Penalised least square in sparse setting with convex penalty and non gaussian errors

Doualeh Abdillahi-Ali, Nourddine Azzaoui, Arnaud Guillin, Guillaume Le Mailloux, Tomoko Matsui

Penalised least square in sparse setting with convex penalty and non gaussian errors

Doualeh Abdillahi-Ali[†], Nourddine Azzaoui[†], Arnaud Guilin[†], Guillaume Le Mailloux[◇], and Tomoko Matsui[⋆]

[†]Laboratoire de Mathematiques Blaise Pascal,UMR CNRS 6620
[◇]Centre Borelli, UMR9010
[⋆]The Institute of Statistical Mathematics Tokyo Japan

21 may 2021

Abstract

This paper considers the penalized least squares estimators with convex penalties or regularisation norms. We provide sparsity oracles inequalities for the prediction error for a general convex penalty and for the particular cases of Lasso and Group Lasso estimators in a regression setting. The main contributions are that our oracle inequalities are established for the more general case where the observations noise is issued from probability measures that satisfy a weak spectral gap (or Poincaré) inequality instead of gaussian distributions, and five easier to verify bounds on compatibility. We Illustrate our results on a heavy tailed example and a sub gaussian one; we especially give the explicit bounds of the oracle inequalities for these two special examples.

Introduction

High-dimensional statistical models have been thoroughly studied in recent research and literatures. In particular, penalized Least Square (LS) estimators have been proposed and extensively investigated; for example the ℓ_1 norm penalized estimator LASSO and its extensions. A common feature of these estimators is the fact that the penalty is a norm satisfying some specific decomposability conditions. As shown in [5], the two main ingredients of the analysis are based on the restricted eigenvalue *compatibility* property, and the empirical process bounding the stochastic error. With this approach, several techniques have been proposed for a unified treatment of LS estimators with decomposable penalties, a wide overview can be found in [13, 15, 8]. Classical results were derived via oracle inequalities depending on unspecified *compatibility* parameters. On the other hand the penalties (and thus, the estimators) depend on the distribution with which the oracle inequality holds. To overcome these problems, a tentative for a general solution were achieved by the small ball method see for instance [10, 12]. Under gaussian noise, many results have been established for the sparsity oracle inequalities for LASSO estimators in different situations: (1) the fixed design case [5, 4, 3, 1, 9, 6, 8] (2) results based on confidence level tied to the tuning parameter see for instance [5, 9, 6, 8] (3) in the case where the noise are *i.i.d.* sub-gaussian see for example [3]. For instance, in [3, 14], sparsity oracle inequalities for the Lasso estimators are obtained in random design regression especially when all entries of the design matrix are *i.i.d.* standard gaussian independent of the observations errors. In this work, we consider the classical general framework of regression model. Using the same notations as in [1, 8] it is expressed by the following :

$$\mathbf{Y} = \mathbf{f} + \xi. \tag{1}$$

We assume in this paper that the distribution of the noise random vector ξ satisfies a weak spectral gap inequality. Following the definition and notations in [2, 7], a probability measure μ satisfies a weak spectral gap (or a weak Poincaré) inequality if there exists a function $\gamma : (0, +\infty) \longrightarrow \mathbb{R}^+$ such that every local function $h : M \longrightarrow \mathbb{R}$ satisfies for all $s > 0$ the inequality:

$$Var_\mu(h) \leq \gamma(s) \int |\nabla h|^2 \, d\mu + s \, Osc(h)^2 \tag{2}$$

Where $Osc(h) = \sup h - \inf h$ is the total oscillation of the function h. We place ourselves in high-dimensional statistics setting, by considering a design matrix $\mathbb{X} \in \mathbb{R}^{n\times p}$ with $p >> n$. We consider to generalise, to Hilbert spaces, the following classical estimation problem of $\mathbf{f}$ by $\mathbb{X}\hat{\beta}$ where:

$$\hat{\beta} \in \underset{\beta\in\mathbb{R}^p}{\arg\min} \|\mathbf{Y} - \mathbb{X}\beta\|_n^2 + F(\beta). \tag{3}$$

The empirical norm $\|\cdot\|_n$ is defined by $\|u\|_n^2 = \frac{1}{n}\sum_{i=1}^n u_i^2$ and $F : \mathbb{R}^p \to \mathbb{R}^+$ is a convex penalty function, being the l_1 norm for the Lasso case for example.

The main result of this paper, is to derive oracle inequalities for the prediction error of the penalized estimator $\hat{\beta}$ solution of (3) under the assumption that the noise random vector ξ in (1) follows a probability measure that satisfies a weak spectral gap inequality [2]. Our result is based on the so-called *compatibility* constant defined as follows.

$$\mu_{c_0}(\beta) = \inf\left\{\mu > 0 : \|\mathcal{P}_\beta u\| \le \mu \|\mathbb{X}u\|_n\,,\ \forall u : \left\|\mathcal{P}_\beta^\perp u\right\| \le c_0 \|\mathcal{P}_\beta u\|\right\}$$

where $\mathcal{P}$ is an operator verifying a *decomposability* condition; as an example we use orthogonal projector for LASSO application. Without loss of generality, we illustrate our result on regression model (1) where the noise vector ξ follows a heavy tailed distribution expressed as the product measure of $d\mu_\alpha(t) = \dfrac{\alpha(1+|t|)^{-1-\alpha}}{2}dt$ for $\alpha > 2$ which satisfies a weak spectral gap inequality with $\gamma(s) = c_\alpha(s/n)^{-2/\alpha}$, $s \in (0, 1/4)$ [2, 7, Example 8]. Furthermore, in the LASSO setting, we establish an explicit lower bound for the compatibility constant depending on the *sparsity* s and κ the maximal correlation between columns of $\mathbb{X}$; namely $\mu_{c_0}(\beta) \le \dfrac{s}{1-\kappa s}$. We then provide the explicit oracle inequality; $\left\|\mathbb{X}\hat{\beta} - \mathbf{f}\right\|_n^2 - \|\mathbb{X}\beta - \mathbf{f}\|_n^2 \le \dfrac{32\left((2pc_\alpha n)^{1/\alpha} + \sqrt{n}\bar{a}\right)^2 s}{n(1-2\sqrt{\kappa s})^2}$. The main difference from works known in litterature is that our oracle inequalities are obtained under non-gaussian distribution satisfying the weak Poincaré inequality, and our compatibility bounds are easier ton handle than in [4, 3, 1] . Indeed, in contrast to classical results, we provide an explicit upper bound of the compatibility constant in the oracle inequality for Lasso and group Lasso estimator.

2 Statement of the problem and preliminary results

Let $\mathcal{H}$ be a Hilbert space with inner product $\langle\cdot,\cdot\rangle$ and its corresponding norm $\|\cdot\|_{\mathcal{H}}$. Let $\mathbb{B}$ a closed convex subset of $\mathcal{H}$. We will be interested in a regression problem as in (1) where $\mathbf{f} \in \mathbb{R}^n$ is an unknown deterministic mean and $\xi \in \mathbb{R}^n$ is a random noise vector. Let $\mathbb{P}$ be the probability distribution of ξ satisfying a weak spectral gap inequality with function γ defined in (2). We focus on estimates of $\mathbf{f}$ having the form $\mathbb{X}\hat{\beta}$ where $\hat{\beta} \in \mathbb{B}$ is data determined like in (3). The matrix $\mathbb{X}$ represents a linear operator from $\mathcal{H} \to \mathbb{R}^n$. We aim to investigate the prediction performances of the estimator $\hat{\beta}$ defined as the solution of the problem minimization :

$$\hat{\beta} \in \underset{\beta\in\mathbb{B}}{\arg\min} \|\mathbf{y} - \mathbb{X}\beta\|_n^2 + F(\beta) \tag{4}$$

where $F : \mathcal{H} \to \mathbb{R}^+$ is a convex penalty function.

