A simple condition for the boundedness of Sign-Perturbed-Sums (SPS) confidence regions ⋆

Algo Carè

Department of Information Engineering, University of Brescia, Italy

Abstract

Sign-Perturbed-Sums (SPS) is a system identification algorithm that, under mild assumptions on the distribution of the noise, constructs confidence regions with finite-sample validity and a user-specified confidence level. For linear regression models, SPS regions are well-shaped in a precise meaning, but it is still possible (though rare in practice) that they are unbounded. In this communication, we provide a reformulation of a technical condition for the boundedness of the SPS regions in terms of a more practical excitation condition. We briefly argue that the simple condition here proposed provides insight to tune the SPS parameters, and even to design refined algorithms that can be guaranteed to deliver bounded regions.

Key words: Identification methods; system identification; estimation theory.

1 Introduction

Consider a system in linear regression form

$$Y_t \;=\; \varphi_t^\top \theta^* + N_t,$$

where t is the discrete time index, Y_t is the (scalar) output, N_t the (scalar) noise, φ_t is a measured d-dimensional input regressor and θ^* is an unknown d-dimensional parameter vector that we want to estimate based on a finite set of n input-output observations $(\varphi_1, Y_1), (\varphi_2, Y_2), \ldots, (\varphi_n, Y_n)$.

The noise sequence $\{N_t\}$ is made up of random variables that are *independent* (not necessarily identically distributed) and *symmetric about zero*. We treat the measured input sequence $\{\varphi_t\}$ as *deterministic*, although the generalisation to the case of random inputs is immediate as soon as $\{\varphi_t\}$ and $\{N_t\}$ are independent.

The least squares estimate (LSE) $\hat{\theta}_n$ is the value of θ that minimises the sum of the squared prediction errors $\{Y_t - \varphi_t^\top \theta\}$, i.e.,

$$\hat{\theta}_n \;\triangleq\; \operatorname*{arg\,min}_{\theta \in \mathbb{R}^d} \sum_{t=1}^{n} (Y_t - \varphi_t^\top \theta)^2. \qquad (1)$$

The LSE, (1), satisfies the normal equation

$$\sum_{t=1}^{n} \varphi_t (Y_t - \varphi_t^\top \theta) = 0,$$

whose solution is unique provided that

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

is invertible. The invertibility of R_n is assumed throughout, so the LSE can always be written as

$$\hat{\theta}_n \;=\; \left(\sum_{t=1}^{n} \varphi_t \varphi_t^\top \right)^{-1} \left(\sum_{t=1}^{n} \varphi_t Y_t \right).$$

The Sign-Perturbed-Sums (SPS) algorithm [6] constructs a confidence region $\widehat{\Theta}_n \subseteq \mathbb{R}^d$ around $\hat{\theta}_n$, with a guaranteed probability of including θ^*.

The construction of the SPS region $\widehat{\Theta}_n$ depends on two integer parameters, q and m, $1 \le q < m$, that are preliminarily set by the user. Sometimes, when it is important

⋆ This paper was not presented at any IFAC meeting.
Email address: `algo.care@unibs.it` (Algo Carè).

to recall the dependence of $\widehat{\Theta}_n$ on q and m, we will write more explicitly $\widehat{\Theta}_{n;(q,m)}$ instead of $\widehat{\Theta}_n$. The SPS region $\widehat{\Theta}_n$ is the set of the candidate parameters θ that satisfy a simple inclusion condition: the values of m functions $v_0(\theta), \dots, v_{m-1}(\theta)$ (which we will define below) are compared and the rank of $v_0(\theta)$ is computed (we say that the rank is one if $v_0(\theta)$ is the smallest value among those in the vector $[v_0(\theta), \dots, v_{m-1}(\theta)]$; two, if it is the second smallest, etc.; in the case of ties, a suitable tie-break rule is introduced, see [6] for details); if the rank of $v_0(\theta)$ is larger than $m - q$, then $\theta \notin \widehat{\Theta}_n$; $\theta \in \widehat{\Theta}_n$ otherwise.

