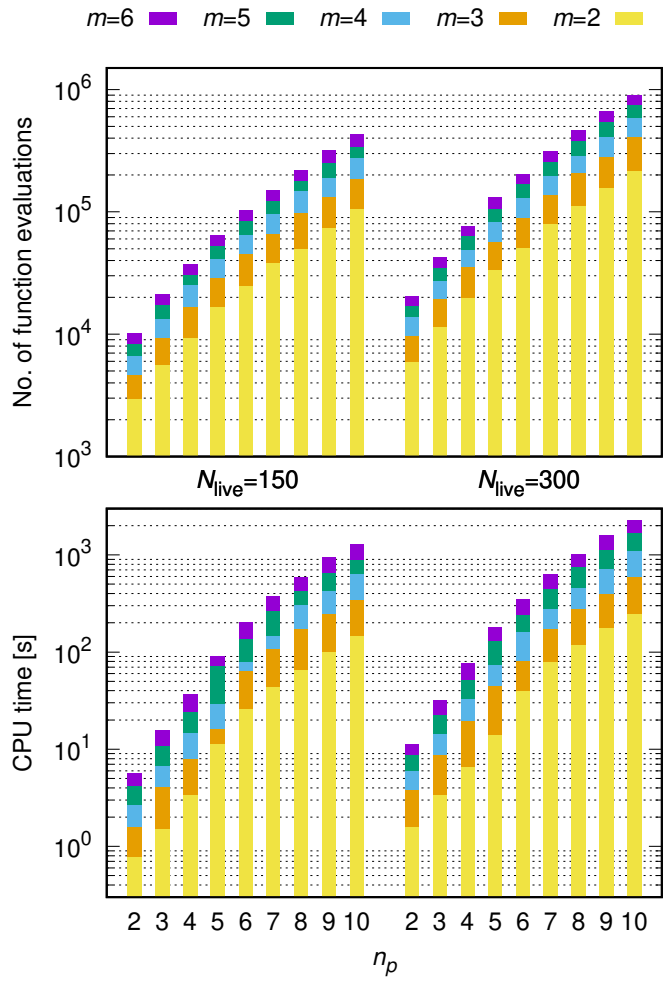


Fig. 2. Number of function evaluations (top) and CPU times (bottom) for various a priori sets $\mathcal{P}_0$.

inner-approximation comprises 496 points and requires ca. 428k function evaluations with $N_{\text{live}} = 150$, whereas it comprises 996 points and requires ca. 904k function evaluations with $N_{\text{live}} = 300$. For comparison, if 10^6 samples were to be generated using Sobol sampling in $\mathcal{P}_0$, none of these point would end up inside $\mathcal{P}_e$. We shall use $N_{\text{live}} := 300$ as a suitable compromise between solution speed and accuracy in the rest of the paper.

To asses the computational attributes of NS-SME, Fig. 2 shows the number of evaluations and CPU time required for various values of m and n_p. The cases with $N_{\text{live}} = 150$ and 300 are reported on the left and right parts of each plot, respectively. Among the factors that make the SME problem more challenging, enlarging the width of the a priori set $\mathcal{P}_0$ is found to have a rather small effect of the solution speed—almost constant effect for a 10^4-fold enlargement at various dimensions n_p. This is a favorable result as the complexity of current set-theoretic SME techniques is far worse. On the other hand, doubling the number of live points (and the number proposal points) has roughly the same effect as either adding three dimensions to the problem or enlarging the search domain by 10,000 times. This is computationally unfavourable and it shows that it would be naïve to think of replacing set-theoretic SME techniques with nested sampling in problems where the full, validated feasible parameter set is required. Finally, we note that the difference in performance of NS-SME with the proposed likelihood function (11) compared to the indicator-based likelihood function (10) is huge—in the 2D case with $m = 2$, for instance, the former terminates after 4 CPU-sec and the latter after 63 CPU-sec.

5.2 Static Nonlinear Estimation

This example is adapted from Jaulin and Walter (1993):

$$\hat{y}(t_i; \boldsymbol{p}) = p_1 \exp(p_2 x(t_i)), \tag{12}$$

The a priori parameter set is chosen as $\mathcal{P}_0 := [-10, 10]^2$. A total of $N = 11$ noisefree measurements $y_{\text{m}}(t_i)$ are considered at equidistant points $x(t_i) = 0, 0.1, \ldots, 1.0$, as generated from (12) using the parameter values $\boldsymbol{p}^* := (1, 1)^\intercal$. The error bound is furthermore set to $\bar{e} = 1$.

