2016 IEEE 55th Conference on Decision and Control (CDC)
ARIA Resort & Casino
December 12-14, 2016, Las Vegas, USA

Sign-Perturbed Sums (SPS) with Asymmetric Noise: Robustness Analysis and Robustification Techniques

Algo Carè* Balázs Csanád Csáji* Marco C. Campi**

***Abstract*— Sign-Perturbed Sums (SPS) is a recently developed finite sample system identification method that can build exact confidence regions for linear regression problems under mild statistical assumptions. The regions are well-shaped, e.g., they are centred around the least-squares (LS) estimate, star-convex and strongly consistent. One of the main assumptions of SPS is that the distribution of the noise terms are symmetric about zero. This paper analyses how robust SPS is with respect to the violation of this assumption and how it could be robustified with respect to non-symmetric noises. First, some alternative solutions are overviewed, then a robustness analysis is performed resulting in a robustified version of SPS. We also suggest a modification of SPS, called LAD-SPS, which builds exact confidence regions around the least-absolute deviation (LAD) estimate instead of the LS estimate. LAD-SPS requires less assumptions as the noise needs only to have a conditionally zero median (w.r.t. the past). Furthermore, that approach can also be robustified using similar ideas as in the LS-SPS case. Finally, some numerical experiments are presented.**

I. Introduction

Estimating parameters of partially unknown systems based on observations corrupted by noise is a fundamental problem in system identification, signal processing, machine learning and statistics, [14], [15], [22]. Standard solutions such as the least squares (LS) method or, more generally, prediction error methods provide *point estimates*. In many situations, for example, when the safety, stability or quality of a process has to be guaranteed, a point estimate should be accompanied with a *confidence region* that certifies the accuracy of the estimate. If the noise is assumed to belong to a known bounded set, set membership approaches can be used to identify the region of parameter values that are consistent with the observed data [16]. We take a probabilistic perspective instead [2]. Standard probabilistic methods for constructing *confidence regions* around point estimates require that much statistical information on the noise be available to the user, which is seldom the case in applications. Indeed, standard probabilistic methods construct confidence regions that are guaranteed only asymptotically. The recent *Sign-Perturbed Sums* (SPS) method [9], [10], [24], [13] can construct confidence regions which have an *exact* coverage probability of the system's true parameter based only on *finite* sample of observations and under *mild* statistical assumptions. The SPS confidence sets are well-shaped, e.g., they are *star convex* with the LS estimate as a star center [10], and *strongly consistent* [7].

A. Problem Setting

In this paper, we consider scalar *linear regression* systems

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

where Y_t is the output, N_t is the noise, φ_t is the regressor, 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 $\varphi_1, \ldots, \varphi_n$ and the outputs $Y_1, \ldots, Y_n$. Following [10], we assume that the regressors are *deterministic*, but all the results here presented can be immediatly generalised to random regressors when they are independent of the noise. We focus on (1) in order to keep notations, algorithms and proofs as simple as possible. However, it is important to note that the arguments developed here can be carried over to more complex systems, in line with, e.g., [8], [9], [18].

The standard assumption in the SPS literature is that the noises, $\{N_t\}$, are *independent* and *symmetric* about zero, but not necessarily identically distributed. This paper aims at studying in a quantitative way the *robustness* of the SPS algorithm when the symmetry assumption is violated, and proposing *robustification* techniques. Some ways of relaxing the symmetry assumption have been already considered in the literature and are now briefly reviewed.

B. Relaxations of the Symmetry Assumption in the Literature

A possible way to circumvent the symmetry assumption on the noise is transferring the symmetry requirement from the noise to the input affecting the regressors. This idea was proposed in [3], [5] for the LSCR (Leave-Out Sign-Dominant Correlation Regions) algorithm, a predecessor of SPS, and has been recently applied to SPS in [20]. In line with this idea, when instrumental variables are available, [23], the symmetry requirement could be transferred to them.

In [4], the authors note that if the noise process is independent and identically but not necessarily symmetrically distributed, then a symmetrically distributed noise sequence can be obtained from it by considering the *difference process*.

The work of A. Carè was supported by the European Research Consortium for Informatics and Mathematics (ERCIM) and the Australian Research Council (ARC) under Discovery Grant DP130104028. 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, pr. no. BO/00217/16/6. The work of M. C. Campi was partly supported by MIUR - Ministero dell'Istruzione, dell'Università e della Ricerca.

*A. Carè and B. Cs. Csáji are with the Institute for Computer Science and Control (SZTAKI), Hungarian Academy of Sciences (MTA), Kende utca 13–17, Budapest, Hungary, H-1111; (email: algocare@gmail.com, balazs.csaji@sztaki.mta.hu)

** M. C. Campi is with the Department of Information Engineering, University of Brescia, Via Branze 38, 25123 Brescia, Italy; (email: marco.campi@unibs.it)

Indeed, defining the difference output process of system (1) as $\Delta Y_t = Y_{2t} - Y_{2t-1}$, $t = 1, 2, \ldots$, we get

$$\Delta Y_t \;=\; (\varphi_{2t} - \varphi_{2t-1})^{\mathrm{T}}\theta^* + (N_{2t} - N_{2t-1}), \qquad (2)$$

which is a process affected by an independent and symmetric noise process $\{N_{2t} - N_{2t-1}\}$ whenever $\{N_t\}$ is i.i.d., so that SPS can be applied rigorously to (2). However, assuming i.i.d. noise could be unrealistic in some real situations. Moreover, the identification of parameter θ^* from the difference process can be poor in the case of slowly changing inputs, due to the bad *signal-to-noise ratio* that occurs when the magnitude of the regressor in (2), $\|\varphi_{2t} - \varphi_{2t-1}\|$, is small.

