2015 IEEE 54th Annual Conference on Decision and Control (CDC)
December 15-18, 2015. Osaka, Japan

Sign-Perturbed Sums (SPS) with Instrumental Variables for the Identification of ARX Systems

Valerio Volpe† Balázs Cs. Csáji§ Algo Carè* Erik Weyer** Marco C. Campi‡

Abstract— **We propose a generalization of the recently developed system identification method called Sign-Perturbed Sums (SPS). The proposed construction is based on the instrumental variables estimate and, unlike the original SPS, it can construct non-asymptotic confidence regions for linear regression models where the regressors contain past values of the output. Hence, it is applicable to ARX systems, as well as systems with feedback. We show that this approach provides regions with exact confidence under weak assumptions, i.e., the true parameter is included in the regions with a (user-chosen) exact probability for any finite sample. The paper also proves the strong consistency of the method and proposes a computationally efficient generalization of the previously proposed ellipsoidal outer-approximation. Finally, the new method is demonstrated through numerical experiments, using both real-world and simulated data.**

I. Introduction

Estimating parameters of partially unknown systems based on observations corrupted by noise is a classic problem in signal processing, system identification, machine learning and statistics [6], [12], [13], [14], [16]. Many standard methods are available which perform point estimations. Given an estimate, it is an intrinsic task to evaluate how close the estimated parameter is to the true one and such evaluation often comes in the form of a confidence region. Confidence regions are especially important for problems where the quality, stability or safety of a process has to be guaranteed.

The Sign-Perturbed Sums (SPS) method was presented in [1], [3], [19], [11]. Implementations of the method based on interval analysis have been proposed in [8], [9], [10], and an application of the method under a different set of assumptions has been presented in [15]. The main feature of the SPS method is that it constructs confidence regions which have an exact probability of containing the system's true parameter based on a finite number of observed data.

The SPS method of [3] and [19] provides exact confidence regions for the true parameter only when the regressors are exogenous (i.e., they do not depend on the noise terms), which is not the case with ARX systems, or, e.g., when feedback is involved. Generalizing the method to the case where the regressors can depend on the noise terms is of high practical importance.

In [1] an SPS method which deals with ARX systems has been given, and even more general systems have been considered in [11], [2]. However, these extensions introduce complications in the simple algorithm of [3] and [19], which make the method more challenging to analyze and more difficult to implement and run. In this paper we follow an alternative path, and show that an instrumental variables approach allows for notable simplifications in the algorithms. This leads, on the one hand, to computationally tractable methods for building regions and, on the other hand, to easy-to-prove, and quite general, strong consistency results.

The paper is organized as follows. In the next section we state the problem setting and our main assumptions. Then, the generalization of the SPS algorithm is presented in Section III, and in Section IV we illustrate the theoretical properties of the constructed confidence regions. Subsequently, we give a simplified construction by way of an outer ellipsoidal approximation algorithm similar to that developed in [3] for the case of exogenous regressors. Finally, in Section VI, we show two applications of the generalized SPS algorithm with numerical experiments, using both real-world and computer generated data. The proofs can be found in the extended version of this paper, [18].

II. Problem setting

This section presents the linear regression problem and introduces our main assumptions.

A. Data generation

The data are generated by the following system

$$Y_t \triangleq \varphi_t^{\mathrm{T}} \theta^* + N_t, \tag{1}$$

where Y_t is the output, N_t is the noise, φ_t is the regressors, and t is the discrete time index. Parameter θ^* is the true parameter to be estimated. The random variables Y_t and N_t are real-valued, while φ_t and θ^* are d-dimensional real vectors. We consider a finite sample of size n which consists of the regressors $\varphi_1, \dots, \varphi_n$ and the outputs $Y_1, \dots, Y_n$.

The work of B. Cs. Csáji was supported by the Hungarian Scientific Research Fund (OTKA), pr. no. 113038, and by the János Bolyai Research Fellowship of the Hungarian Academy of Sciences, pr. no. BO/00683/12/6. The work of A. Carè and E. Weyer was supported by the Australian Research Council (ARC) under Discovery Grant DP130104028. The work of M. C. Campi was partly supported by MIUR - Ministero dell'Istruzione, dell'Università e della Ricerca.