The SPS region $\widehat{\Theta}_{n;(q,m)}$ is guaranteed to include θ^* with probability $1 - \frac{q}{m}$ because the functions $v_0, v_1, \dots, v_{m-1}$ are defined in such a way that the rank of $v_0(\theta^*)$ has uniform distribution over $\{1, 2, \dots, m\}$. More precisely, the "reference" function v_0 is defined as $v_0(\theta) = \|S_0(\theta)\|$ $(= \sqrt{S_0(\theta)^\top S_0(\theta)}\,)$, where

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

($R_n^{-\frac{1}{2}}$ is the inverse of the principal square root of R_n). The functions $v_1, \dots, v_{m-1}$ are sign-perturbed versions of this reference function, namely, $v_i(\theta) = \|S_i(\theta)\|$, where

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

and $\{\alpha_{i,t}\}$ are independent symmetric Bernoulli variables, such that $\mathbb{P}(\alpha_{i,t} = +1) = \mathbb{P}(\alpha_{i,t} = -1) = 0.5$ for all $i = 1, \dots, m-1$ and $t = 1, \dots, n$.

The interested reader is referred to [6,9,12,2] for more details on SPS and its theoretical and computational aspects, and to [3] for an overview of related methods.

1.1 Problem statement and contribution of this study

This study moves from the premise that it is often desirable that $\widehat{\Theta}_n$ be bounded. For example, when $\widehat{\Theta}_n$ is used as an uncertainty set in a robust design context, boundedness enables one to obtain structured approximations of $\widehat{\Theta}_n$ (see, e.g., Section VI.A of [6]) and to apply standard robust optimization techniques, [1]. Boundedness is also precious in uncertainty quantification, where a bounded $\widehat{\Theta}_n$ allows one to simulate different uncertainty scenarios by generating samples according to a uniform distribution over $\widehat{\Theta}_n$ (see, e.g., Section 4.1. in [4]).

The shape of the SPS region is known to satisfy an asymptotic optimality property (Theorem 3 in [12]) and, for finite n, some favourable topological properties were demonstrated in [6]. Nonetheless, the finite sample analysis of [6] leaves open the possibility that unbounded regions could be constructed. Indeed, one can find situations where SPS constructs unbounded regions under the standard condition that the regressors $\varphi_1, \dots, \varphi_n$ are "exciting" in the sense that their linear span is the whole space $\mathbb{R}^d$; furthermore, there exist situations where unbounded regions appear even under the condition that the regressors are "completely exciting" in the sense that *any* d regressors span the whole space $\mathbb{R}^d$. For example, consider having $n = 10$ completely exciting regressors of size $d = 3$: when the SPS algorithm is run with parameters $q = 1$ and $m = 5$, it typically builds an unbounded region in more than one third of the simulation runs (in Section 4.1, we will see that this behaviour is really "typical" as it does not depend on the specific regressors and on the specific distribution of the noise, provided that it has density). On the other hand, experience shows that, with the same $n = 10$ data and the same confidence level $1 - \frac{q}{m}$, increasing the values of q reduces the probability of constructing an unbounded region. The present study aims at throwing some light on these and similar phenomena by formulating a simple condition for boundedness. As a preview of the implications of the results here obtained, we anticipate that, with $n = 10$ completely exciting regressors of size $d = 3$, the probability of occurrence of an unbounded region is necessarily below $3 \cdot 10^{-9}$ if $(q, m) = (100, 500)$. With $(q, m) = (1, 5)$, on the other hand, one needs at least $n = 40$ completely exciting regressors of size $d = 3$ to reduce the probability of building an unbounded region below 10^{-8}.

In the following section, we recall some important facts that are necessary to understand our main result, which is Theorem 1 in Section 3. Some of the implications of Theorem 1 will be discussed in Section 4.