Another approach was pursued in [13], where the authors devise a permutation-based algorithm that relies on the assumption that the noise sequence is exchangeable rather than independent and symmetric. The drawbacks of this method are basically the same that affect the "difference process" idea, as the distribution of $\{N_t\}$ is not allowed to be time-varying and additional requirements on the variability of the regressors are to be met [13, Definition 3].

C. Aim and Structure

The aim of this paper is assessing in a rigorous way the robustness of the SPS approach to *violations of the symmetry assumption* without imposing any further assumption. We also suggest some robustification techniques. Section II begins with a measure-theoretic argument to show that the original SPS algorithm, with no modifications, is expected to exhibit some degree of robustness. An ideal SPS algorithm that makes use of an "oracle", i.e., that has access to information that are not available to the user in real life, is introduced in Section II-A. The oracle-based algorithm builds regions that are guaranteed to contain the true parameter θ^* with an exact user-chosen confidence. In the presence of asymmetries, the shape of the oracle-based regions incurs a graceful degradation as a higher level of asymmetry is introduced, but these regions remain guaranteed with exact, user-chosen confidence. In Section II-B, we show that *outer-approximations* of the oracle-based regions can be built in practice, which only require a bound on the asymmetric deviation of the noises, and which lead to a robust variant of the SPS algorithm. We also show that comparing the oracle-based algorithm and the standard SPS lead to an alternative way to quantify the robustness of the original algorithm.

In the second part of the paper, in Section III, we consider a variant of SPS that builds regions around the *least absolute deviation* (LAD) estimate instead of around the *least squares* estimate as in the standard SPS algorithm. Since this variant is still a fully-fledged "sign-perturbed-sum" method, we took the liberty to name it LAD-SPS. Connections to the econometric literature where similar sign-error approaches have been proposed will also be pointed out. LAD-SPS is more robust than standard SPS against violations of the symmetry assumption. In particular, LAD-SPS builds *exact* regions also in the presence of noise that is not zero-mean, if the requirement of having *conditionally zero-median* noise is met. In Section III-A, by applying the robustness analysis and robustification tools of Section II, we show that robustness of LAD-SPS to asymmetries in the median can also be achieved. Finally, in Section IV, we investigate the presented methods through a series of numerical experiments.

II. Robustness of SPS with Asymmetric Noise

Let us denote by P_t the probability measure of the noise sample N_t at time t, and by B the probability measure over $\{-1, +1\}$ that assigns a probability of $1/2$ to each sign. Theorem 1 in [10] states that, under the assumption that the noise terms are independent and symmetric about zero,

$$\mathbb{P}\{\,\theta^* \in \widehat{\Theta}\,\} \;=\; 1 - \frac{q}{m},$$

i.e., the probability that the SPS region $\widehat{\Theta}$ includes the true parameter is user-chosen through the parameters q and m. Note that the region $\widehat{\Theta}$ depends on the noise and the random sign sequences.[1] Using that the noises form an independent sequence and they are also independent from the i.i.d. signs, the probability $\mathbb{P}$ that measures the event $\{\theta^* \in \widehat{\Theta}\}$ is a product measure of the marginal distributions P_t of N_t, $t = 1, \ldots, n$, and B, namely $\mathbb{P}_n \triangleq \mathrm{P}_1\mathrm{P}_2\cdots\mathrm{P}_n\mathrm{B}^{m\cdot n}$. Thus, a more explicit way of stating Theorem 1 in [10] is:

$$\mathrm{P}_1\mathrm{P}_2\cdots\mathrm{P}_n\mathrm{B}^{m\cdot n}\{\,\theta^* \in \widehat{\Theta}\,\} \;=\; 1 - \frac{q}{m},$$

for all symmetric probability measures $\mathrm{P}_1, \mathrm{P}_2, \cdots, \mathrm{P}_n$, i.e., the result is essentially *distribution-free*.

Assume that the true distribution of the noise at time t is instead $\tilde{\mathrm{P}}_t$, possibly non-symmetric. Define $\tilde{\mathbb{P}}_n \triangleq \tilde{\mathrm{P}}_1\tilde{\mathrm{P}}_2\cdots\tilde{\mathrm{P}}_n\mathrm{B}^{m\cdot n}$ and let $\sigma(N_t)$ be the σ-algebra generated by N_t. Assume also that, for each t, the *total variation* distance [11], [19] between the true distribution $\tilde{\mathrm{P}}_t$ of N_t and *at least one* symmetric distribution P'_t is bounded by ν_t, that is, *there exists* a symmetric P'_t such that

$$\sup_{A\in\sigma(N_t)} |\,\mathrm{P}'_t(A) - \tilde{\mathrm{P}}_t(A)\,| < \nu_t,$$

and $\nu_t \le \nu$ for all t. Then, for the product measure built by these distributions instead, $\mathbb{P}'_n \triangleq \mathrm{P}'_1\mathrm{P}'_2\cdots\mathrm{P}'_n\mathrm{B}^{m\cdot n}$,

$$\sup_{A\in\mathcal{F}} |\,\mathbb{P}'_n(A) - \tilde{\mathbb{P}}_n(A)\,| \le \sum_{t=1}^{n}\nu_t \le n\cdot\nu,$$

where $\mathcal{F}$ is the σ-algebra over the (product) space of the noise and the m random sign sequences of length n, and the inequality follows from the properties of total variation distance, cf. [19]. Hence, we have shown that

$$\tilde{\mathbb{P}}_n\{\,\theta^* \in \widehat{\Theta}\,\} \;\ge\; 1 - \frac{q}{m} - n\nu. \qquad (3)$$

Thus, if for each t there exists a symmetric distribution that is at a distance much smaller than $\frac{1}{n}$ from the (marginal) distribution of noise at time t, then the deterioration of the confidence in the SPS algorithm is negligible. However conservative, this bound is a first indication that the impact of small asymmetries can be kept under control.

