Randomized pick-freeze for sparse Sobol indices estimation in high dimension

Yohann de Castro, Alexandre Janon

▶ **To cite this version:**

Yohann de Castro, Alexandre Janon. Randomized pick-freeze for sparse Sobol indices estimation in high dimension. [Research Report] 2014. hal-00962473

HAL Id: hal-00962473
https://inria.hal.science/hal-00962473

Submitted on 21 Mar 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Randomized pick-freeze for sparse Sobol indices estimation in high dimension

Yohann de Castro[1] and Alexandre Janon[1]

March 21, 2014

Abstract

This article investigates a new procedure to estimate the influence of each variable of a given function defined on a high-dimensional space. More precisely, we are concerned with describing a function of a large number p of parameters that depends only on a small number s of them. Our proposed method is an unconstrained ℓ_1-minimization based on the Sobol's method. We prove that, with only $\mathcal{O}(s \log p)$ evaluations of f, one can find which are the relevant parameters.

1 Introduction

1.1 Context: Sensitivity analysis and Sobol indices

Some mathematical models encountered in applied sciences involve a large number of poorly-known parameters as inputs. It is important for the practitioner to assess the impact of this uncertainty on the model output. An aspect of this assessment is sensitivity analysis, which aims to identify the most sensitive parameters, that is, parameters having the largest influence on the output. The parameters identified as influent have to be carefully tuned (or estimated) by the users of the model. On the other hand, parameters whose uncertainty has a small impact can be set to a nominal value (which can be some special value, for which the model is simpler).

In global (stochastic) variance-based sensitivity analysis (see for example [19] and references therein), the input variables are assumed to be independent random variables. Their probability distributions account for the practitioner's belief about the input uncertainty. This turns the model output into a random variable, whose total variance can be split down into different partial variances (this is the so-called Hoeffding decomposition, also known as functional ANOVA, see [15]). Each of these partial variances measures the uncertainty on the output induced by each input variable uncertainty. By considering the ratio of each partial variance to the total variance, we obtain a measure of importance for each input variable that is called the Sobol index or sensitivity index of the variable [20, 21]; the most sensitive parameters can then be identified and ranked as the parameters with the largest Sobol indices. Each

[1]Laboratoire de Mathématiques d'Orsay, Bâtiment 425, Université Paris-Sud, 91405 Orsay, France

partial variance can be written as the variance of the conditional expectation of the output with respect to each input variable.

Once the Sobol indices have been defined, the question of their effective computation or estimation remains open. In practice, one has to estimate (in a statistical sense) those indices using a finite sample (of size typically in the order of hundreds of thousands) of evaluations of model outputs [23]. Indeed, many Monte Carlo or quasi Monte Carlo approaches have been developed by the experimental sciences and engineering communities. This includes the Sobol pick-freeze (SPF) scheme (see [21, 11]). In SPF a Sobol index is viewed as the regression coefficient between the output of the model and its pick-freezed replication. This replication is obtained by holding the value of the variable of interest (frozen variable) and by sampling the other variables (picked variables). The sampled replications are then combined to produce an estimator of the Sobol index.

1.2 High-dimensional, sparse contexts

The pick-freeze scheme is used on models with a reasonable (typically, less than one thousand) number of inputs. When there is a large number of input parameters (what we call an *high-dimensional* context), this scheme will require a number of model evaluations which is generally too large to be computed in practice. Hence, in high-dimensional contexts, some specific sensitivity analysis methods exist, such as *screening* methods (for instance, Morris' scheme [18]), but they do not target the estimation of Sobol indices. Note that in [24], an interesting method for estimating Sobol indices is proposed and is claimed to be applicable in high-dimensional contexts.

Besides, models with a large number of input parameters often display a so-called *sparsity of effects* property, that is, only a small number of input parameters are actually influent: in other terms, we want to efficiently estimate a sparse vector of Sobol indices. Sparse estimation in high-dimensional contexts is the object of high-dimensional statistics methods, such as the LASSO estimator.

In our frame, we would like to find the most influent inputs of a function that is to be described. This framework is closely related to exact support recovery in high-dimensional statistics. Note exact support recovery using ℓ_1-minimization has been intensively investigated during the last decade, see for instance [28, 10, 25, 16] and references therein. We capitalize on these works to build our estimation procedure. The goal of this paper is to draw a bridge, which, to the best of our knowledge, has not been previously drawn, between Sobol index estimation via pick-freeze estimators and sparse linear regression models. This bridge can be leveraged so as to propose an efficient estimation procedure for Sobol indices in high-dimensional sparse models.

1.3 Organization of the paper

The contribution of this paper is twofold: Section 2 describes a new algorithm to simultaneously estimate the Sobol indices using ℓ_1-relaxation and give elementary error analysis of this algorithm (Theorem 1 and Theorem 2), and Section 3 presents a new result on exact support recovery by Thresholded-Lasso

that do not rely on coherence propriety. In particular, we prove that exact support recovery holds beyond the Welch bound. Appendix A gives the proofs of the results in Section 2. Appendix B.1 gives preliminary results for proving Theorem 3 of Section 3. The remaining appendices apply these results to different designs (leading for Appendix B.4 to Theorem 3); Appendix B.2 and B.3 are rather independent and study Thresholded-Lasso in the frame of random sparse graphs.

2 A convex relaxation of Sobol's method

2.1 Notation and model

Denote by $X_1, \dots, X_p$ the input parameters, assumed to be independent random variables of known distribution. Let Y be the model output of interest:

$$Y = f(X_1, \dots, X_p),$$

where $f : \mathbb{R}^p \to \mathbb{R}$ is so that $Y \in L^2$ and $\mathbb{V}\mathbf{ar}(Y) \neq 0$. Assume that f is *additive*, i.e.

$$f(X_1, \dots, X_p) = f_1(X_1) + \dots + f_p(X_p) \tag{1}$$

for some functions $f_i : \mathbb{R} \to \mathbb{R}$, $i = 1, \dots, p$. We want to estimate the following vector:

$$S = (S_i)_{i=1}^p \quad \text{where} \quad S_i = \frac{\mathbb{V}\mathbf{ar}[\mathbb{E}(Y|X_i)]}{\mathbb{V}\mathbf{ar}(Y)},$$

is the i^{th} Sobol index of Y and quantifies the influence of X_i on Y. In this article we present a new procedure for evaluating the Sobol indices when p is large. We make the assumption that the number of nonzero Sobol indices:

$$s = \#\{i = 1, \dots, p \text{ s.t. } S_i \neq 0\}$$

remains small in comparison to p. Observe our model assumes that we know an upper bound on s. The Sobol indices can be estimated using the so-called *pick-freeze* scheme, also know as *Sobol's method* [20, 21]. Let X' be an independent copy of X and note, for $i = 1, \dots, p$:

$$Y^i = f(X'_1, \dots, X'_{i-1}, X_i, X'_{i+1}, \dots, X'_p). \tag{2}$$

Then we have:

$$S_i = \frac{\mathbb{C}\mathbf{ov}(Y, Y^i)}{\mathbb{V}\mathbf{ar}(Y)}.$$

This identity leads to an empirical estimator of S_i:

$$\widehat{S}_i = \frac{\frac{1}{N}\sum Y_k Y_k^i - \left(\frac{1}{N}\sum \frac{Y_k + Y_k^i}{2}\right)^2}{\frac{1}{N}\sum \frac{(Y_k)^2 + (Y_k^i)^2}{2} - \left(\frac{1}{N}\sum \frac{Y_k + Y_k^i}{2}\right)^2},$$

where all sums are for k from 1 to N, and $\{(Y_k, Y_k^i)\}_{k=1,\dots,N}$ is an iid sample of the distribution of (Y, Y^i) of size N. This estimator has been introduced in [17] and later studied in [14] and [11].

In the high-dimensional frame, the estimation of the p indices using $\widehat{S}_i$ for $i = 1, \ldots, p$ would require $(p+1)N$ evaluations of f so as to generate the realizations of $(Y, Y^1, \ldots, Y^p)$. This may be too much expensive when p is large and/or evaluation of f is costly. Besides, thanks to our sparsity assumption, such an estimation "one variable at a time" will be inefficient, as many computations will be required to estimate zero many times. To the best of our knowledge, this paper is the first to overcome this difficulty introducing a new estimation scheme.

2.2 Multiple pick-freeze

We now generalize definition (2). Let $F \subset \{1, \ldots, p\}$ be a set of indices. Define Y^F by:

$$Y^F = f(X^F) \quad \text{where} \quad \left(X^F\right)_i = \begin{cases} X_i \text{ if } i \in F, \\ X'_i \text{ if } i \in F^c. \end{cases}$$

where $F^c = \{1, \ldots, p\} \setminus F$. The name of the method stems from the fact that, to generate the Y^F variable, all the input parameters whose indices are in F are **F**rozen. In the pick-freeze scheme of the previous subsection, only one variable was frozen at the time, namely $F = \{i\}$. We then define:

$$S_F = \frac{\mathbb{C}\mathbf{ov}(Y, Y^F)}{\mathbb{V}\mathbf{ar}(Y)},$$

which admits a natural estimator:

$$\widehat{S}_F = \frac{\frac{1}{N}\sum Y_k Y_k^F - \left(\frac{1}{N}\sum \frac{Y_k + Y_k^F}{2}\right)^2}{\frac{1}{N}\sum \frac{(Y_k)^2 + (Y_k^F)^2}{2} - \left(\frac{1}{N}\sum \frac{Y_k + Y_k^F}{2}\right)^2}. \tag{3}$$

Under additivity hypothesis (1), one has:

$$S_F = \sum_{i \in F} S_i.$$

Now, let's choose $n \in \mathbb{N}^\star$, subsets $F_1, \ldots, F_n$ of $\{1, \ldots, p\}$, and denote by E the following vector of estimators:

$$E = \left(\widehat{S}_{F_1}, \ldots, \widehat{S}_{F_n}\right). \tag{4}$$

Notice that, once the $F_1, \ldots, F_n$ have been chosen, the E vector can be computed using $(n+1)N$ evaluations of f.