2.1 Preliminary results

We expose here, two propositions giving the key ingredients for the proof of our main result. The first proposition, based on convexity argument and some simple algebra, provide a deterministic bound of the prediction error in terms of $\|\mathbb{X}\beta - \mathbf{f}\|_n^2$ up to an additional random term.

Proposition 2.1. *If $\hat{\beta}$ is a solution of the minimization problem (4), Then $\hat{\beta}$ satisfies, for all $\beta \in \mathbb{B}$ and for all $\boldsymbol{f} \in \mathbb{R}^n$,*

$$\left\|\mathbb{X}\hat{\beta} - \boldsymbol{f}\right\|_n^2 - \|\mathbb{X}\beta - \boldsymbol{f}\|_n^2 \le \frac{2}{n}\xi^T\mathbb{X}(\hat{\beta} - \beta) + F(\beta) - F(\hat{\beta}) - \left\|\mathbb{X}(\hat{\beta} - \beta)\right\|_n^2 \tag{5}$$

Proof. The proof mainly relies on the sub-differentials and optimality condition of convex functions and some simple algebra. For details and a complete proof see for instance [1, Proposition 3.2]. □

The second proposition provides an upper bound of the random quantity $\frac{1}{n}\xi^T\mathbb{X}(\hat{\beta}-\beta)$ that holds with large enough probability.

Proposition 2.2. *Let $\mathbb{P}$ satisfies a weak spectral gap inequality (2) with a function γ. Let $h:\mathcal{H}\to[0,+\infty[$ be a positive homogeneous mapping and let $\tau>0$. Denote the event :*

$$\Omega=\left\{\sup_{u\in\mathcal{H}:h(u)\leq 1}\frac{1}{n}\xi^T\mathbb{X}u\leq\tau\right\}, \tag{6}$$

and assume it to satisfy $\mathbb{P}(\Omega)\geq\frac{1}{2}$. Then, for $k\geq 1, c\geq 0$ and $s\in(0,1/4)$, we have

$$\mathbb{P}\left(\forall u\in\mathcal{H}:\frac{1}{n}\xi^T\mathbb{X}u\leq(\tau+k)\max\left(h(u),c\left\|\mathbb{X}u\right\|_n\right)\right)\geq 1-3\Theta(kc\sqrt{n}) \tag{7}$$

where $\Theta(x)=\inf\left\{s\in(0,1/4);\exp\left(\frac{-x}{4\sqrt{\gamma(s)}}\right)\leq s\right\}$ vanishes when x goes to infinity.

Proof. Define the function $f:\mathbb{R}^n\to\mathbb{R}$ and the subset $T\subset\mathcal{H}$ as follows :

$$f(\xi)=\sup_{u\in T}\frac{1}{n}\xi^T\mathbb{X}u\qquad\text{and}\qquad T=\left\{u\in\mathcal{H}:\max\left(h(u),c\left\|\mathbb{X}u\right\|_n\right)\leq 1\right\}.$$

It is easy to verify that, for every $\xi_1,\,\xi_2\in\mathbb{R}^n$, we have,

$$|f(\xi_1)-f(\xi_2)|\leq\frac{1}{c\sqrt{n}}\left\|\xi_1-\xi_2\right\|_2$$

By the concentration inequality [2, theorem 8] and the fact that f is a $\frac{1}{c\sqrt{n}}$-Lipschitz function, we have with probability at least $1-3\Theta(kc\sqrt{n})$,

$$\begin{aligned}\sup_{u\in T}\frac{1}{n}\xi^T\mathbb{X}u&\leq\mathbf{Med}\left(\sup_{u\in T}\frac{1}{n}\xi^T\mathbb{X}u\right)+k\\&\leq\mathbf{Med}\left(\sup_{u\in\mathcal{H}:h(u)\leq 1}\frac{1}{n}\xi^T\mathbb{X}u\right)+k,\end{aligned}$$

where $\Theta(u)=\inf\left\{s\in(0,1/4);\exp\left(\frac{-u}{4\sqrt{\gamma(s)}}\right)\leq s\right\}$ tends to 0 when u tends to infinity. The notation $\mathbf{Med}(Z)$ stands for the median of the random variable Z.

Assume that $\mathbb{P}\left(\sup_{u\in\mathcal{H}:h(u)\leq 1}\frac{1}{n}\xi^T\mathbb{X}u\leq\tau\right)\geq 1/2$, then the median of $\sup_{u\in\mathcal{H}:h(u)\leq 1}\frac{1}{n}\xi^T\mathbb{X}u$ can be bounded from above by τ. This implies that, with probability at least $1-3\Theta(kc\sqrt{n})$,

$$\forall u\in\mathcal{H},\frac{1}{n}\xi^T\mathbb{X}u\leq(\tau+k)\max\left(h(u),c\left\|\mathbb{X}u\right\|_n\right),$$

this achieves the proof. □

2.2 Main assumptions of decomposability and compatibility

Consider the linear operator $\mathbb{X}:\mathcal{H}\to\mathbb{R}^n$ defined by the relation:

$$\mathbb{X}\beta=\left(\left\langle\beta,X_1\right\rangle,\cdots,\left\langle\beta,X_n\right\rangle\right)^T,\ \forall\beta\in\mathcal{H}, \tag{8}$$

where $X_1,\cdots,X_n$ are deterministic elements of $\mathcal{H}$. The convex penalty $F:\mathcal{H}\to\mathbb{R}$ is taken proportional, with a tunning parameter $\lambda>0$, to a regularization norm :

$$F(\beta)=\lambda\left\|\beta\right\|.$$

The estimator $\hat{\beta}$ introduced in (4) becomes then :

$$\hat{\beta}\in\arg\min_{\beta\in\mathbb{B}}\left\|\mathbb{X}\beta-y\right\|_n^2+\lambda\left\|\beta\right\| \tag{9}$$

Before establishing our main results, we recall two others important ingredients under which oracle inequalities are obtained namely the decomposability assumptions of the regularization norm and the compatibility factor.

- **Decomposability assumption and consequences:** Let A be a subset of $\mathcal{H}$ for which we associate a linear operator $\mathcal{P}_A : \mathcal{H} \to \mathcal{H}$ and $\mathcal{P}_A^{\perp} = \mathcal{I} - \mathcal{P}_A$ where $\mathcal{I}$ is the identity operator.

  Assumption 2.1. *We suppose that there exists a subset $\mathbb{A}$ of $\mathbb{B}$ such that:*

$$\mathcal{P}_A A = A,\ \forall A \in \mathbb{A} \quad \textit{and} \quad \|A\| + \left\|\mathcal{P}_A^{\perp} B\right\| = \|A + \mathcal{P}_A(A - B)\|,\ \forall A \in \mathbb{A},\ \forall B \in \mathcal{H}$$

  The last assumption has been also discussed in [1, Assumption 4.2]. One can see easily that, for the ℓ_1 regularization norm, the decomposability assumption is satisfied when the linear operator $\mathcal{P}_A$ is an orthogonal projector. As a consequence of the decomposability assumption 2.1, we have the following triangular property :

  Corollary 2.1. *The above assumption 2.1 implies the following triangular property:*

$$\mathcal{P}_A A = A,\ \forall A \in \mathbb{A} \qquad \textit{and} \qquad \|A\| - \|B\| \leq \|\mathcal{P}_A(A - B)\| - \left\|\mathcal{P}_A^{\perp} B\right\| \tag{10}$$

  This corollary has been discussed also in [1, Assumption 4.1]. In the case of ℓ_1 regularization norm, the decomposability assumption and triangular property are defined differently in [8, paragraph 2.5].