[1] We omit, without loss of generality, the probability over the $m!$ possible permutations that affect the region only in case of ties.

A. Oracle-Based Algorithm

Given a random variable V distributed with probability $\mathbb{P}_V$, we define the *asymmetric deviation* of V as

$$\eta_V \triangleq \mathbb{E}_V\left[\,\mathrm{sign}(V) \mid \sigma(|V|)\,\right],$$

i.e., the expected value of the sign of V conditional to the σ-algebra generated by $|V|$, $\sigma(|V|)$. Note that the asymmetric deviation is a *random variable* in general and, if $\mathbb{P}_V$ is symmetric, its value is zero with probability one. The value of η_V can be expressed as a function of $|V|$, [21]. For a given number $\delta \in [0,1]$, we define the function $\Phi(V;\delta)$ as

$$\Phi(V;\delta) = \begin{cases} -V & \text{if } \mathrm{sign}(V) = \mathrm{sign}(\eta_V) \text{ and } \delta \leq \lambda(|V|) \\ V & \text{otherwise} \end{cases}$$

where $\lambda(|V|) \triangleq \frac{|\eta_V|}{1+|\eta_V|}$. When δ is randomly sampled, $\Phi(\,\cdot\,;\delta)$ works as a symmetrising operator:

Lemma 1: If δ is a random variable uniformly distributed over $[0,1]$, independent of the random variable V, then the distribution of $\Phi(V;\delta)$ is symmetric about zero.

Proof: The random variable $\Phi(V;\delta)$ is symmetric if and only if $\mathbb{E}_{V,\delta}\left[\,\mathrm{sign}(\Phi(V;\delta)) \mid \sigma(|\Phi(V;\delta)|)\,\right] = 0$ (a.s.). Note that $\mathrm{sign}(\Phi(V;\delta)) = 0$ whenever $V = 0$, so we can restrict the proof to the events where $V \neq 0$. Denote for brevity $\Phi(V;\delta)$ by Φ, the $\mathrm{sign}(\cdot)$ function by $\mathrm{s}(\cdot)$, and the indicator function of an event $\{E\}$ by $\mathbb{1}\{E\}$. Then,

$$\begin{aligned}
&\mathbb{E}_{V,\delta}\left[\mathrm{s}(\Phi)|\sigma(|\Phi|)\right] = \mathbb{E}_{V,\delta}\left[\mathrm{s}(\Phi)|\sigma(|V|)\right] \\
&= \mathbb{E}_{V,\delta}\left[\mathbb{1}\{\mathrm{s}(\Phi) = \mathrm{s}(V)\}\mathrm{s}(\Phi)|\sigma(|V|)\right] \\
&\quad + \mathbb{E}_{V,\delta}\left[\mathbb{1}\{\mathrm{s}(\Phi) \neq \mathrm{s}(V)\}\mathrm{s}(\Phi)|\sigma(|V|)\right] \\
&= \mathbb{E}_{V,\delta}\left[\mathbb{1}\{\mathrm{s}(\Phi) = \mathrm{s}(V)\}\mathrm{s}(V)|\sigma(|V|)\right] \\
&\quad - \mathbb{E}_{V,\delta}\left[\mathbb{1}\{\mathrm{s}(\Phi) \neq \mathrm{s}(V)\}\mathrm{s}(V)|\sigma(|V|)\right] \\
&\quad \left[\text{ by definition of } \Phi\,\right] \\
&= \left(\mathbb{E}_{V,\delta}\left[\mathbb{1}\{\mathrm{s}(V) = \mathrm{s}(\eta_V)\}\mathbb{1}\{U > \lambda(|V|)\}\mathrm{s}(V)|\sigma(|V|)\right]\right. \\
&\quad + \mathbb{E}_{V,\delta}\left[\mathbb{1}\{\mathrm{s}(V) \neq \mathrm{s}(\eta_V)\}\mathrm{s}(V)|\sigma(|V|)\right]) \\
&\quad - \mathbb{E}_{V,\delta}\left[\mathbb{1}\{\mathrm{s}(V) = \mathrm{s}(\eta_V)\}\mathbb{1}\{U \leq \lambda(|V|)\}\mathrm{s}(V)|\sigma(|V|)\right] \\
&= \mathbb{E}_V\left[\mathbb{1}\{\mathrm{s}(V) = \mathrm{s}(\eta_V)\}|\sigma(|V|)\right](1 - \lambda(|V|))\mathrm{s}(\eta_V) \\
&\quad + \mathbb{E}_V\left[\mathbb{1}\{\mathrm{s}(V) \neq \mathrm{s}(\eta_V)\}|\sigma(|V|)\right](-\mathrm{s}(\eta_V)) \\
&\quad - \mathbb{E}_V\left[\mathbb{1}\{\mathrm{s}(V) = \mathrm{s}(\eta_V)\}|\sigma(|V|)\right]\lambda(|V|)\mathrm{s}(\eta_V) \\
&= 0,
\end{aligned}$$

where the last equality follows by definition of $\lambda(|V|)$ and by noting that $\mathbb{E}_V\left[\mathbb{1}\{\mathrm{s}(V) = \mathrm{s}(\eta_V)\}|\sigma(|V|)\right] = \frac{1}{2}(1+|\eta_V|)$ and $\mathbb{E}_V\left[\mathbb{1}\{\mathrm{s}(V) \neq \mathrm{s}(\eta_V)\}|\sigma(|V|)\right] = \frac{1}{2}(1-|\eta_V|)$. ■

The operator $\Phi(\,\cdot\,;\delta)$ will be used in the following in order to symmetrise the noise sample $\{N_t\}$.