2.2.1 Bernoulli Regression model

The choice of $F_1, \ldots, F_n$ can be encoded in a binary matrix Φ with n lines and p columns, so that:

$$\Phi_{ji} = \begin{cases} 1 \text{ if } i \in F_j, \\ 0 \text{ otherwise.} \end{cases} \qquad j = 1, \ldots, n \text{ and } i = 1, \ldots, p. \tag{5}$$

It is clear that $(S_{F_1}, \ldots, S_{F_n}) = \Phi S$, hence:

$$E = \Phi S + \epsilon, \tag{6}$$

where the ϵ vector defined by $\epsilon_j = \widehat{S}_{F_i} - S_{F_i}$ gives the estimation error of ΦS by E. In practice, the E vector and the Φ matrix are known, and one has to estimate S. Thereby, Eq. (6) can be seen as a linear regression model whose coefficients are the Sobol indices to estimate. Moreover, observe that $n \ll p$ and S sparse, hence we are in a high-dimensional sparse linear regression context. The problem (6) has been extensively studied in the context of sparse estimation [16, 2, 7] and compressed sensing [5, 4], and a classical solution is to use the LASSO estimator [22]:

$$\widehat{S} \in \underset{U \in \mathbb{R}^p}{\operatorname{argmin}} \left(\frac{1}{n} \|E - \Phi U\|_2^2 + 2r \|U\|_1 \right), \tag{7}$$

where $r > 0$ is a regularization parameter and:

$$\|v\|_2^2 = \sum_{j=1}^{n} v_j^2, \quad \|u\|_1 = \sum_{i=1}^{p} |u_i|.$$

Many efficient algorithms, such as LARS [9], are available in order to solve the above minimization problem, and to find an appropriate value for r. In high dimensional statistics, one key point for the LASSO procedure is the choice of the Φ matrix. In the Compressed Sensing literature, a random matrix with i.i.d. coefficients often proves to be a good choice, hence we will study possible random choices for Φ.

Summary of the method "Randomized Pick-Freeze" (RPF) for Bernoulli matrices

Our estimation method can be summarized as follows:

1. Choose N (Monte-Carlo sample size), n (number of estimations) and r (regularization parameter).
2. Randomly sample a 0-1 matrix Φ with Bernoulli distribution of parameter μ.
3. Deduce from Φ the $F_1, \ldots, F_n$ subsets using (5).
4. Generate a N-sized sample of $(Y, Y^{F_1}, \ldots, Y^{F_n})$.
5. Use this sample in (3), for $F = F_1, \ldots, F_n$, to obtain the E vector (4).
6. Solve problem (7) to obtain an $\widehat{S}$ which estimates S.

Given the binary constraint on Φ, we will choose a Bernoulli distribution with parameter $\mu \in]0;1[$. In this model, $(\Phi_{ji})_{j,i}$ are independent, with for all i, j:

$$\mathbb{P}(\Phi_{ji} = 1) = \mu = 1 - \mathbb{P}(\Phi_{ji} = 0). \tag{8}$$

Theorem 1 (ℓ^∞ error bound)**.** *Suppose that:*

1. *δ is a real in* $\left]0; \dfrac{1-\mu}{16s}\right[$;
2. *ϵ is a centered Gaussian vector whose covariance matrix has σ^2 as largest eigenvalue;*
3. *$r = A\sigma\sqrt{\mu(1+\delta)}\sqrt{\dfrac{\ln p}{n}}$ for some $A > 2\sqrt{2}$.*

Let:

$$\begin{aligned} t &= \left(\frac{3}{2} + \frac{24(\mu+\delta)}{\frac{1-\mu}{s} - 16\delta}\right)\frac{r}{\mu}; \\ \alpha &= 1 - \left(1 - p^{1-A^2/8}\right)\left(1 - 2\exp\left(-2n\delta^2\mu^2 + \ln p\right)\right) \\ &\quad + \exp\left(-2n\delta^2\mu^2 + 2\ln p\right). \end{aligned}$$

Then, with probability at least $1-\alpha$, any solution $\widehat{S}$ of (7) satisfies:

$$\max_{i=1,\dots,p} |\widehat{S}_i - S_i| \leq t.$$

Proof. A proof can be found in Appendix A.1. □

Remark 1: For the probability above to be greater than zero, it is necessary to have:

$$n \geq \frac{\ln p}{\delta^2\mu^2} \geq \frac{256s^2\ln p}{\mu^2(1-\mu)^2}. \tag{9}$$

Remark 2: The statement of Theorem 1 can be compared to standard results in high-dimensional statistics such as exact support recovery under coherence property [16]. Nevertheless, observe that a standard assumption is that the column norm of the design matrix is $\sqrt{n}$ while in our frame this norm is random with expectation of order $\sqrt{\mu n}$.

Remark 3: In our context, the second hypothesis of the above theorem does not exactly hold; indeed, ϵ is only asymptotically Gaussian (when $N \to +\infty$), see [14] for instance. However, for practical purposes, the observed convergence is fast enough. One can also see a related remark in our proof of this theorem, in Appendix A.1.

Corollary 1 (Support recovery by Thresholded-Lasso)**.** *Let:*

$$S_{min} = \min_{\substack{i=1,\dots,p \\ s.t.\ S_i \neq 0}} S_i.$$

Then, under the same assumptions of Theorem 1, we have, with probability greater than $1-\alpha$ and for all $i = 1, \dots, p$:

$$\widehat{S}_i > t \implies S_i > 0,$$

and:

$$\widehat{S}_i < S_{min} - t \implies S_i = 0.$$

Proof of Corollary 1. For the first point, notice that:

$$|S_i| \geq |\widehat{S}_i| - |\widehat{S}_i - S_i| \geq |\widehat{S}_i| - t > 0 \text{ if } \widehat{S}_i > t.$$

For the second point: if $\widehat{S}_i < S_{min} - t$, we have:

$$|S_i| \leq |S_i - \widehat{S}_i| + |\widehat{S}_i| < t + (S_{min} - t) = S_{min},$$

and $S_i = 0$ by definition of S_{min}. □

Remark 4 (important): Theorem 1 and Corollary 1 show that one can identify the most important inputs of a function as soon as the corresponding Sobol indices are above the threshold t. Recall *Thresholed-Lasso* is a thresholded version of any solution to (7). Moreover, observe that we do not address the issue of estimating the Sobol indices. This can be done using a two-step procedure: estimate the support using Thresholed-Lasso and then estimate the Sobol indices using a standard least squares estimator.

2.2.2 Rademacher Regression model

The choice of $F_1, \ldots, F_n$ can also be encoded in a ± 1 matrix Φ with n lines and p columns, so that:

$$\Phi_{ji} = \begin{cases} 1 & \text{if } i \in F_j, \\ -1 & \text{otherwise.} \end{cases} \qquad j = 1, \ldots, n \text{ and } i = 1, \ldots, p. \tag{10}$$

It is clear that:

$$(S^{\Delta}_{F_1}, \ldots, S^{\Delta}_{F_n}) = \Phi S\,,$$

where $S^{\Delta}_{F_i} = S_{F_i} - S_{F_i^c}$. Hence:

$$E = \Phi S + \epsilon^{\Delta}, \tag{11}$$

where the ϵ vector defined by $\epsilon^{\Delta}_j = \widehat{S}^{\Delta}_{F_i} - S^{\Delta}_{F_i}$ gives the estimation error of ΦS by E. Thus, the problem of estimating S from E has been casted into linear regression which can be tackled by (7).

Summary of the method "Randomized Pick-Freeze" (RPF) for Rademacher matrices

Our estimation method can be summarized as follows:

1. Choose N (Monte-Carlo sample size), n (number of estimations), and r (regularization parameter).
2. Sample a Φ matrix according to a ± 1 symmetric Rademacher distribution.
3. Deduce from Φ the $F_1, \ldots, F_n$ subsets using the correspondance (10).
4. Generate a N-sized sample of $(Y, Y^{F_1}, \ldots, Y^{F_n})$.
5. Use this sample in (3), for $F = F_1, \ldots, F_n$, to obtain the E vector (4).
6. Solve problem (7) to obtain an $\widehat{S}$ which estimates S.

We now consider a different sampling procedure for Φ, which will make it possible to improve on the constants in (9) as it will be stated in (13). Specifically, we sample Φ using a symmetric Rademacher distribution:

$$(\Phi_{ji})_{j,i} \text{ are independent : } P(\Phi_{ji} = 1) = P(\Phi_{ji} = -1) = 1/2. \tag{12}$$

The following theorem is the equivalent of Theorem 1 for Rademacher designs.

Theorem 2 (ℓ^∞ error bound)**.** *Suppose that:*

1. *ϵ is a centered Gaussian vector whose covariance matrix has σ^2 as largest eigenvalue;*
2. *$\delta = \dfrac{1}{7\delta' s}$ for some real $\delta' > 1$;*
3. *$r = A\sigma\sqrt{\dfrac{\ln p}{n}}$ for some $A > 2\sqrt{2}$.*

Let:

$$\begin{aligned} t &= \frac{3}{2}\left(1 + \frac{16}{5(\delta' - 1)}\right) r \\ \alpha &= 1 - \left(1 - p^{1-A^2/8}\right)\left(1 - \exp\left(-n\frac{49\delta^2 s^2}{2} + 2\ln p\right)\right). \end{aligned}$$

Then, with probability at least $1 - \alpha$, any solution $\widehat{S}$ of (7) satisfies:

$$\max_{i=1,\dots,p} |\widehat{S}_i - S_i| \leq t.$$

Proof. A proof can be found in Appendix A.2. □

Remark 1: For the probability above to be greater than zero, it is necessary to have:

$$n \geq Cs^2 \ln p \tag{13}$$

for some constant $C > 0$.
Remark 2: Support recovery property (Corollary 1) also holds in this context.

2.3 Numerical experiments

2.3.1 LASSO convergence paths

In this section, we perform a numerical test of the "Randomized Pick-Freeze" estimation procedure for Bernoulli and Rademacher matrices, summarized respectively on pages 5 and 7. We use the following model:

$$Y = f(X_1, \dots, X_{300}) = X_1^2 + 4X_1 + 4X_2 + 10X_3,$$

hence $p = 300$ and $s = 3$, with $(X_i)_{i=1,\dots,120}$ iid uniform on $[0, 1]$. It is easy to see that, in this model, we have $S_3 > S_1 > S_2 > 0$ and $S_i = 0$ for all $i > 3$. The tests are performed by using $n = 30$. The obtained LASSO regularization paths (ie., the estimated indices for different choices of the penalization parameter r) are plotted in Figures 1 (for Bernoulli design matrix with parameter $\mu = 1/2$) and

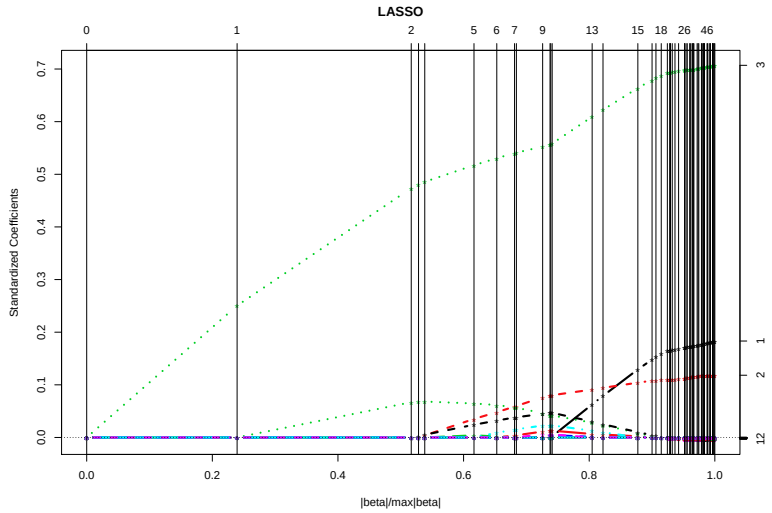

Figure 1: LASSO convergence path for a Bernoulli design.