- **Compatibility factor assumption**

  Another main ingredient for the proof of our main result is the compatibility factor [1]. For any $A \in \mathcal{H}$ and for any constant $c_0 \geq 0$, define the following cone in $\mathbb{B}$ as follows:

$$\mathbb{C}_{A,c_0} = \left\{B \in \mathbb{B} : \left\|\mathcal{P}_A^{\perp} B\right\| \leq c_0 \|\mathcal{P}_A B\|\right\}.$$

  The compatibility factor associated to the cone $\mathbb{C}_{A,c_0}$ is the quantity :

$$\mu_{c_0}(A) = \inf\left\{\mu > 0 : \|\mathcal{P}_A B\| \leq \mu \|\mathbb{X}B\|_n,\ \forall B \in \mathbb{C}_{A,c_0}\right\}. \tag{11}$$

  On of our main contribution is to provide upper bounds on this compatibility factor, with easy to verify assumptions.

3 Main results and Oracle inequalities

The next theorem states our main result in the case of distributions verifying the spectral gap inequality.

Theorem 3.1. *Let $\mathbb{P}$ satisfies a weak spectral gap inequality with function γ. Assume that assumption 2.1 holds. Let $\tau > 0$ and suppose that Ω defined in (6) satisfies $\mathbb{P}(\Omega) \geq \frac{1}{2}$. Let $k \geq 1, c \geq 0$ and $\lambda \geq 2(\tau + k)$. Then, the estimator $\hat{\beta}$ (9) satisfy with probability at least $1 - 3\Theta(kc\sqrt{n})$,*

$$\left\|\mathbb{X}\hat{\beta} - f\right\|_n^2 \leq \inf_{\beta \in A}\left[\|\mathbb{X}\beta - f\|_n^2 + \frac{(\lambda + 2(\tau + k))^2}{4}\mu_{c_0}^2(\beta)\right] + \frac{(\lambda + 2(\tau + k))^2 c^2}{4} \tag{12}$$

where $\Theta(u) = \inf\left\{s \in (0, 1/4); \exp\left(\frac{-u}{4\sqrt{\gamma(s)}}\right) \leq s\right\}$ goes to 0 when u goes to infinity.

Proof. Let us consider the positive simple function $h(u) = \|u\|$. Combining the two inequalities from proposition 2.1 and proposition 2.2, we get with probability at least $1 - 3\Theta(kc\sqrt{n})$

$$\left\|\mathbb{X}\hat{\beta} - f\right\|_n^2 - \|\mathbb{X}\beta - f\|_n^2 \leq 2(\tau + k)\max\left(\|u\|, c\|\mathbb{X}u\|_n\right) + \lambda\|\beta\| - \lambda\left\|\hat{\beta}\right\| - \|\mathbb{X}u\|_n^2,$$

where $u = \hat{\beta} - \beta$. For the rest of the proof we distinguish tree cases :

Case 1. *$\beta \in \mathbb{A}$ is such that $\|u\| < c\|\mathbb{X}u\|_n$. This will imply that,*

$$\left\|\mathbb{X}\hat{\beta} - f\right\|_n^2 - \|\mathbb{X}\beta - f\|_n^2 \leq 2(\tau + k)c\|\mathbb{X}u\|_n + \lambda\|\beta\| - \lambda\left\|\hat{\beta}\right\| - \|\mathbb{X}u\|_n^2$$

using the triangle inequality $\lambda\|\beta\| - \lambda\left\|\hat{\beta}\right\| \leq \lambda\left\|\hat{\beta} - \beta\right\|$ and $\|u\| < c\|\mathbb{X}u\|_n$ imply,

$$\left\|\mathbb{X}\hat{\beta}-f\right\|_n^2-\|\mathbb{X}\beta-f\|_n^2\leq 2(\tau+k)c\,\|\mathbb{X}u\|_n+\lambda c\,\|\mathbb{X}u\|_n-\|\mathbb{X}u\|_n^2$$
$$\leq(2(\tau+k)+\lambda)c\,\|\mathbb{X}u\|_n-\|\mathbb{X}u\|_n^2$$

using inequality $2ab\leq a^2+b^2$ *we obtain,*

$$\left\|\mathbb{X}\hat{\beta}-f\right\|_n^2-\|\mathbb{X}\beta-f\|_n^2\leq\frac{(2(\tau+k)+\lambda)^2c^2}{4}+\|\mathbb{X}u\|_n^2-\|\mathbb{X}u\|_n^2$$
$$\leq\frac{(2(\tau+k)+\lambda)^2c^2}{4}. \tag{13}$$

Thus inequality (12) is satisfied.

Case 2. $\beta\in\mathbb{A}$ *is such that* $\|u\|>c\,\|\mathbb{X}u\|_n$. *Then,*

$$\left\|\mathbb{X}\hat{\beta}-f\right\|_n^2-\|\mathbb{X}\beta-f\|_n^2\leq 2(\tau+k)\,\|u\|+\lambda\,\|\beta\|-\lambda\left\|\hat{\beta}\right\|-\|\mathbb{X}u\|_n^2 \tag{14}$$

Assumption 2.1 with $A=\beta$ *and* $B=\hat{\beta}$ *allow to write,*

$$\|\beta\|-\left\|\hat{\beta}\right\|\leq\left\|\mathcal{P}_\beta(\beta-\hat{\beta})\right\|-\left\|\mathcal{P}_\beta^\perp\hat{\beta}\right\|,$$

and the triangle inequality implies,

$$\left\|\hat{\beta}-\beta\right\|=\left\|\mathcal{P}_\beta(\hat{\beta}-\beta)+\mathcal{P}_\beta^\perp(\hat{\beta}-\beta)\right\|$$
$$\leq\left\|\mathcal{P}_\beta(\hat{\beta}-\beta)\right\|+\left\|\mathcal{P}_\beta^\perp(\hat{\beta}-\beta)\right\|.$$

Combining the last two inequalities and (14), we obtain

$$\left\|\mathbb{X}\hat{\beta}-f\right\|_n^2-\|\mathbb{X}\beta-f\|_n^2\leq(\lambda+2(\tau+k))\,\|\mathcal{P}_\beta u\|-(\lambda-2(\tau+k))\left\|\mathcal{P}_\beta^\perp u\right\|-\|\mathbb{X}u\|_n^2. \tag{15}$$

Case 2.1. $\beta\in\mathbb{A}$ *is such that* $\|u\|>c\,\|\mathbb{X}u\|_n$ *and* $(\lambda+2(\tau+k))\,\|\mathcal{P}_\beta u\|<(\lambda-2(\tau+k))\left\|\mathcal{P}_\beta^\perp u\right\|$. *In view of inequality (15) we get,*

$$\left\|\mathbb{X}\hat{\beta}-f\right\|_n^2-\|\mathbb{X}\beta-f\|_n^2\leq 0$$

and

$$\left\|\mathbb{X}\hat{\beta}-f\right\|_n^2\leq\|\mathbb{X}\beta-f\|_n^2 \tag{16}$$

Which implies that inequality (12) holds

Case 2.2. $\beta\in\mathbb{A}$ *is such that* $\|u\|>c\,\|\mathbb{X}u\|_n$ *and* $(\lambda-2(\tau+k))\left\|\mathcal{P}_\beta^\perp u\right\|<(\lambda+2(\tau+k))\,\|\mathcal{P}_\beta u\|$.

