Automatica 82 (2017) 287–294

Contents lists available at ScienceDirect

Automatica

journal homepage: www.elsevier.com/locate/automatica

Brief paper

Asymptotic properties of SPS confidence regions☆

Erik Weyer [a], Marco Claudio Campi [b], Balázs Csanád Csáji [c]

[a] Department of Electrical and Electronic Engineering, The University of Melbourne, VIC 3010, Australia
[b] Department of Information Engineering, University of Brescia, Via Branze 38, 25123 Brescia, Italy
[c] Institute for Computer Science and Control, Hungarian Academy of Sciences, Kende utca 13-17, 1111 Budapest, Hungary

ARTICLE INFO

Article history:
Received 11 February 2016
Received in revised form
15 December 2016
Accepted 31 March 2017
Available online 23 May 2017

Keywords:
System identification
Parameter estimation
Regression analysis
Asymptotic properties

ABSTRACT

Sign-Perturbed Sums (SPS) is a system identification method that constructs non-asymptotic confidence regions for the parameters of linear regression models under mild statistical assumptions. One of its main features is that, for any finite number of data points and any user-specified probability, the constructed confidence region contains the true system parameter with exactly the user-chosen probability. In this paper we examine the size and the shape of the confidence regions, and we show that the regions are strongly consistent, i.e., they almost surely shrink around the true parameter as the number of data points increases. Furthermore, the confidence region is contained in a marginally inflated version of the confidence ellipsoid obtained from the asymptotic system identification theory. The results are also illustrated by a simulation example.

1. Introduction

Models of dynamical systems are of widespread use in many fields of science and engineering. Such models are often obtained using system identification techniques, that is, the models are estimated from observed data. There will always be uncertainty associated with models of dynamical systems, and an important problem is the uncertainty evaluation. For example, if the model is going to be used for design, the model uncertainty will be one of the factors which determine how much robustness needs to be built into the design. A common way to characterise the uncertainty in the model parameter is to use confidence regions, and in earlier papers (Csáji, Campi, & Weyer, 2012, 2015), we introduced the Sign-Perturbed Sums (SPS) method for the construction of confidence regions for the parameters of linear regression models. The main features of the SPS method are that it constructs confidence regions from a finite number of data points and that the confidence regions contain the true parameter with an exact user-chosen probability. This is in contrast to asymptotic theory of system identification, e.g. Ljung (1999), which delivers confidence ellipsoids which are only guaranteed as the number of data points tends to infinity. SPS has some similarities with the Leave-out Sign-dominant Correlation Regions (LSCR) method (Campi, Ko, & Weyer, 2009; Campi & Weyer, 2005, 2010; Dalai, Weyer, & Campi, 2007) which also generates confidence regions based upon a finite number of data points. However, unlike SPS, LSCR usually only provides an upper bound on the probability that the true parameter belongs to the confidence region. Numerical implementations and further developments in the vein of LSCR and SPS are considered in Granichin (2012), Kieffer and Walter (2013a,b), Kolumbán, Vajk, and Schoukens (2015) and Schoukens, Rolain, Vandersteen, and Pintelon (2013), while other methods and studies of finite sample properties in system identification can be found in Dabbene, Sznaier, and Tempo (2014) and den Dekker, Bombois, and Van den Hof (2008).

Though the main draw card of SPS is the finite sample properties, the asymptotic properties are also of interest, since any reasonable method for uncertainty evaluation should deliver smaller and smaller confidence sets as the information about the system increases. Here, we analyse the asymptotic properties of SPS and we show that

- SPS is strongly consistent (Theorem 2), i.e., its confidence regions shrink around the true parameter and, asymptotically, all parameter values different from the true one will be excluded.
- The SPS confidence regions are contained in marginally inflated versions of the confidence ellipsoids obtained from the asymptotic system identification theory (Theorem 3), where the amount of inflation needed is asymptotically vanishing.

☆ The material in this paper was partially presented at the 53rd IEEE Conference on Decision and Control, December 15–17, 2014, Los Angeles, CA, USA. This paper was recommended for publication in revised form by Associate Editor Alessandro Chiuso under the direction of Editor Torsten Söderström.

E-mail addresses: ewey@unimelb.edu.au (E. Weyer), marco.campi@unibs.it (M.C. Campi), balazs.csaji@sztaki.mta.hu (B.Cs. Csáji).

http://dx.doi.org/10.1016/j.automatica.2017.04.041
0005-1098/

A simulation example is also included which illustrates the behaviour of the SPS confidence region as the number of data points and sign-perturbed sums increase.

A preliminary version of the consistency result was presented in Csáji, Campi, and Weyer (2014) where, however, stronger assumptions were applied. While the practical use of the SPS method is not affected by the results in this paper, they may increase the users' confidence in the method.

The paper is organised as follows. In Section 2 we introduce the system setting and briefly summarise the SPS algorithm. The asymptotic results are given in Section 3, and they are illustrated by a simulation example in Section 4. The proofs can be found in Appendices A–D.

2. Setting

Here we briefly summarise the Sign-Perturbed Sums (SPS) method. For more details, see Csáji et al. (2015). We consider linear regression models of the form

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

where Y_t is the output, N_t is the noise, φ_t is the regressor, θ^* is the true parameter (constant), and t is the time index. Y_t and N_t are scalars, while φ_t and θ^* are d dimensional vectors. We consider a sample of size n which consists of the regressors $\varphi_1, \ldots, \varphi_n$ and the outputs $Y_1, \ldots, Y_n$.

The assumptions on the noise and the regressors are

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

A2 *The regressors* $\{\varphi_t\}$ *are deterministic and*

$$R_n \triangleq \frac{1}{n}\sum_{t=1}^{n}\varphi_t\varphi_t^{\mathrm{T}}$$

is non-singular.

Although it is assumed that $\{\varphi_t\}$ are deterministic, the results in this paper also hold for stochastic regressors as long as they are independent of the noise sequence.

2.1. Main idea of SPS

The least-squares estimate (LSE) of θ^* is given by

$$\hat{\theta}_n \triangleq \arg\min_{\theta\in\mathbb{R}^d}\sum_{t=1}^{n}(Y_t - \varphi_t^{\mathrm{T}}\theta)^2,$$

which can be found by solving the *normal equation*, i.e.,

$$\sum_{t=1}^{n}\varphi_t(Y_t - \varphi_t^{\mathrm{T}}\theta) = 0.$$

The main building block of the SPS algorithm is, as the name suggests, $m-1$ sign-perturbed versions of the normal equation (normalised by $\frac{1}{n}R_n^{-\frac{1}{2}}$). The sign-perturbed sums are defined as

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

$i = 1, \ldots, m-1$, and a reference sum is given by

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

Here, $R_n^{\frac{1}{2}}$ is a matrix[1] that satisfies $R_n = R_n^{\frac{1}{2}}R_n^{\frac{1}{2}\mathrm{T}}$, and $\{\alpha_{i,t}\}$ are independent and identically distributed (i.i.d.) random variables (independent of $\{N_t\}$) that take on the values ± 1 with probability $1/2$ each.

[1] One such matrix $R_n^{1/2}$ can be found from the Cholesky decomposition of R_n. However, the equation $R_n = R_n^{1/2}R_n^{1/2\mathrm{T}}$ admits more than one solution $R_n^{1/2}$, and any solution can be used.

Table 1
Pseudocode: SPS-initialisation.

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 $R_n \triangleq \frac{1}{n}\sum_{t=1}^{n}\varphi_t\varphi_t^{\mathrm{T}}$, and find a factor $R_n^{1/2}$ such that $R_n^{1/2}R_n^{1/2\mathrm{T}} = R_n$;
3. Generate $n\,(m-1)$ i.i.d. random signs $\{\alpha_{i,t}\}$ with $\mathbb{P}(\alpha_{i,t} = 1) = \mathbb{P}(\alpha_{i,t} = -1) = \frac{1}{2}$, for $i \in \{1, \ldots, m-1\}$ and $t \in \{1, \ldots, n\}$;
4. Generate a random permutation π of the set $\{0, \ldots, m-1\}$, where each of the $m!$ possible permutations has the same probability $1/(m!)$.

