Guaranteed Non-Asymptotic Confidence Ellipsoids for FIR Systems

Erik Weyer[†] Balázs Csanád Csáji[‡] Marco C. Campi[*]

Abstract— Recently, a new finite-sample system identification algorithm, called Sign-Perturbed Sums (SPS), was introduced in [2]. SPS constructs finite-sample confidence regions that are centered around the least squares estimate, and are guaranteed to contain the true system parameters with a user-chosen exact probability for any finite number of data points. The main assumption of SPS is that the noise terms are independent and symmetrically distributed about zero, but they do not have to be stationary, nor do their variances and distributions have to be known. Although it is easy to determine if a particular parameter belongs to the confidence region, it is not easy to describe the boundary of the region, and hence to compactly represent the exact confidence region. In this paper we show that an ellipsoidal outer-approximation of the SPS confidence region can be found by solving a convex optimization problem, and we illustrate the properties of the SPS region and the ellipsoidal outer-approximation in simulation examples.

I. INTRODUCTION

Parametric estimation of dynamical systems is one of the fundamental problems of system identification [4], [5]. Classical approaches, such as the least squares (LS) method, typically provide point estimates. Given an estimate, it is an intrinsic task to evaluate how close the estimated model is to the true system. Such evaluation often comes in the form of confidence regions, which are especially important for problems with strict safety, stability or quality guarantees.

While it is possible to construct confidence ellipsoids for standard system identification methods, the evaluation of the confidence is usually based on the central limit theorem, and hence it is only guaranteed asymptotically, as the number of data points tends to infinity. In practice, however, we have a finite number of measurements, and the statistical characteristics of the noise may even change over time. Applying the asymptotic theory in these cases results only in heuristic confidence sets, which do not come with strict theoretical guarantees.

Recently, a new finite-sample system identification algorithm, called Sign-Perturbed Sums (SPS), was introduced in [2]. SPS constructs non-asymptotic confidence regions for parameters of dynamical systems under very mild statistical assumptions. The confidence regions are centered around the LS estimate, and are guaranteed to contain the true parameters with a user-chosen exact probability for any finite data set. The main assumption of SPS is that the noise terms are independent and symmetrically distributed about zero, but they do not have to be stationary, nor do their variances and distributions have to be known.

A practical drawback of SPS is that it defines the confidence regions point-wise, and it is not easy to describe the boundary of the set and hence, for example, to compactly represent it in a computer. In order to overcome this issue, here we present a method to construct confidence ellipsoids by over-bounding the SPS confidence region for finite impulse response (FIR) systems. Unlike the ellipsoids of the asymptotic theory, the constructed confidence ellipsoids have a theoretically guaranteed confidence probability.

We show that the proposed ellipsoidal outer-approximation can be efficiently computed numerically by reformulating it as a convex optimization problem. The method is evaluated in simulations with different types of noise distributions as well as on higher order systems. The results demonstrate that the construction is not too conservative and that the ellipsoidal over-bounds get closer to the exact SPS confidence region, as the number of observations increases.

The structure of the paper is as follows. In Section II the original SPS method for building exact confidence regions for FIR systems is presented. In Section III the construction of ellipsoidal over-bounds are discussed and shown to be obtained as solution of a convex optimization problem. Section IV contains results of numerical experiments, which are followed by some concluding remarks in Section V.

II. THE METHOD OF SIGN-PERTURBED SUMS

In this section we give a short summary of the SPS algorithm for FIR systems, and we also provide a brief account of the asymptotic system identification theory for such systems.

A. Data generating system and model class

We consider FIR systems of the type

$$y_t = b_1^0 u_{t-1} + b_2^0 u_{t-2} + \cdots + b_d^0 u_{t-d} + n_t, \qquad (1)$$

with true parameters $\theta^0 = [\, b_1^0 \ldots b_d^0 \,]^\mathrm{T}$. We assume that

- The order d is known.
- The noise $\{n_t\}$ is a sequence of independent random variables symmetrically distributed about zero.
- The input signal $\{u_t\}$ is independent of the noise.