2 Technical Preliminaries

From the analysis in [6], Section VI and Appendix B, it is clear that the shape of $\widehat{\Theta}_n$ is affected by the Hessians of the functions $\frac{1}{2}(\|S_0(\theta)\|^2 - \|S_i(\theta)\|^2)$, $i = 1, \dots, m-1$, which are equal to

$$K_i \triangleq R_n - Q_i R_n^{-1} Q_i, \; i = 1, \dots, m-1,$$

where Q_i is a sign-perturbed version of R_n, namely

$$Q_i \triangleq \frac{1}{n} \sum_{t=1}^{n} \alpha_{i,t} \varphi_t \varphi_t^\top.$$

Lemma 4 of [6] ensures that $K_i \succeq 0$.[1] Although not stated explicitly in [6], Fact 1 below is immediately implied by the analysis in Appendix B of [6] (a similar fact

[1] The symbols $\succ$ and $\succeq$ refers to the Löwner partial order, i.e., $M \succeq 0$ ($M \succ 0$) means that the matrix M is positive semidefinite (definite).

was stated explicitly for a variant of the SPS algorithm, see Definition 3 and Theorem 4 in [10]).

Fact 1. *Let the SPS parameters q and m, $1 \le q < m$, be arbitrarily chosen. Suppose that $K_i \succ 0$ for at least $m-q$ values of $i \in \{1,\dots,m-1\}$. Then $\widehat{\Theta}_{n;(q,m)}$ is bounded.*

Thus, Fact 1 provides a natural test for the boundedness of $\widehat{\Theta}_n$. When $q=1$, a useful (converse) statement is also valid.

Fact 2. *Let the SPS parameter q be equal to 1, and let $m > 1$ be arbitrarily chosen. Suppose that there exists at least one value of $i \in \{1,\dots,m-1\}$ for which $S_i(\hat{\theta}_n) \neq 0$ and K_i is* not *positive definite. Then $\widehat{\Theta}_{n;(1,m)}$ is unbounded.*

It is remarkable that the conditions $K_i \succ 0$, $i = 1,\dots,m-1$, do not depend on the noise N_t. Hence, drawing conclusions based on $K_i \succ 0$ is a very robust way of proceeding. The main contribution of this paper is formulating the condition $K_i \succ 0$ in terms of a more usual regressor excitation condition.

3 A simple equivalent condition for $K_i \succ 0$

In this section, the index $i \in \{1,\dots,m-1\}$ is kept fixed. Let us consider the time indices corresponding to the occurrence of a sign perturbation in (2): $I_p \triangleq \{t : \alpha_{i,t} = -1\}$. The set of the remaining time indices is denoted by I_u, i.e., $I_u \triangleq \{t : \alpha_{i,t} = +1\}$. We call *perturbed regressors* the regressors indexed by I_p (i.e., those that undergo a sign change in (2)), and we call *unperturbed regressors* the remaining ones, with indices in I_u.

We say that the regressors φ_t indexed by a set $I \subseteq \{1,\dots,n\}$ are *exciting* if $\sum_{t\in I}\varphi_t\varphi_t^\top$ is invertible (or, equivalently, the linear span of these regressors is $\mathbb{R}^d$).

We prove the following theorem.

Theorem 1. $K_i \succ 0$ if and only if both the perturbed regressors and the unperturbed regressors are exciting.

Proof. Note that, by definition of I_p and I_u, $\sum_{t\in I_p}\varphi_t\varphi_t^\top = \sum_{t=1}^n \frac{(1-\alpha_{i,t})}{2}\varphi_t\varphi_t^\top$ and $\sum_{t\in I_u}\varphi_t\varphi_t^\top = \sum_{t=1}^n \frac{(1+\alpha_{i,t})}{2}\varphi_t\varphi_t^\top$. Hence, the perturbed regressors and the unperturbed regressors are both exciting if and only if the matrices $\frac{1}{n}\sum_{t=1}^n(1-\alpha_{i,t})\varphi_t\varphi_t^\top$ and $\frac{1}{n}\sum_{t=1}^n(1+\alpha_{i,t})\varphi_t\varphi_t^\top$ are invertible. By definition, $\frac{1}{n}\sum_{t=1}^n(1-\alpha_{i,t})\varphi_t\varphi_t^\top = R_n - Q_i$, and $\frac{1}{n}\sum_{t=1}^n(1+\alpha_{i,t})\varphi_t\varphi_t^\top = R_n + Q_i$. Since $\varphi_t\varphi_t^\top \succeq 0$, it is true that

