2014 American Control Conference (ACC)
June 4-6, 2014. Portland, Oregon, USA

Exact Confidence Regions for Linear Regression Parameter under External Arbitrary Noise[1]

Alexander Senov, Konstantin Amelin, Natalia Amelina, and Oleg Granichin[2]

***Abstract*— The paper propose new method for identifying non-asymptotic confidence regions for linear regression parameter under external arbitrary noise. This method called Modified Sign-Perturbed Sums (MSPS) method and it is a modification of previously proposed one, called Sign-Perturbed Sums which is applicable only in case of symmetrical centred noise. MSPS algorithm correctness and obtained confidence region convergence are proved theoretically under some additional assumptions. SPS and MSPS methods are compared basing on simulated data. Few advantages of MSPS method in case of biased and asymmetric noise are illustrated.**

I. Introduction

The system identification consists of two major problems: obtaining *mathematical model* of dynamic system and estimation of obtained model parameters or their boundaries from noised measurements [1]. Noise may appear from absolutely different sources. It could be result of difference between complex real world systems and chosen mathematical model which often induce many implicit assumptions and uncertainties. It also could result from the outside of the model, for example, intentionally injected by the opponent. Such types of noise produced by sources not within the model and independent from system inputs we will call *external* noise. It is important to note that external noise is not only measurement model error — it could have any kind of distribution or, moreover, be deterministic. However, it is often rather difficult to determine real nature and true properties of corresponding noise. Therefore, developing methods and procedures for assessment of a model quality is another central issue in system identification [1], [2]. The model parameter's estimation could be done in two ways. The first one is to estimate exact values of the parameter's. The second one is to construct so called *confidence region* which contains true parameter's values with specified probability. We focus on the confidence region construction task.

In this paper we consider only linear regression model. Despite the fact that more complex models could better fit complex systems and hence reduce the uncertainty, such models could also overfit real system. Many types of models could be reduced to linear regression model via different techniques, such as linearisation or redesignation. As an example, auto-regression with exogenous input, finite impulse-response models, and many others models could be reduced to multivariate linear regression. Hence, it is natural to develop almost arbitrary noise-robust confidence region construction method for linear regression model. We formulate linear regression model as following

$$y_i = \phi_i^T \theta^\star + \varepsilon_i, \quad i = 1, \ldots, N, \tag{1}$$

where N is a number of observations, $\phi_i \in \mathbb{R}^d$ is a system input, $y_i \in \mathbb{R}$ is a system output, $\varepsilon_i \in \mathbb{R}$ is an unknown noise and $\theta^\star \in \mathbb{R}^d$ is an unknown system parameter.

In paper we aimed at the second task: for given probability $\alpha \in [0,1]$ construct exact α-confidence region, which we defined as

***Definition** 1:* For given $\alpha \in [0,1]$ Θ_α is *exact α-confidence region* for parameter $\theta^\star$ if

$$P(\theta^\star \in \Theta_\alpha) = \alpha.$$

There are some classical methods for confidence region construction in linear regression model which works well under assumption of i.i.d. Gaussian noises ε_i (see, for example [1]). *Asymptotic-theory* for system identification (see, e.g., [1], [3]) is the standard approach for confidence region construction, if an exact noise distribution is unknown,. The asymptotic-theory have been successfully applied in many practical tasks, but its results are precise only if N tends to infinity. In case of finite and especially small number of observations it could cause erroneous results. However, until recent decades, there were almost no methods for constructing exact (non-asymptotic) confidence regions for linear regression in case of external arbitrary noise.

Surprising, introducing additional randomnicity into considering model could not only increase quality of parameter estimation (for example, see [3]), but also made parameter's exact confidence region construction possible even under almost arbitrary noise. Such techniques, which consist of implantation additional randomnicity into system called *randomization* and successfully applied in variety of areas (for examples, see [4], [5], [6], [7], [8]). Interesting randomized algorithm for linear regression parameter estimation problem under almost arbitrary noise proposed in [9].

Generally, one approach called *Set Membership Identification* could be successfully used under assumptions that noise is arbitrary except its boundaries are known apriori. The provided guaranteed region consist of such values θ which do not violate this apriori boundaries [10], [11], [12]. Besides, prior knowledge of noise boundaries is quite a strong requirement, which is not always true.