[†]E. Weyer is with Department of Electrical and Electronic Engineering, The University of Melbourne, Parkville, VIC 3010, Australia. ewey@unimelb.edu.au

[‡]B. Cs. Csáji is with MTA SZTAKI: The Institute for Computer Science and Control of the Hungarian Academy of Sciences, Kende utca 13-17, H-1111 Budapest, Hungary. balazs.csaji@sztaki.mta.hu

[*]M. C. Campi is with Department of Information Engineering, University of Brescia, Via Branze 38, 25123 Brescia, Italy. marco.campi@ing.unibs.it

The model class is

$$\hat{y}_t(\theta) = b_1 u_{t-1} + b_2 u_{t-2} + \cdots + b_d u_{t-d},$$

which can be written in linear regression form as

$$\hat{y}_t(\theta) = \phi_t^{\mathrm{T}} \theta,$$

with

$$\begin{aligned}
\phi_t &= [\, u_{t-1} \ldots u_{t-d} \,]^{\mathrm{T}}, \\
\theta &= [\, b_1 \ldots b_d \,]^{\mathrm{T}}.
\end{aligned}$$

The observed data are $y_1, \phi_1, y_2, \phi_2, \ldots, y_N, \phi_N$.

B. The least squares estimate

The least squares estimate of the true system parameters is given by

$$\hat{\theta}_N = \arg\min_{\theta} \sum_{t=1}^{N} (y_t - \phi_t^{\mathrm{T}} \theta)^2.$$

The LS estimate $\hat{\theta}_N$ satisfies the normal equation

$$\sum_{t=1}^{N} \phi_t(y_t - \phi_t^{\mathrm{T}} \hat{\theta}_N) = 0,$$

which has the closed form solution

$$\hat{\theta}_N = \left(\frac{1}{N} \sum_{t=1}^{N} \phi_t \phi_t^{\mathrm{T}} \right)^{-1} \left(\frac{1}{N} \sum_{t=1}^{N} \phi_t y_t \right),$$

provided $\frac{1}{N} \sum_{t=1}^{N} \phi_t \phi_t^{\mathrm{T}}$ is nonsingular.

C. Confidence region for the least squares estimate based on asymptotic system identification theory

It is well-known that, under natural conditions, the least squares estimate $\hat{\theta}_N$ converges to the true system parameters θ^0 with probability one as N tends to infinity, and $\sqrt{N}(\hat{\theta}_N - \theta^0)$ converges in distribution to a zero mean Gaussian with covariance matrix $\sigma^2 (\mathbb{E} \, \phi_t \phi_t^{\mathrm{T}})^{-1}$ under the additional assumptions that the input and the noise are stationary and that the latter has variance σ^2 [4], [5].

Based on these asymptotic results, confidence ellipsoids for the least squares estimate are often constructed by including all values of θ which satisfies

$$(\theta - \hat{\theta}_N)^{\mathrm{T}} \frac{1}{\hat{\sigma}} \left(\sum_{t=1}^{N} \phi_t \phi_t^{\mathrm{T}} \right) (\theta - \hat{\theta}_N) \leq \alpha, \qquad (2)$$

where $\hat{\sigma} = \frac{1}{N-d} \sum_{t=1}^{N} (y_t - \phi_t^{\mathrm{T}} \hat{\theta}_N)^2$ and α is the quantile in the χ^2 distribution with d degrees of freedom corresponding to the desired probability of the confidence ellipsoid.

However, as these confidence regions are based on asymptotic theory, they have in general no strictly guaranteed properties for finite N. An exception is when the noise $\{n_t\}$ is a sequence of independent and identically distributed Gaussian random variables, in which case the random variable $\frac{1}{d}(\theta - \hat{\theta}_N)^{\mathrm{T}} \frac{1}{\hat{\sigma}} \left(\sum_{t=1}^{N} \phi_t \phi_t^{\mathrm{T}} \right) (\theta - \hat{\theta}_N)$ follows an f-distribution with d and $N - d$ degrees of freedom.

D. The SPS algorithm for constructing an exact non-asymptotic confidence region

We now present the main idea behind the SPS algorithm. More details can be found in [2].