The top plot of Fig. 3 shows the inner-approximation set computed with Algorithm 1 together with the exact SME constraints as solid lines—the red lines correspond to the constraints $\hat{y}(t_i; \boldsymbol{p}) = -\bar{e}$ and the blue ones to $\hat{y}(t_i; \boldsymbol{p}) = \bar{e}$. We can see that, despite the skewed shape of the SME set, the resulting inner-approximation is very good. The bottom plot shows the rejected proposals (dead points) alongside the inner-approximation (live points). Observe the good spread of the samples over $\mathcal{P}_0$ and how they concentrate towards the desired solution set $\mathcal{P}_e$. The presented solution is obtained within 7 CPU-sec using 7,200 likelihood function evaluations, which is in good agreement with the analysis in Sec. 5.1.

5.3 Dynamic Estimation under Lack of Identifiabilty

This example from Kieffer and Walter (2011) considers a dynamic model with two states $\boldsymbol{x} = (x_1, x_2)^\intercal$ and three uncertain parameters $\boldsymbol{p} = (p_1, p_2, p_3)^\intercal \in \mathcal{P}_0 := [0.01, 1]^3$:

$$\dot{x}_1(t; \boldsymbol{p}) = -(p_1 + p_3)x_1(t; \boldsymbol{p}) + p_2 x_2(t; \boldsymbol{p}), \tag{13a}$$

$$\dot{x}_2(t; \boldsymbol{p}) = p_1 x_1(t; \boldsymbol{p}) - p_2 x_2(t; \boldsymbol{p}), \tag{13b}$$

with the initial conditions $\boldsymbol{x}(0; \boldsymbol{p}) = (1, 0)^\intercal$. The system has a single output variable $\hat{y}(t; \boldsymbol{p}) = x_2(t, \boldsymbol{p})$, with $N = 15$ measurements corresponding to the time instants $t_i = 1, \ldots, 15$. Synthetic experimental data are generated by simulating the model (13) with the parameter values $\boldsymbol{p}^* = (0.6, 0.15, 0.35)^\intercal$, and then rounding the output $\hat{y}(t_i; \boldsymbol{p}^*)$ up or down to the nearest value by retaining two significant digits only. The measurement error range is thus taken as $\pm 5 \times 10^{-3}$.

The inner-approximation set computed with Algorithm 1 is shown in Fig. 4. This result is consistent with the outer-approximation sets presented in Kieffer and Walter

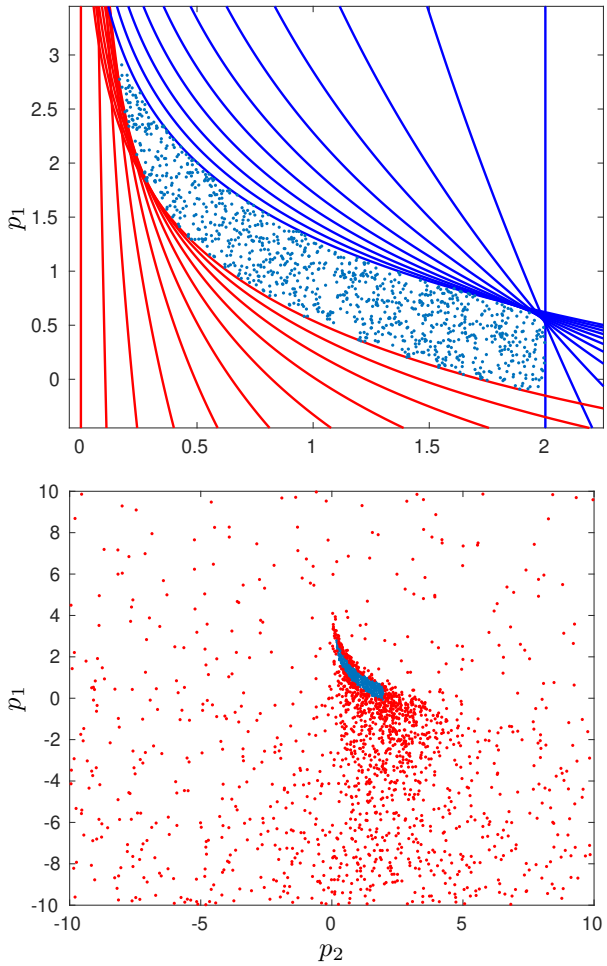


Fig. 3. Inner-approximation of the SME solution compared with the exact SME constraints (top) and with the rejected proposals (bottom).

