A Markov Chain Monte Carlo Approach to Nonlinear Parametric System Identification

Er-Wei Bai, Hideaki Ishii, and Roberto Tempo

***Abstract*—Nonlinear system identification is discussed in a mixed set-membership and statistical setting. A Markov chain Monte Carlo (MCMC) approach is proposed that estimates the feasible parameter set, the minimum volume outer-bounding ellipsoid and the minimum variance estimate. The proposed algorithm is proved to be convergent and enjoys some desirable properties. Further, its computational complexity and numerical accuracy are studied.**

***Index Terms*—Monte Carlo, parameter estimation, system identification.**

I. Introduction

This technical note considers identification of a discrete time scalar nonlinear system parameterized by an unknown parameter vector. Clearly, identification methods and results depend on the characterization of the noise. For traditional probabilistic approaches, the noise is often assumed to have certain probabilistic properties and then asymptotic convergence analysis is carried out. On the other hand, the set-membership approach simply assumes a hard bound on the noise and the goal is to find the set of all parameters that are consistent with the system assumptions and the observed data, which is referred to as the feasible parameter set, e.g., see [3]–[5].

Set-membership identification has a long history and has regenerated interest in recent years [7], [9], [15], [18]. Finding the feasible parameter set is, however, very challenging, in particular for nonlinear systems [24]. An accurate feasible parameter set can be constructed but the computational complexity may grow exponentially [12]. In fact, it was shown in [24] that calculation of the feasible parameter set even for a simply structured nonlinear system, e.g., a Hammerstein or a Weiner system, is NP hard. To this end, one direction of research is to find the minimum volume outer-bounding ellipsoid of the feasible parameter set [10], [11]. Unfortunately, finding the minimum volume outer-bounding ellipsoid is still nonconvex and intractable in a nonlinear setting [8], [13].

Another direction of research is to combine the probabilistic and hard bound approaches, referred to as a mixed approach [5], [9], [16], [22]. The motivation is as follows. In the set-membership approach, the assumption is that only a hard bound, but no other *a priori* information on the noise is available. In some applications, engineers have more information than just a bound. For instance, it is often the case in practice that the noise is assumed to be a truncated Gaussian so not only the bound but the distribution inside the bound is available. In this mixed setting, it is assumed that the unknown sequence is strictly within a known hard bound, but at the same time is a random variable with support provided by the hard bound. Note that there are key differences between the set-membership and the mixed settings. In the set-membership approach, every parameter in the feasible parameter set is possibly the true but unknown parameter and it is impossible to say that one estimate is more likely to be the true but unknown parameter than any other. With a probabilistic distribution of the noise within the bound, it makes sense to ask which estimate is the "best" in some probabilistic sense since not all parameter estimates are equal. If the minimum variance estimate is preferred for a given observed data set, the best estimate is provided by the posterior conditional expectation.

The objective of this technical note is to estimate the feasible parameter set, the minimum volume outer-bounding ellipsoid and the minimum variance estimate and to establish their convergence and complexity properties. These three tasks are non-trivial individually and, moreover, they require different tools. For example, algorithms that aim at finding the minimum volume outer-bounding ellipsoid do not provide any insight for deriving the minimum variance estimator.

In this technical note, we show that these three tasks can be accomplished by one specific algorithm which enjoys some significant convergence properties. In particular, we compute estimates of the feasible parameter set, the minimum volume outer-bounding ellipsoid and the minimum variance estimate following a different direction. We remark that many numerical approaches have been developed with great success, e.g., the Markov chain Monte Carlo (MCMC) algorithm [14], [17], [19], [21], [23]. We follow the same direction and present a sampling based approach.

The contribution of the technical note is threefold. First, an MCMC algorithm to compute the feasible parameter set, its minimum volume outer-bounding ellipsoid and the minimum variance estimate is proposed, and its convergence and computational complexity properties are derived. Secondly, we show that the computational complexity of the proposed minimum volume outer-bounding ellipsoid is polynomial in the dimension, even if the feasible parameter set is nonconvex. The result is new in the context of set-membership identification and has not been discussed in the identification literature on MCMC. Finally, the assumption that the support of the target density is a connected set, which is standard in the existing identification literature on MCMC [17], is no longer required.

We would like to emphasize the differences between the Monte Carlo (MC) method and the MCMC which is adopted in this technical note. To apply the MC method, a set that outer-bounds the unknown feasible parameter set has to be known *a priori*. Then, we randomly generate points in the set according to, e.g., the uniform distribution. The points that satisfy the hard bound constraints are accepted and the others are rejected. This procedure leads us to an estimate of the feasible parameter set. The MCMC works in a different way and

Manuscript received June 30, 2014; revised October 16, 2014 and November 10, 2014; accepted December 1, 2014. Date of publication December 18, 2014; date of current version August 26, 2015. This work was supported in part by the NSF CNS-1329057 and by the European Union Seventh Framework Programme [FP7/2007-2013] under grant agreement 257462 HYCON2 Network of Excellence. Recommended by Associate Editor S. Anderson.