Let

$$\Phi_N = \sum_{t=1}^{N} \phi_t \phi_t^{\mathrm{T}},$$

and let $\Phi_N^{1/2}$ be a matrix such that $\Phi_N^{1/2} \Phi_N^{1/2\mathrm{T}} = \Phi_N$. Next define the reference vector sum (cf. the normal equation)

$$S_0(\theta) = \Phi_N^{-1/2} \sum_{t=1}^{N} \phi_t(y_t - \phi_t^{\mathrm{T}} \theta), \qquad (3)$$

and the $M - 1$ sign-perturbed sums

$$S_i(\theta) = \Phi_N^{-1/2} \sum_{t=1}^{N} \alpha_{i,t} \phi_t(y_t - \phi_t^{\mathrm{T}} \theta), \quad i = 1, \ldots, M-1 \quad (4)$$

where $\alpha_{i,t}, i = 1, \ldots, M - 1, \ t = 1 \ldots, N$ are "random signs" taking on the values 1 and -1 with probability $1/2$ each. A key observation is that corresponding to $\theta = \theta^0$ we have that

$$S_0(\theta^0) = \Phi_N^{-1/2} \sum_{t=1}^{N} \phi_t n_t, \qquad (5)$$

and

$$S_i(\theta^0) = \Phi_N^{-1/2} \sum_{t=1}^{N} \phi_t \alpha_{i,t} n_t, \quad i = 1, \ldots, M - 1 \quad (6)$$

Since n_t is symmetrically distributed about zero, it follows that n_t and $\alpha_{i,t} n_t$ have the same distribution, and hence $S_i(\theta^0), \ i = 0, \ldots, M - 1$, all have the same distribution. Furthermore if we consider the squared norm of the above vectors, i.e.

$$Z_0(\theta) = ||S_0(\theta)||^2, \qquad (7)$$

and

$$Z_i(\theta) = ||S_i(\theta)||^2, \qquad (8)$$

there is no reason, corresponding to the true parameter value, why a particular $Z_i(\theta^0)$ should be larger or smaller than any other $Z_j(\theta^0)$. A confidence region can therefore be obtained by computing $Z_i(\theta)$ for a generic value of θ, and then exclude those values of θ for which $Z_0(\theta)$ are among the q largest ones. The constructed region has exact probability $1 - q/M$ of containing the true parameter values. Moreover, parameters q and M are controlled by the user, and hence exact SPS confidence regions can be built for any (rational) probability.

For the least squares estimate we have that $Z_0(\hat{\theta}_N) = 0$, since $\hat{\theta}_N$ satisfies the normal equation, and hence the least squares estimate is in the confidence set. Moreover, for $\theta \neq \theta^0$, it can be shown that $Z_0(\theta)$ increases faster than the other $Z_i(\theta)$ functions so that θ will be eventually excluded from the confidence region.

A step-by-step procedure for SPS is as follows.

Construction of non-asymptotic confidence region

1) Select q and M such that $p = 1 - q/M$ is the desired probability of the confidence region.
2) Given the observed data $y_1, \phi_1, y_2, \phi_2, \ldots, y_N, \phi_N$, compute the reference sum

$$S_0(\theta) = \Phi_N^{-1/2} \sum_{t=1}^{N} \phi_t(y_t - \phi_t^{\mathrm{T}}\theta). \qquad (9)$$

3) Compute the $M-1$ sign-perturbed sums

$$S_i(\theta) = \Phi_N^{-1/2} \sum_{t=1}^{N} \alpha_{i,t}\phi_t(y_t - \phi_t^{\mathrm{T}}\theta), \qquad (10)$$

where $\alpha_{i,t}, i = 1, \ldots, M-1, \ t = 1 \ldots, N$ are independent and identically distributed random variables taking the values 1 and -1 with probability $1/2$ each.

4) Compute

$$Z_i(\theta) = ||S_i(\theta)||^2 + \varepsilon_i(\theta),$$

for $i = 0, \ldots, M-1$, where $\varepsilon_i(\theta)$ are "small" i.i.d. random variables, introduced in order to break ties[1].

5) Construct the confidence set as

$$\hat{\Theta}_M^q = \{\, \theta \mid \text{at least } q \text{ of the } Z_i(\theta) \text{ functions are}$$
$$\text{larger than } Z_0(\theta) \,\}$$

It can be shown (see [2]) that the constructed confidence region is exact, i.e., it has the following property.

Theorem 2.1: Under the assumptions in Section II-A the probability that $\hat{\Theta}_M^q$ contains θ^0 is exactly $1 - q/M$.

Note that q and M, which determine the probability that θ^0 belongs to the confidence region, are chosen by the user.

III. NUMERICAL ALGORITHMS

Given a value of θ, it is straightforward to figure out whether θ is in the confidence set. All we have to do is to evaluate the M $\{Z_i(\theta)\}$ functions for the particular value of θ and compare them. This means that SPS lends itself nicely to problems where only a finite number of θ's need to be checked, which is for example the case in certain types of fault detection problems. More generally, the confidence region can be constructed by checking each parameter value on a grid, but this approach will suffer from the curse of dimensionality. The aim of this section is to derive a numerical procedure that delivers a compact representation for an ellipsoidal over-bound of the confidence region.