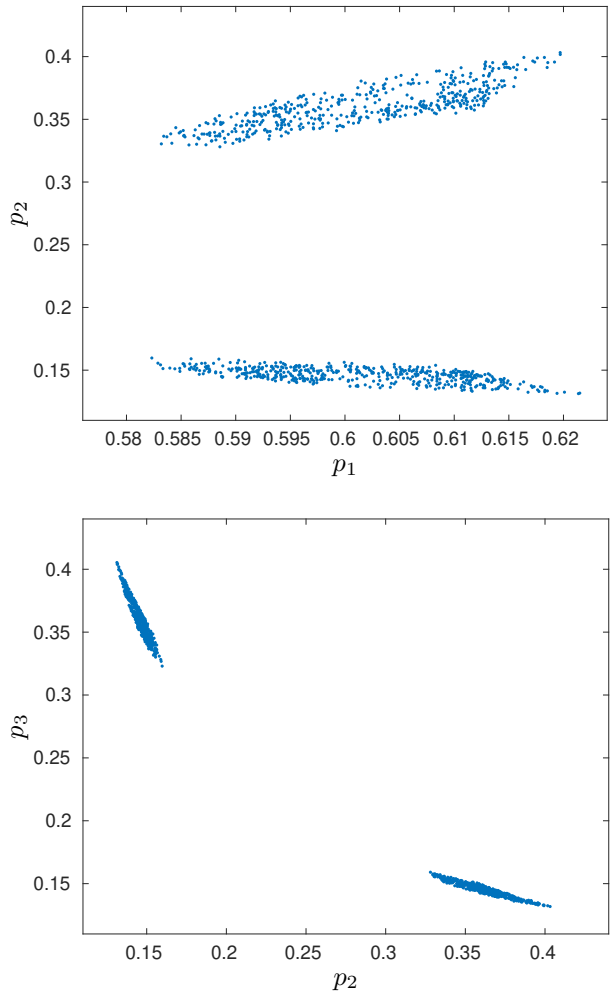


Fig. 4. Inner-approximation of the SME solution set projected onto the (p_1, p_2) subspace (top) and (p_2, p_3) subspace (bottom).

(2011) and Paulen et al. (2016) using set-theoretic SME methods. The disconnectedness of the set stems from a local identifiability problem, since the values of p_2 and p_3 are interchangeable when only x_2 is measured. The algorithm terminates after 477 CPU-sec and requires 267,600 likelihood function evaluations. This shows the challenging nature of this case study, which stems from the shape of the feasible solution set and the rather large initial bounds.

5.4 Joint Parameter/State SME of an Anaerobic digester

The final case study considers a six-state dynamic model describing an anaerobic digester (Bernard et al., 2001). The states represent the concentrations of acidogenic and methanogenic biomass; organic substrates (COD other than VFA); volatile fatty acids (VFA); total alkalinity concentration (TALK); and total inorganic carbon (TIC). The dilution rate profile, inlet concentrations of organic substrate, VFA, TALK and TIC, and other process parameters are given. The estimation problem is to estimate the parameters of the kinetic microbial growth model and the two initial biomass concentrations, as in Paulen et al. (2016). The specific growth rates of acidogenic bacteria and methanogenic bacteria are assumed to follow Monod and Haldane kinetics, respectively. The specific growth rates involve a total of five unknown parameters: two maximum growth rates, two half-saturation constants and an inhibition constant (methanogenic bacteria only).