Then u *belongs to the cone* $\mathbb{C}_{\beta,c_0}=\left\{u\in\mathcal{H}:\left\|\mathcal{P}_\beta^\perp u\right\|\leq c_0\,\|\mathcal{P}_\beta u\|\right\}$ *where* $c_0=\frac{\lambda+2(\tau+k)}{\lambda-2(\tau+k)}$. *We use the compatibility factor associated to* $\mathbb{C}_{\beta,c_0}$,

$\mu_{c_0}(\beta)=\inf\left\{\mu>0:\|\mathcal{P}_\beta u\|\leq\mu\,\|\mathbb{X}u\|_n\,,\ \forall u\in H:0\leq(\lambda+2\,(\tau+k))\,\|\mathcal{P}_\beta u\|-(\lambda-2\,(\tau+k))\left\|\mathcal{P}_\beta^\perp u\right\|\right\}$,

so that $\|\mathcal{P}_\beta u\|\leq\mu_{c_0}(\beta)\,\|\mathbb{X}u\|_n$. *This and inequality (15) yield,*

$$\left\|\mathbb{X}\hat{\beta}-f\right\|_n^2-\|\mathbb{X}\beta-f\|_n^2\leq(\lambda+2\,(\tau+k))\,\|\mathcal{P}_\beta u\|-\|\mathbb{X}u\|_n^2$$
$$\leq(\lambda+2\,(\tau+k))\mu_{c_0}(\beta)\,\|\mathbb{X}u\|_n-\|\mathbb{X}u\|_n^2.$$

Using the obvious inequality $2ab<a^2+b^2$

$$\left\|\mathbb{X}\hat{\beta}-f\right\|_n^2-\|\mathbb{X}\beta-f\|_n^2\leq\frac{(\lambda+2\,(\tau+k))^2}{4}\mu_{c_0}^2(\beta)+\|\mathbb{X}u\|_n^2-\|\mathbb{X}u\|_n^2$$
$$\leq\frac{(\lambda+2\,(\tau+k))^2}{4}\mu_{c_0}^2(\beta). \tag{17}$$

This achieves the proof of inequality (12).

□

In the following we describe direct applications of proposition 2.2 and theorem (3.1), by investigating two examples of distributions verifying the spectral gap inequality (2). The first one is issued from the heavy tailed family obtained from the product of the probability distribution $d\mu_\alpha(t) = \frac{\alpha(1+|t|)^{-1-\alpha}}{2}dt$. The second example is from the sub-exponential family as a product of measures of the form, $d\nu_r = d_r e^{-|t|^r}dt,\ \forall\ r \in (0,1)$.

- **A heavy tailed example:** We consider on $\mathbb{R}^n$, the distribution $\mathbb{P}$ issued from the product probability measure of

$$d\mu_\alpha(t) = \frac{\alpha(1+|t|)^{-1-\alpha}}{2}dt \ \text{ for } \alpha > 2. \tag{18}$$

This measure satisfies a weak spectral gap inequality with,

$$\gamma(s) = c_\alpha \left(\frac{s}{n}\right)^{-2/\alpha}, \quad \text{for } s \in (0, \frac{1}{4})$$

For more details see discussions in [2, example 8]. Let $h : \mathcal{H} \to [0, +\infty[$ be a positive homogeneous mapping and let $\tau > 0$. Assume that the event Ω defined in (6) satisfies $\mathbb{P}(\Omega) \geq \frac{1}{2}$. Then, there exists constants $t_0(\alpha) > e$ and $C(\alpha)$ such that $\frac{tc\sqrt{n}}{n^{\frac{1}{\alpha}}} \geq t_0(\alpha)$ and by applying, proposition 2.2 for $c \geq 0$, we have:

$$\mathbb{P}\left(\forall u \in \mathcal{H} : \frac{1}{n}\xi^T \mathbb{X}u \leq (\tau + t)\max\left(h(u), c\left\|\mathbb{X}u\right\|_n\right)\right) \geq 1 - \frac{1}{2}C(\alpha)\left(\frac{\log(\frac{tc\sqrt{n}}{n^{\frac{1}{\alpha}}})}{\frac{tc\sqrt{n}}{n^{\frac{1}{\alpha}}}}\right)^\alpha.$$

We deduce then, by appying theorem 3.1 that the estimator $\hat{\beta}$ (9) satisfy with probability at least $1 - \frac{1}{2}C(\alpha)\left(\frac{\log(\frac{tc\sqrt{n}}{n^{\frac{1}{\alpha}}})}{\frac{tc\sqrt{n}}{n^{\frac{1}{\alpha}}}}\right)^\alpha$

$$\left\|\mathbb{X}\hat{\beta} - f\right\|_n^2 \leq \inf_{\beta \in A}\left[\left\|\mathbb{X}\beta - f\right\|_n^2 + \frac{(\lambda + 2(\tau + t))^2}{4}\mu_{c_0}^2(\beta)\right] + \frac{(\lambda + 2(\tau + t))^2 c^2}{4} \tag{19}$$

- **A sub-exponential example:** In this case, we consider $\mathbb{P}$ as the product of the probability measure,

$$d\nu_r = d_r e^{-|t|^r}dt \ \text{ for } r \in (0,1). \tag{20}$$

This measure satisfies also a weak spectral gap inequality with:

$$\gamma(s) = k_r\left(\log(\frac{2n}{s})\right)^{(2/r)-s},\ s \in (0, 1/4)$$

For more details see for instance [2, example 9]. Let $h : \mathcal{H} \to [0, +\infty[$ be a positive homogeneous mapping and let $\tau > 0$. We suppose that Ω satisfies $\mathbb{P}(\Omega) \geq \frac{1}{2}$. Then, for $k \geq 0,\ c \geq 0$ and $r \in (0,1)$, we have

$$\mathbb{P}\left(\forall u \in \mathcal{H} : \frac{1}{n}\xi^T \mathbb{X}u \leq (\tau + k)\max\left(h(u), c\left\|\mathbb{X}u\right\|_n\right)\right) \geq 1 - 5\exp\left(\frac{-c_r k c\sqrt{n}}{\max\left((kc\sqrt{n})^r, \log n\right)^{1/r-1}}\right)$$

where c_r is a constant depending only on r. In particular, for a fixed ϵ, a large n, and k verifying $k \geq c_r(\log\frac{10}{\epsilon})(\log n)^{\frac{1}{r}-1}$, we have:

$$\mathbb{P}\left(\forall u \in \mathcal{H} : \frac{1}{n}\xi^T \mathbb{X}u \leq \left(\tau + \frac{k}{c\sqrt{n}}\right)\max\left(h(u), c\left\|\mathbb{X}u\right\|_n\right)\right) \geq 1 - \frac{\epsilon}{2}$$

In this example, the estimator $\hat{\beta}$ given by the minimisation problem (9) satisfy with probability at least $1 - 5\exp\left(\frac{-c_r k c\sqrt{n}}{\max\left((kc\sqrt{n})^r, \log n\right)^{1/r-1}}\right)$,

$$\left\|\mathbb{X}\hat{\beta} - f\right\|_n^2 \leq \inf_{\beta \in A}\left[\left\|\mathbb{X}\beta - f\right\|_n^2 + \frac{(\lambda + 2(\tau + k))^2}{4}\mu_{c_0}^2(\beta)\right] + \frac{(\lambda + 2(\tau + k))^2 c^2}{4}. \tag{21}$$

In particular, for $\lambda \geq 2(\tau + \frac{k}{c\sqrt{n}})$, then with probability at least $1 - \frac{\epsilon}{2}$,

$$\left\|\mathbb{X}\hat{\beta} - f\right\|_n^2 \leq \inf_{\beta \in A} \left[\|\mathbb{X}\beta - f\|_n^2 + \frac{(\lambda + 2\left(\tau + \frac{k}{c\sqrt{n}}\right))^2}{4} \mu_{c_0}^2(\beta) \right] + \frac{(\lambda + 2(\tau + \frac{k}{c\sqrt{n}}))^2 c^2}{4}$$

4 Application to the classical LASSO and group LASSO

We use the same notations as in [1]; we denote $|.|_q$ for the ℓ_q norm of a finite dimensional vector, $1 \leq q \leq \infty$. The support of β will be denoted $supp(\beta) = \{j : \beta_j \neq 0\}$. If $(e_j)_{j=1,\cdots,p}$ is the canonical basis of $\mathbb{R}^p$, then β_S will denote the orhtogonal projection of β onto the linear span of $\{e_j : j \in S\}$ for $S \subset \{1, \cdots, p\}$.