[1]This work was supported by RFBR (projects 13-07-00250, 14-08-01015)
[2]Alexander Senov, Konstantin Amelin, Natalia Amelina, and Oleg Granichin are with Department of Mathematics and Mechanics, Saint Petersburg State University, 198504, Universitetskii pr. 28, St. Petersburg, Russia. Oleg Granichin is also with Laboratory for Analysis and Modeling of Social Processes, Saint Petersburg State University alexander.senov@gmail.com, konstantinamelin@gmail.com, ngranichina@gmail.com, oleg_granichin@mail.ru

A significant breakthrough was done in recent work [13], in which authors M.C. Campi, E. Weyer and B.C. Csaji introduced method called *Sign-Perturbed Sums* (SPS) for construction exact confidence region for linear model parameter in case of noise symmetrically distributed around zero. This method is tied to the concept of *sign-perturbed sum*

$$S_k(\theta) = \sum_{i=1}^{N} a_{k,i}\phi_i(y_i - \phi_i^T\theta), k = 1,\dots,M-1, \qquad (2)$$

$$a_{k,i} = \begin{cases} 1, & \text{with probability } \frac{1}{2} \\ -1, & \text{with probability } \frac{1}{2} \end{cases}$$

where k is just some index and $a_{k,i}$ called *random signs*. This method is based on the idea similar to Leave-out Sign-dominant Correlation Regions (LSCR) method. The idea is that, under several assumptions, sums $S_k(\theta)$ and sum $S_0(\theta) = \sum_{i=1}^{N}\phi_i(y_i - \phi_i^T\theta)$ distributed equally if $\theta = \theta^\star$. Briefly, SPS use symmetry of noise distribution to provide equality of all sign-perturbed sums distributions to use it in further confidence regions construction.

Nevertheless, SPS method have several significant limitations. First, due to the lack of solid analysis, there are no theoretical results so far, that states some $\Theta_{M,q}$ properties, such as statistical mean and variation of its borders and asymptotically $N \to \infty$ behaviour. Second, because of assumptions done in SPS, we could not always assume that noise is symmetrical and unbiased. In many practical cases noise bias is too significant to neglect it [1]. Moreover, noise asymmetry could play big role especially in small data sets [13].

This papers focuses on the second limitation. In this paper we introduce modification of SPS algorithm called *Modified Sign-Perturbed Sums* (MSPS) which use symmetry of system inputs x_i instead of ε_i to provide equality of all MSPS analogues of sign-perturbed sums distribution. So, it pushing the symmetrical distributed role from the noises to system inputs. MSPS is aimed on use in situations when SPS is inapplicable due to asymmetry of noise distribution but symmetry of system inputs distribution around known mean.

This paper organized as follows. At first, we formulate problem statement as confidence region constructing problem under external arbitrary noise in the first section. Then we MSPS algorithm which provide exact confidence regions with user-chosen probability under external arbitrary noise in Section III. We also provide some theoretical results about algorithm correctness and properties of obtained confidence region, collaterally show more effective way for confidence region calculation in one-dimensional case in Section III-A. Finally, in Section IV we present comparative examples on simulated data in section.

II. Problem Statement

The overall problem statement is as follows. In terms of model (1) and under following assumptions

1) $\{\varepsilon_i\}_{i=1}^N$ — an unknown arbitrary noises independent from model inputs $\{\phi\}_{i=1}^N$;
2) $\{\phi\}_{i=1}^N$ — *observable* but *uncontrollable* i.i.d. inputs, symmetrically distributed about known mean $m_\phi = E\phi_i$ $\forall i = 1,\dots,N$

for given probability $\alpha \in [0,1]$ we need to construct exact α-confidence region Θ_α for parameter $\theta^\star$, such as

$$P(\theta^\star \in \Theta_\alpha) = \alpha.$$

One important feature of these assumptions is that the restrictions imposed on the noise distribution are extremely weak. This makes most of known confidence region construction techniques (at least all mentioned in this paper) inapplicable.

III. Modified Sign-Perturbed Sums

Hereinafter we presume that Assumptions II satisfied. In order to describe a Modified Sign-Perturbed Sums algorithm we should introduce *modified sign-perturbed sum* notation.

$$S_k(\theta) = \sum_{i=1}^{N} a_{k,i}\Delta_i(y_i - \phi_i^T\theta) = \qquad (3)$$
$$= \sum_{i=1}^{N} a_{k,i}(\phi_i - m_\phi)(y_i - \phi_i^T\theta),$$

where m_ϕ is a known ϕ_i mean, and we denote $\Delta_i = \phi_i - m_\phi$. This sum is similar to sign-perturbed sum (2) introduced in [13] with only difference in added m_ϕ. The key idea of the subtraction is that $(\phi_i - m_\phi)$ multiplier plays the role which ε_i plays in original SPS method. Since ε_i no longer symmetric about zero this property "transferred" to $(\phi_i - m_\phi)$.

A. MSPS algorithm

The MSPS exact confidence region construction algorithm is quite similar to SPS algorithm and consist of computation of all sums $S_i(\theta)$ for all values θ and returns *TRUE* if θ belongs to confidence regions and *FALSE* otherwise. Algorithm 1 is pseudo code of single MSPS procedure.