Table 2
Pseudocode: SPS-indicator(θ).

1. For a given θ, compute the prediction errors $\varepsilon_t(\theta) \triangleq Y_t - \varphi_t^{\mathrm{T}}\theta$, for $t \in \{1, \ldots, n\}$;
2. Evaluate, for $i \in \{1, \ldots, m-1\}$, functions $S_0(\theta) \triangleq R_n^{-\frac{1}{2}}\frac{1}{n}\sum_{t=1}^{n}\varphi_t\varepsilon_t(\theta)$; $S_i(\theta) \triangleq R_n^{-\frac{1}{2}}\frac{1}{n}\sum_{t=1}^{n}\alpha_{i,t}\varphi_t\varepsilon_t(\theta)$;
3. Order the 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.
5. Return 1 if $\mathcal{R}(\theta) \le m - q$, otherwise return 0.

The key observation is that for $\theta = \theta^*$ one has

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

$$S_i(\theta^*) = R_n^{-\frac{1}{2}}\frac{1}{n}\sum_{t=1}^{n}\alpha_{i,t}\varphi_t N_t.$$

As N_t is an independent and symmetric sequence, there is no reason why $\|S_0(\theta^*)\|^2$ should be bigger or smaller than any other $\|S_i(\theta^*)\|^2$. This property is exploited in the construction of the confidence regions where the values of θ for which $\|S_0(\theta)\|^2$ is among the q largest ones are excluded. As stated in Theorem 1, the confidence region has exact probability $1 - q/m$ of containing the true system parameter. In Csáji et al. (2015) it has also been noted that when $\theta - \theta^*$ is "large", $\|S_0(\theta)\|^2$ tends to be the largest of the m functions, so that θ values far away from θ^* will be excluded from the confidence set.

2.2. Formal construction of the SPS confidence region

The SPS algorithm consists of two parts. The initialisation (Table 1) sets the main global parameters and generates the objects needed for the construction of the confidence region. In the initialisation, the user provides the desired confidence probability p. The second part (Table 2) evaluates an indicator function, which determines if a particular parameter θ belongs to the confidence region.

The random permutation π generated in the initialisation defines a strict total order $\succ_\pi$ which is used to break ties in case two values $\|S_i(\theta)\|^2$ and $\|S_j(\theta)\|^2$, $i \neq j$ are equal. Given m scalars $\{Z_i\}$, $i = 0, \ldots, m-1$, $\succ_\pi$ is

$$Z_k \succ_\pi Z_j \quad \text{if and only if}$$
$$\big(Z_k > Z_j\big) \quad \text{or} \quad \big(Z_k = Z_j \text{ and } \pi(k) > \pi(j)\big).$$

The p-level *SPS confidence region* is given by

$$\widehat{\Theta}_n \triangleq \{\theta : \text{SPS-INDICATOR}(\theta) = 1\}.$$

As it was shown in Csáji et al. (2015), the confidence region $\widehat{\Theta}_n$ contains θ^* with exact probability p as stated in the next theorem.

Theorem 1. *Assuming* A1 *and* A2*, the confidence probability of the constructed confidence region is exactly p,*

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

Note that this probability is w.r.t. both the noises $\{N_t\}$ and the random signs $\{\alpha_{i,t}\}$, i.e., the probability is a product measure. It is known that the LSE, $\hat{\theta}_n$, has the property that $S_0(\hat{\theta}_n) = 0$ (cf. the normal equation). Hence, the LSE is always included in the SPS confidence region (Csáji et al., 2015), provided that it is non-empty. Moreover the confidence region is star convex having the LSE as a star centre, see again Csáji et al. (2015).

3. Asymptotic properties of SPS

In addition to the probability of containing the true parameter, other important aspects are the size and the shape of the confidence regions. In this section we show that, under some additional mild assumptions, as the number of data points gets larger, the confidence regions get smaller. Moreover, as both n and m tend to infinity, the confidence regions are contained in marginally inflated versions of the confidence ellipsoids obtained from using asymptotic system identification results.

3.1. Strong consistency

Our first result shows that SPS is *strongly consistent*, in the sense that the confidence sets shrink around the true parameter as the sample size increases, and eventually exclude any other parameters $\theta' \neq \theta^*$.

The following additional assumptions are needed:

A3 (*nonvanishing excitation*)

$$\liminf_{n\to\infty} \lambda_{\min}(R_n) = \bar{\lambda} > 0$$

where $\lambda_{\min}(\cdot)$ *denotes minimum eigenvalue.*

A4 (*regressor growth rate restriction*)

$$\sum_{t=1}^{\infty} \frac{\|\varphi_t\|^4}{t^2} < \infty.$$

A5 (*noise variance growth rate restriction*)

$$\sum_{t=1}^{\infty} \frac{(\mathbb{E}[N_t^2])^2}{t^2} < \infty.$$

In the theorem below, $B_\varepsilon(\theta^*)$ denotes the Euclidean norm-ball centred at θ^* with radius $\varepsilon > 0$, i.e.

$$B_\varepsilon(\theta^*) \triangleq \{\theta \in \mathbb{R}^d : \|\theta - \theta^*\| \leq \varepsilon\}.$$

Theorem 2 states that the confidence regions $\widehat{\Theta}_n$ will eventually be included in any given norm-ball centred at the true parameter, θ^*.

Theorem 2. *Assume* A1, A2, A3, A4 *and* A5*. Then, for all* $\varepsilon > 0$ *almost surely (a.s.) there exists an* $\bar{N}$ *such that* $\widehat{\Theta}_n \subseteq B_\varepsilon(\theta^*)$ *for all* $n > \bar{N}$.

The proof of Theorem 2 can be found in Appendix A. The actual sample size $\bar{N}$ for which the confidence region will remain inside an ε-ball depends on the noise realisation, that is $\bar{N}$ is stochastic and depends on a generic element of the underlying probability space.

Note also that, for this asymptotic result to hold, the noise terms can be nonstationary and their variances can grow to infinity, as long as their growth-rate satisfies Assumption A5. Also, the magnitude of the regressors can grow without bound, as long as it does not grow too fast, as controlled by Assumption A4.

3.2. Asymptotic shape

Here we analyse the shape of the SPS confidence regions when n and m tend to ∞. Before we present our results, the confidence ellipsoids based on the asymptotic statistical theory, also widespread in system identification, are briefly reviewed, see Ljung (1999) for details.

3.2.1. Confidence ellipsoids of the asymptotic theory

Assuming that $\{N_t\}$ are zero mean and i.i.d. with variance σ^2, under mild conditions $\sqrt{n}\,(\hat{\theta}_n - \theta^*)$ converges in distribution to the Gaussian distribution with zero mean and covariance matrix $\sigma^2 R^{-1}$, where $R = \lim_{n\to\infty} R_n$ assuming the limit exists. As a consequence, $\frac{n}{\sigma^2}(\hat{\theta}_n - \theta^*)^{\mathrm{T}} R\,(\hat{\theta}_n - \theta^*)$ converges in distribution to the χ^2 distribution with $\dim(\theta^*) = d$ degrees of freedom.

An approximate confidence region can be obtained by replacing the matrix R with its estimate R_n,

$$\widetilde{\Theta}_n \triangleq \left\{\theta : (\theta - \hat{\theta}_n)^{\mathrm{T}} R_n\,(\theta - \hat{\theta}_n) \leq \frac{\mu\sigma^2}{n}\right\},$$

where the probability that θ^* is in the confidence region $\widetilde{\Theta}_n$ is *approximately* $p = F_{\chi^2}(\mu)$, where F_{χ^2} is the cumulative distribution function of the χ^2 distribution with d degrees of freedom. In the limit as n tends to infinity θ^* is contained in the set $\widetilde{\Theta}_n$ with probability $F_{\chi^2}(\mu)$, and this result also holds if σ^2 is replaced with its estimate,

$$\widehat{\sigma}_n^2 \triangleq \frac{1}{n-d}\sum_{t=1}^{n}(y_t - \varphi_t^T\hat{\theta}_n)^2.$$

3.2.2. Asymptotic shape of SPS confidence regions

In order to show that the SPS confidence regions asymptotically have similar shapes as the standard confidence ellipsoids, the assumptions on the regressors and the noise terms are strengthened to

A6 (*regressor growth rate restriction*)

$$\limsup_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} \|\varphi_t\|^4 < \infty.$$

A7 *(i.i.d. noise with bounded 4th order moment):* $\{N_t\}$ *is i.i.d. with* $\mathbb{E}[N_t^2] = \sigma^2$ *and* $\mathbb{E}[N_t^4] = \rho < \infty$.