E.-W. Bai is with the Department of Electrical and Computer Engineering, University of Iowa, Iowa City, IA 52242 USA, and also the School of Electronics, Electrical Engineering, Queen's University, Belfast BT9 6AZ, U.K. (e-mail: er-wei-bai@uiowa.edu).

H. Ishii is with the Department of Computational Intelligence and Systems Science, Tokyo Institute of Technology, Yokohama 226-8503, Japan (e-mail: ishii@dis.titech.ac.jp).

R. Tempo is with IEIIT-CNR, Politecnico di Torino, Torino 10129, Italy (e-mail: roberto.tempo@polito.it).

Digital Object Identifier 10.1109/TAC.2014.2380997

actually generates a biased and correlated random sequence with the limiting distribution approaching the true but unknown distribution. It improves the convergence rate compared to the MC and this is the reason why the MCMC is widely used [14], [19], [23]. The MCMC algorithm is in fact cited as the top algorithm having the greatest influence on the development and practice of science and engineering in the 20th century.

The outline of the technical note is as follows. Section II provides the problem formulation. The algorithms and technical results are presented in Section III. Finite data length accuracy and computational complexity issues are discussed in Section IV. Numerical examples and discussions are provided in Section V. Some concluding remarks are given in Section VI.

II. Problem Statement

The discrete time scalar nonlinear system considered is of the form

$$\begin{aligned} y(k) &= f\left(y(k-1),\ldots,y(k-l),u(k-1),\right.\\ &\qquad \left.\ldots,u(k-l),\theta^*\right)+v(k)\\ &= f\left(\phi(k),\theta^*\right)+v(k),\ k=1,2,\ldots,n \end{aligned} \tag{II.1}$$

where $u(\cdot)$'s are the given system inputs, $y(\cdot)$'s the observed outputs and $\phi(k)=((y(k-1)\ \ldots\ y(k-l),u(k-1)\ \ldots\ u(k-l))^T$ the regressor, respectively. The nonlinear function $f(\phi(k),\theta^*):R\rightarrow R$ is a known measurable function which is parameterized by the unknown parameter vector $\theta^*\in R^q$. The sequence $v(k)$ indicates the noise. The purpose of identification is to estimate θ^* for a given input output data set $Z=\{\{y(k)\}_{k=1}^n,\{u(k)\}_{k=1}^n\}$.

Identification methods depend on the characterization of the unknown $v(\cdot)$'s. For the traditional set-membership approach, $v(\cdot)$ is assumed to be unknown but bounded by a known constant $\epsilon>0$,

$$|v(k)|\leq\epsilon,\quad \forall k. \tag{II.2}$$

For a given k, all parameter vectors θ that are consistent with the assumptions on the system (II.1) and the bound (II.2) as well as the observed data $\{y(k),\phi(k)\}$ can be described by

$$S_k=\left\{\theta\in R^q \mid |y(k)-f\left(\phi(k),\theta\right)|\leq\epsilon\right\}.$$

Then, the feasible parameter set is the intersections of S_k's

$$S=\bigcap_{k=1}^n S_k. \tag{II.3}$$

A goal of set-membership identification is to determine S for the given data set. Since determination of the exact S is a computationally difficult problem [24], a common approach is to approximate S by the minimum volume outer-bounding ellipsoid. It is well known that any ellipsoid in R^q can be written as $\{x \mid \|Ax-b\|_2\leq 1\}$ for some matrix $A\in R^{q\times q}$ and $b\in R^q$. It is also well known [6] that the volume of an ellipsoid is proportional to $\det(A^{-1})$. Thus, the minimum volume ellipsoid E containing S is given by

$$E=\{x \mid \|Ax-b\|_2\leq 1\} \tag{II.4}$$

where A and b are the solutions to

$$\min_{A,b}\det(A^{-1}),\ s.t.\ \sup_{x\in S}\|Ax-b\|_2\leq 1.$$

We now make an assumption on the feasible parameter set.

Assumption II.1:

1) The feasible parameter set S is bounded in R^q.
2) For any $\theta\in S$ and any small $\delta>0$, let $B_\delta(\theta)=\{x\in R^q \mid \|x-\theta\|_2\leq\delta\}$ be a δ ball centered at θ. Then, $\int_{x\in B_\delta(\theta)\cap S}dx>0$.

Boundedness on S is quite reasonable and standard because the objective of identification is to reduce uncertainty in the parameters. Clearly, if S is unbounded, the identification setting has to be redesigned or more data need to be collected. The meaning of the second part of the assumption is that any $\theta\in S$ is arbitrarily close to its interior. This assumption prevents pathological cases that S has empty interior.