$$R_n - Q_i \succeq 0 \tag{3}$$

and

$$R_n + Q_i \succeq 0. \tag{4}$$

Therefore, the invertibility of both $\frac{1}{n}\sum_{t=1}^n(1-\alpha_{i,t})\varphi_t\varphi_t^\top$ and $\frac{1}{n}\sum_{t=1}^n(1+\alpha_{i,t})\varphi_t\varphi_t^\top$ is equivalent to positive definiteness ($\succ$) in both (3) and (4). Letting $R_n^{\frac{1}{2}}$ be the principal square root matrix of R_n, we define

$$A \triangleq R_n^{-\frac{1}{2}}(R_n - Q_i)R_n^{-\frac{1}{2}}$$

and

$$B \triangleq R_n^{-\frac{1}{2}}(R_n + Q_i)R_n^{-\frac{1}{2}}.$$

Since $R_n^{-\frac{1}{2}} \succ 0$, the positive definiteness ($\succ$) in both (3) and (4) is equivalent to the condition "$A \succ 0$ and $B \succ 0$". Since A and B are symmetric and commute (i.e., $AB = BA$), they are simultaneously diagonalisable (see, e.g., Theorem 4.1.6. in [8]). Simultaneous diagonalisation reveals that "$A \succ 0$ and $B \succ 0$" if and only if $AB \succ 0$. To sum up, the perturbed regressors and the unperturbed regressors are both exciting if and only if $AB \succ 0$. Finally, the following chain of equalities reveals that $AB \succ 0$ if and only if K_i is positive definite:

$$\begin{aligned}
AB &= I - R_n^{-\frac{1}{2}}Q_iR_n^{-1}Q_iR_n^{-\frac{1}{2}} \\
&= R_n^{-\frac{1}{2}}\left(R_n - Q_iR_n^{-1}Q_i\right)R_n^{-\frac{1}{2}} \\
&= R_n^{-\frac{1}{2}}K_iR_n^{-\frac{1}{2}}.
\end{aligned}$$

□

4 Some implications

We first show how the condition provided by Theorem 1 can help tune the parameters q and m of the SPS algorithm. Then, we briefly discuss how this condition can guide the user in designing SPS-based algorithms that are guaranteed to construct bounded regions. For the sake of brevity, in what follows we work under the simplifying but meaningful assumption that any d regressors are exciting (i.e., the linear span of any d regressors is the whole $\mathbb{R}^d$ space).

4.1 On tuning the SPS parameters

Thanks to Theorem 1, the following bound is easily obtained:

$$\mathbb{P}\{\text{unbounded } \widehat{\Theta}_n\} \le \sum_{j=0}^{m-q-1}\binom{m-1}{j}p_{n,d}^j(1-p_{n,d})^{m-1-j}, \tag{5}$$

where

$$p_{n,d} \triangleq \sum_{j=d}^{n-d}\binom{n}{j}\left(\frac{1}{2}\right)^j\left(1-\frac{1}{2}\right)^{n-j}$$

is the probability that, for a given i, the condition $K_i \succ 0$ is satisfied. [2]