As was already observed, the least squares estimate is in the SPS confidence region[2]. Moreover, it can be shown that if θ' is such that $Z_0(\theta') \leq Z_i(\theta')$ then $Z_0(\theta) \leq Z_i(\theta)$ for all θs on the line segment between the least squares estimate and θ', i.e. the region for which $Z_0(\theta) \leq Z_i(\theta)$ is a star

---

[1]$\{\varepsilon_i(\theta)\}$ can be chosen such that they break ties in case two or more $||S_i(\theta)||^2$ functions are equal, but do not change the order otherwise [2].

[2]Assuming that it is nonempty. It has a positive, but very small probability that the set is empty, e.g., it can happen if some strings with all 1s or all -1s are drawn in point 3, since then the order of the corresponding $\{Z_i(\theta)\}$ functions is determined by the tie breaking variables in point 4.

---

set with the least squares estimate as a star center. As the confidence set is constructed by the values of θ for which $Z_0(\theta) \leq Z_i(\theta)$ for at least q $Z_i(\theta)$ functions, it follows that the confidence region itself is also a star set.

The random variables $\{\varepsilon_i(\theta)\}$ introduced in point 4 are only used to break ties for values of θ for which $||S_0(\theta)||^2$ and $||S_i(\theta)||^2$ are equal. For the purpose of finding an ellipsoidal over-bound we can ignore these variables and consider the set formed by those values of θ for which at least q of the $||S_i(\theta)||^2$ are larger *or equal* to $||S_0(\theta)||^2$. Ignoring variables $\{\varepsilon_i(\theta)\}$, $Z_0(\theta)$ can be written as

$$Z_0(\theta) = ||S_0(\theta)||^2$$
$$= \left(\sum_{t=1}^{N} \phi_t(y_t - \phi_t^{\mathrm{T}}\theta)\right)^{\mathrm{T}} \Phi_N^{-1} \left(\sum_{t=1}^{N} \phi_t(y_t - \phi_t^{\mathrm{T}}\theta)\right)$$
$$= \left(\sum_{t=1}^{N} \phi_t\phi_t^{\mathrm{T}}(\hat{\theta}_N - \theta)\right)^{\mathrm{T}} \Phi_N^{-1} \left(\sum_{t=1}^{N} \phi_t\phi_t^{\mathrm{T}}(\hat{\theta}_N - \theta)\right)$$
$$= (\theta - \hat{\theta}_N)^{\mathrm{T}}\Phi_N(\theta - \hat{\theta}_N), \qquad (11)$$

and hence the confidence region is given by those values of θ that satisfy

$$(\theta - \hat{\theta}_N)^{\mathrm{T}}\Phi_N(\theta - \hat{\theta}_N) \leq r(\theta),$$

where $r(\theta)$ is the qth largest values of the functions $Z_i(\theta), i = 1, \ldots, M-1$. The idea is now to seek an over-bound by replacing $r(\theta)$ with a θ independent r. This over-bound will hence have the same shape as the confidence ellipsoid based on the asymptotic theory, cf. (2), but different volume, and importantly, it will be rigorously guaranteed for a finite number of data points. Moreover, it has a compact representation, since it is characterised in terms of $\hat{\theta}_N, \Phi_N$ and r.

Comparing $Z_0(\theta)$ with one single $Z_i(\theta)$ function, we have

$$\{\theta : Z_0(\theta) \leq Z_i(\theta)\}$$
$$\subseteq \{\theta : Z_0(\theta) \leq \max_{\theta:Z_0(\theta)\leq Z_i(\theta)} Z_i(\theta)\}$$