In this technical note, we consider identification in a mixed set-membership setting. It is assumed that $v(\cdot)$ is i.i.d. (independent and identically distributed) and has a continuous positive density $p_v(\cdot)$ in $[-\epsilon,\epsilon]$ and $p_v(v(k))=0$ if $|v(k)|>\epsilon$ for all k. Note that there is a difference between the deterministic and mixed settings. In the deterministic setting, the answer to the question if a point lies in the feasible set or not is binary, either yes or no. In the mixed setting, the answer is given in probability. For a given noise density, if a subset $Q\subset R^q$ is in the feasible set or not is measured by a (true but unknown) distribution $\pi(Q)$. However, $\pi(Q)=0$ does not guarantee that Q is not in S, but it only says that this event occurs with probability zero.

In this technical note, we also derive the minimum variance estimate. For a given observed data set $Z=\{\{y(k)\}_{k=1}^n,\{u(k)\}_{k=1}^n\}$, the minimum variance estimate is the posterior conditional expectation [17], [19]

$$\theta_b=\mathbf{E}(\theta|Z)=\int_{\theta\in S}\theta\, p(\theta|Z)d\theta \tag{II.5}$$

where the unknown posterior conditional density is given by

$$p(\theta|Z)=\frac{p(Z|\theta)p(\theta)}{p(Z)}$$

and $p(Z)$ is the density of Z that is a constant with respect to θ and $p(\theta)$ is the prior density of θ. In this technical note, we assume that no *a priori* knowledge on the distribution of θ is available and therefore the prior density is uniform in S as

$$p(\theta)=\begin{cases}1/\mathrm{Vol}(S), & \text{if } \theta\in S\\ 0, & \text{if } \theta\notin S\end{cases}$$

where $\mathrm{Vol}(S)$ is the volume of the set S. Note that $p(Z|\theta)$ can be computed as in [17] by

$$p(Z|\theta)=\alpha_1\Pi_{k=1}^n p_v\left(y(k)-f\left(\phi(k),\theta\right)\right)$$

for some constant $\alpha_1>0$. As a consequence, the unknown conditional density may be written as

$$p(\theta|Z)=\frac{p(Z|\theta)p(\theta)}{p(Z)}=\alpha\Pi_{k=1}^n p_v\left(y(k)-f\left(\phi(k),\theta\right)\right) \tag{II.6}$$

for some unknown $\alpha>0$. The computation of θ_b in (II.5) and the evaluation of the conditional density $p(\theta|Z)$ of (II.6) are non-trivial in practice.

To summarize, the purpose of the technical note is, for a given data set and a positive density function $p_v(\cdot)$ over $[-\epsilon,\epsilon]$, to estimate the feasible parameter set S of (II.3), the minimum volume outer-bounding ellipsoid E of (II.4) and the minimum variance estimate θ_b of (II.5). Further, these estimates must be convergent in some sense.

III. Algorithm and Estimates

The computation of $\theta_b = \int_{\theta \in S} \theta p(\theta|Z) d\theta$ is hard because both S and $p(\theta|Z)$ are unknown. To this end, for a given data set Z, define

$$q(\theta|Z) = \begin{cases} 0, \text{ if } |y(k) - f(\phi(k), \theta))| > \epsilon \text{ for some } 1 \le k \le n \\ \Pi_{k=1}^n p_v(y(k) - f(\phi(k), \theta)), \\ \qquad \text{if } |y(k) - f(\phi(k), \theta))| \le \epsilon \text{ for all } 1 \le k \le n. \end{cases} \tag{III.7}$$

The key idea in our approach is a method for generating a random sequence in the unknown S according to the unknown conditional density $p(\theta|Z)$. The MCMC technique can be used because the explicit knowledge of the feasible parameter set S and the conditional density $p(\theta|Z)$ is not required.

For a given data set Z and an integer m, the algorithm which generates a sequence of $\theta_i, i = 1, 2, \ldots, m$, according to the unknown density $p(\theta|Z)$ in the unknown set S is now described.

Algorithm III.1

Step 1: Set $i = 0$.

(i) Generate a random vector $\zeta \in R^q \sim N(0, \sigma_\eta^2 I)$, where $N(0, \sigma_\eta^2 I)$ is the Gaussian distribution with zero mean and fixed variance $\sigma_\eta^2 I$.

(ii) If $q(\zeta|Z) > 0$ as defined in (III.7), set $\theta_0 = \zeta \in S$ and $i = 1$. Then go to Step 2. Otherwise, if $q(\zeta|Z) = 0$, i.e., $\zeta \notin S$, then go to Step 1.