V. Volpe† and M. C. Campi‡ are with Department of Information Engineering, University of Brescia, Via Branze 38, 25123 Brescia, Italy; (email: v.volpe@studenti.unibs.it, marco.campi@unibs.it)

B. Cs. Csáji§ is with Fraunhofer Project Center at the Institute for Computer Science and Control (SZTAKI), Hungarian Academy of Sciences (MTA), Kende utca 13–17, Budapest, Hungary, H-1111; (email: balazs.csaji@sztaki.mta.hu)

A. Carè* and E. Weyer** are with Department of Electrical and Electronic Engineering, Melbourne School of Engineering, The University of Melbourne, 240 Grattan Street, Parkville, Melbourne, Victoria, 3010, Australia; (email: {algo.care,ewey}@unimelb.edu.au)

In addition, we assume that a set of instrumental variables $\{\psi_t\}_{t=1}^n$ is available to the user. The terms in the sequence must be correlated with the data and independent of the noise. Typically, past or filtered past inputs are used as instrumental variables.

B. Examples

There are many examples in signal processing and control of systems taking the form of (1), see [12], [16]. An important example is the widely used ARX model

$$Y_t = \sum_{i=1}^{d_1} a_i^* Y_{t-i} + \sum_{i=1}^{d_2} b_i^* U_{t-i} + N_t$$

where $\varphi_t = [Y_{t-1}, \ldots, Y_{t-d_1}, U_{t-1}, \ldots, U_{t-d_2}]^{\mathrm{T}}$ consists of past outputs and inputs, and the true parameter $\theta^* \in \mathbb{R}^{d_1+d_2}$ is the vector $[a_1^*, \ldots, a_{d_1}^*, b_1^*, \ldots, b_{d_2}^*]^{\mathrm{T}}$. An instrumental variables sequence $\{\psi_t\}$ can be easily obtained from the data. In particular, the instrumental variables vector can be constructed from the regressor φ_t by replacing the (noise-dependent) outputs with some other variables, such as delayed inputs, or noise-free reconstructed output terms, that can be computed using a guess of the true system parameter. The latter approach, in particular, is used and showed in Section VI.

C. Basic assumptions

Our assumptions on the regressors, the instrumental variables and the noise are:

A1 $\{N_t\}$ is a sequence of independent random variables. Each N_t has a symmetric probability distribution about zero.

A2 $\det(V_n) \neq 0$ almost surely, where

$$V_n \triangleq \frac{1}{n} \sum_{t=1}^{n} \psi_t \varphi_t^{\mathrm{T}}.$$

Note that A2 implies that matrix $H_n \triangleq \frac{1}{n}\sum_{t=1}^n \psi_t \psi_t^{\mathrm{T}}$ is (almost surely) invertible.

Like the SPS of [3] the assumptions are rather mild, since there are no moment or density requirements on the noise terms, and their distributions can change with time and need not be known. The strongest assumption on the noise is that it forms an independent sequence, but it can be somehow relaxed with the suitably modified Block SPS [3]. The core assumption is the symmetricity of the noise. Many standard distributions satisfy this property. These weak requirements make the method widely applicable.

III. Sign-Perturbed Sums with Instrumental Variables

In this section we introduce the generalization of SPS using instrumental variables.

A. Intuitive idea

First, recall that the instrumental variables estimate $\hat{\theta}_n$ comes as the solution to a modified version of the normal equations, i.e.,

$$\sum_{t=1}^{n} \psi_t (Y_t - \varphi_t^{\mathrm{T}} \theta) = 0, \tag{2}$$

and the instrumental variables (IV) estimate is

$$\hat{\theta}_n \triangleq \left(\sum_{t=1}^{n} \psi_t \varphi_t^{\mathrm{T}} \right)^{-1} \sum_{t=1}^{n} \psi_t Y_t.$$

Then, referring to the same ideas as in [3] for the construction of the SPS method, we can build $m-1$ sign-perturbed versions of equation (2), and define the sign-*perturbed sums* as

$$S_i(\theta) \triangleq H_n^{-\frac{1}{2}} \frac{1}{n} \sum_{t=1}^{n} \psi_t \alpha_{i,t} (Y_t - \varphi_t^{\mathrm{T}} \theta),$$