Relation $Z_0(\theta) \leq Z_i(\theta)$ can be rewritten as

$$(\theta - \hat{\theta}_N)^{\mathrm{T}}\Phi_N(\theta - \hat{\theta}_N)$$
$$\leq \left(\sum_{t=1}^{N} \alpha_{i,t}\phi_t(y_t - \phi_t^{\mathrm{T}}\theta)\right)^{\mathrm{T}} \Phi_N^{-1} \left(\sum_{t=1}^{N} \alpha_{i,t}\phi_t(y_t - \phi_t^{\mathrm{T}}\theta)\right)$$
$$= \theta^{\mathrm{T}}\Phi_{\alpha,i}\Phi_N^{-1}\Phi_{\alpha,i}\theta - 2\theta^{\mathrm{T}}\Phi_{\alpha,i}\Phi_N^{-1}\bar{\phi}_{\alpha,i}^{y} + \bar{\phi}_{\alpha,i}^{yT}\Phi_N^{-1}\bar{\phi}_{\alpha,i}^{y},$$

where

$$\Phi_{\alpha_i} = \sum_{t=1}^{N} \alpha_{i,t}\phi_t\phi_t^{\mathrm{T}},$$
$$\bar{\phi}_{\alpha,i}^{y} = \sum_{t=1}^{N} \alpha_{i,t}\phi_t y_t.$$

Therefore, using the notation $z = (\Phi_N^{1/2})^T(\theta - \hat{\theta}_N)$, $\max_{\theta:Z_0(\theta)\leq Z_i(\theta)} Z_i(\theta)$ is obtained as the value of the program

$$\text{maximize} \quad ||z||^2$$
$$\text{subject to} \quad z^{\mathrm{T}}A_i z + 2b_i^{\mathrm{T}} z + c_i \leq 0$$

where

$$A_i = I - \Phi_N^{-1/2}\Phi_{\alpha,i}\Phi_N^{-1}\Phi_{\alpha,i}(\Phi_N^{-1/2})^T,$$
$$b_i = \Phi_N^{-1/2}\Phi_{\alpha,i}\Phi_N^{-1}(\bar{\phi}_{\alpha,i}^y - \Phi_{\alpha,i}\hat{\theta}_N),$$
$$c_i = -\bar{\phi}_{\alpha,i}^y \Phi_N^{-1}\bar{\phi}_{\alpha,i}^y + 2\hat{\theta}_N^T\Phi_{\alpha_i}\Phi_N^{-1}\bar{\phi}_{\alpha,i}^y$$
$$-\hat{\theta}_N^T\Phi_{\alpha_i}\Phi_N^{-1}\Phi_{\alpha,i}\hat{\theta}_N,$$

This program is *not convex*. However, it can be shown (cf. Appendix B of [1]) that *strong duality* holds, so that the value of the above optimization program is equal to the value of its dual

$$\begin{aligned}
\text{minimize} \quad & \gamma \\
\text{subject to} \quad & \lambda \geq 0 \\
& \begin{bmatrix} -I + \lambda A_i & \lambda b_i \\ \lambda b_i^T & \lambda c_i + \gamma \end{bmatrix} \succeq 0
\end{aligned}$$
$$(12)$$

where "$\succeq 0$" denotes that a matrix is positive semidefinite. This program is convex, and can be easily solved using, e.g., Yalmip [3] and a solver such as SDPT3.

Letting γ_i^* be the value of program (12), we now have

$$\{\theta : Z_0(\theta) \leq Z_i(\theta)\} \subseteq \{\theta : Z_0(\theta) \leq \gamma_i^*\}$$

Thus, the ellipsoidal over-bound is given by

$$\{(\theta - \hat{\theta}_N)^T\Phi_N(\theta - \hat{\theta}_N) \leq r\}$$

where $r = q$th largest value of γ_i^*, $i = 1, \ldots, m-1$.

A step-by-step procedure to find an ellipsoidal over-bound of the SPS confidence region is therefore.

Construction of ellipsoidal over-bound

1) Compute $\Phi_N = \sum_{t=1}^N \phi_t \phi_t^T$ and the least squares estimate $\hat{\theta}_N = \Phi_N^{-1}\left(\sum_{t=1}^N \phi_t y_t\right)$.
2) For each $i = 1, \ldots, M-1$ solve the optimization problem (12), and let γ_i^* be the optimal value.
3) Let r be the qth largest γ_i^* value.
4) The ellipsoidal over-bound is given by $\{(\theta - \hat{\theta}_N)^T\Phi_N(\theta - \hat{\theta}_N) \leq r\}$.

Remark. The ellipsoidal over-bound can be used as a starting point to find the exact SPS confidence region. As the confidence region is a star set, one can perform a line search along segments from the least squares estimate to the boundary of the ellipsoid, in order to find the exact boundary of the SPS region. This still involves gridding, but the dimension of the problem is one less than the original problem, and moreover the gridding can be restricted to a bounded set. No compact representation is available though since the boundary points needs to be stored.