Step 2: At each $i \ge 1$, let

$$\zeta = \theta_{i-1} + \eta_{i-1},\ \eta_{i-1} \sim N\left(0, \sigma_\eta^2 I\right) \tag{III.8}$$

and compute the acceptance probability

$$\gamma(\zeta|\theta_{i-1}) = \min\left\{1, \frac{q(\zeta|Z)}{q(\theta_{i-1}|Z)}\right\}.$$

Step 3: Draw a random number z from the uniform distribution on [0,1]. Set $\theta_i = \theta_{i-1}$ if $z \ge \gamma(\zeta|\theta_{i-1})$ and set $\theta_i = \zeta$ if $z < \gamma(\zeta|\theta_{i-1})$. Set $i = i + 1$.

Step 4: If $i = m$, stop. Otherwise, go back to Step 2.

The sequence generated by the above algorithm is dense in the unknown set S when $m \to \infty$ and further is distributed according to the unknown $p(\theta|Z)$. Note that the convergence to $p(\theta|Z)$ is needed if statistical properties of S are of interest. Furthermore, the sequence should be dense in the unknown S since otherwise some parts of S could be "missing." Formally, the sequence $\{\theta_i\}_{i=1}^m$ is dense in S in the limit if for each $\theta \in S$, there is a point θ_i arbitrarily close to θ as $m \to \infty$. Here, we consider this notion from the probabilistic perspective. More precisely, on the limiting distribution, let $\pi(Q) = \int_{\theta \in Q} p(\theta|Z) d\theta$ denote the probability of any subset $Q \subset S$ according to the density $p(\theta|Z)$ on S. Let $\mathbf{P}^m$ be the probability distribution according to which the random sequence was generated with the above algorithm. Then, we have the following result.

Theorem III.1: Consider the sequence $\theta_i \in S,\ i = 1, 2, \ldots, m$, generated by the above algorithm. Then

$$\sup_{Q \subset S} |\mathbf{P}^m(Q) - \pi(Q)| \le (1 - \delta)^m \to 0, \text{ as } m \to \infty \tag{III.9}$$

for some small $\delta \in (0, 1)$. This implies that the limiting distribution is $p(\theta|Z)$ and the sequence is dense in S in probability as $m \to \infty$.

Proof: First, notice the relation $q(\zeta|Z)/q(\theta_{i-1}|Z) = p(\zeta|Z)/p(\theta_{i-1}|Z)$. Now, the Gaussian density $p_N(x)$ of $N(0, \sigma_\eta^2 I)$ is strictly positive for all $x \in R^q$ and symmetric, which implies that the sequence of (III.8) is irreducible and aperiodic [23]. Moreover, the support of $p(\theta|Z)$ is S, which is bounded and the support of the Gaussian density $p_N \sim N(0, \sigma_\eta^2 I)$ contains the support of $p(\theta|Z)$. Thus, there exists some small $\delta > 0$ such that

$$p_N(\theta) > \delta p(\theta|Z), \text{ if } \theta \in S.$$

Then, (III.9) follows from [17, Theorem 6.1]. To show that the sequence is dense in S, observe that for every $\theta \in S$, we have $p(\theta|Z) > 0$. This fact and (III.9) imply that the sequence must be dense as $m \to \infty$. This completes the proof.

Consider a random sequence $\theta_i \in S,\ i = 1, 2, \ldots, m$, generated by the above algorithm. We now construct the estimates $\hat{S}_m$, $\hat{\theta}_{b,m}$ and $\hat{E}_m$ of S, θ_b and E, respectively, by

$$\hat{S}_m = \{\theta_1, \theta_2, \ldots, \theta_m\}, \tag{III.10}$$

$$\hat{\theta}_{b,m} = \frac{1}{m} \sum_{i=1}^m \theta_i, \tag{III.11}$$

$$\hat{E}_m = \left\{x \mid \|\hat{A}x - \hat{b}\|_2 \le 1\right\} \tag{III.12}$$

where $\hat{A}$ and $\hat{b}$ are the solutions to

$$\min_{\hat{A}, \hat{b}} \det(\hat{A}^{-1}),\ s.t. \max_{\theta_i \in \hat{S}_m} \|\hat{A}\theta_i - \hat{b}\|_2 \le 1.$$

To quantify the estimation errors, we adopt the standard Hausdorff metric [20] given by

$$h(A, B) = \max\left\{\sup_{x \in A} \inf_{y \in B} \|x - y\|,\ \sup_{y \in B} \inf_{x \in A} \|x - y\|\right\} \tag{III.13}$$

which measures the distance between two sets. The following result shows the convergence of these estimates.