$i \in \{1, \ldots, m-1\}$, where $H_n^{1/2}$ is the principal square root of H_n, which is introduced in order to give a better shape to the confidence regions, and $\{\alpha_{i,t}\}$ are i.i.d. Rademacher variables, i.e., they take on the values ± 1 with probability 1/2 each. Also, without applying sign-perturbations, we can define the *reference sum* as

$$S_0(\theta) \triangleq H_n^{-\frac{1}{2}} \frac{1}{n} \sum_{t=1}^{n} \psi_t (Y_t - \varphi_t^{\mathrm{T}} \theta).$$

An important property of these functions is that corresponding to $\theta = \theta^*$ we have

$$S_0(\theta^*) = H_n^{-\frac{1}{2}} \frac{1}{n} \sum_{t=1}^{n} \psi_t N_t,$$

$$S_i(\theta^*) = H_n^{-\frac{1}{2}} \frac{1}{n} \sum_{t=1}^{n} \alpha_{i,t} \psi_t N_t = H_n^{-\frac{1}{2}} \frac{1}{n} \sum_{t=1}^{n} \pm \psi_t N_t,$$

and such variables are uniformly ordered, i.e., once the values of $\{\|S_i(\theta^*)\|^2\}_{i=0}^{m-1}$ have been sorted according to a particular strict total order, any $\|S_i(\theta^*)\|^2$ has the same probability of being ranked in a given position (see [18, Appendix A]). This observation is crucial to SPS since it builds the confidence regions by excluding those θ for which $\|S_0(\theta)\|^2$ is among the q largest ones, and the so constructed confidence set has exact probability $1 - q/m$ of containing the true parameter[1].

Moreover, when $\|\theta' - \theta^*\|$ is large $\|S_0(\theta')\|^2$ tends to be the largest of the m functions. Therefore, defining π as a *random* permutation of the set $\{0, \ldots, m-1\}$ and the strict total order by[2]

$$Z_j \succ_\pi Z_k \Leftrightarrow (Z_j > Z_k) \vee (Z_j = Z_k \wedge \pi(j) > \pi(k)),$$

where $Z_i = \|S_i(\theta')\|^2$, it happens that values far away from θ^* are excluded from the confidence set.

[1] Notice that many q and m pairs give the same ratio q/m. Refer to [3] for more discussion on the choice of q and m.

[2] The random permutation π is used to break ties in case two different $\|S_i(\theta')\|^2$ variables take on the same value.

B. Formal construction of the confidence region

The pseudocode of the generalized SPS algorithm is presented in two parts. The initialization (Table I) sets the main global parameters and generates the random objects needed for the construction. In the initialization, the user provides the desired confidence probability p. The second part (Table II) evaluates an indicator function, SPS-Indicator(θ), which determines if a particular parameter θ is included in the confidence region.

PSEUDOCODE: SPS-INITIALIZATION

1. Given a (rational) confidence probability $p \in (0,1)$, set integers $m > q > 0$ such that $p = 1 - q/m$;
2. Calculate the outer product
$$H_n \triangleq \frac{1}{n}\sum_{t=1}^{n} \psi_t \psi_t^{\mathrm{T}},$$
and find the principal square root $H_n^{1/2}$, such that
$$H_n^{1/2} H_n^{1/2} = H_n;$$
3. Generate $n\,(m-1)$ i.i.d. random signs $\{\alpha_{i,t}\}$ with
$$\mathbb{P}(\alpha_{i,t} = 1) \,=\, \mathbb{P}(\alpha_{i,t} = -1) \,=\, \tfrac{1}{2},$$
for $i \in \{1, \dots, m-1\}$ and $t \in \{1, \dots, n\}$;
4. Generate a random permutation π of the set $\{0, \dots, m-1\}$, where each of the $m!$ possible permutations has the same probability $1/(m!)$ to be selected.