IV. SIMULATIONS AND ILLUSTRATIONS

In this section we illustrate the proposed ellipsoidal over-bounding algorithm for SPS by comparing it to the original SPS method and also the asymptotic theory. Simulation experiments on 2nd and higher order FIR systems with various noise distributions are presented.

A. Second order FIR system

We consider a second order data generating FIR system

$$y_t = b_1^0 u_{t-1} + b_2^0 u_{t-2} + n_t,$$

where $b_1^0 = 0.7$ and $b_2^0 = 0.3$ are the true system parameters and $\{n_t\}$ is a sequence of independent Gaussian random variables with zero mean. The variances $\sigma_{n_t}^2$ are given by independent and identically distributed random variables which take the values 0.1 with probability 0.8 and 1 with probability 0.2. The input signal is given by

$$u_t = 0.75 u_{t-1} + w_t,$$

where $\{w_t\}$ is an independent and identically distributed Gaussian sequence with zero mean and variance 1.

The model class is

$$\hat{y}_t(\theta) = b_1 u_{t-1} + b_2 u_{t-2},$$

where $\theta = [b_1 \ b_2]$ are the model parameters.

The exact 95% SPS confidence regions together with the ellipsoidal over-bound and the confidence ellipsoid based on asymptotic system identification theory, as given in equation (2), are shown in Figures 1 to 3 for $N = 50$, 200 and 800 observations. During the experiments, we have configured SPS by using parameters $M = 100$ and $q = 5$.

The solid red lines represent the boundaries of the exact SPS confidence regions, and the dash dotted black lines are the ellipsoidal over-bounds. The dashed blue lines show the confidence ellipsoid based on asymptotic system identification theory. The black "x"s are the least squares estimates, while the red "⋄"s show the true parameter values.

The exact SPS confidence regions were found by starting from the overbound, and then finding the boundary of the exact region by a line search along line segments between the LS estimate and the boundary of the ellipsoidal outer approximation with angular increment of $2\pi/1000$. Compared to a standard gridding approach, this approach also involve gridding but in a one-dimensional space rather than a two-dimensional space. Moreover, the gridding can be confined to the bounded set $[0, 2\pi)$.

The exact non-asymptotic SPS regions are similar in size and shape to the asymptotic confidence regions, but have the advantage that they are guaranteed to contain the true parameter values with exact probability 0.95. The experiments demonstrate that as the number of data points increases, the confidence regions gets smaller and the ellipsoidal over-bound gets closer and closer to the exact SPS region.

B. Second order FIR systems. Different types of noise

Next we ran Monte Carlo simulations with different types of noise for $N = 50$ and $N = 200$. In addition to the mixture of Gaussians used in the previous section we also considered the case where $\{n_t\}$ was independent and identically distributed with a uniform and a Laplacian (double-exponential) distribution. In both cases the variance was $\sigma_n^2 = 0.28$. We also considered the situation where the noise was a time varying Gaussian. In this case the noise term was given by
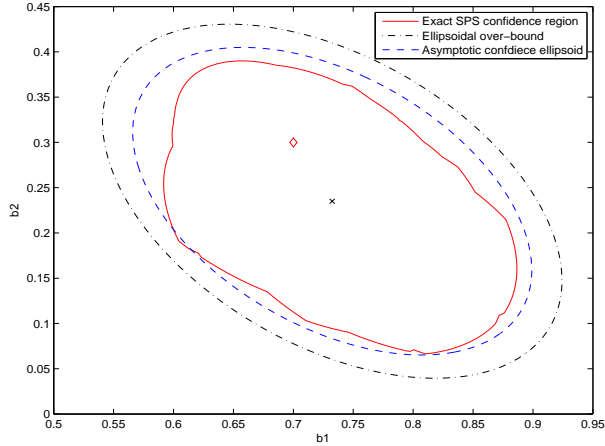
$$n_t = z_t n_{1t},$$
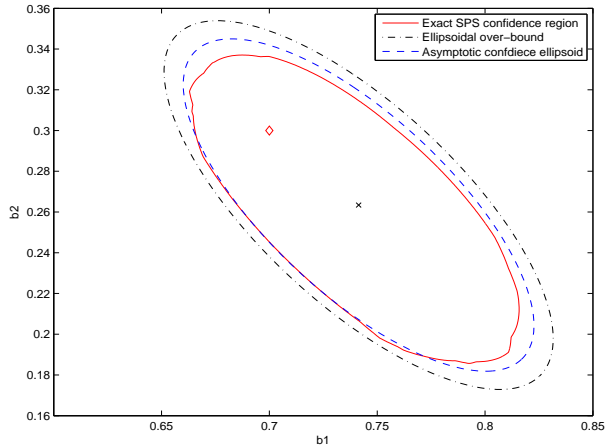
Fig. 1. 95% confidence regions, $N = 50$.