The key idea of the MSPS algorithm is that in case of ϕ_i symmetrically distributed around zero, all sums $S_k(\theta^\star) = \sum_{i=1}^N a_{k,i}\Delta_i\varepsilon_i$ and $S_0(\theta^\star) = \sum_{i=1}^N \Delta_i\varepsilon_i$ are identically distributed due to the fact that random vectors $(a_{k,1}\Delta_1,\dots,a_{k,N}\Delta_N)$ and $(\Delta_1,\dots,\Delta_N)$ identically distributed. Thus, if we denote $Z_k = ||S_k(\theta^\star)||$ for $k = 0,\dots,M-1$ than Z_0 would took any position in sorted row $Z_{(0)},\dots,Z_{(M-1)}$ with equal probability $\frac{1}{M}$. Hence, Z_0 will not be among $q \le M$ greatest values in $\{Z_k\}_{k=0}^{M-1}$ with exact probability $1 - \frac{q}{M}$.

It is worth to note, that in case of $E\phi_i = 0$ MSPS algorithm is identical to SPS one.

The following theorem stands what set of all θ for which algorithm 0 returns true forms an exact $(1 - \frac{q}{M})$-confidence region for parameter $\theta^\star$.

Theorem 1: In Assumptions II, let $M > q \in \mathbb{R}$ be two positive integers, $\{S_k(\theta)\}_i = 1^{M-1}$ be $M-1$ modified sign-perturbed sums and $S_0(\theta) = \sum_{i=1}^N \Delta_i(y_i - (\phi_i - m_\phi)^T\theta)$ be original not perturbed sum. Taking $Z_k(\theta) = ||S_k(\theta)||_2^2$, $k = 0,\dots,M-1$ following set

$$\Theta_{M,q} = \{\theta \in \mathbb{R}^p : |\{k \in \{1,\dots,M-1\} : Z_0(\theta) < Z_k(\theta)\}| \ge 1\}$$

Algorithm 1 MSPS algorithm

```
procedure MSPS({y_i}_{i=1}^N, {φ}_{i=1}^N, m_φ, M, q, θ)
    for i in 1...N do
        n_i ← y_i − φ_i^T · θ
        Δ_i ← φ_i − m_φ
    end for
    for k in 1...M−1 do
        S_k ← 0 ∈ R^p
        for i in 1...N do
            S_k ← S_k + Δ_i · n_i · RANDOMSIGN()
        end for
        Z_k ← ||S_k||_2^2
    end for
    S_0 ← 0 ∈ R^p
    for i in 1...N do
        S_0 ← S_0 + Δ_i · n_i
    end for
    Z_0 ← ||S_0||_2^2
    r ← 0
    for k in 1...M−1 do
        if Z_0 > Z_k then
            r ← r+1
        end if
    end for
    if r >= q then
        return FALSE
    else
        return TRUE
    end if
end procedure
```

would be an exact confidence region for parameter $\theta^\star$ with probability $1-\frac{q}{M}$, that is

$$P_{\varepsilon_i, a_{k,i}}(\theta^\star \in \Theta_{M,q}) = 1 - \frac{q}{M}.$$

Proof can be found in Appendix VI.

This theorem is similar for the corresponding theorem for SPS algorithm [13] as well as modified sign-perturbed sums similar to sign-perturbed sums. It is essential that this theorem considers a much wider class of noises because there is no requirements on its symmetry and mean.

B. Confidence region calculation via SSPS area

In addition, we describe $\Theta_{M,q}$ a bit more thoroughly. In simple words, $\Theta_{M,q}$ consists of such points $\theta \in \mathbb{R}^d$, for which $Z_0(\theta)$ is greater no more then $M-q-1$ different $Z_k(\theta)$. Equivalent, $\Theta_{M,q}$ consists of points θ, for which $Z_0(\theta)$ is lower than at least q different $Z_k(\theta)$ — so that for at least q different k inequality $Z_0(\theta) < Z_k(\theta)$ holds true. Using this interpretation, $\Theta_{M,q}$ could be written as an intersections union.

Remark 1: Note that

$$\Theta_{M,q} = \bigcup_{\mathbb{I} \subset \{1,..,M-1\} : |\mathbb{I}| = q} \left(\bigcap_{k \in \mathbb{I}} \{\theta : Z_0(\theta) < Z_k(\theta)\} \right) \neq \varnothing.$$

The importance of this result is that we reduce the problem of $\Theta_{M,q}$ calculation to problem of calculation $\{\theta : Z_0(\theta) < Z_k(\theta)\}$. We will declare this set as Single Sign-Perturbed Sum (SSPS) area. Some ways and ideas about how to calculate SSPS area will be described further.