The theorem below is given in terms of relaxed asymptotic confidence ellipsoids, which are defined as

$$\widetilde{\Theta}_n(\varepsilon) \triangleq \left\{\theta : (\theta - \hat{\theta}_n)^{\mathrm{T}} R_n (\theta - \hat{\theta}_n) \leq \frac{\mu\,\sigma^2 + \varepsilon}{n}\right\},$$

where $\varepsilon > 0$ is a margin. In the theorem, both n and m (recall that $m - 1$ is the number of sign-perturbed sums) go to infinity, and we use the notation $\widehat{\Theta}_{n,m}$ for the SPS region to explicitly indicate the dependence on n and m. We take $q_m = \lfloor(1-p)\,m\rfloor$, where $\lfloor(1-p)\,m\rfloor$ is the largest integer less than or equal to $(1-p)m$, so that Theorem 1 gives a confidence probability of $1 - \frac{q_m}{m} \triangleq p_m \to p$ from above as $m \to \infty$.

Theorem 3. *Assume* A1, A2, A3, A6 *and* A7*. Then, there exists a doubly-indexed set of random variables* $\{\varepsilon_{n,m}\}$ *such that* $\lim_{m\to\infty}\lim_{n\to\infty}\varepsilon_{n,m} = 0$ *a.s., and*

$$\widehat{\Theta}_{n,m} \subseteq \widetilde{\Theta}_n(\varepsilon_{n,m}).$$

The proof of Theorem 3 can be found in Appendix B.

We know from the Gauss–Markov theorem (Gentle, 2013; Kailath, Sayed, & Hassibi, 2000) that, under the assumptions of Theorem 3, the least-squares estimator is the *best linear unbiased estimator*(BLUE). Theorem 3 demonstrates that in the long run

$\widehat{\Theta}_{n,m}$ is almost surely contained in the asymptotic ellipsoid for the least-squares estimate when the noise variance is increased by a small (asymptotically vanishing) margin.

4. Simulation example

In this section we illustrate the asymptotic properties of the SPS method by a simulation example.

Consider the same second order data generating FIR system as in Csáji et al. (2015), that is,

$$Y_t = b_1^* U_{t-1} + b_2^* U_{t-2} + N_t,$$

where $\theta^* = [b_1^*\ b_2^*]^{\mathrm{T}} = [0.7\ 0.3]^{\mathrm{T}}$ is the true parameter and $\{N_t\}$ is a sequence of i.i.d. Laplacian random variables with zero mean and variance 0.1. The input is

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

where $\{V_t\}$ is a sequence of i.i.d. Gaussian random variables with zero mean and variance 1. The predictor is

$$\widehat{Y}_t(\theta) = b_1 U_{t-1} + b_2 U_{t-2} = \varphi_t^{\mathrm{T}}\theta,$$

where $\theta = [b_1\ b_2]^{\mathrm{T}}$ is the model parameter, and $\varphi_t = [U_{t-1}\ U_{t-2}]^{\mathrm{T}}$ is the regressor at time t.

Initially we construct a 95% confidence region for $\theta^* = [b_1^*\ b_2^*]^{\mathrm{T}}$ based on $n = 25$ data points, namely: $(Y_t, \varphi_t) = (Y_t, [U_{t-1}\ U_{t-2}]^{\mathrm{T}})$, $t = 1, \ldots, 25$.

We compute the shaping matrix

$$R_{25} = \frac{1}{25}\sum_{t=1}^{25}\begin{bmatrix} U_{t-1} \\ U_{t-2}\end{bmatrix}[U_{t-1}\ \ U_{t-2}],$$

and find a factor $R_{25}^{\frac{1}{2}}$ such that $R_{25}^{\frac{1}{2}} R_{25}^{\frac{1}{2}\mathrm{T}} = R_{25}$. Then, we compute the reference sum

$$S_0(\theta) = R_{25}^{-\frac{1}{2}}\frac{1}{25}\sum_{t=1}^{25}\begin{bmatrix} U_{t-1} \\ U_{t-2}\end{bmatrix}(Y_t - b_1 U_{t-1} - b_2 U_{t-2}),$$

and, using $m = 100$ and $q = 5$, we compute the 99 sign-perturbed sums, $i = 1, \ldots, 99$,

$$S_i(\theta) = R_{25}^{-\frac{1}{2}}\frac{1}{25}\sum_{t=1}^{25}\alpha_{i,t}\begin{bmatrix} U_{t-1} \\ U_{t-2}\end{bmatrix}(Y_t - b_1 U_{t-1} - b_2 U_{t-2}),$$

where $\{\alpha_{i,t}\}$ are i.i.d. random signs. The confidence region is formed by those θ's for which at least 5 of the $\|S_i(\theta)\|^2$, $i = 1, \ldots, 99$, values are larger than $\|S_0(\theta)\|^2$. It follows from Theorem 1 that the constructed confidence region contains the true parameter with exact probability $1 - \frac{5}{100} = 95\%$.

The SPS confidence region is shown in Fig. 1 together with the approximate confidence ellipsoid based on asymptotic system identification theory (with the noise variance estimated as $\widehat{\sigma}^2 = \frac{1}{23}\sum_{t=1}^{25}(Y_t - \varphi_t^{\mathrm{T}}\widehat{\theta}_n)^2$).

It can be observed that the non-asymptotic SPS region is similar in size and shape to the asymptotic confidence region, but it has the advantage that it is guaranteed to contain the true parameter with exact probability 95%.

Next, the number of data points was increased to $n = 400$, still with $q = 5$ and $m = 100$, and the confidence region in Fig. 2 was obtained. As can be seen, the SPS confidence region shrinks around the true parameter as n increases in accordance with Theorem 2 (observe the smaller range of the two axes in Fig. 2). This is further illustrated in Fig. 3 where the number of data points has been increased to 4000. When $q = 5$ and $m = 100$, we can still observe a difference between the SPS confidence region and the confidence ellipsoid based on the asymptotic theory, but when $q = 200$, $m = 4000$ is used, there is very little difference between the SPS confidence region and the confidence ellipsoid based on the asymptotic theory demonstrating the convergence result established in Theorem 3.

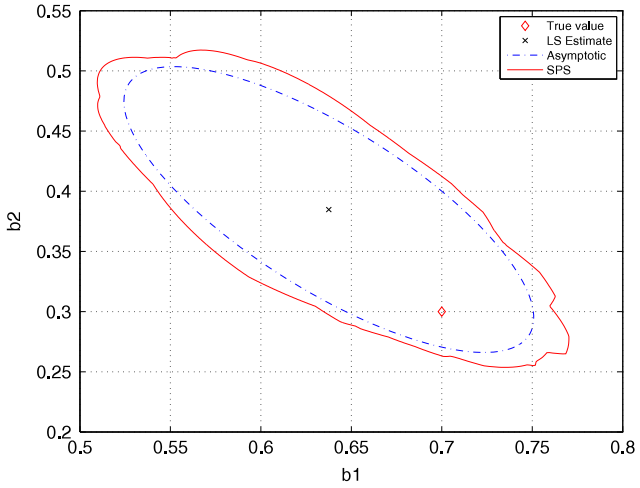

Fig. 1. 95% confidence regions, $n = 25$, $m = 100$.