The oracle-based SPS algorithm is obtained by modifying the original SPS algorithm [10, Tables I and II] as follows:

- A new step is added to the *initialisation* [10, Table I].

> 5. Generate a random sequence of numbers $\{\delta_1,\ldots,\delta_n\}$, where $\delta_t, t = 1,\ldots,n$, are independently and uniformly drawn from $[0,1]$

- The $\{S_i(\theta)\}$ functions in the pseudo-code of the SPS *indicator* [10, Table II] are re-defined as

> 2. $S'_0(\theta) \triangleq R_n^{-\frac{1}{2}} \frac{1}{n} \sum_{t=1}^{n} \varphi_t \Phi(\varepsilon_t(\theta), \delta_t)$,
> $S'_i(\theta) \triangleq R_n^{-\frac{1}{2}} \frac{1}{n} \sum_{t=1}^{n} \alpha_{i,t}\, \varphi_t \Phi(\varepsilon_t(\theta), \delta_t)$,
> for all $i \in \{1,\ldots,m-1\}$.

  Recall that $\varepsilon_t(\theta) \triangleq Y_t - \varphi_t^{\mathrm{T}}\theta$, $R_n \triangleq \frac{1}{n}\sum_{t=1}^{n} \varphi_t\varphi_t^{\mathrm{T}}$, and R_n ("shaping matrix") is assumed to be *invertible* [10].

Giving the name of "Oracle-SPS-Indicator(θ)" to this version of "SPS-Indicator(θ)", the corresponding oracle-based confidence region is defined as follows

$$\widehat{\Theta}_{\text{oracle}} \triangleq \left\{\, \theta \in \mathbb{R}^d : \text{Oracle-SPS-Indicator}(\theta) = 1 \,\right\}.$$

Theorem 1 (Exact Confidence for Arbitrary Marginals): If $N_1,\ldots,N_n$ are independent, but may be non-symmetric,

$$\mathbb{P}\{\, \theta^* \in \widehat{\Theta}_{\text{oracle}} \,\} = 1 - \frac{q}{m}.$$

Proof: Note that $\Phi(\varepsilon_t(\theta^*), \delta_t) = \Phi(N_t, \delta_t)$, and $\{\Phi(N_t,\delta_t)\}$ is a sequence of independent, symmetric random variables (Lemma 1). Then, the proof is as the proof of Theorem 1, [10], where $N_1,\ldots,N_n$ (which now need not be a sequence of symmetric variables) is replaced by $\Phi(N_1,\delta_1),\ldots,\Phi(N_n,\delta_n)$, which satisfies the required independence and symmetry conditions of [10]. ■

B. Robustifying SPS

Since the oracle-based algorithm must compute $\Phi(\,\cdot\,;\delta_t)$, it can be implemented as-it-is only when knowledge on η_{N_t}, the asymmetric deviation of the noise, is available.

However, if a constant $\bar{\eta} \in [0,1)$ that bounds the absolute asymmetric deviation, i.e., such that $|\eta_{N_t}| \leq \bar{\eta}$ with probability one for all t, is available, then the oracle-based algorithm can be mimicked to build guaranteed confidence regions $\widehat{\Theta}_{\bar{\eta}}$ that satisfy $\mathbb{P}\{\,\theta \in \widehat{\Theta}_{\bar{\eta}}\,\} \geq 1 - \frac{q}{m}$, and such that, if $\bar{\eta}$ is small, then $\widehat{\Theta}_{\bar{\eta}} \approx \widehat{\Theta}_{\text{oracle}}$ or even $\widehat{\Theta}_{\bar{\eta}} = \widehat{\Theta}_{\text{oracle}}$ with high probability. In what follows, we assume that *only the bound $\bar{\eta}$ is known*, and we describe how to modify the oracle-based algorithm to construct an outer-approximation of its regions and thus to get a robustified variant of SPS.

Given the random sequence $\delta_1,\ldots,\delta_n$, define the index set $\mathcal{K} = \{t \in \{1,\ldots,n\} : \delta_t > \frac{\bar{\eta}}{1+\bar{\eta}}\}$, and let $\mathcal{U}$ be its complementary with respect to $\{1,\ldots,n\}$. Then, we know that, in the oracle-based algorithm, $S'_0(\theta)$ is equal to

$$S'_0(\theta) = R_n^{-\frac{1}{2}} \frac{1}{n} \sum_{t\in\mathcal{K}}^{n} \varphi_t\varepsilon_t(\theta) + R_n^{-\frac{1}{2}} \frac{1}{n} \sum_{t\in\mathcal{U}}^{n} \varphi_t \Phi(\varepsilon_t(\theta), \delta_t),$$

where the first sum is known, while the second one is unknown due to the fact that the exact values taken by η_{N_t},