C. One-dimensional SSPS area

We will examine one-dimensional case ($d=1$) because of its simplicity and interpretability.

Lemma 1: Let $q=1$. Then

$$\{\theta : Z_0(\theta) < Z_k(\theta)\} = (B_k^{min}, B_k^{max}), \quad \text{where}$$

$$B_k^{(1)} = \frac{\sum_{i=1}^N (1-a_{k,i})\phi_i y_i}{\sum_{i=1}^N (1-a_{k,i})\phi_i^2}, \quad B_k^{(2)} = \frac{\sum_{i=1}^N (1+a_{k,i})\phi_i y_i}{\sum_{i=1}^N (1+a_{k,i})\phi_i^2}$$

$$B_k^{min} = \min(B_k^{(1)}, B_k^{(2)}), \qquad B_k^{max} = \max\left(B_k^{(1)}, B_k^{(2)}\right)$$

Proof can be found in Appendix VI

It is important to note that by using formulas for Single Sign-Perturbed Sum area $\{\theta : Z_0(\theta) < Z_k(\theta)\}$ we can significantly improve one-dimensional MSPS confidence region compared to origin point-wise calculation through algorithm 0. Since B_k forming $\Theta_{M,q}$ boundaries, it would be interesting to find out some of its statistical properties.

Proposition 1: Let B_k be either $B_k^{(1)}$ or $B_k^{(2)}$ and in addition for assumption II following is true

- $\{\phi_i\}$ second and fourth moments exists;
- $\{\varepsilon_i\}$ is either a random variables which raw second moments exists and bounded by constant $C < \infty$, or determined sequence of values bounded $\sqrt{C}$.

Then

$$E_{\varepsilon,\phi,a}[B_k] = \theta^\star,$$

$$E_{\varepsilon,\phi,a}[B_k - E_{\varepsilon,\phi,a}[B_k]]^2 \xrightarrow[N\to\infty]{} 0,$$

where $E_{\varepsilon,\phi,a}$ is mathematical expectation of joint distribution of random variables $\{\varepsilon_i\}_{i=1}^N, \{\phi_i\}_{i=1}^N, \{a_i\}_{i=1}^N$.

Proof can be found in Appendix VI.

This proposition brings two important facts. First, the mean of $\Theta_{M,q}$ boundaries B_k is equal to true parameter value $\theta^\star$, so boundaries just fluctuate around it. Second, this boundaries tend to $\theta^\star$ with N increased. Summing up, the confidence region converges to the parameter $\theta^\star$ when $N \to \infty$. By analogy with estimator, the $\Theta_{M,q}$ confidence region could be called *unbiased* and *consistent*.

IV. Simulated Examples

In this section we provide illustrated comparison of SPS, MSPS and asymptotic confidence regions construction method on simulated data. Since MSPS method is aimed to replace SPS than the second one is inapplicable we will illustrate such kind of situation. Additionally, we will use traditional asymptotic confidence regions construction to illustrate it inapplicability in such situations. At first, we briefly describe a method for constructing asymptotic confidence region under assumption of asymptotic normality of the noise.

A. Asymptotic confidence regions

One of the standard methods for $\theta^\star$ parameter estimation is least squares estimator, which finds the best solution in terms of least squares error $\sum_{i=1}^{N}||y_i - \phi_i\theta||_2^2$. Obtained estimate is called *least squares estimate* and could be calculated by following formula

$$\hat{\theta}_{LSE} = \arg\min_{\theta\in\mathbb{R}^d} \sum_{i=1}^{N} ||y_i - \phi_i^T\theta||.$$

Useful fact, that under some moment conditions on the noise sequence the least square estimate $\hat{\theta}_{LSE}$ is asymptotically normal with $cov(\hat{\theta}_{LSE} - \theta^\star)$ converges to $\Xi_N = \hat{\sigma}^2\sum_{i=1}^{N}(\phi_i\phi_i^T)^{-1}$, where $\hat{\sigma}^2$ is a ε_i variance estimate [1] and $\theta^\star$ asymptotic confidence region could be calculated as

$$\Theta_\alpha = \{\theta : (\theta - \theta_{LSE})^T\Xi_N^{-}1(\theta - \theta_{LSE}) \leq \chi_\alpha(1)\hat{\sigma}^2/N\}, \quad (4)$$

where $\chi_\alpha(1)$ — is an χ^2 distribution with 1 degree of freedom α-quantile. However, such confidence regions provide very rough estimate under Assumptions II and could cause mistakes and incorrect results.