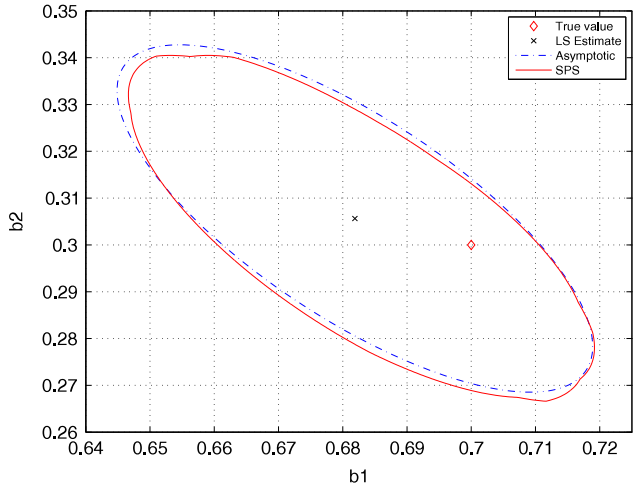

Fig. 2. 95% confidence regions, $n = 400$, $m = 100$.

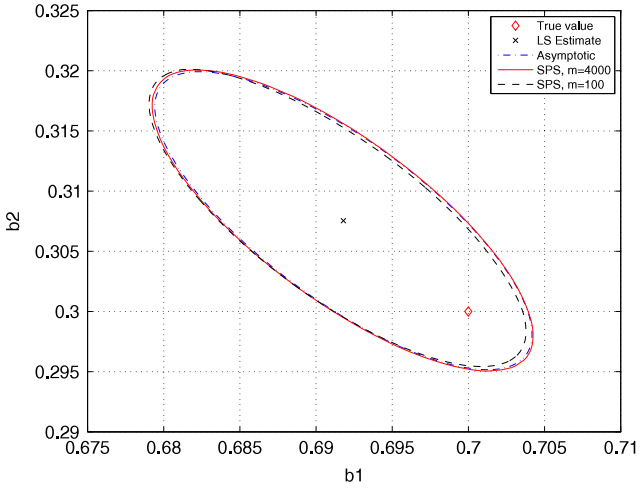

Fig. 3. 95% confidence regions, $n = 4000$, $m = 100$ and $m = 4000$.

5. Summary and conclusion

In this paper we have investigated the asymptotic properties of the SPS method, which constructs confidence regions for the parameters of linear regression models. It was shown that SPS is strongly consistent in the sense that its confidence regions become

smaller and smaller as the number of data points increases, and any parameter value different from θ^* will eventually be excluded. Moreover, as both the number of data points and the number of sign-perturbed sums tend to infinity, the confidence regions are included in the confidence ellipsoids from classical system identification theory when the noise variance is slightly increased. This shows that, in addition to its attractive finite sample properties, SPS has also very desirable asymptotic properties.

Acknowledgements

The work of E. Weyer was supported by the Australian Research Council (ARC) under Discovery Grants DP0986162 and DP130104028. The work of M.C. Campi was partly supported by the H&W program of the University of Brescia under the project "Classificazione della fibrillazione ventricolare a supporto della decisione terapeutica"—CLAFITE. B.Cs. Csáji was partially supported by the ARC grant DE120102601, the János Bolyai Research Fellowship, BO/00217/16/6, and the Hungarian Scientific Research Fund (OTKA), Grant No. 113038.

Appendix A. Proof of Theorem 2: strong consistency

We will prove that, for any $\varepsilon > 0$, there is an n such that $\|S_0(\theta)\|^2$ becomes the largest element in the ordering for all θ that are outside the ball $B_\varepsilon(\theta^*)$, so that all these θ's are excluded from the confidence region as $n \to \infty$.

Introduce the notations

$$\psi_n \triangleq \frac{1}{n}\sum_{t=1}^{n} \varphi_t N_t,$$

$$\gamma_{i,n} \triangleq \frac{1}{n}\sum_{t=1}^{n} \alpha_{i,t}\,\varphi_t N_t, \tag{A.1}$$

$$\Gamma_{i,n} \triangleq \frac{1}{n}\sum_{t=1}^{n} \alpha_{i,t}\,\varphi_t \varphi_t^{\mathrm{T}}. \tag{A.2}$$

We prove that ψ_n, $\gamma_{i,n}$, and $\Gamma_{i,n}$ are almost surely vanishing as $n \to \infty$.

The almost sure convergence to zero of ψ_n follows from a component-wise application of the Kolmogorov's strong law of large numbers (Theorem 8 in Appendix D). Indeed, by using the Cauchy–Schwarz inequality as well as A4 and A5, we have ($\varphi_{t,k}$ is the kth component of φ_t)

$$\sum_{t=1}^{\infty} \frac{\mathbb{E}[\varphi_{t,k}^2 N_t^2]}{t^2} \le \sum_{t=1}^{\infty} \frac{\|\varphi_t\|^2}{t}\,\frac{\mathbb{E}[N_t^2]}{t} \le \sqrt{\sum_{t=1}^{\infty} \frac{\|\varphi_t\|^4}{t^2}}\sqrt{\sum_{t=1}^{\infty} \frac{(\mathbb{E}[N_t^2])^2}{t^2}} < \infty,$$

which shows that Kolmogorov's condition is satisfied. Therefore, $\psi_n \xrightarrow{\text{a.s.}} 0$, as $n \to \infty$. The almost sure convergence to zero of $\gamma_{i,n}$ is proven similarly since the variance of $\alpha_{i,t}\,\varphi_t N_t$ is the same as the variance of $\varphi_t N_t$ and, hence, $\gamma_{i,n} \xrightarrow{\text{a.s.}} 0$, as $n \to \infty$. The result $\Gamma_{i,n} \xrightarrow{\text{a.s.}} 0$, as $n \to \infty$, is obtained by applying the Kolmogorov's strong law of large numbers to each element of the matrix and by noting that the Kolmogorov's condition holds in view of A4 since

$$\sum_{t=1}^{\infty} \frac{\mathbb{E}[\alpha_{i,t}^2\,[\varphi_t\varphi_t^{\mathrm{T}}]_{j,k}^2]}{t^2} = \sum_{t=1}^{\infty} \frac{\varphi_{t,j}^2\varphi_{t,k}^2}{t^2} \le \sum_{t=1}^{\infty} \frac{\|\varphi_t\|^4}{t^2} < \infty.$$

Based on these convergence results, we can now make a comparison between $\|S_0(\theta)\|^2$ and $\|S_i(\theta)\|^2$, $i = 1, \ldots, m-1$. Note that

$$\begin{aligned} S_0(\theta) &= R_n^{-\frac{1}{2}}\frac{1}{n}\sum_{t=1}^{n}\varphi_t(Y_t - \varphi_t^{\mathrm{T}}\theta) \\ &= R_n^{\frac{1}{2}\mathrm{T}}\tilde{\theta} + R_n^{-\frac{1}{2}}\psi_n, \end{aligned}$$

where $\tilde{\theta} \triangleq \theta^* - \theta$ and, for $i = 1, \ldots, m-1$,

$$\begin{aligned} S_i(\theta) &= R_n^{-\frac{1}{2}}\frac{1}{n}\sum_{t=1}^{n}\alpha_{i,t}\varphi_t(Y_t - \varphi_t^{\mathrm{T}}\theta) \\ &= R_n^{-\frac{1}{2}}\Gamma_{i,n}\tilde{\theta} + R_n^{-\frac{1}{2}}\gamma_{i,n}. \end{aligned}$$