Proof. We first show that $\mathbb{P}\{K_i \succ 0\} = p_{n,d}$. By Theorem 1, $\mathbb{P}\{K_i \succ 0\} = \mathbb{P}\{|I_p| \geq d \text{ and } |I_u| \geq d\}$ ($|\cdot|$ denotes cardinality). By construction, the event that a given regressor φ_t is a perturbed regressor is an independent Bernoulli trial with probability of success equal to $\mathbb{P}\{\alpha_{i,t} = -1\} = \frac{1}{2}$. Since $|I_p| + |I_u| = n$, we have that $\mathbb{P}(|I_p| \geq d \text{ and } |I_u| \geq d) = \mathbb{P}(d \leq |I_p| \leq n-d) = p_{n,d}$. Then, Fact 1 ensures that $\mathbb{P}\{\text{unbounded } \widehat{\Theta}_n\} \leq \mathbb{P}\{K_i \succ 0 \text{ for less than } m-q \text{ values of } i\}$, and (5) follows by observing that, for each $i \in \{1, \ldots, m-1\}$, the outcome $K_i \succ 0$ is a Bernoulli trial with probability of success equal to $p_{n,d}$, independent of the other $m-2$ outcomes. □

Inequality (5) can help choose the SPS parameters q and m in order to favour the construction of bounded regions. For instance, the numerical evaluations in the example provided in Section 1.1 are obtained by directly substituting the values of n, d, m and q into (5).

In concluding, we also remark that Theorem 1 justifies the claim in Section 1.1 that, in a typical simulation with $n = 10$, $d = 3$, $m = 5$ and $q = 1$, an unbounded region is constructed in *more* than 1/3 of the cases. To see this, note that, if the random variables N_t are distributed with density, then, with probability one, the LSE $\hat{\theta}_n$ (which satisfies $S_0(\hat{\theta}_n) = 0$ by definition) satisfies $S_i(\hat{\theta}_n) = 0$ for $i \neq 0$ if and only if $\alpha_{i,1} = \ldots = \alpha_{i,n}$. Thus, using Theorem 1, we have that, for every $i \neq 0$, $\mathbb{P}\{S_i(\hat{\theta}_n) = 0 \text{ or } K_i \succ 0\} = \mathbb{P}\{|I_p| = 0 \text{ or } |I_p| = n \text{ or } d \leq |I_p| \leq n-d\} = 2^{1-n} + p_{n,d}$, and Fact 2 ensures that $\mathbb{P}\{\text{bounded } \widehat{\Theta}_n\} \leq \mathbb{P}\{\forall i \in \{1, \ldots, m-1\}\ S_i(\hat{\theta}_n) = 0 \text{ or } K_i \succ 0\} = (2^{1-n} + p_{n,d})^{m-1} \approx 0.63$.

4.2 On designing refined SPS algorithms

Let us consider the following simple idea: if we could force any i-th sign string $\alpha_{i,1}, \ldots, \alpha_{i,n}$ to satisfy the condition $d \leq |I_p| \leq n-d$, then we would have $K_i \succ 0$ for all i (by Theorem 1), and the SPS region would be certainly bounded (by Fact 1). In what follows, we consider two implementations of this idea that build on a variant of SPS, known under the name of *Block SPS* (Block SPS was introduced in [6] with the aim of making SPS more robust against non-independent noise). However, the same idea could be beneficial for studies beyond the scope of this short communication, for example, for the analysis and the design of regularised versions of SPS, [5,7,11,4].

[2] In MATLAB, provided that $n \geq 2d$, $p_{n,d}$ can be computed by the command `binocdf(n-d, n, 0.5)-binocdf(d-1, n, 0.5)`, and the right-hand side of (5) by `binocdf(m-q-1,m-1,binocdf(n-d, n, 0.5)-binocdf(d-1, n, 0.5))`.

4.2.1 Two variants of Block SPS

Block SPS is like SPS with the difference that a sequence $\alpha_{i,1}, \ldots, \alpha_{i,n}$ is kept to the same value "+" or "−" for blocks of λ_b consecutive time instants before the sign is randomly drawn. If the block length is $\lambda_b \geq d$ (the last block can be longer than λ_b when n is not a multiple of λ_b), then the boundedness condition $K_i \succ 0$ is satisfied in all the cases except when the sign string turns out to be made entirely of "+" or of "−", which happens with probability $2^{1-\nu_b}$, where ν_b is the number of blocks. To rule out this event, we can slightly modify the Block SPS algorithm.