2 (for Rademacher design matrix). The Monte-Carlo sample size used are $N = 3000$ and $N = 2000$, respectively for Bernoulli and Rademacher designs. This difference in sample sizes accounts for the increase in the number of required evaluations of the f function when a Rademacher design is used (as, in this case, each replication is a difference of two pick-freeze estimators on the same design).

We observe that the Rademacher design seems to perform better (as LASSO convergence is faster) than the Bernoulli design, in accordance with the remarks made in the beginning of Section 3. Both designs perfectly recover the active variables (the support of S), as well as the ordering of indices. Note that the proposed method requires only $30 \times 2 \times 3000 = 180000$ evaluations of the f function to estimate the 300 Sobol indices, while a classic one-by-one pick-freeze estimation with the same Monte-Carlo sample size would require $3000 \times (300 + 1) = 903000$ evaluations of f.

2.3.2 Illustration of ℓ^∞ error bounds

We now present a synthetic example which shows the performance of the Rademacher RPF algorithm, used with Theorem 2 and the support recovery corollary.

Suppose that we work on a model with $p = 30000$ inputs, with only $s = 3$ of them have a nonzero Sobol index. We postulate that all the $\widehat{S}_i$ estimators, as well as the $\widehat{S}_F^\Delta$ have standard Gaussian distribution. By using $N = 10^6$ and $n = 100$ in Theorem 2, we get that the t error bound given in this theorem is $t = 0.03$, with probability greater than $1 - \alpha = 95\%$. Hence, by doing calling $3Nn = 3 \times 10^8$ to the f function, one can correctly identify parameters whose Sobol indices are greater than 0.03.

On the other hand, when using classical one-by-one Sobol index estimation, one has to perform $p = 30000$ independent estimations of Sobol index

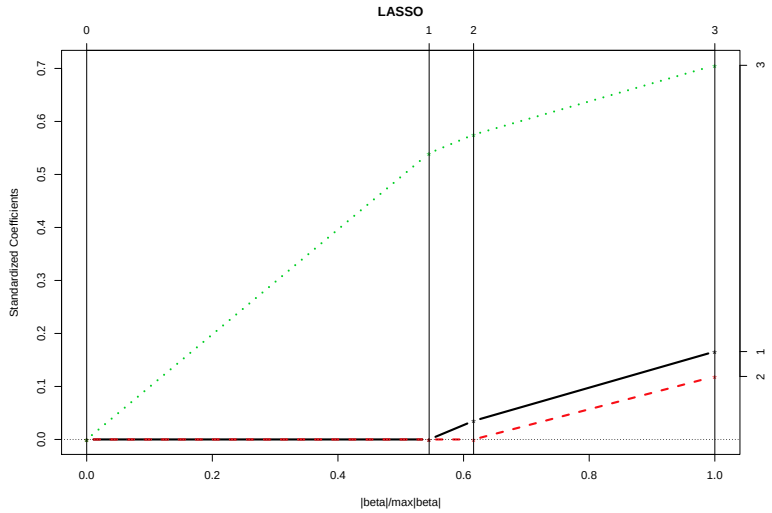

Figure 2: LASSO convergence path for a Rademacher design.

confidence intervals, at level $1 - 0.95^{1/30000} = 1.71 \times 10^{-6}$ (by using ÅăidÃąk correction). From the quantiles of the Gaussian distribution, the length of these intervals is $9.568/\sqrt{N}$. Hence, getting confidence intervals of width 0.03 require $N' = (9.568/0.03)^2 \approx 101720$ sample size. Hence, the total cost for this method is $2N'(p+1) = 6103200000 \approx 61 \times 10^8$ calls to the f function.

3 Breaking the square-root bottleneck

In the beginning of this paper, we have showed results that are limited by the constraint $n \geq Cs^2 \log p$ for some constant C. This limitation is due to the use of the mutual incoherence property in the proofs, which is heuristically bounded by Welch's bound [26]. We now present a new approach, based on Universal Distortion Property [7] and a relaxed version of the coherence (see Lemma 3 in Appendix B.1) which enables to break this "bottleneck" for Rademacher designs. Note that applying this approach for Bernoulli designs leads to a new proof of the above stated results. For sake of completeness, we give these proofs in Appendix B.2. This appendix covers the frame of exact support recovery using Thresholded-Lasso using adjacency matrix as design. In this section we focus on Rademacher designs defined by (12), namely $(\Phi_{ji})_{j,i}$ are independent and for all i, j, $\mathbb{P}(\Phi_{ji} = \pm 1) = 1/2$.

Theorem 3 (Exact recovery with Rademacher designs)**.** *There exists universal constants $C_1, C_2, C_3 > 0$ such that the following holds. Let $c > 1$ and $\Phi \in \{\pm 1\}^{n \times p}$ a Rademacher matrix drawn according to* (12) *with:*

- $n \geq n_0 := C_1 s \log(C_2 p)$,
- $s \geq 6(2+c)/C_1$,

- $\epsilon \sim \mathcal{N}(0, \Sigma_n)$ *and the covariance diagonal entries enjoy* $\Sigma_{i,i} \leq \sigma^2$.

Let $\hat{S}$ *be any solution to* (7) *with regularizing parameter:*

$$r \geq r_1 := 45\,\sigma\Big[\frac{c\log p}{n}\Big]^{1/2},$$

Then, with a probability greater than $1 - 3p^{-c} - 2\exp(-C_3 n)$,

$$\|\hat{S} - S\|_\infty \leq \sigma\sqrt{\frac{n_0}{n}}\Big[\frac{r}{r_1}\Big]\Big[C_1' + \frac{C_2'}{\sqrt{s}}\Big]\sqrt{s}\,, \tag{14}$$

where $C_1' = 35869(c(2+c))^{1/2}/C_1$ *and* $C_2' = 46.31c^{1/2}/C_1^{1/2}$.

Proof. • Invoke Lemma 10 to get that:

$$\max_{1\leq k\neq l\leq p}\frac{1}{n}|\sum_{j=1}^{n}\Phi_{j,k}\Phi_{j,l}| \leq \Big[\frac{(2+c)8}{3C_1}\Big]^{1/2}\frac{1}{\sqrt{s}}\,,$$

with probability greater than $1 - 2p^{-c}$.

• Set $r_0 := \sigma(2c\log p/n)^{1/2}$ and $Z_i = (1/n)\Phi^\top\epsilon$. Observe that Z_i is centered Gaussian random variable with variance less than σ^2/n. Taking union bounds, it holds:

$$\mathbb{P}[(1/n)\|\Phi^\top\epsilon\|_\infty > r_0] \leq \sum_{i=1}^{p}\mathbb{P}[|Z_i| > \sqrt{2c}\sqrt{\log p}\,\sigma/\sqrt{n}] \leq p^{1-c}\,,$$

using $\|\Phi_i\|_2^2 = n$ and the fact that, for $\sqrt{2c}\sqrt{\log p} \geq \sqrt{2\log 2}$, we have:

$$\mathbb{P}[|\mathcal{N}(0,1)| > \sqrt{2c}\sqrt{\log p}] \leq \frac{1}{\sqrt{\pi\log 2}}\exp(-c\log p) \leq p^{-c}\,.$$

• From Lemma 8 and Lemma 9 with $\delta = 9/50$ and $\kappa = 4/9$, it holds that, with a probability greater than $1 - 2\exp(-C_3 n)$, for all $\gamma \in \mathbb{R}^p$ and for $T \subseteq \{1,\ldots,p\}$ such that $|T| \leq s$,

$$\|\gamma_T\|_1 \leq 4.4128\Big(\frac{s}{n}\Big)^{1/2}\|\Phi\gamma\|_2 + \frac{4}{9}\|\gamma\|_1\,.$$

Observe that $C_1 = 5/(c_1\delta^2)$, $C_2 = c_2/\delta^2$ and $C_3 = c_3 C_1$ where c_1, c_2, c_3 are universal constants appearing in Lemma 8 and $\delta = 9/50$.

• Invoke Lemma 3 with parameters $\rho = 4.4128/\sqrt{n}$, $\kappa = 4/9$, $\theta_2 = 1$ and $\theta_1 = ((2+c)8/(3C_1 s))^{1/2}$, to get that for all regularizing parameter $r \geq r_1 := 31.74 r_0$,

$$\|\hat{S} - S\|_\infty \leq \Big(1.0316 + 799\Big(\frac{2+c}{C_1}\Big)^{1/2}\sqrt{s}\Big)r\,,$$

on the event $\{(1/n)\|\Phi^\top\epsilon\|_\infty \leq r_0\}$. □

Remark. *Observe that* (14) *reads:*

$$\|\hat{S} - S\|_\infty \leq [\frac{r}{r_1}][C_1' + \frac{C_2'}{\sqrt{s}}]\,\sigma\sqrt{\frac{C_1 s^2\log(C_2 p)}{n}}\,.$$

where $C_1, C_2, C_1', C_2' > 0$ are constants. It shows that, for all $\alpha > 0$, Thresholded-lasso exactly recovers the true support if the non-zero coefficients are above a threshold that is proportional to $\sigma s^{\frac{1-\alpha}{2}}$ from $n = \mathcal{O}(s^{1+\alpha} \log p)$ observations. Hence, we have tackled the regime $0 < \alpha < 1$ where the elementary analysis of Theorem 2 fails to be applicable.

4 Conclusions

We have presented a new and performant method for estimating Sobol indices in high-dimensional additive models. We have shown that this method can lead to very good results in terms of computational costs. Besides, the error analysis of our algorithm led us to propose the results in Section 3, which are also of interest outside of the Sobol indices context, and which gives support recovery property for thresholded LASSO that are, to our best knowledge, greatly improving the results of the literature.

A Proof of the theorems

A.1 Proof of Theorem 1

We capitalize on [16, 3, 28] to prove sup-norm error bound when the design matrix has Bernoulli distribution.