Based on the above expressions, for any $\theta \notin B_\varepsilon(\theta^*)$, i.e., for any θ such that $\|\tilde{\theta}\| > \varepsilon$, we have

$$\begin{aligned} &\|S_0(\theta)\|^2 - \|S_i(\theta)\|^2 \\ &= \tilde{\theta}^{\mathrm{T}}R_n\tilde{\theta} + \psi_n^{\mathrm{T}}R_n^{-1}\psi_n + 2\psi_n^{\mathrm{T}}\tilde{\theta} \\ &\quad - \tilde{\theta}^{\mathrm{T}}\Gamma_{i,n}^{\mathrm{T}}R_n^{-1}\Gamma_{i,n}\tilde{\theta} - \gamma_{i,n}^{\mathrm{T}}R_n^{-1}\gamma_{i,n} - 2\gamma_{i,n}^{\mathrm{T}}R_n^{-1}\Gamma_{i,n}\tilde{\theta} \\ &= \tilde{\theta}^{\mathrm{T}}\left(R_n - \Gamma_{i,n}^{\mathrm{T}}R_n^{-1}\Gamma_{i,n}\right)\tilde{\theta} + 2\left(\psi_n^{\mathrm{T}} - \gamma_{i,n}^{\mathrm{T}}R_n^{-1}\Gamma_{i,n}\right)\tilde{\theta} \\ &\quad + \left(\psi_n^{\mathrm{T}}R_n^{-1}\psi_n - \gamma_{i,n}^{\mathrm{T}}R_n^{-1}\gamma_{i,n}\right) \\ &\ge \|\tilde{\theta}\|^2\lambda_{\min}\left(R_n - \Gamma_{i,n}^{\mathrm{T}}R_n^{-1}\Gamma_{i,n}\right) \\ &\quad - 2\|\tilde{\theta}\|\cdot\|\psi_n^{\mathrm{T}} - \gamma_{i,n}^{\mathrm{T}}R_n^{-1}\Gamma_{i,n}\|\frac{\|\tilde{\theta}\|}{\varepsilon} \\ &\quad - |\psi_n^{\mathrm{T}}R_n^{-1}\psi_n - \gamma_{i,n}^{\mathrm{T}}R_n^{-1}\gamma_{i,n}| \\ &\ge \|\tilde{\theta}\|^2\Bigg[\lambda_{\min}\left(R_n - \Gamma_{i,n}^{\mathrm{T}}R_n^{-1}\Gamma_{i,n}\right) \\ &\quad - 2\frac{\|\psi_n^{\mathrm{T}} - \gamma_{i,n}^{\mathrm{T}}R_n^{-1}\Gamma_{i,n}\|}{\varepsilon}\Bigg] - |\psi_n^{\mathrm{T}}R_n^{-1}\psi_n - \gamma_{i,n}^{\mathrm{T}}R_n^{-1}\gamma_{i,n}|. \end{aligned}$$

Since ψ_n, $\gamma_{i,n}$, and $\Gamma_{i,n}$ asymptotically vanish (a.s.), and $\liminf_{n\to\infty}\lambda_{\min}(R_n) = \bar{\lambda} > 0$ (Assumption A3), we obtain that there exists (a.s.) an n_i such that, for any $\theta \notin B_\varepsilon(\theta^*)$, $\|S_0(\theta)\|^2 - \|S_i(\theta)\|^2$ becomes positive from that n_i on. Hence, by the construction of $\widehat{\Theta}_n$, we have that $\widehat{\Theta}_n \subseteq B_\varepsilon(\theta^*)$, for all $n \ge \max_{i\in\{1,\ldots,m-1\}} n_i$. □

Appendix B. Proof of Theorem 3: asymptotic shape

We first give a characterisation of an outer approximation of the SPS confidence region (cf. Eq. (B.3)). Then, we show that this outer approximation can be interpreted (as $n \to \infty$) as the set of θ's for which $n\|S_0(\theta)\|^2$ is smaller than the q_mth largest value of m independently drawn χ^2 distributed random variables (a consequence of Lemma 1), and, finally, we show that as $m \to \infty$ this set is included in a confidence ellipsoid obtained from asymptotic system identification theory.

Let $P_i(\theta) = n\cdot\|S_i(\theta)\|^2$, $i = 0, \ldots, m-1$. Hence,

$$P_0(\theta) = \sqrt{n}(\theta - \widehat{\theta}_n)^{\mathrm{T}}R_n\sqrt{n}(\theta - \widehat{\theta}_n),$$

and, for $i = 1, \ldots, m-1$,

$$\begin{aligned} P_i(\theta) &= (\theta^* - \theta)^{\mathrm{T}}\sqrt{n}\Gamma_{i,n}R_n^{-1}\sqrt{n}\Gamma_{i,n}(\theta^* - \theta) \\ &\quad + \sqrt{n}\gamma_{i,n}^{\mathrm{T}}R_n^{-1}\sqrt{n}\gamma_{i,n} + 2\sqrt{n}\gamma_{i,n}^{\mathrm{T}}R_n^{-1}\sqrt{n}\Gamma_{i,n}(\theta^* - \theta), \end{aligned}$$

where $\gamma_{i,n}$ and $\Gamma_{i,n}$ are given by (A.1) and (A.2).

Let $\bar{P}(\theta) = [P_1(\theta)\cdots P_{m-1}(\theta)]^{\mathrm{T}}$. The SPS confidence set is contained in the set of θ's for which

$$P_0(\theta) \overset{q_m}{\le} \bar{P}(\theta),$$

where $P_0(\theta) \overset{q_m}{\leq} \bar{P}(\theta)$ means that $P_0(\theta)$ is less than or equal to q_m or more of the elements in the vector on the right-hand side. $\bar{P}(\theta)$ can be written as

$$\bar{P}(\theta) = s_1(\theta) + s_2 + s_3(\theta),$$

where $s_1(\theta) = [s_{1,1}(\theta) \cdots s_{1,m-1}(\theta)]^{\mathrm{T}}$, $s_2 = [s_{2,1} \cdots s_{2,m-1}]^{\mathrm{T}}$ and $s_3(\theta) = [s_{3,1}(\theta) \cdots s_{3,m-1}(\theta)]^{\mathrm{T}}$, and, for $i = 1, \ldots, m-1$,

$$\begin{aligned} s_{1,i}(\theta) &= (\theta^* - \theta)^{\mathrm{T}} \sqrt{n}\Gamma_{i,n} R_n^{-1} \sqrt{n}\Gamma_{i,n} (\theta^* - \theta), \\ s_{2,i} &= \sqrt{n}\gamma_{i,n}^{\mathrm{T}} R_n^{-1} \sqrt{n}\gamma_{i,n}, \\ s_{3,i}(\theta) &= 2\sqrt{n}\gamma_{i,n}^{\mathrm{T}} R_n^{-1} \sqrt{n}\Gamma_{i,n} (\theta^* - \theta). \end{aligned}$$

Furthermore, let

$$\begin{aligned} \tilde{s}_{1,i} &= \sqrt{n}\Gamma_{i,n} R_n^{-1} \sqrt{n}\Gamma_{i,n}, \\ \tilde{s}_{3,i} &= 2\sqrt{n}\gamma_{i,n}^{\mathrm{T}} R_n^{-1} \sqrt{n}\Gamma_{i,n}, \end{aligned}$$

and let $\tilde{s}_1 = [\|\tilde{s}_{1,1}\| \cdots \|\tilde{s}_{1,m-1}\|]^{\mathrm{T}}$ and $\tilde{s}_3 = [\|\tilde{s}_{3,1}\| \cdots \|\tilde{s}_{3,m-1}\|]^{\mathrm{T}}$.

The confidence set can be written as

$$\begin{aligned} \widehat{\Theta}_{n,m} &= \widehat{\Theta}_{n,m} \cap \widehat{\Theta}_{n,m} \\ &= \left\{\theta : P_0(\theta) \overset{q_m}{\leq} \bar{P}(\theta) = s_1(\theta) + s_2 + s_3(\theta)\right\} \cap \widehat{\Theta}_{n,m} \\ &\subseteq \left\{\theta : P_0(\theta) \overset{q_m}{\leq} \|\theta^* - \theta\|^2 \tilde{s}_1 + s_2 + \|\theta^* - \theta\| \tilde{s}_3\right\} \cap \widehat{\Theta}_{n,m}. \quad \text{(B.1)} \end{aligned}$$

As we are taking the intersection with $\widehat{\Theta}_{n,m}$, we can restrict the considered values of θ in the first set of (B.1) to $\widehat{\Theta}_{n,m}$ thus obtaining the outer bound

$$\widehat{\Theta}_{n,m} \subseteq \left\{\theta : P_0(\theta) \overset{q_m}{\leq} \sup_{\theta \in \widehat{\Theta}_{n,m}} \|\theta^* - \theta\|^2 \tilde{s}_1 + s_2 + \sup_{\theta \in \widehat{\Theta}_{n,m}} \|\theta^* - \theta\| \tilde{s}_3\right\}.$$