TABLE I

PSEUDOCODE: SPS-INDICATOR (θ)

1. For the given θ, compute the prediction errors for $t \in \{1, \dots, n\}$
$$\varepsilon_t(\theta) \,\triangleq\, Y_t - \varphi_t^{\mathrm{T}} \theta;$$
2. Evaluate
$$S_0(\theta) \triangleq H_n^{-\frac{1}{2}} \frac{1}{n} \sum_{t=1}^{n} \psi_t \varepsilon_t(\theta),$$
$$S_i(\theta) \triangleq H_n^{-\frac{1}{2}} \frac{1}{n} \sum_{t=1}^{n} \alpha_{i,t}\, \psi_t \varepsilon_t(\theta),$$
for $i \in \{1, \dots, m-1\}$;
3. Order scalars $\{\|S_i(\theta)\|^2\}$ according to $\succ_\pi$;
4. Compute the rank $\mathcal{R}(\theta)$ of $\|S_0(\theta)\|^2$ in the ordering where $\mathcal{R}(\theta) = 1$ if $\|S_0(\theta)\|^2$ is the smallest in the ordering, $\mathcal{R}(\theta) = 2$ if $\|S_0(\theta)\|^2$ is the second smallest, and so on;

6. Return 1 if $\mathcal{R}(\theta) \leq m - q$, otherwise return 0.

TABLE II

Using this construction, we can define the p-level SPS *confidence region* as follows
$$\widehat{\Theta}_n \triangleq \left\{ \theta \in \mathbb{R}^d : \text{SPS-Indicator}(\theta) = 1 \right\}.$$

Note that, corresponding to the instrumental variables estimate $\hat{\theta}_n$, it holds that $S_0(\hat{\theta}_n) = 0$. Therefore, with exception of pathological cases, $\hat{\theta}_n$ is included in the SPS confidence region, and the set is built around $\hat{\theta}_n$.

IV. THEORETICAL RESULTS

A. Exact confidence

The most important property of the SPS method is that the generated regions have *exact* confidence probabilities for any *finite* sample. The following theorem holds.

Theorem 1: Assuming A1 and A2, the confidence probability of the constructed confidence region is exactly p, that is,
$$\mathbb{P}\big(\theta^* \in \widehat{\Theta}_n\big) \,=\, 1 - \frac{q}{m} \,=\, p.$$

The proof of the theorem, which is along the lines of the proof of Theorem 1 of [3], can be found in [18]. Since the confidence probability is exact, no conservatism is introduced. Moreover, the statistical assumptions imposed on the noise are rather weak. Indeed the noise distribution can change during time, and there are no moment or density requirements whatsoever.

B. Strong consistency

An important aspect of the confidence region is its size. Clearly for any finite sample the size of the region depends much on the statistical properties of the noise. However, we show that asymptotically the SPS regions become smaller and smaller, shrinking to the true parameter. Indeed the SPS algorithm is *strongly consistent*, under the following (rather mild) assumptions.

A3 There exists a positive definite matrix H such that
$$\lim_{n \to \infty} H_n = H, \text{ almost surely.}$$

A4 There exists an invertible matrix V such that
$$\lim_{n \to \infty} V_n = V, \text{ almost surely.}$$

A5 (*regressor growth rate restriction*):
$$\sum_{t=1}^{\infty} \frac{\|\varphi_t\|^4}{t^2} < \infty, \text{ almost surely.}$$

A6 (*instruments growth rate restriction*):
$$\sum_{t=1}^{\infty} \frac{\|\psi_t\|^4}{t^2} < \infty, \text{ almost surely.}$$

A7 (*noise variance growth rate restriction*):
$$\sum_{t=1}^{\infty} \frac{\mathbb{E}[N_t^2]^2}{t^2} < \infty.$$

The following theorem holds.

Theorem 2: Assuming A1, A2, A3, A4, A5, A6 and A7, $\forall \varepsilon > 0$ there almost surely exists an N such that $\forall n > N, \hat{\Theta}_n \subseteq \{\theta \in \mathbb{R}^d : \|\theta - \theta^*\| \leq \varepsilon\}$.