Step 1: Rescaling

We rewrite (6) as $\tilde{E} = \tilde{\Phi} S + \tilde{\epsilon}$ where:

$$\tilde{E} = \frac{1}{\sqrt{\mu}} E, \quad \tilde{\Phi} = \frac{1}{\sqrt{\mu}} \Phi, \quad \tilde{\epsilon} = \frac{1}{\sqrt{\mu}} \epsilon.$$

Note $\widehat{S}$ satisfies:

$$\widehat{S} \in \underset{U \in \mathbb{R}^p}{\operatorname{argmin}} \left(\frac{1}{n} \left\| \tilde{E} - \tilde{\Phi} U \right\|_2^2 + 2\tilde{r} \left\| U \right\|_1 \right)$$

with

$$\tilde{r} = r/\mu = A\sigma \sqrt{\frac{1+\delta}{\mu}} \sqrt{\frac{\ln p}{n}}. \tag{15}$$

Step 2: Expectation and concentration

We define:

$$\Psi = \frac{1}{n} \tilde{\Phi}^T \tilde{\Phi} = \frac{1}{n\mu} \Phi^T \Phi.$$

Thanks to the rescaling above, we have, for all $i = 1, \ldots, p$:

$$\mathbb{E}(\Psi_{ii}) = \frac{1}{n} \sum_{k=1}^{n} \mathbb{E}(\tilde{\Phi}_{ki}^2) = 1,$$

and, for all $j = 1, \dots, p$, $j \neq i$:

$$\mathbb{E}(\Psi_{ij}) = \frac{1}{n} \sum_{k=1}^{n} \mathbb{E}(\tilde{\Phi}_{ki} \tilde{\Phi}_{kj}) = \mu.$$

Besides, Hoeffding's inequality [13] gives that for all $i = 1, \dots, p$ and any $\delta > 0$,

$$\mathbb{P}\left(|\Psi_{ii} - 1| \geq \delta\right) = \mathbb{P}\left(\left|\frac{1}{n} \sum_{k=1}^{n} (\Phi_{ki}^2 - \mu)\right| \geq \delta\mu\right) \leq 2\exp(-2n\delta^2\mu^2),$$

and, similarly, for any $j \neq i$,

$$\mathbb{P}\left(|\Psi_{ij} - \mu| \geq \delta\right) \leq 2\exp\left(-2n\delta^2\mu^2\right).$$

Thus, by union bound:

$$\mathbb{P}\left(\max_{i=1,\dots,p} |\Psi_{ii} - 1| \geq \delta\right) \leq 2\exp\left(-2n\delta^2\mu^2 + \ln p\right),$$

and:

$$\begin{aligned}\mathbb{P}\left(\max_{\substack{i=1,\dots,p \\ j=1,\dots,p \\ j\neq i}} |\Psi_{ij} - \mu| \geq \delta\right) &\leq 2\exp\left(-2n\delta^2\mu^2 + \ln\frac{p(p-1)}{2}\right), \\ &\leq \exp\left(-2n\delta^2\mu^2 + 2\ln p\right).\end{aligned}$$

Step 3: Noise control

We proceed as in the proof of Lemma 1 of [16]. We define, for $i = 1, \dots, p$:

$$Z_i = \frac{1}{n} \sum_{j=1}^{n} \tilde{\Phi}_{ji} \tilde{\epsilon}_j = \frac{1}{n} \left(\tilde{\Phi}^T \tilde{\epsilon}\right)_i.$$

We define the following event:

$$\mathcal{B} = \left\{\max_{i=1,\dots,p} |\Psi_{ii} - 1| \leq \delta\right\}.$$

For a given Φ, we denote by $\Sigma = \Sigma(\Phi)$ the covariance matrix of ϵ, hence the covariance matrix of $\tilde{\epsilon}$ is Σ/μ. Note that, as a function of Φ, Σ is also a random variable. We also denote by $\mathbb{V}\mathbf{ar} Z_i$ the variance of Z_i for a fixed Φ, which is

also a Φ-mesurable random variable. Conditionally on $\mathcal{B}$, we have:

$$\begin{aligned}
\mathbb{V}\mathbf{ar}Z_i &= \frac{1}{n^2}\mathbb{V}\mathbf{ar}\left[\left(\tilde{\Phi}^T\tilde{\epsilon}\right)_i\right] \\
&= \frac{1}{\mu n^2}e_i^T\left(\tilde{\Phi}^T\Sigma\tilde{\Phi}\right)e_i \text{ where } (e_i)_k = \begin{cases} 1 \text{ if } i=k \\ 0 \text{ else} \end{cases} \\
&= \frac{1}{\mu n^2}(\tilde{\Phi}e_i)^T\Sigma(\tilde{\Phi}e_i) \\
&\leq \frac{1}{\mu n^2}\sigma^2\left\|\tilde{\Phi}e_i\right\|_2^2 \\
&= \frac{1}{n\mu}\sigma^2 e_i^T\Psi e_i \\
&\leq \frac{1}{n\mu}\sigma^2(1+\delta) \text{ as } \mathcal{B} \text{ holds.}
\end{aligned}$$

Now consider the following event:

$$\mathcal{A} = \bigcap_{i=1}^{p}\{|Z_i| \leq \frac{\tilde{r}}{2}\}.$$

We have:

$$\mathbb{P}(\mathcal{A}\cap\mathcal{B}) = \mathbb{P}(\mathcal{A}|\mathcal{B})\mathbb{P}(\mathcal{B}).$$

From union bound and standard results on Gaussian tails, we get:

$$\begin{aligned}
\mathbb{P}(\mathcal{A}|\mathcal{B}) &\geq 1 - p\exp\left(-\frac{n\mu}{2\sigma^2(1+\delta)}\left(\frac{\tilde{r}}{2}\right)^2\right) \\
&\geq 1 - p^{1-\frac{A^2}{8}}
\end{aligned}$$

by using (15). Hence, step 2 gives:

$$\mathbb{P}(\mathcal{A}\cap\mathcal{B}) \geq \left(1-p^{1-A^2/8}\right)\left(1-2\exp\left(-2n\delta^2\mu^2+\ln p\right)\right).$$

Remark: following Remark 3 (given after the statement of the proven theorem), one can precisely account for the non-gaussianity of the ϵ noise by subtracting a correction term to minor $\mathbb{P}(\mathcal{A}|\mathcal{B})$, by using the Berry-Esseen theorem for the $\hat{S}$ estimator given in [11].

Now suppose that $\mathcal{A}\cap\mathcal{B}$ is realized. We have:

$$\frac{1}{n}\left\|\tilde{\Phi}^T\epsilon\right\|_\infty \leq \frac{\tilde{r}}{2},$$

where

$$\|v\|_\infty = \max|v_i|.$$

Set $\Delta = S - \widehat{S}$. We have:

$$\begin{aligned}
\|\Psi\Delta\|_\infty &= \frac{1}{n}\left\|\tilde{\Phi}^T\tilde{\Phi}\Delta\right\|_\infty, \\
&= \frac{1}{n}\left\|\tilde{\Phi}^T\tilde{\Phi}S - \tilde{\Phi}^T\tilde{\Phi}\widehat{S}\right\|_\infty, \\
&= \frac{1}{n}\left\|\tilde{\Phi}^T\tilde{E} - \tilde{\Phi}^T\epsilon - \tilde{\Phi}^T\tilde{\Phi}\widehat{S}\right\|_\infty, \\
&\leq \frac{1}{n}\left\|\tilde{\Phi}^T\left(\tilde{E} - \tilde{\Phi}\widehat{S}\right)\right\|_\infty + \frac{1}{n}\left\|\tilde{\Phi}^T\epsilon\right\|_\infty.
\end{aligned}$$

As the Dantzig constraint:

$$\left\|\frac{1}{n}\tilde{\Phi}^T\left(\tilde{E} - \tilde{\Phi}\widehat{S}\right)\right\|_\infty \leq \tilde{r}$$

holds, see [16], we have:

$$\|\Psi\Delta\|_\infty \leq \frac{3\tilde{r}}{2}. \tag{16}$$

Step 4: Control of $\|\Delta\|_1$

Step 4a: Majoration of $\Delta^T\Psi\Delta$. We have, on the event $\mathcal{A}\cap\mathcal{B}$:

$$\begin{aligned}
\left|\Delta^T\Psi\Delta\right| &\leq \|\Psi\Delta\|_\infty \|\Delta\|_1 \\
&\leq \frac{3\tilde{r}}{2}\left(\|\Delta_J\|_1 + \|\Delta_{J^c}\|_1\right),
\end{aligned}$$

by introducing the Δ_J and Δ_{J^c} vectors defined by:

$$(\Delta_J)_i = \begin{cases} \Delta_i \text{ if } S_i \neq 0 \\ 0 \text{ else} \end{cases} \qquad (\Delta_{J^c})_i = \begin{cases} 0 \text{ if } S_i \neq 0 \\ \Delta_i \text{ else} \end{cases}$$

We recall that $\|\Delta_{J^c}\|_1 \leq 3\|\Delta_J\|_1$ (see [16], Lemma 1, (9)). Hence, on $\mathcal{A}\cap\mathcal{B}$,

$$\left|\Delta^T\Psi\Delta\right| \leq 6\tilde{r}\|\Delta_J\|_1. \tag{17}$$

Step 4b: Minoration of $\Delta^T\Psi\Delta$. Let's introduce the circulant matrix M:

$$M = \begin{pmatrix} 1 & \mu & \cdots & \mu \\ \mu & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mu \\ \mu & \cdots & \mu & 1 \end{pmatrix}$$

whose smallest eigenvalue is $1-\mu$ (see [12]). Hence:

$$\begin{aligned}
\Delta^T\Psi\Delta &= \Delta^T M\Delta + \Delta^T(\Psi - M)\Delta \\
&\geq (1-\mu)\|\Delta\|_2^2 - |\Delta^T(\Psi - M)\Delta| \\
&\geq (1-\mu)\|\Delta_J\|_2^2 - |\Delta^T(\Psi - M)\Delta| \\
&\geq \frac{1-\mu}{s}\|\Delta_J\|_1^2 - |\Delta^T(\Psi - M)\Delta|,
\end{aligned}$$

since Δ_J has s nonzero components. We have:

$$|\Delta^T(\Psi - M)\Delta| \leq \|\Delta\|_1 \|(\Psi - M)\Delta\|_\infty \leq 4 \left\|\Delta_J\right\|_1 \|(\Psi - M)\Delta\|_\infty . \tag{18}$$

Now define the event:

$$\mathcal{C} = \left\{ \max_{\substack{i=1,\dots,p \\ j=1,\dots,p \\ j\neq i}} |\Psi_{ij} - \mu| \geq \delta \right\}.$$

It is clear that, on $\mathcal{B} \cap \mathcal{C}$, all entries of $\Psi - M$ are absolutely bounded by δ. Hence, on $\mathcal{B} \cap \mathcal{C}$,

$$\|(\Psi - M)\Delta\|_\infty \leq \delta \|\Delta\|_1 \leq 4\delta \left\|\Delta_J\right\|_1 ,$$

and, by (18):

$$|\Delta^T(\Psi - M)\Delta| \leq 16\delta \left\|\Delta_J\right\|_1^2 ,$$

which gives:

$$\Delta^T \Psi \Delta \geq \left(\frac{1-\mu}{s} - 16\delta \right) \left\|\Delta_J\right\|_1^2 . \tag{19}$$