Let $\widehat{\mu}_{n,m}\sigma^2$ be the value of the q_mth largest entry among the $m-1$ entries of the vector

$$\sup_{\theta \in \widehat{\Theta}_{n,m}} \|\theta^* - \theta\|^2 \tilde{s}_1 + s_2 + \sup_{\theta \in \widehat{\Theta}_{n,m}} \|\theta^* - \theta\| \tilde{s}_3 \quad \text{(B.2)}$$

Hence, $\widehat{\Theta}_{n,m}$ is included in a set characterised by

$$\widehat{\Theta}_{n,m} \subseteq \left\{\theta : P_0(\theta) \leq \widehat{\mu}_{n,m}\sigma^2\right\}. \quad \text{(B.3)}$$

or, equivalently,

$$\widehat{\Theta}_{n,m} \subseteq \left\{\theta : (\theta - \widehat{\theta}_n)^{\mathrm{T}} R_n (\theta - \widehat{\theta}_n) \leq \frac{\mu\sigma^2}{n} + \frac{(\widehat{\mu}_{n,m} - \mu)\sigma^2}{n}\right\},$$

where $F_{\chi^2}(\mu) = p$ and F_{χ^2} is the cumulative distribution function of the χ^2 distribution with d degrees of freedom. Let $\varepsilon_{n,m} = (\widehat{\mu}_{n,m} - \mu)\sigma^2$. In order to prove the theorem, we must show that $\lim_{m\to\infty} \lim_{n\to\infty} \widehat{\mu}_{n,m} = \mu$ a.s.

The next lemma characterises the convergence in distribution of (B.2) as $n \to \infty$.

Lemma 1. *For a fixed m,*

$$\sup_{\theta \in \widehat{\Theta}_{n,m}} \|\theta^* - \theta\|^2 \tilde{s}_1 + s_2 + \sup_{\theta \in \widehat{\Theta}_{n,m}} \|\theta^* - \theta\| \tilde{s}_3 \xrightarrow{d} \sigma^2 \cdot \chi^2_{m-1}$$

as $n \to \infty$, where χ^2_{m-1} is a vector of $m-1$ independent χ^2 distributed random variables with d degrees of freedom.

Proof. See Appendix C.

Based on Lemma 1, we can argue as follows to conclude the proof of Theorem 3. From Lemma 1 the expression in (B.2) (divided by σ^2) converges in distribution as $n \to \infty$ to a vector of $m-1$ independent χ^2 distributed variables. The function selecting the q_mth largest element in a vector is a continuous function, and hence by Lemma 4 $\widehat{\mu}_m \triangleq \lim_{n\to\infty} \mu_{n,m}$ has the same distribution as the q_mth largest element of $m-1$ independent χ^2 distributed random variables. We next show that $\widehat{\mu}_m$ converges a.s. to μ as $m \to \infty$, and this concludes the proof.

Given $m-1$ values $x_1, \ldots, x_{m-1}$ extracted from $m-1$ independent χ^2 distributed random variables with d degrees of freedom, consider the following empirical estimate for the cumulative χ^2 distribution function

$$\widehat{F}_m(z) = \frac{1}{m-1} \sum_{i=1}^{m-1} \mathbb{I}(x_i \leq z),$$

where $\mathbb{I}$ is the indicator function. From the Glivenko–Cantelli Theorem (Theorem 6 in Appendix D), we have

$$\sup_z |\widehat{F}_m(z) - F_{\chi^2}(z)| \to 0 \quad \text{a.s. as } m \to \infty. \quad \text{(B.4)}$$

By construction, $\widehat{F}_m(\widehat{\mu}_m) = 1 - \frac{q_m - 1}{m-1} = p_m \to p$, and $F_{\chi^2}(\mu) = p$. Since F_{χ^2} is continuous and strictly monotonically increasing, in view of (B.4) this implies that $\lim_{m\to\infty} \widehat{\mu}_m = \mu$ almost surely. □

Appendix C. Proof of Lemma 1

We first present two technical lemmas which are needed in the proof of Lemma 1.

Lemma 2.

$$\begin{bmatrix} R_n^{-\frac{1}{2}} \sqrt{n}\gamma_{1,n} \\ R_n^{-\frac{1}{2}} \sqrt{n}\gamma_{2,n} \\ \vdots \\ R_n^{-\frac{1}{2}} \sqrt{n}\gamma_{m,n} \end{bmatrix} \xrightarrow{d} \mathcal{N}(0, \sigma^2 I_{md}),$$

where $\mathcal{N}$ denotes the normal distribution.

Proof. We only prove the result for $m = 2$. The case $m > 2$ follows with obvious modifications. The main tools in the proof are the Cramer–Wold Theorem (Theorem 4 in Appendix D) and the Central limit theorem (Theorem 7 in Appendix D) using the Lyapunov condition (D.1).

We first show that, for any $2d$-vector $[a_1^{\mathrm{T}}\ a_2^{\mathrm{T}}] \neq 0$,

$$[a_1^{\mathrm{T}}\ a_2^{\mathrm{T}}] \begin{bmatrix} \sqrt{n}R_n^{-\frac{1}{2}} \gamma_{1,n} \\ \sqrt{n}R_n^{-\frac{1}{2}} \gamma_{2,n} \end{bmatrix} \xrightarrow{d} \mathcal{N}(0, (a_1^{\mathrm{T}} a_1 + a_2^{\mathrm{T}} a_2)\sigma^2).$$

Note that

$$[a_1^{\mathrm{T}}\ a_2^{\mathrm{T}}] \begin{bmatrix} \sqrt{n}R_n^{-\frac{1}{2}} \gamma_{1,n} \\ \sqrt{n}R_n^{-\frac{1}{2}} \gamma_{2,n} \end{bmatrix} = [a_1^{\mathrm{T}}\ a_2^{\mathrm{T}}] \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \begin{bmatrix} \alpha_{1,t} R_n^{-\frac{1}{2}} \varphi_t N_t \\ \alpha_{2,t} R_n^{-\frac{1}{2}} \varphi_t N_t \end{bmatrix},$$

and let $\xi_t = [a_1^{\mathrm{T}}\ a_2^{\mathrm{T}}] \begin{bmatrix} \alpha_{1,t} R_n^{-\frac{1}{2}} \varphi_t N_t \\ \alpha_{2,t} R_n^{-\frac{1}{2}} \varphi_t N_t \end{bmatrix}$. We have $\mathbb{E}[\xi_t] = 0$ and

$$\begin{aligned} D_n^2 &= \sum_{t=1}^{n} \mathbb{E}[\xi_t^2] \\ &= \sum_{t=1}^{n} \mathbb{E}\left[\left(a_1^{\mathrm{T}} R_n^{-\frac{1}{2}} \varphi_t \alpha_{1,t} + a_2^{\mathrm{T}} R_n^{-\frac{1}{2}} \varphi_t \alpha_{2,t}\right)^2\right] \mathbb{E}[N_t^2] \end{aligned}$$

$$= \sum_{t=1}^{n} \left(\left(a_1^{\mathrm{T}} R_n^{-\frac{1}{2}} \varphi_t \right)^2 + \left(a_2^{\mathrm{T}} R_n^{-\frac{1}{2}} \varphi_t \right)^2 \right) \sigma^2$$
$$= n(a_1^{\mathrm{T}} a_1 + a_2^{\mathrm{T}} a_2)\sigma^2, \tag{C.1}$$

and

$$\sum_{t=1}^{n} \mathbb{E}[\xi_t^4] = \sum_{t=1}^{n} \mathbb{E}\left[\left(a_1^{\mathrm{T}} R_n^{-\frac{1}{2}} \varphi_t \alpha_{1,t} + a_2^{\mathrm{T}} R_n^{-\frac{1}{2}} \varphi_t \alpha_{2,t} \right)^4 \right] \mathbb{E}[N_t^4]$$
$$= \sum_{t=1}^{n} \left(\left(a_1^{\mathrm{T}} R_n^{-\frac{1}{2}} \varphi_t \right)^4 + 6 \left(a_1^{\mathrm{T}} R_n^{-\frac{1}{2}} \varphi_t \right)^2 \left(a_2^{\mathrm{T}} R_n^{-\frac{1}{2}} \varphi_t \right)^2 + \left(a_2^{\mathrm{T}} R_n^{-\frac{1}{2}} \varphi_t \right)^4 \right) \rho = o(n^2),$$