Block SPS with rejection

A first option is modifying the Block SPS algorithm as follows: for i going from 1 to $m-1$, when the sign string $\alpha_{i,1}, \ldots, \alpha_{i,n}$ (to be used in $S_i(\theta)$ according to formula (2)) is drawn, it is accepted only if it is different from all the sign strings that have already be employed in $S_0(\theta), S_1(\theta), \ldots, S_{i-1}(\theta)$ and it is different from the opposite values of such strings. If the sign string is equal to an already employed string or to its opposite, then it is rejected and redrawn, until a different string is obtained. A string and its opposite are considered as the same string because they have the same effect on the value of $\|S_i(\theta)\|$. The string made entirely of "+" (or entirely of "−") coincides with the sign string that is employed in the definition of $S_0(\theta)$, therefore is always rejected.

Deterministic Block SPS

A second option is building the functions $S_1(\theta), \ldots, S_{m-1}(\theta)$ according to *all* the possible sequences of block perturbations, i.e., in an exhaustive way and without resorting to random sampling. Namely, under the assumption that $\lambda_b \geq d$ and $n \geq 2d$, denoting by ν_b the number of blocks, we can take $m = 2^{\nu_b - 1}$, and define each function S_i, for $i = 1, \ldots, m-1$, by using a different (deterministic) string $\{\alpha_{i,t}\}$ of block perturbations in (2) (the sign strings to be used to construct these functions are $2^{\nu_b-1} - 1$ because a string and its opposite are considered as the same string and the string made entirely of "+" is already used in S_0).

Proof of exactness

In appendix A, we provide a sketch of the proof that, like SPS and standard Block SPS, the two variants of Block SPS here proposed construct regions that include θ^* with an exact, user-chosen probability.

5 Conclusions

In this communication, we have stated a simple condition for the boundedness of SPS regions. By using the standard concept of regressor excitation, we have reformulated a technical condition that was available in the literature. We have also argued that, in this form, the condition provides a tool to tune the SPS parameters and also to guide the design of new SPS-like algorithms with guarantees on the boundedness of the regions. We have offered examples of both these possibilities.

Acknowledgements

The author is grateful to Marco Campi for reading a preliminary version of this work and to Giacomo Baggio for a discussion that motivated the present study.

References

[1] Ben-Tal Aharon, Laurent El Ghaoui, and Arkadi Nemirovski. *Robust Optimization.* Princeton Series in Applied Mathematics. Princeton University Press, October 2009.

[2] Algo Carè, Marco C. Campi, Balázs Cs. Csáji, and Erik Weyer. Facing undermodelling in Sign-Perturbed-Sums system identification. *Systems & Control Letters*, 153:104936, 2021.

[3] Algo Carè, Balázs Cs. Csáji, Marco C. Campi, and Erik Weyer. Finite-sample system identification: An overview and a new correlation method. *IEEE Control Systems Letters*, 2(1):61–66, Jan 2018.

[4] Algo Carè, Gianluigi Pillonetto, and Marco C Campi. Uncertainty bounds for kernel-based regression: A Bayesian SPS approach. In *IEEE 28th International Workshop on Machine Learning for Signal Processing (MLSP)*. IEEE, 2018.

[5] Balázs Cs. Csáji. Non-asymptotic confidence regions for regularized linear regression estimates. In *Progress in Industrial Mathematics at ECMI 2018*, pages 605–611. Springer, 2019.

[6] 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.

[7] Balázs Cs. Csáji and Krisztián B. Kis. Distribution-free uncertainty quantification for kernel methods by gradient perturbations. *Machine Learning*, 108(8-9):1677–1699, 2019.

[8] Roger A. Horn and Charles R. Johnson. *Matrix Analysis, Second Edition.* Cambridge University Press, 2013.

[9] Michel Kieffer and Eric Walter. Guaranteed characterization of exact non-asymptotic confidence regions as defined by LSCR and SPS. *Automatica*, 49:507–512, 2013.

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

[11] Gianluigi Pillonetto, Algo Carè, and Marco C. Campi. Kernel-based SPS. *IFAC-PapersOnLine*, 51(15):31 – 36, 2018.