Theorem III.2: Consider the system (II.1) and the estimates $\hat{S}_m$, $\hat{\theta}_{b,m}$ and $\hat{E}_m$ in (III.10), (III.11) and (III.12), respectively, under Assumption II.1. Let $\theta_i \in S,\ i = 1, 2, \ldots, m$, be a random sequence generated by the algorithm discussed previously. Then, in probability as $m \to \infty$,

$$h(S, \hat{S}_m) \to 0,\ h(\hat{E}_m, E) \to 0,\ \hat{\theta}_{b,m} \to \theta_b. \tag{III.14}$$

Proof: First, notice that $\sup_{y \in \hat{S}_m} \inf_{x \in S} \|x - y\| = 0$ is obvious since $\hat{S}_m \subset S$. Then, $\sup_{x \in S} \inf_{y \in \hat{S}_m} \|x - y\| \to 0$ follows from Assumption II.1 and the fact that $\theta_i \in S$ is dense in S. Consequently, $h(S, \hat{S}_m) \to 0$. This also implies that $h(\hat{E}_m, E) \to 0$. Finally, we obtain $\hat{\theta}_{b,m} \to \theta_b$ from the MCMC theory [14], [17], [19], [23] as

$$\hat{\theta}_{b,m} = \frac{1}{m} \sum_{i=1}^m \theta_i \to \int_{\theta \in S} \theta p(\theta|Z) d\theta = \theta_b.$$

This completes the proof.

We now make a few comments concerning the algorithm.

- The estimates of the unknown feasible parameter set, the minimum volume outer-bounding ellipsoid and the minimum variance estimate can be obtained by the proposed MCMC approach. Further, all three estimates can be calculated by one specific algorithm.
- Among the three estimates, S and E are defined deterministically and θ_b is defined over a probability space.
- The length of the observed data is n and the convergence of the algorithm is with respect to m, not n, i.e., $\hat{S}_m \to S$, $\hat{\theta}_{b,m} \to \theta_b$ for any given n as m gets larger. This fact is important since one of the advantages of the set-membership identification, or the minimum variance estimate, is that the results apply for finite n even when n is small.
- In the proposed algorithm, there is a design variable $\sigma_\eta^2 > 0$. If σ_η^2 is too small, most of $\zeta = \theta_{i-1} + \eta_{i-1}$ are accepted and this implies that θ_i "slowly" moves in S resulting in slow convergence. However, if σ_η^2 is too large, most of $\zeta = \theta_{i-1} + \eta_{i-1}$ are rejected

making the algorithm inefficient. There is a large literature on the selection of σ_η^2 and the interested readers may find the details and references in [14], [19], [23].

IV. Finite Data Length Accuracy and Computational Complexity

The exact calculation of a nonconvex feasible parameter set S is known to be NP hard as the dimension q or the length of data n increases. Thus, the minimum volume outer bounding ellipsoid E is often used in practice [10], [11]. However, the calculation of E is also expensive since S is unknown and it is hard to compute it. Even in the case when S is available, the calculation of E is still expensive because it is an optimization over a nonconvex set S.

In this technical note, an estimate $\hat{E}_m$ of E is proposed. Further, it is shown that $\hat{E}_m \to E$ as a consequence of the Hausdorff metric $h(\hat{E}_m, E) \to 0$ as $m \to \infty$, which is an asymptotic result. A more interesting question is related to the estimation accuracy for a large but finite m. To answer this question, we first observe that convergence of $\hat{E}_m$ to E is not crucial. Secondly, we observe that $\hat{E}_m$ is actually smaller than E in the sense that $\hat{E}_m$ is the minimum volume ellipsoid outer bounding $\hat{S}_m$ and E is the minimum volume ellipsoid outer bounding S, and $\hat{S}_m \subset S$. Therefore, for a finite m, there could be some θ's that are in S but not in $\hat{E}_m$. Thirdly, the data Z, θ is distributed in S according to the unknown $p(\theta|Z)$.

To quantify the estimation accuracy with its associated probability, we define the ϵ approximation.

Definition IV.1: For any $\epsilon \in (0,1)$, $\hat{E}_m$ is said to be an ϵ approximation of S if $\mathrm{Prob}\{\theta \in S \text{ and } \theta \notin \hat{E}_m\} \le \epsilon$, where the probability is with respect to $p(\theta|Z)$.

In this definition, $\hat{E}_m$ depends on a random sequence generated by the MCMC algorithm and the probability can change with each realization. Hence, to guarantee the desired confidence, we next modify the above definition as follows.

Definition IV.2: For any $\epsilon, \delta \in (0,1)$, $\hat{E}_m$ is said to be an ϵ approximation with confidence $1-\delta$ if

$$\mathrm{Prob}\left\{ \mathrm{Prob}\{\theta \in S \text{ and } \theta \notin \hat{E}_m\right\} \le \epsilon \right\} \ge 1-\delta. \qquad \text{(IV.15)}$$

We note that for a pair $\epsilon, \delta \in (0,1)$, the above two-level probability is always satisfied as $m \to \infty$ since $S \subset E \leftarrow \hat{E}_m$. The question is what the minimum m is so that $\hat{E}_m$ satisfies the above two level probability. To this end, by an argument similar to that in [8], we obtain the following result.