4.1 Application to LASSO

In this case $\mathcal{H} = \mathbb{B} = \mathbb{R}^p$ equipped with the Euclidean norm $\|.\|_{\mathcal{H}} = |.|_2$. The penalty function is given by $\|.\|$ the ℓ_1 norm. Then the estimator $\hat{\beta}$ defined in (9) is the LASSO estimator

$$\hat{\beta} \in \arg\min_{\beta \in \mathbb{R}^p} \|\mathbb{X}\beta - y\|_n^2 + \lambda |\beta|_1 \tag{22}$$

where $\lambda > 0$ is a tuning parameter. For $\beta \in \mathbb{R}^p$, we consider the decomposability conditions with $\mathcal{P}_\beta$, the orthogonal projection operator onto the linear span of $\{e_j : j \in supp(\beta)\}$. Following a similar argument as in [1] or [3, 8], one can show, using the duality between ℓ_1 and ℓ_∞ norms, that for LASSO setting the set Ω defined in (6), becomes:

$$\Omega = \left\{ \sup_{u \in \mathbb{R}^p : |u|_1 \leq 1} \frac{1}{n} \xi^T \mathbb{X} u \leq \tau \right\} = \left\{ \sup_{u \in \mathbb{R}^p : |u|_1 \leq 1} \frac{1}{n} (\mathbb{X}^T \xi)^T u \leq \tau \right\} = \left\{ \frac{1}{n} \left|\mathbb{X}^T \xi\right|_\infty \leq \tau \right\}$$

In the following we recall that, in the LASSO setting, the assumption 2.1 holds for the orthogonal projector operator.

4.2 Upper bound of the compatibility factor

In this section we state a close form for the upper bound of the compatibility factor $\mu_{c_0}(\beta)$. For this purpose, we need the following results:

Lemma 4.1. *In the LASSO setting we have the following:*

1. *For the orthogonal projector operator $\mathcal{P}_\beta$, the assumption'2.1 is satisfied.*
2. *The compatibility factor can be expressed as follows:*

$$\mu_{c_0}(\beta) = \frac{\lambda - a}{\lambda + a} \sup_{u \in \mathbb{R}^p : \left|\mathcal{P}_\beta^\perp u\right|_1 \leq c_0 |\mathcal{P}_\beta u|_1} \frac{c_0 |\mathcal{P}_\beta u|_1 - \left|\mathcal{P}_\beta^\perp u\right|_1}{\|\mathbb{X}u\|_n} \tag{23}$$

   where $a = 2(\tau + t)$ and $c_0 = \frac{\lambda + a}{\lambda - a}$.
3. *We assume the empirical norm of the columns $\mathbb{X}e_j$ are normalized(ie. $\|\mathbb{X}e_j\|_n = 1, \forall\ 1 \leq j \leq p$), and we write $\kappa = \sup_{1 \leq i \neq j \leq p} \frac{|\langle \mathbb{X}e_i, \mathbb{X}e_j \rangle|}{n}$, the maximal correlation between columns in $\mathbb{X} \in \mathbb{R}^{n \times p}$. For all $u \in \mathbb{R}^p$, we have:*

$$\|\mathbb{X}u\|_n^2 \geq |u|_2^2 - \kappa |u|_1^2 .$$

Proof. We use a similar argument as in [3, 1] to verify the decomposability assumptions 2.1.

1. For $\beta \in \mathbb{R}^p$, by definition $\mathcal{P}_\beta = \text{Proj}_{\{ej : j \in supp(\beta)\}}$ where $supp(\beta) = \{j : \beta_j \neq 0\}$, we have

$$\mathcal{P}_\beta \beta = \sum_{j=1}^p \beta_j \text{Proj}_{\{ej : j \in supp(\beta)\}} e_j = \sum_{j=1}^p \beta_j e_j \mathbb{1}_{\{j \in supp(\beta)\}} = \sum_{j \in supp(\beta)} \beta_j e_j = \beta$$

consider $\acute{\beta} = \sum_{j=1}^{p} \acute{\beta_j} e_j$. We have,

$$|\beta|_1 - \left|\acute{\beta}\right|_1 = \left|\sum_{j\in supp(\beta)} \beta_j e_j\right|_1 - \left|\sum_{j=1}^{p} \acute{\beta_j} e_j\right|_1 = \sum_{j\in supp(\beta)} |\beta_j| - \sum_{j\in supp(\beta)} \left|\acute{\beta_j}\right| - \sum_{j\in supp^c(\beta)} \left|\acute{\beta_j}\right|$$

using the triangular inequality $\sum_{j\in supp(\beta)} |\beta_j| - \sum_{j\in supp(\beta)} \left|\acute{\beta_j}\right| \leq \sum_{j\in supp(\beta)} \left|(\beta - \acute{\beta})_j\right|$, we have

$$|\beta|_1 - \left|\acute{\beta}\right|_1 \leq \left|\mathcal{P}_\beta(\beta - \acute{\beta})\right|_1 - \left|\mathcal{P}_\beta^\perp \acute{\beta}\right|_1$$

2. Taking $c_0 = \frac{\lambda+a}{\lambda-a}$ and we recal that:

$$\mu_{c_0}(\beta) = \inf\left\{\mu > 0 : |\mathcal{P}_\beta u|_1 \leq \mu \|\mathbb{X}u\|_n \,,\ \forall u \in \mathbb{R}^p : \left|\mathcal{P}_\beta^\perp u\right|_1 \leq c_0 |\mathcal{P}_\beta u|_1\right\}$$

Using the fact that $\frac{1}{c_0}\left|\mathcal{P}_\beta^\perp u\right|_1 \leq |\mathcal{P}_\beta u|_1 \leq \mu \|\mathbb{X}u\|_n \Rightarrow \frac{c_0 |\mathcal{P}_\beta u|_1 - \left|\mathcal{P}_\beta^\perp u\right|_1}{c_0 \|\mathbb{X}u\|_n} \leq \mu$. We can write, $\mu_{c_0}(\beta) = \inf\left\{\mu > 0 : \frac{1}{c_0} \times \frac{c_0 |\mathcal{P}_\beta u|_1 - \left|\mathcal{P}_\beta^\perp u\right|_1}{\|\mathbb{X}u\|_n} \leq \mu,\ \forall u \in \mathbb{R}^p : \left|\mathcal{P}_\beta^\perp u\right|_1 \leq c_0 |\mathcal{P}_\beta u|_1\right\}$. which implies equality (23).

3. A simple algebric shows the result as follows:

$$\begin{aligned}
\|\mathbb{X}u\|_n^2 &= \frac{1}{n}\langle u, \mathbb{X}^T\mathbb{X}u\rangle \\
&= \frac{1}{n}\sum_{1\leq i,j\leq p} \langle \mathbb{X}e_i, \mathbb{X}e_j\rangle u_i u_j \\
&= \sum_{i=1}^{p} \|\mathbb{X}e_i\|_n^2 u_i^2 + \sum_{1\leq i\neq j\leq p} \frac{\langle \mathbb{X}e_i, \mathbb{X}e_j\rangle}{n} u_i u_j \\
&\geq |u|_2^2 - \kappa \sum_{i=1}^{p}\sum_{j\neq i} |u_j||u_i| \\
&\geq |u|_2^2 - \kappa |u|_1^2 .
\end{aligned}$$

□

Let us denote $\zeta = \frac{|P_\beta^\perp u|_1}{|P_\beta u|_1}$. Using the expression (23) of the compatibility factor, inequaltiy of lemma 4.1, homogeneity of ℓ_1 norm and equality $|u|_1 = |P_\beta u|_1 + \left|P_\beta^\perp u\right|_1 = (1+\zeta)|P_\beta u|_1$, we get the following upper bound for the compatibility factor:

$$\mu_{c_0}(\beta)^2 \leq \frac{(\lambda-a)^2}{(\lambda+a)^2} \sup_{0\leq\zeta\leq c_0} \sup_{\substack{|P_\beta^\perp u|_1 = \zeta |P_\beta u|_1 \\ |u|_2 = 1}} \frac{(c_0-\zeta)^2}{1-\kappa(1+\zeta)^2 |P_\beta u|_1^2} |P_\beta u|_1^2 \tag{24}$$

In the following proposition we give, for a fixed ζ, the value of the supremum of the right hand side of inequality (24).