[12] Erik Weyer, Marco C. Campi, and Balázs Cs. Csáji. Asymptotic properties of SPS confidence regions. *Automatica*, 82:287–294, 2017.

A Technical argument for the exactness of the two variants of Block SPS

We sketch, in some detail, the proof of the fact that the variants of Block SPS proposed in Section 4.2.1 construct regions that include θ^* with probability $1-\frac{q}{m}$. The key fact to be proven is that the rank of $\|S_0(\theta^*)\|$ among $\|S_0(\theta^*)\|,\dots,\|S_{m-1}(\theta^*)\|$ is no larger than $m-q$ with probability $1-\frac{q}{m}$.

In what follows, we consider a sequence $\alpha_{i,1},\dots,\alpha_{i,n}$ and its opposite $-\alpha_{i,1},\dots,-\alpha_{i,n}$ as the same sequence.

It is convenient to start with the proof of the exactness of the *Deterministic Block SPS* algorithm.

Let $\mathbf{N}$ be the random vector $[N_1,N_2,\dots,N_n]$ and consider the vector function

$$V(\mathbf{N}) = [\|S_0(\theta^*)\|,\dots,\|S_{m-1}(\theta^*)\|].$$

For any m-dimensional vector W, we denote by $\mathcal{R}(W)$ the rank of the first component of the vector W with respect to the values of all the m components (the case of ties is discussed briefly later). With this notation, the key fact to be proven is that $\mathbb{P}(\mathcal{R}(V(\mathbf{N}))\le m-q) = 1-\frac{q}{m}$.

Note that, because of the symmetry of the noise, $\mathbf{N}$ is distributed like its randomly block-perturbed version defined as $\tilde{\mathbf{N}} = [\alpha_1 N_1,\alpha_1 N_2,\dots,\alpha_1 N_{\lambda_b},\alpha_2 N_{\lambda_b+1},\dots,\alpha_2 N_{2\lambda_b},\dots,\alpha_{\nu_b} N_n]$, where $\alpha_1,\dots,\alpha_{\nu_b}$ are i.i.d. signs with equal probability of being positive or negative. This fact implies that the random vector $V(\mathbf{N})$ has the same distribution of the random vector $V(\tilde{\mathbf{N}})$.

Now, we condition on a given realisation of $\mathbf{N}$ and let just the block perturbation $\alpha_1,\dots,\alpha_{\nu_b}$ be random. In particular, we condition on a realisation of $\mathbf{N}$ such that $V(\mathbf{N})$ is a vector of m *different* values (no ties) and we study $V(\tilde{\mathbf{N}})$ as a function of the $2^{\nu_b-1}=m$ possible realisations of the block perturbation (we recall that we have identified a block perturbation with its opposite). Since the functions $S_i(\theta)$ $i=1,\dots,m-1$ are, by construction, obtained by setting in all the possible ways the signs of the blocks of $\alpha_{i,1},\dots,\alpha_{i,n}$ in (2), each block-perturbed version $\tilde{\mathbf{N}}$ of $\mathbf{N}$, when plugged into $V(\cdot)$, yields a different permutation of $V(\mathbf{N})$. In particular, each realisation of the block perturbation $\alpha_1,\dots,\alpha_{\nu_b}$ yields a different value of the first component of $V(\tilde{\mathbf{N}})$. As each block perturbation has probability $1/m$, the first component of $V(\tilde{\mathbf{N}})$ assumes each value among the m possible ones with equal probability. We conclude that, conditioning on $\mathbf{N}$, the distribution of $\mathcal{R}(V(\tilde{\mathbf{N}}))$ is uniform over $\{1,\dots,m\}$. When, on the other hand, we condition on a realisation of $\mathbf{N}$ for which there are repeated values in $V(\mathbf{N})$, a tie-break rule must be used to define the rank in a suitable manner.