The proof of the theorem can be found in [18]. The claim states that the confidence regions $\{\hat{\Theta}_n\}$ will eventually be included (almost surely) in any norm-ball centered at θ^* as the sample size increases. Although the regions generated by the generalization of SPS introduced in this paper have no theoretical guarantee of being bounded, they normally are, and, moreover, the strong consistency result implies that they are bounded with probability 1 asymptotically.

V. Ellipsoidal approximation algorithm

The purpose of the SPS-Indicator function is to check whether a given θ belongs to the confidence region or not. In particular, it computes the $\{\|S_i(\theta)\|^2\}_{i=0}^{m-1}$ functions for that specific θ and compares them. This way the SPS region can be constructed by decomposing the space of interest in a grid, possibly very dense, and checking whether the points in the grid belongs to the region. However, this approach is *computationally demanding*, and it gets slower and slower as the dimensions increase. Here, we introduce a generalization of the ellipsoidal outer approximation algorithm previously introduced for the SPS of [3], [19]. The algorithm leads to an ellipsoidal over-bound that can be efficiently computed in polynomial time.

In particular, referring to the same ideas and procedure discussed in detail in [3] and [19], with slight and straightforward modifications, we can build the sought over-bound region as

$$\hat{\hat{\Theta}}_n \triangleq \left\{ \theta \in \mathbb{R}^d \colon (\theta - \hat{\theta}_n)^{\mathrm{T}} V_n^{\mathrm{T}} H_n^{-1} V_n (\theta - \hat{\theta}_n) \leq r \right\},$$

where r is defined as the qth largest solution of the following convex semi-definite programming problems[3], for $i = 1, \ldots, m-1$,

$$\begin{array}{ll} \text{minimize} & \gamma \\ \text{subject to} & \lambda \geq 0 \\ & \begin{bmatrix} -I + \lambda A_i & \lambda b_i \\ \lambda b_i^{\mathrm{T}} & \lambda c_i + \gamma \end{bmatrix} \succeq 0, \end{array} \tag{3}$$

where "$\succeq 0$" denotes that a matrix is positive semidefinite, and

$$\begin{aligned} A_i &\triangleq I - H_n^{\frac{1}{2}\mathrm{T}} V_n^{-\mathrm{T}} Q_i^{\mathrm{T}} H_n^{-1} Q_i V_n^{-1} H_n^{\frac{1}{2}}, \\ b_i &\triangleq H_n^{\frac{1}{2}\mathrm{T}} V_n^{-\mathrm{T}} Q_i^{\mathrm{T}} H_n^{-1} (\rho_i - Q_i \hat{\theta}_n), \\ c_i &\triangleq -\rho_i^{\mathrm{T}} H_n^{-1} \rho_i + 2\hat{\theta}_n^{\mathrm{T}} Q_i^{\mathrm{T}} H_n^{-1} \rho_i - \hat{\theta}_n^{\mathrm{T}} Q_i^{\mathrm{T}} H_n^{-1} Q_i \hat{\theta}_n, \\ Q_i &\triangleq \frac{1}{n} \sum_{t=1}^{n} \alpha_{i,t} \psi_t \varphi_t^{\mathrm{T}}, \\ \rho_i &\triangleq \frac{1}{n} \sum_{t=1}^{n} \alpha_{i,t} \psi_t Y_t. \end{aligned}$$

[3]Any of these problem can be easily solved in polynomial time using, e.g., MATLAB and a toolbox such as CVX [7].

Since $\hat{\hat{\Theta}}_n$ is an overbound of the SPS region $\hat{\Theta}_n$, i.e., $\hat{\Theta}_n \subseteq \hat{\hat{\Theta}}_n$, it clearly holds that

$$\mathbb{P}\big(\theta^* \in \hat{\hat{\Theta}}_n\big) \geq 1 - \frac{q}{m} = p,$$

for any finite n.

The pseudocode for computing $\hat{\hat{\Theta}}_n$ is given in table III.