Step 4c: Majoration of $\|\Delta\|_1$. By using (17) and (19), we get that on $\mathcal{A} \cap \mathcal{B} \cap \mathcal{C}$:

$$\left\|\Delta_J\right\|_1 \leq \frac{6\tilde{r}}{\frac{1-\mu}{s} - 16\delta},$$

hence:

$$\|\Delta\|_1 \leq \frac{24\tilde{r}}{\frac{1-\mu}{s} - 16\delta}. \tag{20}$$

Step 5: Majoration of $\|\Delta\|_\infty$

On $\mathcal{A} \cap \mathcal{B} \cap \mathcal{C}$, we have:

$$\begin{aligned}
\|\Delta\|_\infty &\leq \|\Psi\Delta\|_\infty + \|\Psi\Delta - \Delta\|_\infty \\
&\leq \frac{3\tilde{r}}{2} + \|(\Psi - \mathrm{Id})\Delta\|_\infty \quad \text{by using (16)} \\
&\leq \frac{3\tilde{r}}{2} + (\mu + \delta) \|\Delta\|_1 \quad \text{since each entry in } \Psi - \mathrm{Id} \text{ is less than } \mu + \delta \\
&\leq \left(\frac{3}{2} + \frac{24(\mu+\delta)}{\frac{1-\mu}{s} - 16\delta} \right) \tilde{r} \quad \text{by using (20)}
\end{aligned}$$

To finish, it is easy to see, using step 2, that $\mathbb{P}(\mathcal{A} \cap \mathcal{B} \cap \mathcal{C}) \geq 1 - \alpha$. □

A.2 Proof of Theorem 2

We rely on the result of [16]. Observe that:

$$\Psi = \frac{1}{n}\Phi^T\Phi.$$

We have for all $j = 1, \ldots, p$, $j \neq i$:

$$\mathbb{E}(\Psi_{ij}) = \frac{1}{n}\sum_{k=1}^{n}\mathbb{E}(\tilde{\Phi}_{ki}\tilde{\Phi}_{kj}) = 0.$$

Hence, for any $\delta > 0$, Hoeffding's inequality and union bound give:

$$\mathbb{P}\left(\max_{\substack{i=1,\ldots,p\\ j=1,\ldots,p\\ j\neq i}} |\Psi_{ij}| \geq \delta\right) \leq \exp\left(-n\frac{\delta^2}{2} + 2\ln p\right).$$

We also notice that $\Psi_{ii} = 1$ for all i. Hence, Assumptions 1 and 2 of Theorem 1 in [16] are satisfied with probability as described in the statement of the theorem.

B Exact support recovery using Thresholded-Lasso

B.1 A new result

We begin with two lemmas.

Lemma 1 (Lemma A.2 in [7]). *Let $r > r_0 > 0$ and $\hat{S}$ a solution to (7) with regularizing parameter r. Set $\Delta = \hat{S} - S$. On the event $\{(1/n)\|\Phi^\top\epsilon\|_\infty \leq r_0\}$, it holds that for all $T \subseteq \{1, \ldots, p\}$ such that $|T| \leq s$,*

$$\frac{1}{2r}\left[\frac{1}{2n}\|\Phi\Delta\|_2^2 + (r - r_0)\|\Delta\|_1\right] \leq \|\Delta_T\|_1 + \|S_{T^c}\|_1. \tag{21}$$

Proof. By optimality in (7), we get:

$$\frac{1}{2n}\|E - \Phi\hat{S}\|_2^2 + r\|\hat{S}\|_1 \leq \frac{1}{2n}\|\epsilon\|_2^2 + r\|S\|_1.$$

It yields,

$$\frac{1}{2n}\|\Phi\Delta\|_2^2 - \frac{1}{n}\langle\Phi^\top\epsilon, \Delta\rangle + r\|\hat{S}\|_1 \leq r\|S\|_1.$$

Let $T \subseteq \{1, \ldots, p\}$ such that $|T| \leq s$. We assume that $(1/n)\|\Phi^\top\epsilon\|_\infty \leq r_0$. Invoking Hölder's inequality, we have:

$$\frac{1}{2n}\|\Phi\Delta\|_2^2 + r\|\hat{S}_{J^c}\|_1 \leq r(\|S_J\|_1 - \|\hat{S}_J\|_1) + r\|S_{J^c}\|_1 + r_0\|\Delta\|_1.$$

Adding $r\|S_{J^c}\|_1$ on both sides, observe that:

$$\frac{1}{2n}\|\Phi\Delta\|_2^2 + (r - r_0)\|\Delta_{J^c}\|_1 \leq (r + r_0)\|\Delta_J\|_1 + 2r\|S_{J^c}\|_1. \tag{22}$$

Adding $(r - r_0)\|\Delta_J\|_1$ on both sides, we conclude the proof. □

Lemma 2 (Theorem 2.1 in [7]). *Assume that for all $\gamma \in \mathbb{R}^p$, for all $T \subseteq \{1, \dots, p\}$ such that $|T| \leq s$,*

$$\|\gamma_T\|_1 \leq \rho\sqrt{s}\,\|\Phi\gamma\|_2 + \kappa\,\|\gamma\|_1\,. \tag{23}$$

where $\rho > 0$ and $1/2 > \kappa > 0$. Moreover, assume that the regularizing parameter r of the convex program (7) enjoys $r > r_0/(1-2\kappa)$. Then, on the event $\{(1/n)\|\Phi^\top\epsilon\|_\infty \leq r_0\}$, any solution $\hat{S}$ to (7) satisfies:

$$\|\hat{S} - S\|_1 \leq \frac{2rn\rho^2 s}{1 - (r_0/r) - 2\kappa}\,.$$

Proof. Assume that $(1/n)\|\Phi^\top\epsilon\|_\infty \leq r_0$. Using (23) and (21) with $J = T$, the support of S, we get:

$$\frac{1}{2r}\Big[\frac{1}{2n}\|\Phi\Delta\|_2^2 + (r - r_0)\|\Delta\|_1\Big] \leq \rho\sqrt{s}\,\|\Phi\Delta\|_2 + \kappa\,\|\Delta\|_1\,,$$

where $\Delta = \hat{S} - S$. It yields,

$$\Big[\frac{1}{2}(1 - \frac{r_0}{r}) - \kappa\Big]\|\Delta\|_1 \leq -\frac{1}{4rn}\|\Phi\Delta\|_2^2 + \rho\sqrt{s}\,\|\Phi\Delta\|_2 \leq rn\rho^2 s\,,$$

using the fact that the polynomial $x \mapsto -1/(4rn)x^2 + \rho\sqrt{s}x$ is not greater than $rn\rho^2 s$. □

We deduce the following new result on exact support recovery when using Thresholded-Lasso.

Lemma 3 (Exact support recovery with Thresholded-Lasso). *Assume that for all $\gamma \in \mathbb{R}^p$, for all $T \subseteq \{1, \dots, p\}$ such that $|T| \leq s$,*

$$\|\gamma_T\|_1 \leq \rho\sqrt{s}\,\|\Phi\gamma\|_2 + \kappa\,\|\gamma\|_1\,.$$

where $\rho > 0$ and $1/2 > \kappa > 0$. Moreover, assume that:

$$\max_{1 \leq k \neq l \leq p} \frac{1}{n}\Big|\sum_{j=1}^n \Phi_{j,k}\Phi_{j,l}\Big| \leq \theta_1 \quad \text{and} \quad \forall i,\ \frac{1}{n}\|\Phi_i\|_2^2 \geq \theta_2\,,$$

where Φ_i denotes the columns of Φ. Let $r_0 > 0$ and suppose that the regularizing parameter r of the convex program (7) enjoys:

$$r > \frac{r_0}{1 - 2\kappa}\,.$$

Then, on the event $\{(1/n)\|\Phi^\top\epsilon\|_\infty \leq r_0\}$, any solution $\hat{S}$ to (7) satisfies:

$$\|\hat{S} - S\|_\infty \leq \frac{1}{\theta_2}\Big[1 + \frac{r_0}{r} + \frac{2n\theta_1\rho^2 s}{1 - (r_0/r) - 2\kappa}\Big]r\,.$$

Proof. The first order optimality conditions of the convex program (7) shows that there exists $\tau \in \mathbb{R}^p$ such that $\|\tau\|_\infty \leq 1$ and:

$$\frac{1}{n}\Phi^\top(E - \Phi\hat{S}) = r\tau\,.$$

Set $\Delta = \hat{S} - S$ and $\Psi = (1/n)\,\Phi^\top\Phi$. We assume that $(1/n)\|\Phi^\top\epsilon\|_\infty \leq r_0$. It holds:

$$\|\Psi\Delta\|_\infty \leq r + r_0\,. \tag{24}$$

Moreover, Lemma 2 shows that:

$$\|\Delta\|_1 \leq \frac{2nr\rho^2 s}{1 - r_0/r - 2\kappa}\,. \tag{25}$$

Since each entry in the matrix $\Psi - \mathrm{Diag}(\|\Phi_1\|_2^2/n, \ldots, \|\Phi_p\|_2^2/n)$ is less than θ_1, we deduce that:

$$\begin{aligned}
\theta_2\|\Delta\|_\infty &\leq \|\Psi\Delta\|_\infty + \|(\Psi - \mathrm{Diag}(\|\Phi_1\|_2^2/n, \ldots, \|\Phi_p\|_2^2/n))\Delta\|_\infty\,,\\
&\leq r + r_0 + \theta_1\|\Delta\|_1\,,\\
&\leq \Big[1 + \frac{r_0}{r} + \frac{2n\theta_1\rho^2 s}{1 - r_0/r - 2\kappa}\Big]r\,,
\end{aligned}$$

using (24) and (25). □

B.2 Expander graphs

This subsection is devoted to a new proof of support recovery of Thresholded-Lasso when using adjacency matrices. Given the binary constraint, we choose Φ as the adjacency matrix of a bi-partite simple graph $G = (A, B, E)$ where $A = \{1, \ldots, p\}$, $B = \{1, \ldots, n\}$ and $E \subseteq A \times B$ denotes the set of edges between A and B. In this model, $(\Phi_{ji})_{j,i}$ is equal to 1 if there exists an edge between $j \in B$ and $i \in A$, and 0 otherwise. Assume that G is left regular with degree d, i.e. Φ has exactly d ones per column. Consider unbalanced expander graphs defined as follows.

Definition 1 ((s,e)-unbalanced expander)**.** *A (s,e)-unbalanced expander is a bi-partite simple graph $G = (A, B, E)$ with left degree d such that for any $I \subset A$ with $\#I \leq s$, the set of neighbors $N(I)$ of I has size:*

$$\#N(I) \geq (1 - e)\,d\,\#I\,. \tag{26}$$

The parameter e is called the expansion constant.