B. Simulation description

In context of this paper cause of Assumptions II we interested in situation when

- noise has asymmetrical and even biased distribution, because this is a field of MSPS application,
- data sample size is relatively small, because it confine the application of asymptotic methods.

We consider that ϕ_i belongs to $\mathbb{R}^2$ because it is much easier to illustrate it on a plane. Furthermore, ϕ_i is simulated as Gaussian random vector

$$\phi_i \sim N(\mu_\phi, \Sigma_\phi), \text{ where } \quad \mu_\phi = (1,-1)^T, \quad \Sigma_\phi = \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{2} \end{pmatrix}$$

and true parameter $\theta^\star = (-1,2)^T$ always. The only thing that varies in the examples below is the noise ε_i distribution.

We divide this section into two parts: the first one is dedicated to the case of relatively big number of observations ($N = 50$) and the second one is dedicated to the case of relatively small number of observations ($N = 15$).

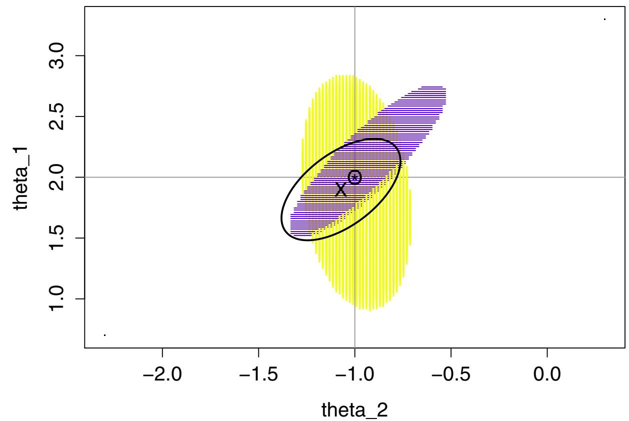
At each figure we provide three confidence regions: SPS-based confidence region, MSPS-based confidence region and confidence region obtained by the asymptotic formula (4). *Notation remark*, by $N(0,1)$ we refer to univariate normal distribution with zero mean, and by $Exp(\lambda = 1)$ we refer to exponential distribution with rate parameter equal to one.

C. Big sample size

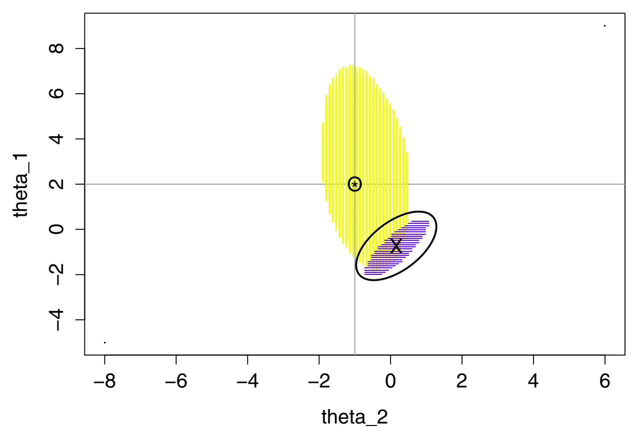
Here we will consider two types of noise: *unbiased symmetric* noise and *biased asymmetric* noise.

1) $\varepsilon_i \sim N(0,1)$
2) $\varepsilon_i \sim Exp(\lambda = 1) + 4$

It is expected that MSPS will perform slightly worse than both SPS and 4 in case of symmetric noise, but will outperform them in case of noise asymmetry. It could be seen from the figure 1a, that corresponds to normally distributed



(a) Confidence regions obtained by SPS (blue horizontal lines), MSPS (yellow vertical lines) and asymptotic one (ellipsoid). True parameter is marked by circle with dot inside, least squares estimate is marked by cross



(b) Confidence regions obtained by SPS (blue horizontal lines), MSPS (yellow vertical lines) and asymptotic one (ellipsoid). True parameter is marked by circle with dot inside, least squares estimate is marked by cross

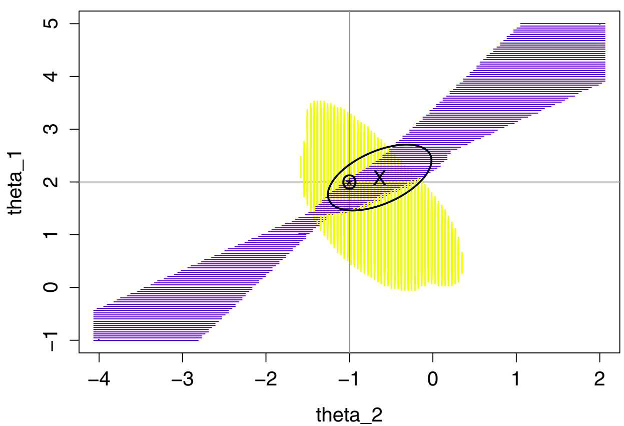
zero mean (and hence symmetric) noise 1, all confidence regions contain both true parameter and least square estimate. Noticeably that MSPS and SPS-confidence regions are much bigger than the asymptotic one.