PSEUDOCODE: SPS-OUTER-APPROXIMATION
1. Compute the instrumental variables estimate $\hat{\theta}_n = \left(\sum_{t=1}^{n} \psi_t \varphi_t^{\mathrm{T}} \right)^{-1} \sum_{t=1}^{n} \psi_t Y_t;$
2. For $i \in \{1, \ldots, m-1\}$, solve the optimization problem (3), and let γ_i^* be the optimal value (or ∞ if the problem is infeasible);
3. Let r be the qth largest γ_i^* value;
4. The outer approximation of the SPS confidence region is given by the ellipsoid $\hat{\hat{\Theta}}_n = \big\{ \theta \in \mathbb{R}^d \colon (\theta - \hat{\theta}_n)^{\mathrm{T}} V_n^{\mathrm{T}} H_n^{-1} V_n (\theta - \hat{\theta}_n) \leq r \big\}.$

TABLE III

VI. Numerical experiments

In this section we illustrate SPS with numerical experiments. Firstly, we apply the method to a simple first-order ARX system. Then, SPS is applied to a real-world identification problem, with the purpose of showing that the method is robust against the assumptions from which the guarantees provided in this paper are established.

A. Simulation example

We consider the following data generating ARX system

$$Y_t = a^* Y_{t-1} + b^* U_t + N_t,$$

where $a^* = 0.7, b^* = 1$, and $\{U_t\}$ is a sequence of random inputs generated as

$$U_t = 0.75 U_{t-1} + V_t,$$

being $\{V_t\}$ a sequence of i.i.d. Gaussian random variables $N(0,1)$. $\{N_t\}$ is a sequence of i.i.d. Laplacian random variables with zero mean and variance 1. We consider a finite sample of size n, that consists of couples $\{(Y_t, \varphi_t)\}_{t=1}^{n}$.

The instrumental variables $\{\psi_t\}_{t=1}^{n}$ are constructed from the data. In particular, we replace the autoregressive components of the regressors φ_t, for $t = 2, \ldots, n$, with reconstructed outputs. Firstly we find an estimate $\hat{\theta}_{\mathrm{LS}}$ of the true parameter via least squares on $\{(Y_t, \varphi_t)\}_{t=1}^{n}$, and then we use such estimate[4] to build the noise-free sequence $\{\tilde{Y}_t\}_{t=1}^{n}$ using the following recursive procedure

$$\tilde{Y}_t = \hat{a} \tilde{Y}_{t-1} + \hat{b} U_t,$$

[4]We could also use a *guess* (even imprecise) of the true parameter coming from some *a-priori* knowledge.

where $\hat{\theta}_{\text{LS}} = [\hat{a}, \hat{b}]^{\text{T}}$, and we use Y_1 as initialization value. Finally, the instrumental variables are

$$\psi_t \triangleq [\tilde{Y}_{t-1}, U_t]^{\text{T}}.$$

Note that, rigorously speaking, these instrumental variables are not completely independent of the noise, due to the presence of the noise realization in the least squares estimate. However, in $\hat{\theta}_{\text{LS}}$, the noise is *averaged out*, so that the effect of the noise is toned down. If the least squares estimate were built from a set independent of the one used by SPS then the constructed regions would be rigorous. Yet, the difference would be minimal, thus, for the sake of simplicity, we used just one data set.

Based on $n = 25$ data points $\{(Y_t, \varphi_t)\}_{t=1}^{25}$ we want to find a 95% confidence region for θ^*. We build 99 sign-perturbed sums (m is set to 100), and the confidence region is constructed as the values of θ for which at least $q = 5$ of the $\|S_i(\theta)\|^2$, $i = 1, \ldots, 99$, functions are "larger"[5] than $\|S_0(\theta)\|^2$. An example of constructed confidence region is illustrated in figure 1. The solid red line has been obtained by evaluating the SPS-Indicator(θ) function in table II on a very fine grid.

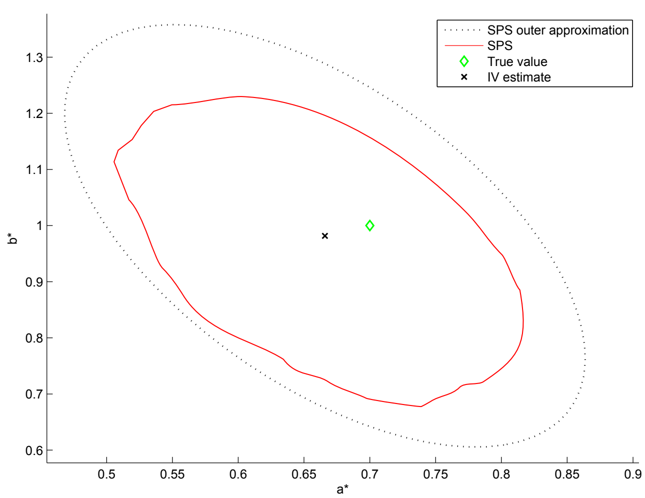

Fig. 1. 95% confidence region, $n = 25, m = 100$.