We recall that expander graphs satisfy the UDP property, see the following lemma.

Lemma 4. *Let $\Phi \in \mathbb{R}^{n\times p}$ be the adjacency matrix of a $(2s,e)$-unbalanced expander with an expansion constant $e < 1/2$ and left degree d. If the quantities $1/e$ and d are smaller than p then Φ satisfies for all $\gamma \in \mathbb{R}^p$ and for all $T \subseteq \{1, \ldots, p\}$ such that $|T| \leq s$,*

$$\|\gamma_T\|_1 \leq \frac{\sqrt{s}}{(1 - 2e)\sqrt{d}}\,\|\Phi\gamma\|_2 + \frac{2e}{1 - 2e}\,\|\gamma\|_1\,.$$

Proof. For sake of completeness, we present the proof given in [8]. Without loss of generality, we can assume that T consists of the largest, in magnitude, coefficients of γ. We partition the coordinates into sets T_0, T_1, T_2, ... ,T_q, such that the coordinates in the set T_l are not larger than the coordinates in T_{l-1},

$l \geq 1$, and all sets but the last one T_q have size s. Observe that we can choose $T_0 = T$. Let Φ' be a sub matrix of Φ containing rows from $N(T)$, the set of neighbors of T. Using Cauchy-Schwartz inequality, it holds

$$\sqrt{sd}\,\|\Phi\gamma\|_2 \geq \sqrt{sd}\,\|\Phi'\gamma\|_2 \geq \frac{\sqrt{sd}}{\sqrt{|N(T)|}}\,\|\Phi'\gamma\|_1 \geq \|\Phi'\gamma\|_1 .$$

From [1], we know that:

$$\|\Phi\gamma_T\|_1 \geq d(1-2e)\,\|\gamma_T\|_1 , \tag{27}$$

Moreover, Eq. (27) gives:

$$\begin{aligned}
\sqrt{sd}\,\|\Phi\gamma\|_2 &\geq \|\Phi'\gamma\|_1 , \\
&\geq \|\Phi'\gamma_T\|_1 - \sum_{l\geq 1}\sum_{(i,j)\in E, i\in T_l, j\in N(T)} |\gamma_i| , \\
&\geq d(1-2e)\,\|\gamma_T\|_1 - \sum_{l\geq 1} |E\cap(T_l\times N(T))|\min_{i\in T_{l-1}}|\gamma_i| , \\
&\geq d(1-2e)\,\|\gamma_T\|_1 - \frac{1}{s}\sum_{l\geq 1} |E\cap(T_l\times N(T))|\,\|\gamma_{T_{l-1}}\|_1 .
\end{aligned}$$

From the expansion property (26), it follows that, for $l \geq 1$, we have:

$$|N(T\cup T_l)| \geq d(1-e)|T\cup T_l| .$$

Hence at most $de2s$ edges can cross from T_l to $N(T)$, and so:

$$\begin{aligned}
\sqrt{sd}\,\|\Phi\gamma\|_2 &\geq d(1-2e)\,\|\gamma_T\|_1 - de2\sum_{l\geq 1}\|\gamma_{T_{l-1}}\|_1 /s, \\
&\geq d(1-2e)\,\|\gamma_T\|_1 - 2de\,\|\gamma\|_1 .
\end{aligned}$$

□

Observe the columns Φ_i of the adjacency matrix Φ have small ℓ_2-norm compared to the ℓ_2-norm of the noise, namely:

$$\|\Phi_i\|_2^2 = d \ll \sigma^2 n = \mathbb{E}(\|\epsilon\|_2^2) .$$

A standard hypothesis in the exact recovery frame [2, 3] is that the signal-to-noise ratio is close to one. This hypothesis is often presented as the empirical covariance matrix has diagonal entries equal to 1. However, in our setting, the signal-to-noise ratio goes to zero and eventually we observe only noise. To prevent this issue, we use a noise model adapted to the case of sparse designs. Hence, we assume subsequently that the noise level is comparable to the signal power:

$$\forall i \in \{1,\dots,n\}, \quad \epsilon_i \text{ is Gaussian and } \mathbb{V}\mathbf{ar}(\epsilon_i) \leq \tilde{\sigma}^2\frac{\|\Phi_i\|_2^2}{n} , \tag{28}$$

so that $\|\Phi_i\|_2^2/\mathbb{E}(\|\epsilon\|_2^2) \geq 1/\tilde{\sigma}^2$.

Theorem 4 (Exact recovery with expander graphs)**.** *Let $A > \sqrt{2}$ and $\Phi \in \{0,1\}^{n\times p}$ be the adjacency matrix of a $(2s,e)$-expander graph with expansion constant $1/p < e < 1/6$ and left degree d. Assume that* (28) *holds. Let $\hat{S}$ be any solution to* (7) *with regularizing parameter:*

$$r \geq r_1 := 2A\tilde{\sigma}\Big[\frac{1-2e}{1-6e}\Big]\Big[\frac{d(\log p)^{1/2}}{n^{3/2}}\Big]\,,$$

Then, with probability greater than $1-p^{1-A^2/2}$, it holds:

$$\|\hat{S}-S\|_\infty \leq A\tilde{\sigma}\Big[\frac{\log p}{n}\Big]^{1/2}\Big[1+\frac{2(1-2e)}{1-6e}+\frac{16es}{(1-6e)^2}\Big]\frac{r}{r_1}\,.$$

Proof. Lemma 4 shows that for all $\gamma \in \mathbb{R}^p$ and for all $T \subseteq \{1,\ldots,p\}$ such that $|T| \leq s$,

$$\|\gamma_T\|_1 \leq \frac{\sqrt{s}}{\sqrt{d}(1-2e)}\,\|\Phi\gamma\|_2 + \frac{2e}{1-2e}\,\|\gamma\|_1\,.$$

Moreover, the expansion property implies:

$$\max_{1\leq k\neq l\leq p}\frac{1}{n}|\sum_{j=1}^{n}\Phi_{j,k}\Phi_{j,l}| \leq \frac{2de}{n}\,.$$

Lemma 3 with $1/\rho = (1-2e)\sqrt{d}$, $\kappa = 2e/(1-2e)$, $\theta_1 = 2de/n$ and $\theta_2 = d/n$, shows that for all regularizing parameter $r \geq 2r_0(1-2e)/(1-6e)$,

$$\|\hat{S}-S\|_\infty \leq \frac{n}{d}\Big[1+\frac{1-6e}{2(1-2e)}+\frac{8es}{(1-2e)(1-6e)}\Big]r\,,$$

on the event $\{(1/n)\|\Phi^\top\epsilon\|_\infty \leq r_0\}$. Finally, set $r_0 = A\tilde{\sigma}d(\log p)^{1/2}/n^{3/2}$ and $Z_i = (1/n)\Phi^\top\epsilon$. Observe that Z_i is centered Gaussian random variable with variance less than $\tilde{\sigma}^2\|\Phi_i\|_2^4/n^3$. Taking union bounds, it holds:

$$\begin{aligned}
\mathbb{P}[(1/n)\|\Phi^\top\epsilon\|_\infty > r_0] &\leq \mathbb{P}[(1/n)\|\Phi^\top\epsilon\|_\infty > A\tilde{\sigma}d(\log p)^{1/2}/n^{3/2}]\,,\\
&\leq \sum_{i=1}^{p}\mathbb{P}[|Z_i| > A\tilde{\sigma}d(\log p)^{1/2}/n^{3/2}]\,,\\
&= \sum_{i=1}^{p}\mathbb{P}[|Z_i| > (\tilde{\sigma}\|\Phi_i\|_2^2/n^{3/2})\,A\sqrt{\log p}]\,,\\
&\leq p^{1-A^2/2}\,,
\end{aligned}$$

using $\|\Phi_i\|_2^2 = d$ and the fact that, for $A\sqrt{\log p} \geq \sqrt{2\log 2}$, we have:

$$\mathbb{P}[|\mathcal{N}(0,1)| > A\sqrt{\log p}] \leq \frac{1}{\sqrt{\pi\log 2}}\exp(-c\log p) \leq p^{-A^2/2}\,.$$

□

Note that, with high probability, a random bi-partite simple graph is a (s,e)-unbalanced expander. As a matter of fact, we have the following result using Chernoff bounds and Hoeffding's inequality, see [27] for instance.

Proposition 1. *Consider $e > 0$, $c > 1$ and $p \geq 2s$. Then, with probability greater than $1 - s\exp(-c\log p)$, a uniformly chosen bi-partite simple graph $G = (A, B, E)$ with $|A| = p$, left degree d such that:*

$$d \leq C_1(c,e)\log p\,, \tag{29}$$

and number of right side vertices, namely $n = |B|$, such that:

$$n \geq C_2(c,e)\,s\log p\,, \tag{30}$$

where $C_1(c,e), C_2(c,e)$, do not depend on s but may depend on e, is a (s,e)-unbalanced expander graph.

Hence we deduce the following corollary of Theorem 4.

Corollary 2. *Consider $c > 1$, $p \geq 4s$ and choose $e = 1/12$. Let $\Phi \in \{0,1\}^{n\times p}$ be drawn uniformly according to Proposition 1 so that $d \leq C_1 \log p$ and:*

$$n \geq n_0 := C_2\,s\log p\,, \tag{31}$$

with C_1, C_2 constants that depend only on c. Let $A > [\min(C_1, 2)]^{1/2}$. Let $\hat{S}$ be any solution to (7) *with regularizing parameter:*

$$r \geq r_1 := 3.34A\tilde{\sigma}\Big[\frac{\log p}{n}\Big]^{3/2}\,,$$

Then, with probability greater than $1 - p^{1-A^2/2} - 2s\exp(-c\log p)$, it holds:

$$\|\hat{S} - S\|_\infty \leq 51.7\,A\,C_2^{-1/2}\,\tilde{\sigma}\,\Big[\frac{r}{r_1}\Big]\,\Big[\frac{n_0}{n}\Big]^{1/2}\sqrt{s}\,. \tag{32}$$

Remark. *Observe that* (32) *is also consistent with the regime $n = \mathcal{O}(s^2\log p)$. In this case, we uncover that $\|\hat{S} - S\|_\infty \leq$ (cst) $\tilde{\sigma}$. Namely, the thresholded lasso faithfully recovers the support of entries whose magnitudes are above the noise level.*

B.3 Bernoulli designs

We can relax the hypothesis on the left-regularity using a Bernoulli design that mimics the uniform probability on d-regular graphs. This model is particularly interesting since one can easily generate a design matrix Φ.