$t \in \mathcal{U}$, are unknown. Thus, the number of unknown sign-changes is bounded by $|\mathcal{U}|$, the size of $\mathcal{U}$, and we know that, for a given θ, $S'_0(\theta)$ and $S'_i(\theta)$ are equal to

$$\tilde{S}_0(\theta) = R_n^{-\frac{1}{2}} \frac{1}{n} \sum_{t\in\mathcal{K}}^{n} \varphi_t \varepsilon_t(\theta) + R_n^{-\frac{1}{2}} \frac{1}{n} \sum_{t\in\mathcal{U}}^{n} \beta_t \varphi_t \varepsilon_t(\theta)$$

and

$$\tilde{S}_i(\theta) = R_n^{-\frac{1}{2}} \frac{1}{n} \sum_{t\in\mathcal{K}}^{n} \alpha_{i,t} \varphi_t \varepsilon_t(\theta) + R_n^{-\frac{1}{2}} \frac{1}{n} \sum_{t\in\mathcal{U}}^{n} \beta_t \alpha_{i,t} \varphi_t \varepsilon_t(\theta),$$

for one of the $2^{|\mathcal{U}|}$ possible assignments of $\beta_t \in \{-1, 1\}$, $t \in \mathcal{U}$. Note that the correct sign assignment depends on θ. The indicator function of the oracle-based region is based on the *rank* $R(\theta)$ of $\|S'_0(\theta)\|$ in the ordering of $\|S'_0(\theta)\|, \|S'_i(\theta)\|$, $i = 1, \ldots, m-1$, e.g., $R(\theta) = 1$ means that $\|S'_0(\theta)\|$ is the smallest one, and so on. We will use the notation

$$R(\theta) \triangleq \text{rank}\left(\|S'_0(\theta)\| : \|S'_0(\theta)\|, \ldots, \|S'_{m-1}(\theta)\| \right),$$

and the oracle-based region is defined as the set of θ parameters for which $R(\theta) \leq m - q$. Now, define

$$\bar{R}(\theta) \triangleq \min_{\beta_t\in\{-1,1\},\, t\in\mathcal{U}} \text{rank}\left(\|\tilde{S}_0(\theta)\| : \{\|\tilde{S}_i(\theta)\|\}_{i=0}^{m-1} \right),$$

and

$$\widehat{\Theta}_{\bar{\eta}} \triangleq \{\, \theta : \bar{R}(\theta) \leq m - q \,\}.$$

Clearly, $\widehat{\Theta}_{\text{oracle}} \subseteq \widehat{\Theta}_{\bar{\eta}}$, and an outer-approximation of the oracle region is so obtained. Hence, $\mathbb{P}\{\, \theta^* \in \widehat{\Theta}_{\bar{\eta}} \,\} \geq 1 - \frac{q}{m}$.

C. Alternative Robustness Bound for SPS

Note that when the index set of unknown sign-changes, $\mathcal{U}$, defined in the previous section, is empty, the robustified algorithm builds the same region as the oracle-based and the standard SPS algorithms. This happens with high probability when $\bar{\eta} \ll \frac{1}{n}$, in fact, defining $\bar{\lambda} = \frac{\bar{\eta}}{1+\bar{\eta}}$, it holds that

$$\mathbb{P}\{\, \widehat{\Theta} = \widehat{\Theta}_{\bar{\eta}} \,\} \;\geq\; \mathbb{P}\{\, \mathcal{U} = \emptyset \,\} \;=\; (1 - \bar{\lambda})^n.$$

This leads immediately to another lower-bound, alternative to (3), on the robustness of the standard SPS algorithm.

Theorem 2: If for every t, the asymmetric deviation of the noise satisfies $|\eta_{N_t}| \leq \bar{\eta}$ (a.s.), the standard SPS algorithm with parameters q and m builds a region $\widehat{\Theta}$ with

$$\mathbb{P}\{\, \theta^* \in \widehat{\Theta} \,\} \;\geq\; \frac{1}{(1+\bar{\eta})^n} - \frac{q}{m}\,.$$

Proof: It holds that $\mathbb{P}\{\, \theta^* \notin \widehat{\Theta} \,\} = \mathbb{P}\{\, \theta^* \notin \widehat{\Theta}_{\text{oracle}} \text{ and } \widehat{\Theta}_{\text{oracle}} = \widehat{\Theta} \,\} + \mathbb{P}\{\, \theta^* \notin \widehat{\Theta} \text{ and } \widehat{\Theta}_{\text{oracle}} \neq \widehat{\Theta} \} \leq \mathbb{P}\{\, \theta^* \notin \widehat{\Theta}_{\text{oracle}} \,\} + \mathbb{P}\{\, \widehat{\Theta}_{\text{oracle}} \neq \widehat{\Theta} \,\}$. Taking the complement and using Theorem 1, we get $\mathbb{P}\{\theta^* \in \widehat{\Theta}\} \geq 1 - \frac{q}{m} - \mathbb{P}\{\widehat{\Theta}_{\text{oracle}} \neq \widehat{\Theta}\} = \mathbb{P}\{\widehat{\Theta}_{\text{oracle}} = \widehat{\Theta}\} - \frac{q}{m}$, and the statement follows by noting that the event $\{\, \mathcal{U} = \emptyset \,\}$ has probability $(1 - \bar{\lambda})^n$, where $\bar{\lambda} = \bar{\eta}/(1 + \bar{\eta})$ is an upper bound on the probabilities that the oracle makes sign-changes; and $\{\, \mathcal{U} = \emptyset \,\}$ implies $\{\, \widehat{\Theta}_{\text{oracle}} = \widehat{\Theta} \,\}$. ∎

In practice, we expect that the lower-bound given by (3) and Theorem 2 are pessimistic. In fact, the robust region, which is guaranteed with probability *at least* $1 - \frac{q}{m}$, and the classic SPS region are similar if $|\mathcal{U}|$ is small (and we expect that their performance is similar, as well, as is demonstrated by Monte Carlo trials in the experimental results part).

The distribution of $|\mathcal{U}|$ is known a-priori and it is a binomial distribution, $\mathbb{P}\{\, |\mathcal{U}| \leq \ell \,\} = \sum_{i=0}^{\ell} \binom{n}{i} \bar{\lambda}^i (1-\bar{\lambda})^{n-i}$.