that is, the last term multiplied by $1/n^2$ tends to zero, a fact due to Assumption A6. Using (C.1), the Lyapunov condition (D.1) with $\delta = 2$ holds. Hence,

$$\frac{\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left(a_1^{\mathrm{T}} R_n^{-\frac{1}{2}} \varphi_t \alpha_{1,t} N_t + a_2^{\mathrm{T}} R_n^{-\frac{1}{2}} \varphi_t \alpha_{2,t} N_t \right)}{\sigma \sqrt{a_1^{\mathrm{T}} a_1 + a_2^{\mathrm{T}} a_2}} \xrightarrow{d} \mathcal{N}(0, 1),$$

assuming a_1 and a_2 are not simultaneously null, and so

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} (a_1^{\mathrm{T}} R_n^{-\frac{1}{2}} \varphi_t \alpha_{1,t} N_t + a_2^{\mathrm{T}} R_n^{-\frac{1}{2}} \varphi_t \alpha_{2,t} N_t)$$
$$\xrightarrow{d} \mathcal{N}(0, \sigma^2 (a_1^{\mathrm{T}} a_1 + a_2^{\mathrm{T}} a_2)).$$

Now, from the Cramer–Wold theorem (Theorem 4 in Appendix D), it follows that

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \begin{bmatrix} \alpha_{1,t} R_n^{-\frac{1}{2}} \varphi_t N_t \\ \alpha_{2,t} R_n^{-\frac{1}{2}} \varphi_t N_t \end{bmatrix} \xrightarrow{d} \mathcal{N}\left(0, \sigma^2 \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix} \right),$$

from which the lemma immediately follows. □

Lemma 3. *For a fixed m, each component of the terms $\sup_{\theta \in \widehat{\Theta}_{n,m}} \|\theta^* - \theta\|^2 \tilde{s}_1$ and $\sup_{\theta \in \widehat{\Theta}_{n,m}} \|\theta^* - \theta\| \tilde{s}_3$ converges to zero in probability as $n \to \infty$.*

Proof. We consider $\sup_{\theta \in \widehat{\Theta}_{n,m}} \|\theta^* - \theta\|^2 \tilde{s}_1$ first. We need to show that

$$\mathbb{P}\{ \sup_{\theta \in \widehat{\Theta}_{n,m}} \|\theta^* - \theta\|^2 \cdot \|\tilde{s}_{1,i}\| > \epsilon \} \to 0 \quad \text{as } n \to \infty$$

for every $\epsilon > 0$. Let $\beta_n = \sup_{\theta \in \widehat{\Theta}_{n,m}} \|\theta^* - \theta\|^2$. Since

$$\|\tilde{s}_{1,i}\| \le \left\| \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \alpha_{i,t} \varphi_t \varphi_t^{\mathrm{T}} \right\| \cdot \|R_n^{-1}\| \cdot \left\| \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \alpha_{i,t} \varphi_t \varphi_t^{\mathrm{T}} \right\|,$$

the result follows if

$$\mathbb{P}\left\{ \beta_n^{1/3} \cdot \left\| \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \alpha_{i,t} \varphi_t \varphi_t^{\mathrm{T}} \right\| > \epsilon^{1/3} \right\} \to 0, \tag{C.2}$$

and

$$\mathbb{P}\{ \beta_n^{1/3} \cdot \left\| R_n^{-1} \right\| > \epsilon^{1/3} \} \to 0, \tag{C.3}$$

as $n \to \infty$. (C.3) follows from Theorem 2 and Assumption A3. Next we show (C.2). From Chebyshev's inequality we have

$$\mathbb{P}\left\{ \left\| \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \alpha_{i,t} \varphi_t \varphi_t^{\mathrm{T}} \right\| > K \right\} \le \frac{\mathbb{E}\left[\left\| \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \alpha_{i,t} \varphi_t \varphi_t^{\mathrm{T}} \right\|^2 \right]}{K^2}.$$

On the other hand,

$$\mathbb{E}\left[\left\| \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \alpha_{i,t} \varphi_t \varphi_t^{\mathrm{T}} \right\|^2 \right]$$
$$\le \operatorname{trace} \mathbb{E}\left[\left(\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \alpha_{i,t} \varphi_t \varphi_t^{\mathrm{T}} \right) \left(\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \alpha_{i,t} \varphi_t \varphi_t^{\mathrm{T}} \right) \right]$$
$$= \operatorname{trace} \left(\frac{1}{n} \sum_{t=1}^{n} \varphi_t \varphi_t^{\mathrm{T}} \varphi_t \varphi_t^{\mathrm{T}} \right) = \frac{1}{n} \sum_{t=1}^{n} \|\varphi_t\|^4,$$

which is bounded by a constant C in view of Assumption A6. Hence, $\mathbb{P}\{\| \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \alpha_{i,t} \varphi_t \varphi_t^{\mathrm{T}} \| > K\} \le C/K^2$, $\forall n$, which is an arbitrarily small number provided K is large enough. (C.2) now easily follows from Theorem 2 since it implies that $\mathbb{P}\{\beta_n^{1/3} > \epsilon^{1/3}/K\} \to 0$ as $n \to 0$.

We next investigate the term $\sup_{\theta \in \widehat{\Theta}_{n,m}} \|\theta^* - \theta\| \tilde{s}_{3,i}$. We have $\|s_{3,i}\| = \|2 \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \alpha_{i,t} \varphi_t \varphi_t^{\mathrm{T}} R_n^{-1} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \alpha_{i,t} \varphi_t N_t\|$. The result follows provided that

$$\mathbb{P}\left\{ \beta_n^{1/6} \cdot \left\| \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \alpha_{i,t} \varphi_t \varphi_t^{\mathrm{T}} \right\| > \epsilon^{1/3} \right\} \to 0, \tag{C.4}$$

$$\mathbb{P}\{ \beta_n^{1/6} \cdot \|R_n^{-1}\| > \epsilon^{1/3} \} \to 0, \tag{C.5}$$

and

$$\mathbb{P}\left\{ \beta_n^{1/6} \cdot \left\| \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \alpha_{i,t} \varphi_t N_t \right\| > \epsilon^{1/3} \right\} \to 0, \tag{C.6}$$

as $n \to \infty$. Results (C.4) and (C.5) are essentially the same as (C.2) and (C.3). Result (C.6) can be established along the same lines as (C.2) by noting that

$$\mathbb{E}\left[\left\| \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \alpha_{i,t} \varphi_t N_t \right\|^2 \right] = \frac{1}{n} \sum_{t=1}^{n} \|\varphi_t\|^2 \sigma^2,$$

which is bounded by Assumption A6. □

Proof of Lemma 1. By Lemmas 2 and 4 $\frac{1}{\sigma^2} s_2$ converges in distribution to a vector of independent χ^2 distributed random variables with d degrees of freedom. Lemma 1 now follows from Slutsky's Theorem (see Appendix D) since $\sup_{\theta \in \widehat{\Theta}_{n,m}} \|\theta^* - \theta\|^2 \tilde{s}_1$ and $\sup_{\theta \in \widehat{\Theta}_{n,m}} \|\theta^* - \theta\| \tilde{s}_3$ converge to zero in probability by Lemma 3. □

Appendix D. Main theoretical tools of the proofs

Let X_n and X be random vectors in $\mathbb{R}^s$, and let $\xrightarrow{d}$ denote convergence in distribution. The following results can be found in, e.g., van der Vaart (1998) or Shiryaev (1995).

Theorem 4 (*Cramer–Wold Theorem*). *$X_n \xrightarrow{d} X$ if and only if $a^{\mathrm{T}} X_n \xrightarrow{d} a^{\mathrm{T}} X$ $\forall a \in \mathbb{R}^s$.*

Lemma 4. *Let f be a continuous function from $\mathbb{R}^s$ to $\mathbb{R}^l$. If $X_n \xrightarrow{d} X$, then $f(X_n) \xrightarrow{d} f(X)$.*

The next theorem follows from Lemma 4.