Lemma IV.1: Consider S and $\hat{E}_m$. Then, for any given $\epsilon, \delta \in (0,1)$, the two level probability (IV.15) is satisfied, or $\hat{E}_m$ is an ϵ approximation of S with confidence $1-\delta$, if

$$m \ge \frac{e}{2(e-1)} \frac{1}{\epsilon}\left(q^2 + 3q + 2\ln\frac{1}{\delta}\right)$$

where q is the dimension of the unknown parameter vector and $e \approx 2.71828$ is the Euler number.

This result implies that, for given $\epsilon, \delta \in (0,1)$, $\hat{E}_m$ based on a random sequence with length m generated by the MCMC algorithm is an ϵ approximation of S with confidence $1-\delta$. We remark that the same constant $e/(e-1)$ given in the above lemma also appeared in other bounds for quite different problems (e.g., [1], [2]).

We now discuss the computational complexity of $\hat{E}_m$, which is the minimum volume ellipsoid containing $\hat{S}_m = \{\theta_1, \ldots, \theta_m\}$ generated by the MCMC algorithm. The problem of computing an ellipsoid containing a given finite number of points is a well-studied topic [6]. Let $\xi \in (0,1)$ be the level of accuracy for calculation, where we are interested in finding $(1+\xi)$ volume approximation of $\hat{E}_m$. Then, the computational complexity bound on the calculation of $\hat{E}_m$ based on $\{\theta_1, \ldots, \theta_m\}$ is $O(mq^3/\xi)$ [13]. Thus, for a given approximation probability ϵ, the confidence interval $1-\delta$ and the accuracy ξ, the computational complexity of calculating $\hat{E}_m$ is bounded by

$$O\left(\frac{mq^3}{\xi}\right) = O\left(\frac{e}{2(e-1)}\frac{1}{\epsilon}\left(q^2+3q+2\ln\frac{1}{\delta}\right)\frac{q^3}{\xi}\right)$$

which is $O(q^5)$ for a given triple $\epsilon, \delta, \xi \in (0,1)$.

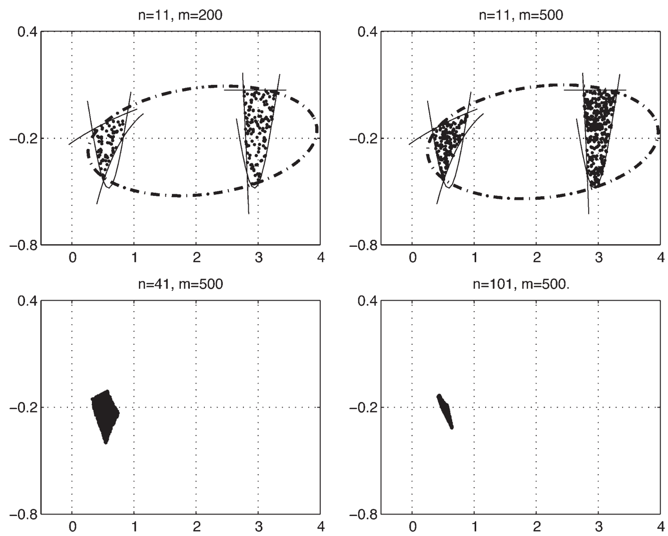

Fig. 1. The three sets S, $\hat{S}_m$, and $\hat{E}_m$.

V. Numerical Simulation

In this section, we investigate the feasible parameter set as a function of the data length. The first example is two-dimensional, which is easy to show visually. Note a major difficulty of nonlinear set-membership identification is that the feasible parameter set may not be necessarily connected. To test the algorithm, we consider the following example.

Example 1: Consider the nonlinear system described by

$$y(k) = \sin\left(a \cdot y(k-1)\right) + \exp\left(b \cdot u(k-1)\right) + v(k).$$

A simulation was conducted with $\theta = (a \ \ b)^T = (0.5 \ \ -0.2)^T$ and $u(k) = \cos(k/3) + \sin(k) + 1$. The unknown $v(\cdot)$ was i.i.d. in $[-1.5, 1.5]$ with the density

$$p_v(x) = \begin{cases} 0, & |x| \ge 1.5 \\ \frac{4}{90}x + \frac{11}{30}, & -1.5 < x < 0 \\ -\frac{4}{90}x + \frac{11}{30}, & 0 \le x < 1.5. \end{cases}$$

First, we consider $n = 11$ with $v(k), k = 1, \ldots, 11$, given by

$$\{-1.2282, -1.0203, -1.3625, 1.2555, -0.9576, 1.0038, 0.8109, 1.3315, -1.2472, 1.3763, 0.3508\}.$$

Two random sequences of θ_i's, $i = 1, \ldots, m$, were generated by the algorithm with $m = 200$ and 500, respectively, both with $\sigma_\eta^2 = 0.5^2$. In the top two diagrams of Fig. 1, the boundaries of the actual S are shown by solid lines, the estimates $\hat{S}_{200}$ and $\hat{S}_{500}$ by dots, and the corresponding minimum volume ellipsoids $\hat{E}_{200}$ and $\hat{E}_{500}$ by dash-dot lines. The unknown feasible parameter set S when $n = 11$ is a union of two disjoint sets. From the figure, the estimate $\hat{S}_m$ even at $m = 200$ represents the unknown and disjoint S reasonably well. The difference between S and $\hat{S}_m$ is smaller as m increases.