Figure 1b corresponds to biased noise and shows MSPS-based confidence region advantage. Despite the fact that MSPS confidence region increased in size, it still contains true parameter value, while both asymptotic and SPS-based confidence regions a shifted because of noise bias and asymmetry.

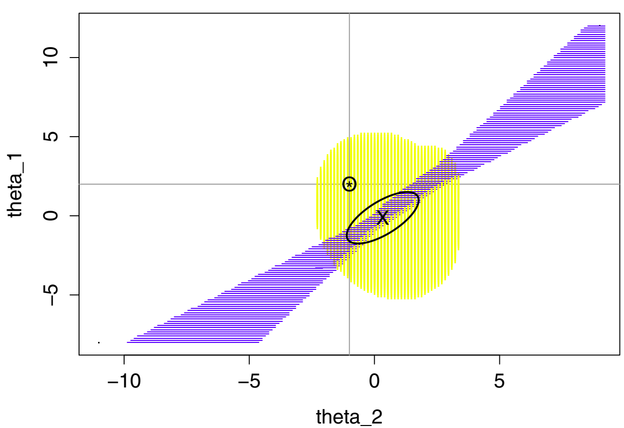
D. Small sample size

Here we will consider the two types of noise: *unbiased asymmetric* and *biased asymmetric*

1) $\varepsilon_i \sim Exp(\lambda = 1) - 1$,
2) $\varepsilon_i \sim Exp(\lambda = 1) + 4$.



(c) Confidence regions obtained by SPS (blue horizontal lines), MSPS (yellow vertical lines) and asymptotic one (ellipsoid). True parameter is marked by circle with dot inside, least squares estimate is marked by cross



(d) Confidence regions obtained by SPS (blue horizontal lines), MSPS (yellow vertical lines) and asymptotic one (ellipsoid). True parameter is marked by circle with dot inside, least squares estimate is marked by cross

The small sample size case is rather important because the noise asymmetry plays higher role there. As soon as noise variables are i.i.d they are asymptotically normal due to Central Limit Theorem (CLT) and hence symmetrical about zero in case of zero mean. Obviously, SPS should outperform both opponents in both cases. Indeed, figure 1c which corresponds to unbiased but asymmetrical noise 1 shows that only MSPS-based confidence region contains significant $\theta^\star$ neighborhood, whereas is shifted due to noise asymmetry and SPS-based region becomes unbounded which is unacceptable. Similar situation is shown on figure 1d where bias added to asymmetrical noise 1 with asymptotic confidence interval shifted even further from true parameter $\theta^\star$.

V. Conclusion

In this paper we proposed a new method called *Modified Sign-Perturbed Sums* (MSPS) for construction exact confidence region of linear regression parameter. This algorithm is a modified version of previously proposed *Sign-Perturbed Sums* (SPS) method [13]. The theoretically proven advantage of the MSPS method is that it could construct exact confidence regions even under *external arbitrary noise*.

From the theoretical point of view, this method is based on the same idea as SPS which consists of obtaining modified sing-perturbed sums from the observation by using *random signs*. It is also shown for one-dimensional case that confidence regions obtained by MSPS method converge to the true parameter value under some additional assumptions.

From the practical point of view, it is shown that for biased and asymmetrically distributed noise MSPS confidence region outperforms both SPS confidence regions and confidence region obtained by classical asymptotic theory, especially in case of small number of measurements then big number of measurements can not neglect the noise asymmetry impact.

In future work we plan to use above described theoretical results in our practical project: multi-agents control system for the group of UAVs. One of the important challenging problems for the UAVs system design is the development of an optimization flight algorithm. We use the randomized algorithm to estimate the wind direction [14] and to determine the center of thermal updrafts [15].

VI. Appendix

Here we provide the sketch of proof of Theorem 1 and some of the propositions.

To prove Theorem 1 we first give particular lemma follows from theorem, proved by B. C. Csaji, M. C. Campi and E. Weyer in [13] which stands for SPS method.