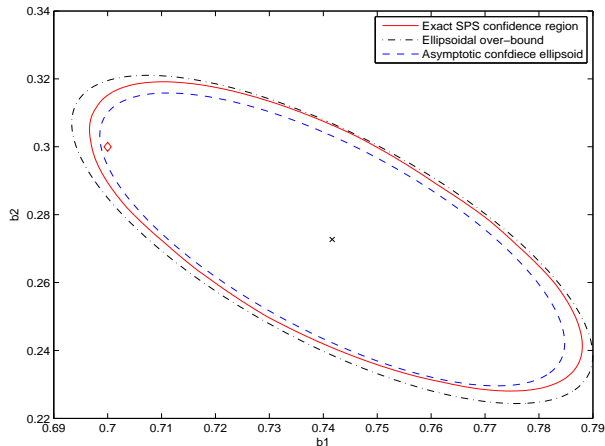Fig. 2. 95% confidence regions, $N = 200$

Fig. 3. 95% confidence regions, $N = 800$

Noise	$M=20$ Over-b	$M=20$ Exact	$M=100$ Over-b	$M=100$ Exact	Asymp
Gaussian	0.1706	0.1036	0.1390	0.0863	0.0753
Uniform	0.1841	0.1119	0.1488	0.0916	0.0767
Laplacian	0.1630	0.1011	0.1373	0.0843	0.0742
Time-var	0.1637	0.1027	0.1400	0.0870	0.0764

TABLE I

AVERAGE AREA. $N = 50$.

where $\{v_{1t}\}$ was an i.i.d. Gaussian with variance one and $\{z_t\}$ was the output of the first order system

$$z_t = az_{t-1} + e_t,$$

where $a = 0.8$. Random variable e_t was a zero mean Gaussian with variance $(1 - a^2)\sigma_n^2$ such that $\mathbb{E} z_t^2 = \sigma_n^2$. In this case $\{n_t\}$ is conditionally independent (and symmetrically distributed) given $\{z_t\}$, and this property is enough for the SPS theory to hold.

For each type of noise we ran 500 Monte Carlo simulations and we computed the 95% exact SPS confidence regions using parameters $M = 20$, $q = 1$ and $M = 100$, $q = 5$, together with the ellipsoidal over-bounds and the confidence ellipsoids based on asymptotic system identification theory, cf. equation (2). The average area of the confidence regions and the empirical probabilities that the true system parameters belonged to the corresponding regions are presented in Tables I to IV.

We observe that the larger value of M gives on average a smaller confidence region, and that the areas and the empirical probabilities do not change much with the type of noise. Even though one can generate different confidence regions for different values of M, the user cannot try out different values of M and then cherry-pick the value which gives the smallest region, since the probability that the true parameter belongs to the region is no longer guaranteed.

Note that the values of parameters M and q must be chosen independently of the observed data in order to generate a region with a guaranteed confidence probability.

The empirical probabilities of the ellipsoidal over-bounds are closer to the empirical probabilities of the exact SPS region for $N = 200$. This is in agreement with the intuition coming from observing Figures 1-3 where the ellipsoidal outer-approximation got closer to the exact SPS confidence region with increasing N.

The exact SPS regions are slightly larger than the confidence ellipsoids based on asymptotic theory, but they have the considerable advantage that their confidence probabilities are theoretically guaranteed for finite data sets.

C. Higher order system

During this experiment, the true data generating system was an eight order FIR system

$$
\begin{aligned}
y_t =\ & b_1^0 u_{t-1} + b_2^0 u_{t-2} + b_3^0 u_{t-3} + b_4^0 u_{t-4} + b_5^0 u_{t-5} + \\
& b_6^0 u_{t-6} + b_7^0 u_{t-7} + b_8^0 u_{t-8} + n_t,
\end{aligned}
$$

	M=20	M=20	M=100	M=100	
Noise	Over-b	Exact	Over-b	Exact	Asymp
Gaussian	0.986	0.948	0.984	0.950	0.934
Uniform	0.990	0.956	0.994	0.954	0.944
Laplacian	0.986	0.960	0.994	0.948	0.948
Time-var	0.988	0.948	0.990	0.946	0.938

TABLE II

EMPIRICAL PROBABILITIES. $N = 50$.