An important fact is that the feasible parameter set is a function of the total data length n. The shape and size of S can change drastically as n increases. Consider the same nonlinear system, but with a longer data length for $n = 41$ and 101. The bottom two diagrams in Fig. 1

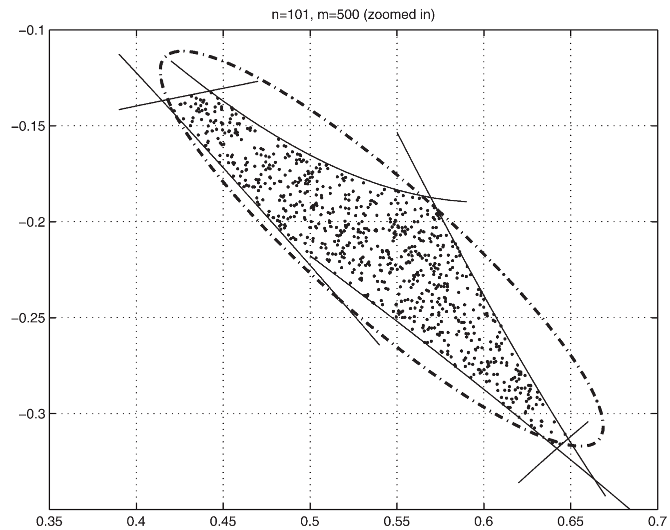

Fig. 2. Zoomed-in figure for $n = 101$ and $m = 500$.

TABLE I
TRUE VALUE AND THE MINIMUM VARIANCE ESTIMATES (EXAMPLE 1)

$\theta = \begin{pmatrix} 0.5 \\ -0.2 \end{pmatrix}$
$\hat{\theta}_b = \begin{pmatrix} 0.53 \\ -0.23 \end{pmatrix}$ when $n = 41, m = 500$
$\hat{\theta}_b = \begin{pmatrix} 0.52 \\ -0.21 \end{pmatrix}$ when $n = 101, m = 500$

TABLE II
TRUE VALUE AND THE MINIMUM VARIANCE ESTIMATES (EXAMPLE 2)

$\theta = \begin{pmatrix} 0.5 \\ 0.5 \\ -0.3 \end{pmatrix}$	$\hat{\theta}_{b,500} = \begin{pmatrix} 0.49 \\ 0.49 \\ -0.3 \end{pmatrix}$	$\hat{\theta}_{b,1000} = \begin{pmatrix} 0.50 \\ 0.51 \\ -0.29 \end{pmatrix}$

show the estimated feasible parameter set S for these cases obtained by the method proposed in the technical note. In these diagrams, we observe that S is smaller and is no longer a disjoint set. Fig. 2 is the zoomed-in diagram of the actual S, the estimated $\hat{S}_{500}$, and the minimum volume ellipsoid $\hat{E}_{500}$ for $n = 101$. The true θ and the minimum variance estimates $\hat{\theta}_{b,m}$ for $m = 500$ and $n = 41, 101$ are shown in Table I. Even with $m = 500$, the estimates are reasonable. As expected, for larger m, the estimates become closer to the true ones.

Example 2: Consider the nonlinear system described by

$$y(k) = a \cdot y(k-1) + b\,(1 - c \cdot y(k-1) \cdot u(k-1))^2 \cdot u(k-1) + v(k)$$

where $\theta = (a, b, c)^T = (0.5, 0.5, -0.3)^T$ is the unknown parameter vector to be estimated. The input and noise were taken to be the same as before. The minimum variance estimates by the algorithm are given in Table II for $n = 100$, $m = 500$, 1000, respectively. We observe that $m = 500$ seems to be sufficient for this three dimensional example.

Note that both examples above illustrate the low-dimensional setting for simple visualization and demonstration. The method however can be applied to higher dimensional systems.

VI. CONCLUDING REMARKS

An MCMC approach is proposed in this technical note for nonlinear system identification in a mixed set-membership and statistical setting. The proposed algorithm defines estimates of the feasible parameter set, the outer-bounding ellipsoid and the minimum variance estimate by an MCMC generated sequence. Calculations of these estimates are simple summations or convex problems which are relatively easy to solve numerically. In future research, we would like to further pursue the approach to address issues related to convergence rates of the algorithm and to study from the viewpoint of the probabilistic approach.