III. LAD-SPS

SPS builds regions that are guaranteed to contain the least-squares (LS) estimate. In this section we consider a variant of the SPS algorithm that builds regions that are guaranteed to contain the *least-absolute-deviation* (LAD) estimate

$$\hat{\theta}_{\text{LAD}} \triangleq \arg\min_{\theta} \sum_{t=1}^{n} |\, y_t - \varphi_t^{\mathrm{T}} \theta \,|\,.$$

The LAD estimate is more robust than the LS estimate (e.g., against outliers), but harder to compute and can have several solutions. For more information on the LAD estimator, in particular for a proof of its asymptotic normality, see [17].

We call our algorithm LAD-SPS, and define it by replacing the S_0 and $\{S_i\}$ functions in the SPS algorithm with

$$Z_0(\theta) \;\triangleq\; R_n^{-\frac{1}{2}} \frac{1}{n} \sum_{t=1}^{n} \varphi_t \,\text{sign}(\varepsilon_t(\theta)), \tag{4}$$

$$Z_i(\theta) \;\triangleq\; R_n^{-\frac{1}{2}} \frac{1}{n} \sum_{t=1}^{n} \alpha_{i,t}\, \varphi_t \,\text{sign}(\varepsilon_t(\theta)), \tag{5}$$

for $i = 1, \ldots, m-1$, all the rest remains the same.

Note that we even use the same "shaping matrix" $R_n^{-\frac{1}{2}}$, since the (scaled) errors of LAD estimates are also asymptotically normal [17] with covariance matrix (up to a constant) R^{-1}, where $R = \lim_{n\to\infty} R_n$, if it exists and is invertible.

Since $\frac{1}{n}\sum_{t=1}^{n} \varphi_t \,\text{sign}(\varepsilon_t(\theta))$ is the (sub)gradient of the mean-of-the-absolute-deviation error $\frac{1}{n}\sum_{t=1}^{n} |y_t - \varphi_t^{\mathrm{T}}\theta|$, [12], it holds that $Z_0(\hat{\theta}_{\text{LAD}}) = 0$. Hence, $\|Z_0(\hat{\theta}_{\text{LAD}})\| = 0$ cannot be larger than $\{\|Z_i(\hat{\theta}_{\text{LAD}})\|\}$, and $\hat{\theta}_{\text{LAD}}$ will be included in the LAD-SPS confidence region, which is built by evaluating the ranking of $\|Z_0\|$ among $\{\|Z_i\|\}$. It is easy to see that LAD-SPS builds a confidence region, which we denote by $\widehat{\Theta}_{\text{LAD}}$, such that $\mathbb{P}\{\, \theta^* \in \widehat{\Theta}_{\text{LAD}} \,\} = 1 - \frac{q}{m}$ under the standard SPS assumptions (following the exact confidence proof of standard SPS [10]). However, with LAD-SPS, the assumptions on the noise distribution can be relaxed significantly. In $Z_0(\theta^*)$ and $\{Z_i(\theta^*)\}$, in fact, the modified sequence $\text{sign}(N_1), \ldots, \text{sign}(N_n)$ is used in place of $N_1, \ldots, N_n$. The modified sequence can be symmetric and i.i.d. also when $N_1, \ldots, N_n$ is not. In particular, for $\text{sign}(N_1), \ldots, \text{sign}(N_n)$ to be i.i.d. and symmetric, so that the exact confidence result holds true, it is enough that $\{N_t\}$ is a *mediangale*, precisely, that the process satisfies

$$\mathbb{E}\left[\, \text{sign}(N_t) \mid \mathcal{F}_t \,\right] \;=\; 0 \;\; \text{(a.s.)}$$

for all t, where $\mathcal{F}_t$ is the σ-algebra generated by the past of the process (the noise sequence up to and including $t-1$ and the randomised sign sequences), see [6] for more details.

Remark 1 (On the Signal-to-Noise Ratio): It is important to note that even if the $\mathrm{sign}(\cdot)$ function does not depend on the magnitude of its argument, LAD-SPS does *not* discard the information on the magnitude of the noise, and exploits this information in the construction of the confidence region.

In order to see this, note that the term $\mathrm{sign}(\varepsilon_t(\theta)) = \mathrm{sign}(N_t - \varphi_t^{\mathrm{T}}(\theta^* - \theta))$ actually *depends* on how large the noise, N_t, is with respect to the size of the error in the tested parameter $\|\theta^* - \theta\|$.

Remark 2 (LAD-SPS and Sign-Tests in Econometrics): The LAD variant of SPS connects finite sample methods in the system identification literature (SPS [10], LSCR [4], dataset-perturbed methods [13]) and an independent, rich thread in econometrics [1], [6] which deserves the attention of the system identification community. In the terminology of econometrics, LAD-SPS is a confidence set construction method that relies on Monte Carlo sign-based joint tests, [6].

A. Robustifying LAD-SPS

In the previous section we have seen that LAD-SPS is more robust than standard SPS to asymmetries in the noise distribution, since it only requires the zero-median property. If the zero-median assumption is not exactly satisfied, the LAD-SPS can be easily robustified to match the required confidence level in line with the analysis in Sections II-A and II-B. To this purpose, define $\tilde{N}_t = \mathrm{sign}(N_t)$ and let the sequence $\tilde{N}_1, \dots \tilde{N}_n$ play the role of $N_1, \dots, N_n$ in the oracle-based algorithm of Section II-A: the resulting algorithm is a LAD-SPS oracle-based algorithm which builds exact confidence regions. Significantly, for nonzero noise, the asymmetric deviation of $\tilde{N}_t$, $\eta_{\tilde{N}_t}$, is a constant, i.e., it does not depend on $|N_t|$ anymore. Thus, in order to build an exact confidence region, we only need information on the (unconditional) probability that $\mathrm{sign}(N_t)$ is positive, which is a much simpler piece of information than in the standard oracle-based SPS case. As soon as a bound on this probability is available, a robustified version of LAD-SPS can be immediately implemented in line with the discussion in Section II-B.