***Lemma** 2:* B. C. Csaji, M. C. Campi and E. Weyer [13] Let g_0 be a random variable

$$g_0 = || \sum_{i=1}^{N} t_i x_i ||_2^2,$$

where $\{t_i\}_{i=1}^N$ are independent random variables symmetrically distributed about zero and $\{x_i\}_{i=1}^N$ is some variables, possibly deterministic, independent from $\{t_i\}_{i=1}^N$. If $a_{k,i}$ are random signs and random variables $\{g_k\}_{k=1}^{M-1}$ defined as

$$g_k = || \sum_{i=1}^{N} a_{k,i} t_i x_i ||_2^2,$$

then $\{g_k\}_{k=0}^{M-1}$ are i.i.d. Furthermore, probability of g_0 not being among of q largest values from $\{g_k\}_{k=0}^{M-1}$ is equal to $1 - \frac{q}{M}$.

Having this lemma we now ready to prove Theorem 1.

Proof: Of the theorem 1

To prove the theorem we need to show that for $\theta^\star \in \Theta_{M,q}$ with probability $1 - \frac{q}{M}$, or equivalently that $Z_0(\theta^\star)$ is less than at least q different $Z_k(\theta^\star)$ from $\{Z_k(\theta^\star)\}_{k=1}^{M-1}$ with this probability. Since $Z_0(\theta^\star)$ and $Z_k(\theta^\star)$ could be written as

$$Z_k(\theta^\star) = ||\sum_{i=1}^{N} a_{k,i}\Delta_i\varepsilon_i||, \quad Z_0(\theta^\star) = ||\sum_{i=1}^{N} \Delta_i\varepsilon_i||.$$

Here, denoting $x_i = \varepsilon_i$ and $t_i = \Delta_i$ for $i = 1,\ldots,N$ we can use Lemma 2 which states that all random (through Δ_i the randomness) values $\{Z_k(\theta^\star)\}_{k=0}^{M-1}$ are i.i.d and $Z_0(\theta^\star)$ is not among q largest values from $\{Z_k(\theta^\star)\}_{k=1}^{M-1}$ with exact probability $1-\frac{q}{M}$. ∎

Proof: of Lemma 1

Lets consider inequality $Z_0(\theta^\star) < Z_k(\theta^\star)$

$$(\sum_{i=1}^{N} \Delta_i(y_i - \phi_i\theta))^2 < (\sum_{i=1}^{N} a_{k,i}\Delta_i(y_i - \phi_i\theta))^2$$

$$(\sum_{i=1}^{N} \Delta_i(y_i - \phi_i\theta))^2 - (\sum_{i=1}^{N} a_{k,i}\phi_i(y_i - \phi_i\theta))^2 < 0$$

$$(\sum_{i=1}^{N} (1-a_{k,i})\Delta_i(y_i - \phi_i\theta))(\sum_{i=1}^{N} (1+a_{k,i})\Delta_i(y_i - \phi_i\theta)) < 0$$

$$\left(\frac{\sum_{i=1}^{N}(1-a_{k,i})\Delta_i(\varepsilon_i+\Delta_i\theta^\star)}{\sum_{i=1}^{N}(1-a_{k,i})\Delta_i^2} - \theta\right)\cdot$$

$$\cdot\left(\frac{\sum_{i=1}^{N}(1+a_{k,i})\Delta_i(\varepsilon_i+\Delta_i\theta^\star)}{\sum_{i=1}^{N}(1+a_{k,i})\Delta_i^2} - \theta\right) < 0.$$

∎

Proof: of Proposition 1

$$E_{\varepsilon,\phi,a}B_{\{a_{k,i}\}} = E_{\varepsilon,\phi,a}\left[\frac{\sum_{i=1}^{N}(1\pm a_{k,i})\Delta_i(\varepsilon_i+\Delta_i\theta^\star)}{\sum_{i=1}^{N}(1\pm a_{k,i})\Delta_i^2}\right] =$$

$$= \theta^\star + E_{\varepsilon,\phi,a}\left[\frac{\sum_{i=1}^{N}(1\pm a_{k,i})\Delta_i\varepsilon_i}{\sum_{i=1}^{N}(1\pm a_{k,i})\Delta_i^2}\right].$$

Since Δ_i independent from ε_i and $E\Delta_i = 0$

$$E_{\varepsilon,\phi,a}\left[\frac{\sum_{i=1}^{N}(1\pm a_{k,i})\Delta_i\varepsilon_i}{\sum_{i=1}^{N}(1\pm a_{k,i})\Delta_i^2}\right] = \sum_{i=1}^{N} E_{\varepsilon,\phi,a}\left[\frac{(1\pm a_{k,i})\Delta_i\varepsilon_i}{\sum_{j=1}^{N}(1\pm a_{k,j})\Delta_j^2}\right] \leq$$

$$\leq \sum_{i=1}^{N} E_{\varepsilon,\phi,a}\left[\frac{(1\pm a_{k,i})\Delta_i\varepsilon_i}{\sum_{j\neq i}^{N}(1\pm a_{k,j})\Delta_j^2}\right] = 0.$$