Proposition 4.1. *Fix $\zeta \in (0,1)$. For every $\beta \in \mathbb{R}^p$ such that $|supp(\beta)| = |\{j :\ \beta_j \neq 0\}| = s$ for some integer $1 \leq s \leq p$, we have:*

$$\sup_{\substack{|P_\beta^\perp u|_1 = \zeta |P_\beta u|_1 \\ |u|_2 = 1}} |P_\beta u|_1 = \sqrt{\frac{s(p-s)}{p-s+\zeta^2 s}}.$$

Proof. The proof relies on constrained optimization and Lagrange multiplier methods. Assume $u_j > 0$ and that the support of β constitutes the first s coordinates. The lagrangien of $u \in \mathbb{R}^p,\ \nu_1 \in \mathbb{R},\ \nu_2 \in \mathbb{R}$:

$$\mathcal{L}(u, \nu_1, \nu_2) = \sum_{j=1}^{s} u_j - \nu_1 \left(\sum_{j=1}^{p} u_j^2 - 1 \right) - \nu_2 \left(\sum_{j=s+1}^{p} u_j - \zeta \sum_{j=1}^{s} u_j \right)$$

Differentiating $\mathcal{L}(u, \nu_1, \nu_2)$ with respect to u_j, $1 \leq j \leq p$ yields

$$\begin{aligned} & \partial_j \mathcal{L} = 1 - 2\nu_1 u_j - \nu_2 \left(\mathbb{1}_{\{j>s\}} - \zeta \mathbb{1}_{\{j \leq s\}} \right) = 0 \\ \iff \quad & 2\nu_1 u_j = 1 - \nu_2 \left(\mathbb{1}_{\{j>s\}} - \zeta \mathbb{1}_{\{j \leq s\}} \right) \end{aligned} \tag{25}$$

Let $\widetilde{s} = \#\{j \leq s \mid u_j \neq 0\}$ and $\widetilde{p-s} = \#\{s+1 \leq j \leq p \ : \ u_j \neq 0\}$. Then the constraints give the following equations:

$$\begin{cases} |u|_2^2 = 1 \\ \left| P_\beta^\perp u \right|_1 = \zeta \left| P_\beta u \right|_1 \end{cases} \iff \begin{cases} (2\nu_1)^2 = \widetilde{s}(\nu_2\zeta + 1)^2 + \widetilde{p-s}(\nu_2 - 1)^2 \\ \widetilde{p-s}(\nu_2 - 1) = -\zeta\widetilde{s}(\nu_2\zeta + 1) \end{cases}$$

Combining the two equalities (25) and $(2\nu_1)^2 = \widetilde{s}(\nu_2\zeta + 1)^2 + \widetilde{p-s}(\nu_2 - 1)^2$, we get

$$u_j^2 = \frac{(\nu_2\zeta + 1)^2}{(2\nu_1)^2} = (\widetilde{s} + \frac{\zeta^2\widetilde{s}^2}{\widetilde{p-s}})^{-1}, j \leq s,$$

we then deduce the objective function $|P_\beta u|_1$

$$|P_\beta u|_1 = \frac{\widetilde{s}}{\sqrt{\widetilde{s} + \frac{\zeta^2\widetilde{s}^2}{\widetilde{p-s}}}} = \sqrt{\frac{\widetilde{sp-s}}{\widetilde{p-s} + \zeta^2\widetilde{s}}}.$$

The map, $(\widetilde{s}, \widetilde{p-s}) \longmapsto \frac{\widetilde{sp-s}}{\widetilde{p-s}+\zeta^2\widetilde{s}} = (\frac{\zeta^2}{\widetilde{p-s}} + \frac{1}{\widetilde{s}})^{-1}$ is maximal for $\widetilde{s} = s$ and $\widetilde{p-s} = p - s$. □

We give an upper bound of $\mu_{c_0}(\beta)$ for all s-sparse vector $\beta \in \mathbb{R}^p$ under condition on the matrix $\mathbb{X}$ from the following proposition.

Proposition 4.2. *For $\beta \in \mathbb{R}^p$ s-sparse(ie.$|supp(\beta)| \leq s$) and κ the maximal correlation between two columns of matrix $\mathbb{X}$, under the condition $\kappa s(1 + c_0)^2 \leq 1 + \frac{c_0^2 s}{p-s}$, we have:*

$$\mu_{c_0}(\beta)^2 \leq \frac{s}{1 - \kappa s} \tag{26}$$

Proof. By proposition 4.1 and inequality (24), we get the following upper bound of the compatibility factor,

$$\mu_{c_0}(\beta)^2 \leq \frac{(\lambda - a)^2}{(\lambda + a)^2} \sup_{0 \leq \zeta \leq c_0} \frac{(c_0 - \zeta)^2 s(p-s)}{(p-s) + \zeta^2 s - \kappa(1+\zeta)^2 s(p-s)}.$$

The upper bound of the function $\zeta \longmapsto \frac{(c_0-\zeta)^2 s(p-s)}{(p-s)+\zeta^2 s-\kappa(1+\zeta)^2 s(p-s)}$ is finite if the denominator is positif when $\zeta = c_0$ namely under the condition $\kappa s(1 + c_0)^2 \leq 1 + \frac{c_0^2 s}{p-s}$. □

4.3 Appliaction to special distributions

In this section we establish explicit oracle inequalities for the tow examples of heavy tailed and sub-exponential distributions discussed bellow. In order to show the main results we need to find, for each distribution family, τ such that $\mathbb{P}(\Omega) > 1/2$.

Lemma 4.2. *Let $\mathbb{X}$ be a deterministic matrix $\in \mathbb{R}^{n \times p}$, the we have:*

1. *Suppose that the probability distribution $\mathbb{P}$ is issued from the product of(18). In this case if*

$$\tau \geq \frac{(2pc_\alpha n)^{1/\alpha} \left\| \mathbb{X}e_j \right\|_n + \sqrt{n}\bar{a}}{\sqrt{n}}$$

where $\bar{a} = t_0(\alpha) + \int_{t_0(\alpha)}^{\infty} C(\alpha) \left(\frac{\log(t)}{t} \right)^\alpha \mathrm{d}t$. Then $\mathbb{P}(\Omega) \geq 1/2$.

2. *Suppose that the probability distribution $\mathbb{P}$ is issued from the product of(20), then if*

$$\tau \geq \frac{\ln(20p)}{c_r} \max\left(\frac{\ln(20p)}{c_r}, \log(n)\right)^{1/r-1} \frac{\|\mathbb{X}e_j\|_n}{\sqrt{n}} + \bar{b}$$

where $\bar{b} = \int_0^\infty 10 \exp\left(\frac{-c_r k}{\max\left(k^r, \log n\right)^{1/r-1}}\right) \mathrm{d}t$ *and* c_r *is a quantity depending only on* r*. then* $\mathbb{P}(\Omega) \geq 1/2$

Proof. Define $f : \mathbb{R}^n \to \mathbb{R}$ by :

$$f(z) = \frac{1}{n}(\mathbb{X}e_j)^T z$$

Then, for every $z_1,\ z_2 \in \mathbb{R}^n$, $|f(z_1) - f(z_2)| \leq \frac{\|\mathbb{X}e_j\|_n}{\sqrt{n}} |z_1 - z_2|_2$. Therefore, f is a Lipschitz function with Lipschitz constant $\frac{\|\mathbb{X}e_j\|_n}{\sqrt{n}}$.