IV. Numerical Experiments

We consider a second order data generating FIR system

$$Y_t \;=\; b_1^*\, U_{t-1} + b_2^*\, U_{t-2} + N_t^{(\gamma)},$$

where $b_1^* = 0.7$ and $b_2^* = 0.3$ are the true system parameters, and $N_t^{(\gamma)}$ is a sequence of i.i.d. random variables with density

$$f(x) = \begin{cases} (1-\gamma)\,\frac{1}{\sqrt{2}\bar{\sigma}}\exp\left(-\frac{\sqrt{2}|x|}{\bar{\sigma}}\right) & \text{if } x \le 0, \\ (1+\gamma)\,\frac{1}{\sqrt{2}\bar{\sigma}}\exp\left(-\frac{\sqrt{2}|x|}{\bar{\sigma}}\right) & \text{if } x > 0, \end{cases}$$

where $\bar{\sigma} = 0.1$. A simple computation shows that, in this case, the asymmetric deviation $\eta_{N_t^{(\gamma)}}$ of $N_t^{(\gamma)}$ is constant and is equal to γ. Note that, when $\gamma = 0$, $N_t^{(\gamma)}$ has a (symmetric) Laplacian distribution with standard deviation $\bar{\sigma} = 0.1$.

The input signal is given by

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

where $\{V_t\}$ is a sequence of i.i.d. Gaussian random variables with zero mean and variance 1 (i.e., standard normal).

We set $1 - \frac{q}{m} = 90\ \%$, and several Monte Carlo tests were performed for different values of γ: $n = 50$ output samples $Y_1, \dots, Y_{50}$ were generated $2,000,000$ times with $U_0 = U_{-1} = 0$ as initial conditions, and we evaluated the coverage of (b_1^*, b_2^*) by various SPS algorithms.

First, we tested the inclusion of the true parameter (b_1^*, b_2^*) in the regions built by the standard SPS algorithm with various asymmetric deviations. Results are reported in the first line of Table I. Then, we considered the case where $\gamma = 10\%$. In this case, standard SPS, which is *not* guaranteed because the noise is not symmetric, scored 89.7%. On the other hand, if $\bar{\eta}$ in Robust-SPS is set as $\bar{\eta} = \gamma = 10\%$, Robust-SPS is guaranteed to include the true parameter with a probability *at least* of 90%. In our test, Robust-SPS scored 96.1%. An instance of the regions built by the standard SPS and Robust-SPS is shown in Fig.1. In the picture, also the regions built by LAD-SPS (which scored 89.5%), its robustified version (which scored 97.2%), as well as the ellipsoid built according to the asymptotic theory [10, Section III] (which scored 88.6%) are included. Note that Robust-SPS and Robust-LAD-SPS build outer approximations of the SPS and LAD-SPS regions, respectively. Also, the SPS and LAD-SPS regions, and consequently their robustified variants, contain the LS and LAD estimates, respectively. The volume of the SPS and LAD-SPS regions are comparable to that of the asymptotic ellipsoid. Furthermore, the size of the robustified versions are only moderately larger than the size of their standard versions indicating a reasonable trade-off between robustness and volume. For the robust regions in the picture, $|\mathcal{U}|$ turned out to be equal to 5, which is a typical outcome in the present setting since, from its binomial distribution, $\mathbb{P}\{\,|\mathcal{U}| \le 5\,\} \approx 70\ \%$.

The correct value of $\bar{\eta}$ to be used in Robust-SPS is rarely known in applications, so we evaluated what happens under misspecification. In Table I, we reported the empirical coverage of Robust-SPS when we used $\bar{\eta} = 10\%$ while the *true* value of the asymmetric deviation γ varies from 0% to 60%. For comparison, also the performance of LAD-SPS and its robustified version (always with $\bar{\eta} = 10\%$) are reported, and they show a good degree of robustness.

The robustness properties of LAD-SPS can be better appreciated on another set of simulations, where each noise

Asymmetric Deviation (γ)	**0 %**	**10 %**	**40 %**	**60 %**
SPS	90.0 %	89.7 %	85.6%	80.2 %
Robust SPS $\bar{\eta} = 10\%$	96.2 %	96.1 %	94.1 %	90.9 %
LAD-SPS	90.0 %	89.5 %	82.2 %	73.4 %
Robust LAD-SPS $\bar{\eta} = 10\%$	97.4 %	97.2 %	94.2 %	88.9 %
Asymptotic Ellipsoid	88.9 %	88.6 %	83.9 %	77.9 %

TABLE I

Empirical Coverages w.r.t. various asymmetric deviations.

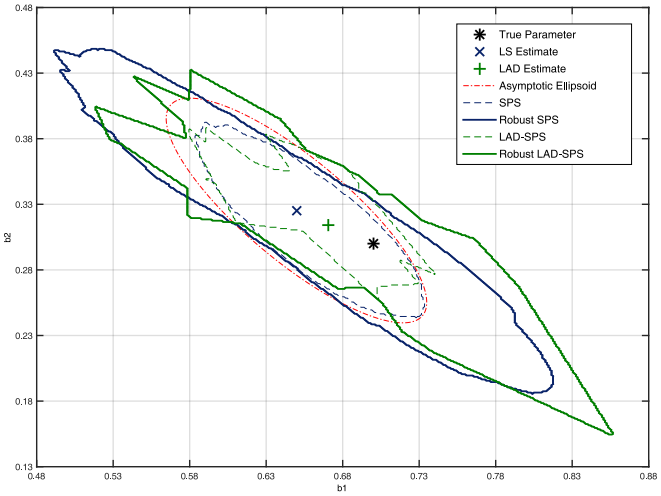

Fig. 1. 90% confidence regions ($n = 50, m = 100$) in case of asymmetric noises (perturbed Laplacian) with asymmetric deviation 0.1 (10 %).

sample $N_t^{(\zeta)}$ was independently generated as $\exp(G_\zeta) - 1$, where G_ζ is a Gaussian random variable with zero mean and ζ standard deviation. In this case, $N_t^{(\zeta)}$ is not symmetric, e.g., its expected value is $\exp(\frac{\zeta^2}{2}) - 1$, so that standard SPS is *not* guaranteed. However, since the median of $N_t^{(\zeta)}$ is zero for all ζ, the theory guarantees that LAD-SPS always delivers an exact confidence region. In Table II, results for various ζ values and $\bar{\eta}$ always set to 10% are reported.