Here we got first statement. Two obtain the second one, note that for any $\{x_i\}_1^N$ inequality $2\sum_1^N x_i^2 \geq (\sum_1^N x_i)^2$ is true. Hence,

$$E_{\varepsilon,\phi,a}[B_{\{a_{k,i}\}} - E_{\varepsilon,\phi,a}B_{\{a_{k,i}\}}]^2 = E_{\varepsilon,\phi,a}\left[\frac{(\sum_{i=1}^{N}(1\pm a_{k,i})\Delta_i\varepsilon_i)^2}{(\sum_{i=1}^{N}(1\pm a_{k,i})\Delta_i^2)^2}\right] \leq$$

$$\leq CE_{\varepsilon,\phi,a}\left[\frac{(\sum_{i=1}^{N}(1\pm a_{k,i})\Delta_i\varepsilon_i)^2}{(\sum_{i=1}^{N}(1\pm a_{k,i})\Delta_i)^4}\right] \xrightarrow[N\to\infty]{} 0,$$

Indeed since ε_i independent with Δ_i we can take mathematical expectation for them separately. Taking expectation by ε_i and using their boundedness we obtain

$$E_{\varepsilon,\phi,a}[B_{\{a_{k,i}\}} - E_{\varepsilon,\phi,a}B_{\{a_{k,i}\}}]^2 \leq CE_{\varepsilon,\phi,a}\left[\frac{(\sum_{i=1}^{N}(1\pm a_{k,i})\Delta_i)^2}{(\sum_{i=1}^{N}(1\pm a_{k,i})\Delta_i)^4}\right] =$$

$$= CE_{\varepsilon,\phi,a}\left[\frac{1}{(\sum_{i=1}^{N}(1\pm a_{k,i})\Delta_i)^2}\right] \xrightarrow[N\to\infty]{} 0.$$

Thus, the proposition proved.

∎

REFERENCES

[1] L. Ljung, *System identification (2nd ed.): theory for the user*. Upper Saddle River, NJ, USA: Prentice Hall PTR, 1999.

[2] V. Vapnik, *Estimation of Dependencies Based on Empirical Data*. Springer-Verlag, 1982.

[3] O. Granichin and B. Polyak, *Randomized Algorithms of an Estimation and Optimization Under Almost Arbitrary Noises*. Moscow, Russia: Nauka, 2003.

[4] K. Amelin and O. N. Granichin, “Randomized controls for linear plants and confidence regions for parameters under external arbitrary noise,” *American Control Conference*, pp. 851–856, June 2012.

[5] M. Campi and E. Weyer, “Non-asymptotic confidence sets for the parameters of linear transfer functions,” *IEEE Trans. Autom. Control*, vol. 55, pp. 2708–2720, December 2010. 12.

[6] A. Vakhitov, O. Granichin, and V. Vlasov, “Adaptive control of siso plant with time-varying coefficients based on random test perturbation,” *American Control Conference*, pp. 4004–4009, December 2010.

[7] O. N. Graichin and V. Fomin, “Adaptive control using test signals in the feedback channel,” vol. 47, pp. 238–248, 1986. 2.

[8] K. Amelin, N. Amelina, O. Granichin, and O. Granichina, “Combined procedure with randomized controls for the parameters confidence region of linear plant under external arbitrary noise,” *51st IEEE Conference on Decision and Control*, pp. 2134–2139, December 2012.

[9] O. Granichin, “Linear regression and filtering under nonstandard assumptions (arbitrary noise),” *IEEE Trans. Autom. Control*, vol. 49, pp. 1830–1835, October 2004.

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

[11] A. Garulli, L. Giarre, and G. Zappa, “Identification of approximated hammerstein models in a worst-case setting,” *IEEE Trans. Autom. Control*, vol. 47, pp. 2046–2050, December 2002. 7.

[12] B. Polyak and P. Sherbakov, *Robust Stability and Control*. Moscow, Russia: Nauka, 2002.

[13] C. Balazs, M. Campi, and E. Weyer, “Non-asymptotic conidence regions for the least-squares estimate,” *Proceedings of the 16th IFAC Symposium on System Identification (SYSID 2012)*, pp. 227–232, July 2012.

[14] K. Amelin, “Randomized controls for the optimization of small uav flight under unknown arbitrary wind distrurdances,” *Cybernetics and Physics*, vol. 1, pp. 79–89, October 2012. 2.

[15] K. Amelin, N. Amelina, O. Granichin, O. Granichina, and B. Andrievsky, “Randomized algorithm for uavs group flight optimization,” *Proc. of 11th IFAC International Workshop on Adaptation and Learning in Control and Signal Processing*, pp. 205–208, July 2013.