Recall we consider a Bernoulli distribution with parameter $\mu \in (0,1)$ and $(\Phi_{ji})_{j,i}$ are independently drawn with respect to this distribution, with for all i, j, it holds $\mathbb{P}(\Phi_{ji} = 1) = \mu = 1 - \mathbb{P}(\Phi_{ji} = 0)$. We begin with some preliminaries lemmas.

Lemma 5. *Let $p, n > 0$. Let $c > 1$. Let $\Phi \in \{0,1\}^{n\times p}$ a Bernoulli matrix drawn according to* (8) *with:*

$$\mu = 799(1+c)\frac{\log p}{n}\,.$$

If $n \geq 799(c+1)\log p$ then Φ satisfies for all $i \in \{1,\dots,p\}$,

$$759(1+c)\log p \leq \|\Phi_i\|_0 = \|\Phi_i\|_2^2 \leq 828(1+c)\log p\,, \tag{33}$$

and

$$\max_{1\leq k\neq l\leq p}\frac{1}{n}|\sum_{j=1}^{n}\Phi_{j,k}\Phi_{j,l}|\leq 879(1+c)\frac{\log p}{n}\,,$$

with a probability greater than $1-(1+2p)p^{-c}$.

Proof. Let $i\in\{1,\dots,p\}$ and consider $Y_i=\Phi_{1,i}+\ldots+\Phi_{n,i}$. Observe that Chernoff bound reads:

$$\mathbb{P}(Y_i\geq n(\mu+\delta))\leq\exp(-n\,\mathbf{H}(\mu+\delta\|\mu))$$

where $\mathbf{H}(a\|b)$ denotes the Kullback-Leibler divergence between two Bernoulli random variables with parameter a and b, namely:

$$\mathbf{H}(a\|b)=a\log(a/b)+(1-a)\log((1-a)/(1-b))\,.$$

Observe that the second derivative of $x\mapsto\mathbf{H}(\mu+x\|\mu)$ is equal to $1/((\mu+x)(1-\mu-x))$ and is bounded from below by $1/(\mu+\delta)$ on $[\mu,\mu+\delta]$. Therefore,

$$\mathbf{H}(\mu+\delta\|\mu)\geq\frac{\delta^2}{2(\mu+\delta)}\,.\tag{34}$$

Using union bound, we get that:

$$\mathbb{P}[\forall i\,,\;Y_i\leq 1.036009\,n\mu]\geq 1-\exp[\log p-0.001252\,n\mu]\geq 1-p^{-(c-1)}\,,$$

as desired. Similarly, one get that:

$$\mathbf{H}(\mu-\delta\|\mu)\leq\frac{\delta^2}{2\mu}\,,\tag{35}$$

and so:

$$\mathbb{P}[\forall i\,,\;Y_i\geq 0.05004\,n\mu]\geq 1-\exp[\log p-0.001252\,n\mu]\geq 1-p^{-(c-1)}\,.$$

The second inequality follows from the same analysis:

$$\mathbb{P}[\forall k\neq l\,,\;\frac{1}{n}\sum_{j=1}^{n}\Phi_{j,k}\Phi_{j,l}\geq\mu^2+0.1\mu)]\leq\exp[\log[\frac{p(p-1)}{2}]-\frac{n}{200(1+1/(10\mu))}]\,.$$

Observe that $\log(p(p-1)/2)\leq 2\log p$, $1+1/(10\mu)\leq 1.01/\mu$ and $\mu^2+0.1\mu\leq 1.01\mu$. Therefore,

$$\mathbb{P}[\forall k\neq l\,,\;\frac{1}{n}\sum_{j=1}^{n}\Phi_{j,k}\Phi_{j,l}\leq 1.1\mu)]\geq 1-\exp(-1.5(1+c)\log p)\,.$$

□

Lemma 6. *Let* $p,s>0$ *and* $c>1$. *Let* $\Phi\in\{0,1\}^{n\times p}$ *a Bernoulli matrix drawn according to* (8) *with* $\mu=799(1+c)\log p/n$ *and:*

$$n\geq 6491(1+c)s\log p\,.$$

Then, with a probability greater than $1 - p^{-cs}$*, the matrix* Φ *satisfies the following vertex expansion property:*

$$\#\{\mathrm{Supp}(\Phi \mathbb{1}_U)\} \geq (6/7)\, \mathrm{d}_{max} \#U\,, \tag{36}$$

where U *is a subset of* $\{1, \ldots, p\}$ *of size* s*,* $\mathbb{1}_U \in \mathbb{R}^p$ *denotes the vector with entry* 1 *on* U *and* 0 *elsewhere, and* $\mathrm{d}_{max} = 828(1+c)\log p$ *is the maximal support size of one column of* Φ *as shown in* (33)*.*

Proof. The number of subsets of size s can be upper bounded by $\exp(s \log p)$. Observe that the left hand side of (36) is a random variable N_n with the same law as:

$$N_n \overset{d}{=} \sum_{i=1}^{n} Z_i \quad \text{where } Z_i \overset{i.i.d}{\sim} \mathcal{B}(\nu)\,,$$

where the Bernoulli parameter $\nu = 1 - (1-\mu)^s$. Using (35), we get that:

$$\mathbb{P}(N_n \leq n(\nu - \delta)) \leq \exp(-n\delta^2/(2\nu))\,.$$

Set $\delta := 1 - 0.8883 s\mu - \exp(-s\mu)$ and observe that it holds $s\mu \leq 0.1231$, $\nu \geq 1 - \exp(-s\mu)$, and $\delta \geq s\mu(0.1117 - 0.5 s\mu) \geq 0.0501 s\mu$. We deduce that:

$$\mathbb{P}(N_n \leq 0.8883 ns\mu) \leq \exp(-0.001255 ns\mu) \leq \exp(-(c+1)s\log p)\,,$$

using $\delta \geq 0.0501 s\mu$ and $\nu \leq s\mu$. □

Lemma 7. *Let* $p > 7$*,* $s > 0$ *and* $c > 1$*. Let* $\Phi \in \{0,1\}^{n \times p}$ *a Bernoulli matrix drawn according to* (8) *with:*

- $\mu = 799(1+c)\dfrac{\log p}{n}$,
- $n \geq n_0 := 12982(1+c)s\log p$.

Then, with a probability greater than $1 - (1 + 2p + (1-p^{-c})^{-1})p^{-c}$*, the matrix* Φ *satisfies for all* $\gamma \in \mathbb{R}^p$ *and for all* $T \subseteq \{1, \ldots, p\}$ *such that* $|T| \leq s$*,*

$$\|\gamma_T\|_1 \leq 0.0551 \Big[\frac{s}{(1+c)\log p}\Big]^{1/2} \|\Phi\gamma\|_2 + 0.4529\, \|\gamma\|_1\,.$$

Proof. From Lemma 6, we get that Φ is the adjacency matrix of a $(2s, 1/7)$-expander graph with left degree d enjoying (33), namely $\mathrm{d}_{min} \leq d \leq \mathrm{d}_{max}$ with $\mathrm{d}_{min} = 759(1+c)\log p$ and $\mathrm{d}_{max} = 828(1+c)\log p$. Observe that the left degree d may depend on the vertex considered. However, note that the proof of Lemma 4 can be extended to this case. Following the lines of Lemma 9 in [1], one can check that:

$$\|\Phi\gamma_T\|_1 \geq \mathrm{d}_{min}(1 - 2(\mathrm{d}_{max}/\mathrm{d}_{min})e)\, \|\gamma_T\|_1\,.$$

Similarly, one can check from the proof of Lemma 4 that for all $\gamma \in \mathbb{R}^p$ and for all $T \subseteq \{1, \ldots, p\}$ such that $|T| \leq s$,

$$\|\gamma_T\|_1 \leq \frac{\sqrt{s}\sqrt{\mathrm{d}_{max}}}{\mathrm{d}_{min}(1 - 2(\mathrm{d}_{max}/\mathrm{d}_{min})e)} \|\Phi\gamma\|_2 + \frac{2\mathrm{d}_{max}e}{\mathrm{d}_{min}(1 - 2(\mathrm{d}_{max}/\mathrm{d}_{min})e)} \|\gamma\|_1\,,$$

where $e = 1/7$. □

We deduce the following result for Thresholded-Lasso using Bernoulli design matrices.

Theorem 5 (Exact recovery with Bernoulli designs)**.** *Let $p > 7$, $s > 0$ and $c > 1$. Let $\Phi \in \{0,1\}^{n\times p}$ a Bernoulli matrix drawn according to* (8) *with:*

- $\mu = 799(1+c)\log p/n$,
- $n \geq 12982(1+c)s\log p$,
- $\epsilon \sim \mathcal{N}(0,\Sigma_n)$ *and the covariance diagonal entries enjoy* $\Sigma_{i,i} \leq \sigma^2$.

Let $\hat{S}$ be any solution to (7) *with regularizing parameter:*

$$r \geq r_1 := 9692\,\sigma\,(1+c)\frac{\log p}{n}\,,$$

Then, with a probability greater than $1-3p^{1-c}$,

$$\|\hat{S}-S\|_\infty \leq 775.36\left[\frac{r}{r_1}\right]\sigma\, s\,.$$

Proof. • Invoke Lemma 5 to get that for all $i \in \{1,\dots,p\}$,

$$759(1+c)\log p \leq \|\Phi_i\|_0 = \|\Phi_i\|_2^2 \leq 828(1+c)\log p\,,$$

and:

$$\max_{1\leq k\neq l\leq p}\frac{1}{n}|\sum_{j=1}^{n}\Phi_{j,k}\Phi_{j,l}| \leq 879(1+c)\frac{\log p}{n}\,.$$

• Set $r_0 = 6\sigma(46c(1+c))^{1/2}\log p/n$ and $Z_i = (1/n)\Phi^\top\epsilon$. Note Z_i is centered Gaussian random variable with variance less than $\sigma^2\|\Phi_i\|_2^2/n^2$. Taking union bounds, it holds:

$$\begin{aligned}
\mathbb{P}[(1/n)\|\Phi^\top\epsilon\|_\infty > r_0] &\leq \mathbb{P}[(1/n)\|\Phi^\top\epsilon\|_\infty > 6\sigma\sqrt{46c(1+c)}\log p/n]\,,\\
&\leq \sum_{i=1}^{p}\mathbb{P}[|Z_i| > 6\sigma\sqrt{46c(1+c)}\log p/n]\,,\\
&\leq \sum_{i=1}^{p}\mathbb{P}[|Z_i| > \sqrt{2c\log p}\,\sigma\|\Phi_i\|_2/n]\,,\\
&\leq p^{1-c}\,,
\end{aligned}$$

using $\|\Phi_i\|_2^2 \leq 828(1+c)\log p$ and the fact that, for $\sqrt{2c\log p} \geq \sqrt{2\log 2}$, we have:

$$\mathbb{P}[|\mathcal{N}(0,1)| > \sqrt{2c\log p}] \leq \frac{1}{\sqrt{\pi\log 2}}\exp(-c\log p) \leq p^{-c}\,.$$