V. Conclusions

In this paper we have shown that the Sing-Perturbed Sums (SPS) method is robust to small asymmetries in the noise. We have also proposed a robustification technique that builds on an oracle-based algorithm. The shape of the region built by the robust SPS algorithm degrades when higher levels of asymmetries are accepted, while the confidence remains guaranteed and user-chosen. We have also considered LAD-SPS, a variant of the SPS approach, which builds exact, non-asymptotic confidence regions around the *least-absolute-deviation* (LAD) estimate and which works under milder assumptions on the noise distribution. Particularly, the shape of the noise distribution is irrelevant for LAD-SPS, as only a mediangale property is crucial, i.e., the conditional *median* of the noise must be zero. We also showed how to apply the newly developed robustification technique to LAD-SPS, in order to hedge against asymmetries in the median. Finally, the presented robustness analysis and robustification techniques were supported by numerical experiments with different kinds and degrees of noise asymmetries.

Scale Parameter (ζ)	**0.1**	**0.5**	**1**	**5**
SPS	89.9 %	88.3 %	85.3 %	84.4 %
Robust SPS $\bar{\eta} = 10\%$	96.5 %	95.5 %	93.3 %	89.9 %
LAD-SPS	90.0 %	90.0 %	90.0 %	90.0 %
Robust LAD-SPS $\bar{\eta} = 10\%$	97.4 %	97.4 %	97.4 %	97.4 %
Asymptotic Ellipsoid	88.8 %	86.6 %	83.7 %	87.2 %

TABLE II

Empirical Coverages for non-symmetric mediangale noises.

References

[1] Michail V. Boldin, Galina I. Simonova, and Yuri Nikolaevich Tyurin. *Sign-based methods in linear statistical models*, volume 162. American Mathematical SoC., 1997.

[2] Marco C. Campi, Balázs Cs. Csáji, Simone Garatti, and Erik Weyer. Certified system identification: towards distribution-free results. In *Procs. of the 16th IFAC Symposium on System Identification*, pages 245–255, 2012.

[3] Marco C. Campi, S. Ko, and E. Weyer. Non-asymptotic confidence regions for model parameters in the presence of unmodelled dynamics. *Automatica*, 45:2175–2186, 2009.

[4] Marco C. Campi and Erik Weyer. Guaranteed non-asymptotic confidence regions in system identification. *Automatica*, 41:1751–1764, 2005.

[5] Marco C. Campi and Erik Weyer. Non-asymptotic confidence sets for the parameters of linear transfer functions. *IEEE Transactions on Automatic Control*, 55:2708–2720, 2010.

[6] Elise Coudin and Jean-Marie Dufour. Finite-sample distribution-free inference in linear median regressions under heteroscedasticity and non-linear dependence of unknown form. *The Econometrics Journal*, 12(s1):S19–S49, 2009.

[7] Balázs Cs. Csáji, Marco C. Campi, and Erik Weyer. Strong consistency of the Sign-Perturbed Sums method. In *Procs. of the 53th IEEE Conference on Decision and Control (CDC)*, pages 3352–3357, 2014.

[8] Balázs Cs. Csáji, Marco C. Campi, and Erik Weyer. A method for constructing exact finite-sample confidence regions for general linear systems. In *Procs. of the 51st IEEE Conference on Decision and Control*, pages 7321–7326, 2012.

[9] Balázs Cs. Csáji, Marco C. Campi, and Erik Weyer. Non-asymptotic confidence regions for the least-squares estimate. In *Procs. of the 16th IFAC Symposium on System Identification*, pages 227–232, 2012.

[10] 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. *IEEE Transactions on Signal Processing*, 63(1):169–181, 2015.

[11] Alison L. Gibbs and Francis Edward Su. On choosing and bounding probability metrics. *Intern. Statistical Review*, 70(3):419–435, 2002.

[12] Roger Koenker and Gilbert Bassett. Tests of linear hypotheses and l1 estimation. *Econometrica: Journal of the Econometric Society*, pages 1577–1583, 1982.

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

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

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

[16] Mario Milanese, John Norton, Hélène Piet-Lahanier, and Éric Walter. *Bounding approaches to system identification*. Springer Science & Business Media, 2013.

[17] David Pollard. Asymptotics for least absolute deviation regression estimators. *Econometric Theory*, 7(02):186–199, 1991.

[18] Balázs Cs. Csáji and E. Weyer. Closed-loop applicability of the Sign-Perturbed Sums method. In *Procs. of the 54th IEEE Conference on Decision and Control, Osaka, Japan*, pages 1441–1446, 2015.

[19] Helmut Rieder. *Robust asymptotic statistics*, volume 1. Springer Science & Business Media, 2012.

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

[21] Albert N. Shiryaev. *Probability*. Springer, 2nd edition, 1995.

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

[23] 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. In *Proceedings of the 54th IEEE Conference on Decision and Control (CDC)*, 2015.

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