Pseudo-experimental data are generated in the same way as Paulen et al. (2016) by simulating the model with the parameter values $\boldsymbol{p}^* = (1.2, 7.1, 0.74, 9.28, 256)^\intercal$ and initial states $\boldsymbol{x}(0)^* = (0.5, 1.0, 1.0, 5.0, 40.0, 50.0)^\intercal$. For the estimation, the a priori ranges ($\mathcal{P}_0$) are taken, in order, as [0.5, 1.5], [5.5, 8.0], [0.735, 0.745], [9.1, 9.35], [250.0, 265.0], [0.3, 0.7], [0.8, 1.2], [0.8, 1.2], [4.0, 6.0], [38.0, 42.0], [48.0, 52.0].

Three outputs are considered in the estimation problem, namely the organic substrate concentration, the volatile fatty acids concentration, and the total alkalinity concentration, with measurements every 4 hours. In order to simulate the effect of measurement noise, uniformly distributed random noise is added such that the maximal measurement error magnitudes are $\bar{\boldsymbol{e}} := (0.01, 0.1, 0.1)^\intercal$.

The inner-approximation set computed with Algorithm 1 is shown in Fig. 5 using various 2D projections. This result is consistent with the outer-approximation sets computed in Paulen et al. (2016), although the set-theoretic method used therein was not able to tackle a problem of this size.

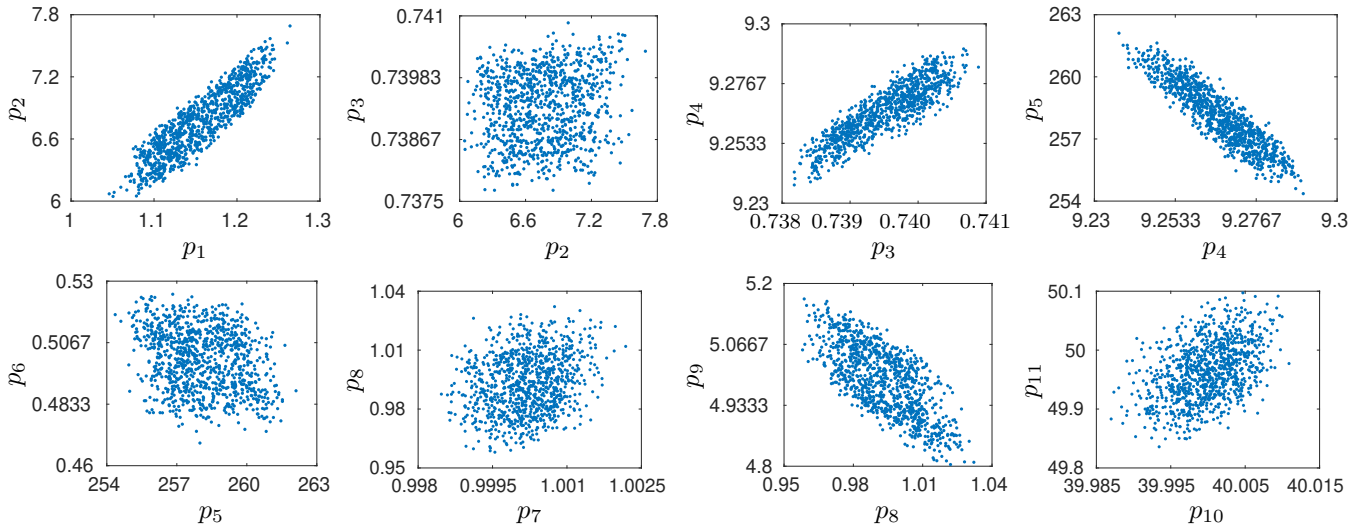


Fig. 5. Inner-approximations of the SME solution set.

The samples show a good spread and it is furthermore possible to analyze correlations between the parameters and/or initial conditions on this basis. The algorithm terminates after 3,729 CPU-sec and requires 578,250 likelihood function evaluations. The total number of function evaluations is in good agreement with the analysis in Sec. 5.1. By contrast, the computational overhead is due to the repeated numerical integration of the dynamic system using CVODE/Sundials.

6. CONCLUSIONS

This paper has presented an approach based on nested sampling to inner-approximate the feasible parameter set in set-membership estimation. The presented case studies confirm the viability of this approach as an alternative to state-of-the-art, outer-approximation approaches based on set-theoretic concepts. The combination of nested sampling and set-theoretic concepts into a hybrid algorithm constitutes a promising direction for future research.