REFERENCES

[1] T. Alamo, R. Tempo, and A. Luque, "On the samplecomplexity of probabilistic analysis and design methods," in *Perspectives Mathematical Syst. Theory, Control, Signal Process.*, J. C. Willems, S. Hara, Y. Ohta, and H. Fujioka, Eds. Berlin, Germany: Springer, 2010, pp. 34–50.

[2] T. Alamo, R. Tempo, A. Luque, and D. R. Ramirez, "Randomized methods for design of uncertain systems: Sample complexity and sequential algorithms," *Automatica*, vol. 51, 2015.

[3] E. W. Bai, Y. Ye, and R. Tempo, "Bounded error parameter estimation: A sequential analytic center approach," *IEEE Trans. Autom. Control*, vol. 39, pp. 1107–1117, 1999.

[4] E. W. Bai, K. M. Nagpal, and R. Tempo, "Bounded-error parameter estimation: Noise models and recursive algorithms," *Automatica*, vol. 32, pp. 985–999, 1996.

[5] E. W. Bai, R. Tempo, and C. Cho, "Membership set estimation: Size, optimal input, complexity, and relations with least squares," *IEEE Trans. Circuits Syst. I*, vol. 42, pp. 266–277, 1995.

[6] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York: Cambridge University Press, 2004.

[7] V. Cerone, D. Piga, D. Regruto, and R. Toth, "Minimal LPV state-space realization driven set-membership identification," in *Proc. American Control Conf.*, Montreal, QC, Canada, 2012, pp. 3421–3426.

[8] F. Dabbene, C. Lagoa, and P. Shcherbakov, "On the complexity of randomized approximations of nonconvex sets," in *Proc. IEEE Multi-Conf. Systems and Control*, Yokohama, Japan, 2010, pp. 1564–1569.

[9] F. Dabbene, M. Sznaier, and R. Tempo, "Probabilistic optimal estimation with uniformly distributed noise," *IEEE Trans. Autom. Control*, vol. 59, pp. 2113–2127, 2014.

[10] J. Deller Jr and Y. F. Huang, "Set membership identification and filtering for signal processing applications," *Circuits Syst. Signal Process.*, vol. 21, pp. 69–82, 2002.

[11] E. Fogel and Y. F. Huang, "The value of informationin system identification—bounded noise case," *Automatica*, vol. 18, pp. 229–238, 1982.

[12] L. Granvilliers and F. Benhamou, "Algorithm 852: Realpaver: An interval solver using constraint satisfaction techniques," *ACM TOMS*, vol. 32, pp. 138–156, 2006.

[13] P. Kumar and E. A. Yildirim, "Minimum volume enclosing ellipsoids and core sets," *J. Optimiz. Theory and Applic.*, vol. 126, pp. 1–21, 2005.

[14] S. Meyn and R. Tweedie, *Markov Chains and Stochastic Stability*, 3rd ed. London, U.K.: Springer-Verlag, 1996.

[15] M. Milanese and C. Novara, "Unified set membership theory for identification, prediction and filtering of nonlinear systems," *Automatica*, vol. 41, pp. 2019–2032, 2011.

[16] B. Ninness and G. Goodwin, "Rapproachment between bounded-error and stochastic estimation theory," *Int. J. Adapt. Control and Signal Process.*, vol. 9, pp. 107–132, 2007.

[17] B. Ninness and S. Henriksen, "Bayesian system identification via Markov chain Monte Carlo techniques," *Automatica*, vol. 46, pp. 40–51, 2010.

[18] N. Ozay, M. Sznaier, C. Lagoa, and O. Camps, "A sparsification approach to set membership identification of switched affine systems," *IEEE Trans. Autom. Control*, vol. 57, pp. 634–648, 2012.

[19] J. Spall, *Introduction to Stochastic Search and Optimization*. Hoboken, NJ, USA: Wiley, 2003.

[20] P. Szeptycki and F. S. Van Vleck, "Centers and nearest points of sets," in *Proc. American Mathematical Society*, 1982, vol. 85, pp. 27–31.

[21] R. Tempo, G. Calafiore, and F. Dabbene, *Randomized Algorithms for Analysis and Control of Uncertain Systems: With Applications*, 2nd ed. London, U.K.: Springer-Verlag, 2012.

[22] F. Tjarnstrom and A. Garulli, "A mixed probabilistic/bounded-error approach to parameter estimation in the presence of amplitude bounded white noise," in *Proc. IEEE Conf. Decision and Control*, Las Vegas, 2002, pp. 3422–3427.

[23] L. Tierney, "Markov chain for exploring posteriordistributions," *Ann. Statistics*, vol. 22, pp. 1701–1762, 1994.

[24] M. Sznaier, "Computational complexity analysis of set membership identification of Hammerstein and Wiener systems," *Automatica*, vol. 45, pp. 701–705, 2009.