By resorting to the same total order $\succ_\pi$ defined in [6], it is possible to conclude that the rank of $V(\tilde{\mathbf{N}})$ is uniformly distributed also in the presence of ties, and that $\mathbb{P}(\mathcal{R}(V(\tilde{\mathbf{N}})) \leq m-q \,|\, \mathbf{N}) = 1-\frac{q}{m}$ holds true irrespective of $\mathbf{N}$. By integrating this conditional probability with respect to $\mathbf{N}$, we get $\mathbb{P}(\mathcal{R}(V(\tilde{\mathbf{N}})) \leq m-q) = 1-\frac{q}{m}$. Recalling that $V(\tilde{\mathbf{N}})$ is distributed as $V(\mathbf{N})$, the key fact is proven for the *Deterministic Block SPS* algorithm.

The exactness of the *Block SPS with rejection* algorithm can be proven by interpreting this algorithm as a randomised version of *Deterministic Block SPS*, where a pre-determined number δ_m of functions among $\|S_1(\theta)\|$, $\ldots, \|S_{m-1}(\theta)\|$ are randomly neglected during the evaluation of the rank of $\|S_0(\theta)\|$. Let us denote by $\mathcal{R}_m$ the rank of $\|S_0(\theta^*)\|$ when all the $m = 2^{\nu_b - 1}$ functions are considered (i.e., $\mathcal{R}_m = \mathcal{R}(V(\mathbf{N}))$, with the notation that we used in the study of *Deterministic Block SPS*), and let us denote by $\mathcal{R}_{m-\delta_m}$ the rank of $\|S_0(\theta^*)\|$ among the values of $\|S_0(\theta^*)\|$ itself and of the unneglected $(m-1-\delta_m)$ functions. We prove below that $\mathcal{R}_{m-\delta_m}$ has uniform distribution over $\{1, \ldots, m-\delta_m\}$.

Let us condition on the event $\mathcal{R}_m = r$, and denote by L the list of the $r-1$ values among $\|S_1(\theta^*)\|,\ldots,\|S_{m-1}(\theta^*)\|$ that are smaller than $\|S_0(\theta^*)\|$. Then, the probability that $\mathcal{R}_{m-\delta_m} = r'$ is the probability that the number of unneglected values among those in L is exactly $r'-1$, i.e., for $r' = 1, \ldots, m-\delta_m$, we have

$$
\begin{aligned}
&\mathbb{P}(\mathcal{R}_{m-\delta_m} = r' \,|\, \mathcal{R}_m = r) \\
&= \begin{cases} \frac{\binom{r-1}{r'-1}\binom{m-r}{m-\delta_m-r'}}{\binom{m-1}{m-\delta_m-1}}, & \text{if } r-\delta_m \leq r' \leq r \\ 0, & \text{otherwise.} \end{cases}
\end{aligned}
$$

From the analysis of *Deterministic Block SPS* we know that $\mathbb{P}(\mathcal{R}_m = r) = \frac{1}{m}$ for $r = 1, \ldots, m$. Then, for $r' = 1, \ldots, m-\delta_m$, we get

$$
\begin{aligned}
&\mathbb{P}(\mathcal{R}_{m-\delta_m} = r') \\
&= \sum_{r=1}^{m} \mathbb{P}(\mathcal{R}_{m-\delta_m} = r' \,|\, \mathcal{R}_m = r)\mathbb{P}(\mathcal{R}_m = r) \\
&= \frac{1}{m}\frac{1}{\binom{m-1}{m-\delta_m-1}} \sum_{r=r'}^{r'+\delta_m} \binom{r-1}{r'-1}\binom{m-r}{m-\delta_m-r'} \\
&= \frac{1}{m}\frac{1}{\binom{m-1}{m-\delta_m-1}} \sum_{j=r'-1}^{r'-1+\delta_m} \binom{j}{r'-1}\binom{(m-1)-j}{(m-\delta_m-1)-(r'-1)} \\
&= \frac{1}{m}\frac{1}{\binom{m-1}{m-\delta_m-1}} \binom{m}{m-\delta_m} = \frac{1}{m-\delta_m}.
\end{aligned}
$$