B. Real-world data experiment

Working with real-world data is almost always a challenge. Usually, the user can only presume the nature of the *best* mathematical representation of the system, and most of the times the real system does not lie in the model class. Moreover, the knowledge on the noise characteristics is limited. All these issues make the identification process much more complicated. Nevertheless, we still want to apply SPS in such a scenario, and even though the theoretical results cannot be expected to hold rigorously, since, e.g., the real system does not lie in the model class, we hope that they hold approximately.

[5]According to the strict total order $\succ_\pi$, with a random permutation π.

Our real-world data set comes from the photovoltaic energy production measurements of a prototype energy-positive public lighting microgrid (E+Grid) system [4]. In particular, the available data contain the hourly historical progression of the amount of energy produced.

The model class is an ARX(5, 4), i.e.,

$$Y_t = \sum_{i=1}^{5} a_i Y_{t-i} + \sum_{i=1}^{4} b_i U_{t-i+1} + N_t = \varphi_t^{\text{T}} \theta + N_t,$$

where Y_t is the amount of produced energy and U_t is an auxiliary input given by the clear-sky predictions of the amount of energy produced (see [4] for more details).

To carry out our tests, we first estimated via least squares a "true parameter" $\hat{\theta}^*$ based on the first half of the large (more than 4200 observations) data set available. After $\hat{\theta}^* \triangleq [\hat{a}^*, \hat{b}^*]^{\text{T}}$ was found, the residuals $\varepsilon_t = Y_t - \sum_{i=1}^{5} \hat{a}_i^* Y_{t-i} - \sum_{i=1}^{4} \hat{b}_i^* U_{t-i+1}$ were tested with the Durbin-Watson algorithm, [5], which returned a p-value bigger than 95% for the uncorrelation hypothesis, supporting the choice of the orders 5 and 4 [17].

Then, SPS was used with the second half of the data set. The instrumental variables $\{\psi_t\}$ were built from the data by replacing the autoregressive components of the regressor with a reconstructed noise-independent trajectory of the output $\{\tilde{Y}_t\}$, similarly to what has been done in the previous example. The estimate of the "true parameter" used to build such a sequence was obtained via least squares on an extra subset of data consisting of 100 samples, which was not used later.

Finally, we evaluated the empirical probability with which $\hat{\theta}^*$ belonged to the SPS regions that were built using many (1000) different data subsets, in a Monte Carlo approach. Each subset was constructed with pairs $\{(Y_t, \varphi_t)\}$ drawn randomly (non-sequentially) from the second half of the global data set. The size of each subset varied from 75 to 250 observations, and the parameter m, q were always set, respectively, to 100 and 10, looking for a region of (desired) confidence probability equal to 90%.

The final results, illustrated in table IV, show a good adherence between theory and empirical results.

n	Empirical confidence
75	0.886
100	0.900
150	0.886
200	0.906
250	0.910

TABLE IV

VII. Concluding Remarks

A new SPS algorithm has been proposed in this paper that, unlike the original version of SPS, can be used when the regressors contain past values of the system output, which makes it suitable for the identification of ARX systems. The algorithm makes use of instrumental variables (IV). However, it has to be noted that the reason for using an IV with SPS is quite different from other IV system identification methods. Particularly, in this version of SPS the IV does not counteract the presence of correlated noise, as it is in other IV approaches, and in fact the noise terms are supposed to form an independent pattern in this paper. Instead, the IV is introduced to ease the implementation of the method which is explained by noting that the IV only contains exogenous variables that are not affected by the system noise so that no noise sign perturbation is required in the IV when the sign-perturbed functions are constructed. Along an alternative approach, one may consider using the initial regressor φ_t in place of the IV, which might give better shaped regions. However, this would require a more cumbersome implementation of the algorithm for the sign perturbation of the regressor, as it is done in [1]. An evaluation of the pros and cons of these two approaches will be the subject of future investigations.