Theorem 5 (*Slutsky's Theorem*). *Let f be a continuous function from $\mathbb{R}^{s+k}$ to $\mathbb{R}^l$. If $X_n \xrightarrow{d} X$ and $Y_n = [Y_{n,1} \dots Y_{n,k}]^{\mathrm{T}}$ converges in probability to a constant vector $c = [c_1 \dots c_k]^{\mathrm{T}}$, then $f(X_n, Y_n) \xrightarrow{d} f(X, c)$.*

Theorem 6 (*Glivenko–Cantelli Theorem*). *Let $x_1, \ldots, x_n$ be i.i.d. random variables with cumulative distribution function $F(z) = \mathbb{P}\{x_1 \leq z\}$. Let $F_n(z)$ be the empirical estimate of $F(z)$: $F_n(z) = \frac{1}{n}\sum_{t=1}^{n} \mathbb{I}(x_t \leq z)$, where $\mathbb{I}$ is the indicator function. Then,*

$$\lim_{n\to\infty} \sup_{z\in R} |F(z) - F_n(z)| = 0 \quad a.s..$$

Theorem 7 (*Central Limit Theorem*). *Let $\xi_1, \xi_2, \ldots$ be independent random variables with finite second moments. Let $m_t = \mathbb{E}[\xi_t]$, $\sigma_t^2 = \mathbb{E}[(\xi_t - m_t)^2] > 0$, $S_n = \sum_{t=1}^{n} \xi_t$, $D_n^2 = \sum_{t=1}^{n} \sigma_t^2$ and let $F_t(x)$ be the cumulative distribution function of ξ_t. If, for every $\epsilon > 0$, the following Lyapunov condition is satisfied for a $\delta > 0$,*

$$\frac{1}{D_n^{2+\delta}} \sum_{t=1}^{n} \mathbb{E}[|\xi_t - m_t|^{2+\delta}] \to 0, \quad \text{as } n \to \infty, \tag{D.1}$$

then

$$\frac{S_n - \mathbb{E}[S_n]}{D_n} \xrightarrow{d} G(0, 1).$$

Theorem 8 (*Strong Law of Large Numbers*). *Let $\xi_1, \xi_2, \ldots$ be a sequence of independent random variables with finite second moments, and let $S_n = \sum_{t=1}^{n} \xi_t$. Assume that*

$$\sum_{t=1}^{\infty} \frac{\mathbb{E}[(\xi_t - \mathbb{E}[\xi_t])^2]}{t^2} < \infty,$$

then

$$\lim_{n\to\infty} \frac{S_n - \mathbb{E}[S_n]}{n} = 0. \quad \text{(a.s.)}$$

References

Campi, M. C., Ko, S., & Weyer, E. (2009). Non-asymptotic confidence regions for model parameters in the presence of unmodelled dynamics. *Automatica, 45*, 2175–2186.

Campi, M. C., & Weyer, E. (2005). Guaranteed non-asymptotic confidence regions in system identification. *Automatica, 41*, 1751–1764.

Campi, M. C., & Weyer, E. (2010). Non-asymptotic confidence sets for the parameters of linear transfer functions. *IEEE Transactions on Automatic Control, 55*, 2708–2720.

Csáji, B.Cs., Campi, M.C., & Weyer, E. (2012). Non-asymptotic confidence regions for the least-squares estimate. In *Proceedings of the 16th IFAC symposium on system identification* (pp. 227–232).

Csáji, B.Cs., Campi, M.C., & Weyer, E. (2014). Strong consistency of the Sign-Perturbed Sums method. In *Proceedings of the 53st IEEE conference on decision and control* (pp. 3352–3357).

Csáji, B. Cs., Campi, M. C., & Weyer, E. (2015). 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*, 169–181.

Dabbene, F., Sznaier, M., & Tempo, R. (2014). Probabilistic optimal estimation with uniformly distributed noise. *IEEE Transactions on Automatic Control, 59*, 2113–2127.

Dalai, M., Weyer, E., & Campi, M. C. (2007). Parameter identification for non-linear systems: guaranteed confidence regions through LSCR. *Automatica, 43*, 1418–1425.

den Dekker, A.J., Bombois, X., & Van den Hof, P.M.J. (2008). Finite sample confidence regions for parameters in prediction error identification using output error. In *IFAC world congress* (pp. 5024–5029).

Gentle, J. E. (2013). *Theory of statistics*. George Mason University.

Granichin, O. N. (2012). The nonasymptotic confidence set for parameters of a linear control object under an arbitrary external disturbance. *Automation and Remote Control, 73*(1), 20–30.

Kailath, T., Sayed, A. H., & Hassibi, B. (2000). *Linear estimation*. Prentice Hall.

Kieffer, M., & Walter, E. (2013a). Guaranteed characterization of exact non-asymptotic confidence regions as defined by LSCR and SPS. *Automatica, 49*, 507–512.

Kieffer, M., & Walter, E. (2013b). Guaranteed characterization of exact non-asymptotic confidence regions in nonlinear parameter estimation. In *9th IFAC symposium on nonlinear control systems* (pp. 56–61).

Kolumbán, S., Vajk, I., & Schoukens, J. (2015). Perturbed datasets methods for hypothesis testing and structure of corresponding confidence sets. *Automatica, 51*, 326–331.

Ljung, L. (1999). *System identification: theory for the user* (2nd ed.). Upper Saddle River: Prentice-Hall.

Schoukens, J., Rolain, Y., Vandersteen, G., & Pintelon, R. (2013). Study of small data set efficiency losses in system identification: The FIR case. In *Proceedings of the 11th IFAC international workshop on adaptation and learning in control and signal processing* (pp. 68–73).

Shiryaev, A. N. (1995). *Probability* (2nd ed.). Springer.

van der Vaart, A. W. (1998). *Asymptotic statistics*. Cambridge University Press.

Erik Weyer is a professor in the Department of Electrical and Electronic Engineering at the University of Melbourne. He received the Siv. Ing. degree in 1988 and the Ph.D. in 1993, both from the Norwegian Institute of Technology, Trondheim, Norway. From 1994 to 1996 he was a Research Fellow at the University of Queensland, and since 1997 he has been with the Department of Electrical and Electronic Engineering at the University of Melbourne. He has held visiting positions at the University of Brescia, Italy, the Technical University of Vienna, Austria, and Politecnico di Milano, Italy. From 2010 to 2012 he was an associate editor of IEEE Transactions of Automatic Control, and he is currently an associate editor of Automatica. His research interests are in the areas of system identification and control, with particular emphasis on finite sample properties of system identification methods, and modelling and control of irrigation channels and rivers. He was a co-recipient of the IEEE CSS Control System Technology Award in 2014.

Marco Claudio Campi is professor of Automatic Control at the University of Brescia, Italy. He is the chair of the Technical Committee IFAC on Modelling, Identification and Signal Processing (MISP) and has been in various capacities on the Editorial Board of Automatica, Systems and Control Letters and the European Journal of Control. Marco Campi is a recipient of the "Giorgio Quazza" prize, and, in 2008, he received the IEEE CSS George S. Axelby outstanding paper award for the article "The Scenario Approach to Robust Control Design". He has delivered plenary and semi-plenary addresses at major conferences including SYSID, MTNS, and CDC. Currently he is a distinguished lecturer of the Control Systems Society. Marco Campi is a Fellow of IEEE, a member of IFAC, and a member of SIDRA.

Balázs Csanád Csáji is a senior research fellow at MTA SZTAKI, the Institute for Computer Science and Control of the Hungarian Academy of Sciences, Budapest, Hungary. He defended his Ph.D. in Computer Science (2008) at the Eötvös Loránd University (ELTE), Budapest, Hungary. Previously, he received Master's degrees in Computer Science combined with Mathematics (2001) as well as in Philosophy (2006), also from ELTE. During his studies he spent semesters and internships at the Eindhoven University of Technology, Netherlands (2001), British Telecom, UK (2002), and Johannes Kepler University, Austria (2003). He was a Postdoctoral Researcher at the Université catholique de Louvain, Belgium (2008–2009), and a Research Fellow at the University of Melbourne, Australia (2009–2012). His main research interests revolve around stochastic models and related statistical problems, especially in the fields of machine learning and system identification.