1. By the concentration inequality [2] applied to (18), there exists constants $t_0(\alpha) > e$ and $C(\alpha)$ such that for all $t > t_0(\alpha)$,

$$\mathbb{P}\left(\left|\frac{1}{n}(\mathbb{X}e_j)^T\xi - Med[\frac{1}{n}(\mathbb{X}e_j)^T\xi]\right| \geq \frac{tn^{\frac{1}{\alpha}} \|\mathbb{X}e_j\|_n}{\sqrt{n}}\right) \leq \varphi(t)$$

where $\varphi(t) = C(\alpha)\left(\frac{\log(t)}{t}\right)^\alpha$. Replacing median by mean, we have

$$\mathbb{P}\left(\left|\frac{1}{n}(\mathbb{X}e_j)^T\xi - \mathbb{E}\left(\frac{1}{n}(\mathbb{X}e_j)^T\xi\right)\right| \geq \frac{tn^{\frac{1}{\alpha}} \|\mathbb{X}e_j\|_n}{\sqrt{n}} + \bar{a}\right) \leq \varphi(t)$$

where $\bar{a} = t_0(\alpha) + \int_{t_0(\alpha)}^\infty C(\alpha)\left(\frac{\log(t)}{t}\right)^\alpha \mathrm{d}t$. We know that $\mathbb{E}\left((\mathbb{X}e_j)^T\xi\right) = \sum_{i=1}^n x_{ij}\mathbb{E}(\xi_i) = 0$. We deduce then that, for all $t > t_0(\alpha)$,

$$\mathbb{P}\left(\left|\frac{1}{n}(\mathbb{X}e_j)^T\xi\right| \geq \frac{tn^{\frac{1}{\alpha}} \|\mathbb{X}e_j\|_n}{\sqrt{n}} + \bar{a}\right) \leq C(\alpha)\left(\frac{\log t}{t}\right)^\alpha$$

It follows that,

$$\begin{aligned}\mathbb{P}\left(\max_{j=1,\dots,p}\left|\frac{1}{n}(\mathbb{X}e_j)^T\xi\right| \geq \frac{tn^{\frac{1}{\alpha}} \|\mathbb{X}e_j\|_n}{\sqrt{n}} + \bar{a}\right) &\leq \sum_{j=1}^p \mathbb{P}\left(\left|\frac{1}{n}(\mathbb{X}e_j)^T\xi\right| \geq \frac{tn^{\frac{1}{\alpha}} \|\mathbb{X}e_j\|_n}{\sqrt{n}} + \bar{a}\right) \\ &\leq pC(\alpha)\left(\frac{\log t}{t}\right)^\alpha\end{aligned}$$

and

$$\begin{aligned}\mathbb{P}\left(\max_{j=1,\dots,p}\left|\frac{1}{n}(\mathbb{X}e_j)^T\xi\right| \leq \frac{tn^{\frac{1}{\alpha}} \|\mathbb{X}e_j\|_n}{\sqrt{n}} + \bar{a}\right) &\geq 1 - pC(\alpha)\left(\frac{\log t}{t}\right)^\alpha \\ &\geq 1 - \frac{pc_\alpha}{t^\alpha}\end{aligned}$$

This probability is greater than $1/2$ if $t \geq (2pc_\alpha)^{1/\alpha}$ and then,

$$\tau \geq \frac{(2pc_\alpha n)^{1/\alpha} \|\mathbb{X}e_j\|_n + \sqrt{n}\bar{a}}{\sqrt{n}}$$

2. The concentration inequality [2, example 9] applied to (20), leads to,

$$\mathbb{P}\left(\left|\frac{1}{n}(\mathbb{X}e_j)^T\xi - Med[\frac{1}{n}(\mathbb{X}e_j)^T\xi]\right| \geq k\frac{\|\mathbb{X}e_j\|_n}{\sqrt{n}}\right) \leq 10\exp\left(\frac{-c_r k}{\max\left(k^r, \log n\right)^{1/r-1}}\right)$$

Replacing median by mean, we have

$$\mathbb{P}\left(\left|\frac{1}{n}(\mathbb{X}e_j)^T\xi - \mathbb{E}\left(\frac{1}{n}(\mathbb{X}e_j)^T\xi\right)\right| \geq k\frac{\|\mathbb{X}e_j\|_n}{\sqrt{n}} + \bar{b}\right) \leq 10\exp\left(\frac{-c_r k}{\max\left(k^r, \log n\right)^{1/r-1}}\right)$$

where $\bar{b} = \displaystyle\int_0^\infty 10\exp\left(\frac{-c_r k}{\max\left(k^r, \log n\right)^{1/r-1}}\right)\mathrm{d}t$. Since $\mathbb{E}\left((\mathbb{X}e_j)^T\xi\right) = 0$, we have,

$$\mathbb{P}\left(\left|\frac{1}{n}(\mathbb{X}e_j)^T\xi\right| \geq k\frac{\|\mathbb{X}e_j\|_n}{\sqrt{n}} + \bar{b}\right) \leq 10\exp\left(\frac{-c_r k}{\max\left(k^r, \log n\right)^{1/r-1}}\right)$$

We deduce that

$$\begin{aligned}\mathbb{P}\left(\max_{j=1,\dots,p}\left|\frac{1}{n}(\mathbb{X}e_j)^T\xi\right| \geq k\frac{\|\mathbb{X}e_j\|_n}{\sqrt{n}} + \bar{b}\right) &\leq \sum_{j=1}^p \mathbb{P}\left(\left|\frac{1}{n}(\mathbb{X}e_j)^T\xi\right| \geq k\frac{\|\mathbb{X}e_j\|_n}{\sqrt{n}} + \bar{b}\right)\\ &\leq 10p\exp\left(\frac{-c_r k}{\max\left(k^r, \log n\right)^{1/r-1}}\right)\end{aligned}$$

and

$$\mathbb{P}\left(\max_{1\leq j\leq p}\left|\frac{1}{n}(\mathbb{X}e_j)^T\xi\right| \leq k\frac{\|\mathbb{X}e_j\|_n}{\sqrt{n}} + \bar{b}\right) \geq 1 - 10p\exp\left(\frac{-c_r k}{\max\left(k^r, \log n\right)^{1/r-1}}\right)$$

if $k^r \geq \log(n)$, this probability is greater than $1/2$. If $k \geq \left(\frac{\ln(20p)}{c_r}\right)^{1/r}$, this implies that $\tau \geq \left(\frac{\ln(20p)}{c_r}\right)^{1/r}\frac{\|\mathbb{X}e_j\|_n}{\sqrt{n}} + \bar{b}$. Otherwise if $k^r \leq \log(n)$, this probability is greater than $1/2$ if $k \geq (\log n)^{1/r-1}\frac{\ln(20p)}{c_r}$, then we have $\tau \geq (\log n)^{1/r-1}\frac{\ln(20p)}{c_r}\frac{\|\mathbb{X}e_j\|_n}{\sqrt{n}} + \bar{b}$.

□

Combining equation (19) of theorem 3.1, proposition 4.2 and lemma 4.2, we have an oracle inequality for the Lasso estimator with tuning parameter that can be explicitly lower bounded in the case of the heavy tailed and the sub-exponential examples discussed above.