	M=20	M=20	M=100	M=100	
Noise	Over-b	Exact	Over-b	Exact	Asymp
Gaussian	0.0272	0.0214	0.0235	0.0185	0.0178
Uniform	0.0276	0.0218	0.0243	0.0191	0.0178
Laplacian	0.0267	0.0214	0.0235	0.0187	0.0179
Time-var	0.0264	0.0210	0.0232	0.0184	0.0176

TABLE III

AVERAGE AREA. $N = 200$.

Data points	Ellipsoidal over-bound	Asymp
$N = 800$	0.989	0.946
$N = 3200$	0.965	0.935

TABLE V

EMPIRICAL PROBABILITIES. 8TH ORDER SYSTEM.

Data points	Ellipsoidal over-bound	Asymptotic theory
$N = 800$	$0.4408 \cdot 10^{-17}$	$0.0484 \cdot 10^{-17}$
$N = 3200$	$0.1873 \cdot 10^{-22}$	$0.0548 \cdot 10^{-22}$

TABLE VI

VOLUMES. 8TH ORDER SYSTEM.

with $\theta^0 = [\,0.7, 0.3, 0.21, 0.2, 0.15, 0.25, 0.1, 0.05\,]^\mathrm{T}$. Processes $\{v_t\}$ and $\{u_t\}$ were the same as in Section IV-A.

We ran 1000 Monte Carlo simulations with $N = 800$ and $N = 3200$, and computed the ellipsoidal over-bound for the 95% confidence region using $M = 100$ and $q = 5$. The computational time for computing a single ellipsoidal over-bound was around 30 seconds on a standard laptop using Yalmip and SDPT3.

We compared the empirical probabilities and volumes for the ellipsoidal over-bound and the asymptotic confidence ellipsoid. Table V gives the empirical probabilities, and Table VI gives $\det A^{-1}$ which is proportional to the volume of the ellipsoids, where $(\theta - \hat{\theta}_{LS})^T A (\theta - \hat{\theta}_{LS}) \leq 1$ is the expression for the confidence ellipsoids. The confidence ellipsoids based on asymptotic theory are smaller than the ones based on the ellipsoidal overbound, but the difference gets smaller as N increases, and more importantly the probability is guaranteed using SPS while it is not when using asymptotic theory.

V. CONCLUSIONS

In this paper we have shown that an ellipsoidal over-bound for the exact SPS confidence region can be found by solving a convex optimization problem. The proposed ellipsoidal outer-approximation can be easily computed using standard software tools, and it has the same shape as the confidence region based on asymptotic system identification theory, and it can be represented in a compact manner. Simulations showed that the ellipsoidal over-bound gets closer to the exact SPS region as the number of data points increases, and the average areas and empirical probabilities are insensitive to the particular type of noise affecting the system.

REFERENCES

[1] Boyd, S. and L. Vandenberghe (2004). *Convex optimization*. Cambridge University Press.
[2] Csáji, B. Cs., M. C. Campi, and E. Weyer (2012). "Non-Asymptotic Confidence Regions for the Least-Squares Estimate" Proceedings of *16th IFAC Symposium on System Identification*, pp. 227-232, Brussels, Belgium, July 2012.
[3] Löfberg, J. (2004). YALMIP : A Toolbox for Modeling and Optimization in MATLAB. In *Proceedings of the CACSD Conference*, Taipei, Taiwan, 2004.
[4] Ljung, L. (1999). *System Identification: Theory for the User*. Prentice-Hall. Upper Saddle River. 2nd edition.
[5] Söderström, T. and P. Stoica (1989). *System Identification*. Prentice Hall. ISBN 0-13-881236-5.

	M=20	M=20	M=100	M=100	
Noise	Over-b	Exact	Over-b	Exact	Asymp
Gaussian	0.980	0.962	0.990	0.966	0.960
Uniform	0.956	0.934	0.978	0.948	0.952
Laplacian	0.970	0.944	0.972	0.936	0.942
Time-var	0.970	0.936	0.972	0.932	0.940

TABLE IV

EMPIRICAL PROBABILITIES. $N = 200$.