• From Lemma 7, it holds that for all $\gamma \in \mathbb{R}^p$ and for all $T \subseteq \{1,\dots,p\}$ such that $|T| \leq s$,

$$\|\gamma_T\|_1 \leq 0.0551\Big[\frac{s}{(1+c)\log p}\Big]^{1/2}\|\Phi\gamma\|_2 + 0.4529\,\|\gamma\|_1\,.$$

• Invoke Lemma 3 with parameters $\rho = 0.0551/\sqrt{(1+c)\log p}$, $\kappa = 0.4529$, $\theta_1 = 879(1+c)\log p/n$ and $\theta_2 = 759(1+c)\log p/n$, to get that for all regularizing parameter $r > 10.616\, r_0$,

$$\|\hat{S} - S\|_\infty \leq \frac{n}{759(1+c)\log p}\Big[1 + \frac{r_0}{r} + \frac{5.338\, s}{0.0942 - r_0/r}\Big] r\,,$$

on the event $\{(1/n)\|\Phi^\top \epsilon\|_\infty \leq r_0\}$. Finally, observe that $r_1 \geq 238.1 r_0$. □

Corollary 3 (Exact recovery under constant SNR hypothesis)**.** *Let $p > 7$, $s > 0$ and $c > 1$. Let $\Phi \in \{0,1\}^{n\times p}$ a Bernoulli matrix drawn according to* (8) *with:*

- $\mu = 799(1+c)\log p/n$,
- $n \geq n_0 := 12982(1+c)s\log p$,
- *assume that* (28) *holds, namely $\epsilon \sim \mathcal{N}(0, \Sigma_n)$ and the covariance diagonal entries enjoy $\Sigma_{i,i} \leq 759\,\tilde{\sigma}^2\,(1+c)\log p/n$ with $\tilde{\sigma} > 0$.*

Let $\hat{S}$ be any solution to (7) *with regularizing parameter:*

$$r \geq 0.1886\,\tilde{\sigma}\Big[\frac{12982(1+c)\log p}{n}\Big]^{3/2},$$

Then, with a probability greater than $1 - 3p^{1-c}$,

$$\|\hat{S} - S\|_\infty \leq 195.82\,\tilde{\sigma}\left[\frac{r}{r_1}\right]\left[\frac{n_0}{n}\right]^{1/2}\sqrt{s}\,.$$

Proof. Eq. (33) shows that:

$$759(1+c)\log p \leq \|\Phi_i\|_2^2 \leq 828(1+c)\log p\,,$$

and so $\Sigma_{i,i} \leq \tilde{\sigma}^2\|\Phi_i\|_2^2/n$, with high probability. □

B.4 Rademacher Designs

The result and the proof given on Page 3 rely on the following lemmas.

Lemma 8 (Rademacher designs satisfy RIP)**.** *There exists universal constants c_1, c_2, c_3 such that the following holds. Let $\delta \in (0,1)$ and $p, n, s' > 0$ such that:*

$$s' = \Big\lfloor \frac{c_1\delta^2 n}{\log(c_2 p/(\delta^2 n))} \Big\rfloor,$$

then, with probability at least $1 - 2\exp(-c_3 n)$, a matrix $\Phi \in \{\pm 1\}^{n\times p}$ drawn according to the Rademacher model (12) *enjoy the RIP property, namely for all $\gamma \in \mathbb{R}^p$ such that $\|\gamma\|_0 \leq s'$,*

$$n(1-\delta)^2\|\gamma\|_2^2 \leq \|\Phi\gamma\|_2^2 \leq n(1+\delta)^2\|\gamma\|_2^2\,.$$

Proof. Numerous authors have proved this result, see Example 2.6.3 and Theorem 2.6.5 in [6] for instance. □

Lemma 9 (Rademacher designs satisfy UDP)**.** *There exists universal constants c_1, c_2, c_3 such that the following holds. Let $\delta \in (0, \sqrt{2}-1)$ and $s > 0$ such that:*

$$5s \leq s' := \Big\lfloor \frac{c_1 \delta^2 n}{\log(c_2 p/(\delta^2 n))} \Big\rfloor ,$$

then, with probability at least $1 - 2\exp(-c_3 n)$, a matrix $\Phi \in \{\pm 1\}^{n \times p}$ drawn according to the Rademacher model (12) *enjoy for all $\gamma \in \mathbb{R}^p$ and for all $T \subseteq \{1, \ldots, p\}$ such that $|T| \leq s$,*

$$\|\gamma_T\|_1 \leq \rho\sqrt{s}\,\|\Phi\gamma\|_2 + \kappa\,\|\gamma\|_1 .$$

where:

- $1/2 > \kappa > \big(1 + 2((1-\delta)/(1+\delta))^{\frac{1}{2}}\big)^{-1}$,
- $\sqrt{n}\rho = \big(\sqrt{1-\delta} + (\kappa_0 - 1)/(2\kappa_0)\sqrt{1+\delta}\big)^{-1}$.

Proof. The proof follows from Lemma 8 and Proposition 3.1 in [7]. □

Lemma 10. *Let $c, C_1 > 0$ and $p, n, s > 0$ such that $n \geq C_1 s \log p$ and $s \geq 3(2+c)/C_1$. Then, with probability greater than $1 - 2p^{-c}$, it holds for all $k \neq l \in \{1, \ldots, p\}$,*

$$\frac{1}{n}|\sum_{i=1}^{n} \Phi_{i,k}\Phi_{i,l}| \leq \Big[\frac{(2+c)8}{3C_1}\Big]^{1/2} \frac{1}{\sqrt{s}}.$$

Proof. Let $k \neq l \in \{1, \ldots, p\}$. Set $X_i = \Phi_{i,k}\Phi_{i,l}$ and observe that X_i are independent Rademacher random variables. From Bernstein's inequality, it holds for all $0 < t < 1$,

$$\mathbb{P}\Big(|\frac{1}{n}\sum_{i=1}^{n} X_i| \geq t\Big) \leq 2\exp\Big(-\frac{3n}{8}t^2\Big).$$

Set $t = ((2+c)8/(3sC_1))^{1/2}$ and observe $\#\{k \neq l\} \leq \exp(2\log p)$. □

References

[1] R. Berinde, A. C. Gilbert, P. Indyk, H. Karloff, and M. J. Strauss. Combining geometry and combinatorics: A unified approach to sparse signal recovery. In *Communication, Control, and Computing, 2008 46th Annual Allerton Conference on*, pages 798–805. IEEE, 2008.

[2] P. J. Bickel, Y. Ritov, and A. B. Tsybakov. Simultaneous analysis of lasso and dantzig selector. *The Annals of Statistics*, 37(4):1705–1732, 2009.

[3] P. L. Bühlmann and S. A. van de Geer. *Statistics for high-dimensional data*. Springer, 2011.

[4] E. J. Candès and Y. Plan. Near-ideal model selection by âĎŞ1 minimization. *The Annals of Statistics*, 37(5A):2145–2177, 2009.

[5] E. J. Candes, J. K. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Communications on pure and applied mathematics*, 59(8):1207–1223, 2006.

[6] D. Chafaı, O. Guédon, G. Lecué, and A. Pajor. Interactions between compressed sensing, random matrices, and high dimensional geometry. *to appear in "Panoramas et Synthèses" (SMF)*, 2013.

[7] Y. de Castro. A remark on the lasso and the dantzig selector. *Statistics and Probability Letters*, 2012.

[8] Y. de Castro. Optimal designs for lasso and dantzig selector using expander codes. *Arxiv preprint arXiv:1010.2457v5*, 2013.

[9] B. Efron, T. Hastie, I. Johnstone, and R. J. Tibshirani. Least angle regression. *The Annals of statistics*, 32(2):407–499, 2004.

[10] J-J Fuchs. On sparse representations in arbitrary redundant bases. *Information Theory, IEEE Transactions on*, 50(6):1341–1344, 2004.

[11] F. Gamboa, A. Janon, T. Klein, A. Lagnoux-Renaudie, and C. Prieur. Statistical inference for sobol pick freeze monte carlo method. *arXiv preprint arXiv:1303.6447*, 2013.

[12] R. Gray. Toeplitz and circulant matrices: A review. *http://ee.stanford.edu/~gray/toeplitz.pdf*, 2002.

[13] W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, March 1963.

[14] A. Janon, T. Klein, A. Lagnoux, M. Nodet, and C. Prieur. Asymptotic normality and efficiency of two sobol index estimators. Preprint available at `http://hal.inria.fr/hal-00665048/en`, 2012.

[15] R. Liu and A. B. Owen. *Estimating mean dimensionality*. Department of Statistics, Stanford University, 2003.

[16] K. Lounici. Sup-norm convergence rate and sign concentration property of lasso and dantzig estimators. *Electronic Journal of statistics*, 2:90–102, 2008.

[17] H. Monod, C. Naud, and D. Makowski. Uncertainty and sensitivity analysis for crop models. In D. Wallach, D. Makowski, and J. W. Jones, editors, *Working with Dynamic Crop Models: Evaluation, Analysis, Parameterization, and Applications*, chapter 4, pages 55–99. Elsevier, 2006.

[18] M. D. Morris. Factorial sampling plans for preliminary computational experiments. *Technometrics*, 33(2):161–174, 1991.

[19] A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, and S. Tarantola. *Global sensitivity analysis: the primer*. Wiley Online Library, 2008.

[20] I. M. Sobol. Sensitivity estimates for nonlinear mathematical models. *Math. Modeling Comput. Experiment*, 1(4):407–414 (1995), 1993.

[21] I.M. Sobol. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and Computers in Simulation*, 55(1-3):271–280, 2001.

[22] R. J. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288, 1996.

[23] J.-Y. Tissot and C. Prieur. Bias correction for the estimation of sensitivity indices based on random balance designs. *Reliability Engineering & System Safety*, 107:205–213, 2012.

[24] J.-Y. Tissot and C. Prieur. Estimating sobol'indices combining monte carlo estimators and latin hypercube sampling. 2012.

[25] J. A. Tropp. Just relax: Convex programming methods for identifying sparse signals in noise. *Information Theory, IEEE Transactions on*, 52(3):1030–1051, 2006.

[26] L. Welch. Lower bounds on the maximum cross correlation of signals (corresp.). *Information Theory, IEEE Transactions on*, 20(3):397–399, 1974.

[27] W. Xu and B. Hassibi. Further results on performance analysis for compressive sensing using expander graphs. In *Signals, Systems and Computers, 2007. ACSSC 2007. Conference Record of the Forty-First Asilomar Conference on*, pages 621–625. IEEE, 2007.

[28] P. Zhao and B. Yu. On model selection consistency of lasso. *The Journal of Machine Learning Research*, 7:2541–2563, 2006.