ACKNOWLEDGEMENTS

Radoslav Paulen gratefully acknowledges the contribution of the European Commission under grant 790017 (IF GuEst), the contribution of the Scientific Grant Agency of the Slovak Republic under grant 1/0004/17, and the contribution of the Slovak Research and Development Agency under project APVV 15-0007. This work is also a partial result of the Research & Development Operational Programme for the project University Scientific Park STU in Bratislava, ITMS 26240220084, supported by the Research 7 Development Operational Programme funded by the ERDF. Part of the work was conducted when Lucian Gomoescu was a Marie Skłodowska-Curie early stage researcher at Process Systems Entreprise Ltd enrolled in the European Union's Horizon 2020 research and innovation program under grant agreement 675585 (ITN SyMBioSys).

REFERENCES

Bai, E., Ishii, H., and Tempo, R. (2015). A Markov Chain Monte Carlo Approach to Nonlinear Parametric System Identification. *IEEE Transactions on Automatic Control*, 60(9), 2542–2546.

Bernard, O., Hadj-Sadok, Z., Dochain, D., Genovesi, A., and Steyer, J.P. (2001). Dynamical model development and parameter identification for an anaerobic wastewater treatment process. *Biotechnology & Bioengineering*, 75(4), 424–438. doi:10.1021/ie0707725.

Feroz, F., Hobson, M.P., and Bridges, M. (2009). MultiNest: an efficient and robust Bayesian inference tool for cosmology and particle physics. *Monthly Notices of the Royal Astronomical Society*, 398(4), 1601–1614.

Gottu Mukkula, A.R. and Paulen, R. (2017). Model-based design of optimal experiments for nonlinear systems in the context of guaranteed parameter estimation. *Computers & Chemical Engineering*, 99, 198 – 213.

Hast, D., Findeisen, R., and Streif, S. (2015). Detection and Isolation of Parametric Faults in Hydraulic Pumps Using a Set-Based Approach and Quantitative & Qualitative Fault Specifications. *Control Engineering Practice*, 40, 61–70.

Jaulin, L. and Walter, E. (1993). Set inversion via interval analysis for nonlinear bounded-error estimation. *Automatica*, 29(4), 1053–1064.

Kieffer, M. and Walter, E. (2011). Guaranteed estimation of the parameters of nonlinear continuous-time models: contributions of interval analysis. *International Journal of Adaptive Control & Signal Processing*, 25(3), 191–207.

Paulen, R., Villanueva, M.E., and Chachuat, B. (2016). Guaranteed parameter estimation of non-linear dynamic systems using high-order bounding techniques with domain and cpu-time reduction strategies. *IMA Journal of Mathematical Control and Information*, 33(3), 563–587. doi:10.1093/imamci/dnu055.

Pelleg, D. and Moore, A. (2000). X-means: Extending k-means with efficient estimation of the number of clusters. In *In Proceedings of the 17th International Conf. on Machine Learning*, 727–734. Morgan Kaufmann.

Raissi, T., Ramdani, N., and Candau, Y. (2004). Set membership state and parameter estimation for systems described by nonlinear differential equations. *Automatica*, 40, 1771–1777.

Skilling, J. (2004). Nested sampling. *AIP Conference Proceedings*, 735(1), 395–405. doi:\doi{10.1063/1.1835238}.

Streif, S., Savchenko, A., Rumschinski, P., Borchers, S., and Findeisen, R. (2012). ADMIT: A Toolbox for Guaranteed Model Invalidation, Estimation, and Qualitative-Quantitative Modeling. *Bioinformatics*, 28, 1290–1291.

Streif, S., Strobel, N., and Findeisen, R. (2013). Inner approximations of consistent parameter sets by constraint inversion and mixed-integer programming. *IFAC Proceedings Volumes*, 46(31), 321–326. 12th IFAC Symposium on Computer Applications in Biotechnology.

Walter, E. (ed.) (1990). *Parameter Identifications with Error Bound*, volume 32 of *Mathematics & Computers in Simulation*. Elsevier.

Walz, O., Djelassi, H., Caspari, A., and Mitsos, A. (2018). Bounded-error optimal experimental design via global solution of constrained min-max program. *Computers & Chemical Engineering*, 111, 92–101.