REFERENCES

[1] Balázs Cs. Csáji, Marco C. Campi, and Erik Weyer. Non-asymptotic confidence regions for the least-squares estimate. In *Proc. IFAC SYSID*, pages 227–232, 2012.

[2] Balázs Cs. Csáji, Marco C. Campi, and Erik Weyer. Sign-Perturbed Sums (SPS): A method for constructing exact finite-sample confidence regions for general linear systems. In *CDC*, pages 7321–7326, 2012.

[3] Balázs Cs. Csáji, Marco C. Campi, and Erik Weyer. Sign-Perturbed Sums: A new system identification approach for constructing exact non-asymptotic confidence regions in linear regression models. *Signal Processing, IEEE Transactions on*, 63(1):169–181, Jan 2015.

[4] Balázs Cs. Csáji, András Kovács, and József Váncza. Adaptive aggregated predictions for renewable energy systems. In *Adaptive Dynamic Programming and Reinforcement Learning (ADPRL), 2014 IEEE Symposium on*, pages 1–8. IEEE, 2014.

[5] James Durbin and Geoffrey S. Watson. Testing for serial correlation in least squares regression: I. *Biometrika*, pages 409–428, 1950.

[6] Michel Gevers. A personal view of the development of system identification: a 30-year journey through an exciting field. *Control Systems, IEEE*, 26(6):93–105, 2006.

[7] Michael Grant, Stephen Boyd, and Yinyu Ye. CVX: Matlab software for disciplined convex programming, 2008.

[8] Michel Kieffer and Eric Walter. Guaranteed characterization of exact confidence regions for FIR models under mild assumptions on the noise via interval analysis. In *Decision and Control (CDC), 2013 IEEE 52nd Annual Conference on*, pages 5048–5053. IEEE, 2013.

[9] Michel Kieffer and Eric Walter. Guaranteed characterization of exact nonasymptotic confidence regions in nonlinear parameter estimation. In *Proc. IFAC NOLCOS*, pages 56–61, 2013.

[10] Michel Kieffer and Eric Walter. Guaranteed characterization of exact non-asymptotic confidence regions as defined by LSCR and SPS. *Automatica*, 50(2):507–512, 2014.

[11] Sándor Kolumbán, István Vajk, and Johan Schoukens. Perturbed datasets methods for hypothesis testing and structure of corresponding confidence sets. *Automatica*, 51:326–331, 2015.

[12] Lennart Ljung. *System Identification: Theory for the User*. Prentice-Hall, Upper Saddle River, 2nd edition, 1999.

[13] Lennart Ljung. Perspectives on system identification. *Annual Reviews in Control*, 34(1):1–12, 2010.

[14] Lennart Ljung and Torkel Glad. *Modeling of Dynamic Systems*. Prentice Hall, 1994.

[15] Alexander Senov, Konstantin Amelin, Natalia Amelina, and Oleg Granichin. Exact confidence regions for linear regression parameter under external arbitrary noise. In *American Control Conference (ACC), 2014*, pages 5097–5102. IEEE, 2014.

[16] Torsten Söderström and Petre Stoica. *System Identification*. Prentice Hall International, Hertfordshire, UK, 1989.

[17] Valerio Volpe. *Identification of dynamical systems with finitely many data points*. University of Brescia, M. Sc. Thesis, March 2015.

[18] Valerio Volpe, Balázs Cs. Csáji, Algo Carè, Erik Weyer, and Marco C. Campi. Sign-Perturbed Sums (SPS) with Instrumental Variables for the Identification of ARX Systems - Extended Version. `http://arxiv.org/abs/1509.04774`, 2015.

[19] Erik Weyer, Balázs Cs. Csáji, and Marco C. Campi. Guaranteed non-asymptotic confidence ellipsoids for FIR systems. In *Proceedings of the 52st IEEE Conference on Decision and Control*, pages 7162–7167, 2013.