Theorem 4.1. *Assume that $\xi \rightsquigarrow \mathbb{P}$ where $\mathbb{P}$ is the n-fold product of $d\mu_\alpha(t) = \dfrac{\alpha(1+|t|)^{-1-\alpha}}{2}dt$ and that $\mathbb{X}$ is deterministic and all the diagonal elements of the matrix $\frac{1}{n}\mathbb{X}^T\mathbb{X}$ are equal to 1. Let κ the maximal correlation between two columns of matrix $\mathbb{X}$. Let $p \geq 2$, $\alpha > 2$, $c_\alpha > 0$ and $s \in \{1, \dots, p\}$. Consider the Lasso estimator $\hat{\beta}$ defined by (22) with tunning parameter*

$$\lambda \geq \frac{4\left((2pc_\alpha n)^{1/\alpha} + \sqrt{n}\bar{a}\right)}{(1-2\sqrt{\kappa s})\sqrt{n}}$$

where $\bar{a} = t_0(\alpha) + \displaystyle\int_{t_0(\alpha)}^\infty C(\alpha)\left(\frac{\log(t)}{t}\right)^\alpha \mathrm{d}t$.

Then, with probability at least $1 - \dfrac{c_\alpha}{\frac{2\left(\left((2pc_\alpha n)^{1/\alpha} + \sqrt{n}\bar{a}\right)\sqrt{s}\right)^\alpha}{n}}$, *we have*

$$\left\|\mathbb{X}\hat{\beta} - \boldsymbol{f}\right\|_n^2 - \|\mathbb{X}\beta - \boldsymbol{f}\|_n^2 \leq \frac{32\left((2pc_\alpha n)^{1/\alpha} + \sqrt{n}\bar{a}\right)^2 s}{n(1-2\sqrt{\kappa s})^2}$$

Proof. Let the matrix $\mathbb{X} \in \mathbb{R}^{n\times p}$ be deterministic such that $\|\mathbb{X}e_j\|_n = 1$ and $c^2 = \dfrac{s}{1-\kappa s}$. Let λ such that $\kappa s(1+c_0)^2 \leq 1 + \dfrac{c_0^2 s}{p-s}$ where $c_0 = \dfrac{\lambda + a}{\lambda - a}$, we get $\lambda = \dfrac{a}{1-2\sqrt{\kappa s}}$ where $a = 2(\tau + t)$.

With $\tau = t = \dfrac{(2pc_\alpha n)^{1/\alpha} + \sqrt{n}\bar{a}}{\sqrt{n}}$, we have $a = \dfrac{4\left((2pc_\alpha n)^{1/\alpha} + \sqrt{n}\bar{a}\right)}{\sqrt{n}}$. By lemma 4.2 $\mathbb{P}(\Omega) \geq 1/2$. Using theorem 2 we obtain :

$$\begin{aligned}\left\|\mathbb{X}\hat{\beta} - \mathbf{f}\right\|_n^2 - \|\mathbb{X}\beta - \mathbf{f}\|_n^2 &\leq \frac{2}{4}\left(4t + \frac{4t}{1-2\sqrt{\kappa s}}\right)^2 \frac{s}{1-\kappa s} \\ &\leq \frac{2}{4} \times \frac{64t^2\left(1-\sqrt{\kappa s}\right)^2}{\left(1-2\sqrt{\kappa s}\right)^2} \times \frac{s}{1-\kappa s} \\ &\leq \frac{32t^2 s}{\left(1-2\sqrt{\kappa s}\right)^2}\end{aligned}$$

□

We have a similar result for the sub-exponential example that we state in the following theorem.

Theorem 4.2. *Assume that $\xi \rightsquigarrow \mathbb{P}$ where $\mathbb{P}$ is the n-fold product of $d\nu_r = d_r e^{-|t|^r} dt$ for $r \in (0,1)$ and that $\mathbb{X}$ is deterministic and all the diagonal elements of the matrix $\frac{1}{n}\mathbb{X}^T\mathbb{X}$ are equal to 1. Let κ the maximal correlation between two columns of matrix $\mathbb{X}$. Let $p \geq 2$ and $s \in \{1, \ldots, p\}$. Consider the Lasso estimator $\hat{\beta}$ defined by (22) with tunning parameter*

$$\lambda \geq \frac{4\left(\ln(20p)\max\left(\frac{\ln(20p)}{c_r}, \log(n)\right)^{1/r-1} + \sqrt{n}c_r\bar{b}\right)}{(1-2\sqrt{\kappa s})\sqrt{n}c_r}$$

Then, with probability at least $1 - 5\exp\left(\dfrac{-\sqrt{s}\left(\ln(20p)\max\left(\frac{\ln(20p)}{c_r}, \log(n)\right)^{1/r-1} + \sqrt{n}c_r\bar{b}\right)}{\max\left(\left(\frac{\ln(20p)\max\left(\frac{\ln(20p)}{c_r}, \log(n)\right)^{1/r-1} + \sqrt{n}c_r\bar{b}}{c_r\sqrt{(1-\kappa s)}}\sqrt{s}\right)^r, \log(n)\right)^{1/r-1}}\right)$,

we have

$$\left\|\mathbb{X}\hat{\beta} - \boldsymbol{f}\right\|_n^2 - \|\mathbb{X}\beta - \boldsymbol{f}\|_n^2 \leq \frac{32\left(\ln(20p)\max\left(\frac{\ln(20p)}{c_r}, \log(n)\right)^{1/r-1} + c_r\sqrt{n}\bar{b}\right)^2 s}{nc_r^2(1-2\sqrt{\kappa s})^2}$$

where $\bar{b} = \displaystyle\int_0^\infty 10\exp\left(\frac{-c_r k}{\max\left(k^r, \log n\right)^{1/r-1}}\right) \mathrm{d}t$ and c_r is a quantity depending only on r.

Proof. We use the same technique as for theorem 4.1. Let λ such that $\kappa s(1+c_0)^2 = 1 \leq 1 + \dfrac{c_0^2 s}{p-s}$ where $c_0 = \dfrac{\lambda + a}{\lambda - a}$, we get $\lambda = \dfrac{a}{1-2\sqrt{\kappa s}}$ where $a = 2(\tau + t)$.
With $\tau = k = \dfrac{\left(\ln(20p)\max\left(\frac{\ln(20p)}{c_r}, \log(n)\right)^{1/r-1} + \sqrt{n}c_r\bar{b}\right)}{\sqrt{n}c_r}$, we have
$a = \dfrac{4\left(\ln(20p)\max\left(\frac{\ln(20p)}{c_r}, \log(n)\right)^{1/r-1} + \sqrt{n}c_r\bar{b}\right)}{\sqrt{n}c_r}$ and by lemma 4.2 we get $\mathbb{P}(\Omega) \geq 1/2$. Using theorem 3.1 we obtain :

$$\begin{aligned}\left\|\mathbb{X}\hat{\beta} - \mathbf{f}\right\|_n^2 - \|\mathbb{X}\beta - \mathbf{f}\|_n^2 &\leq \frac{2}{4}\left(4k + \frac{4k}{1-2\sqrt{\kappa s}}\right)^2 \frac{s}{1-\kappa s} \\ &\leq \frac{2}{4} \times \frac{64k^2\left(1-\sqrt{\kappa s}\right)^2}{\left(1-2\sqrt{\kappa s}\right)^2} \times \frac{s}{1-\kappa s} \\ &\leq \frac{32k^2 s}{\left(1-2\sqrt{\kappa s}\right)^2}.\end{aligned}$$

□

4.4 Application to Group LASSO in case of the heavy tailed example

In order to shorten the length of the paper, we will discuss, in this sub-section, only the case where ξ is having a heavy tailed distribution issued from (18). A similar result can be obtained for the other sub-exponential example. We begin by introducing the notations usually adopted in literature for the group LASSO as follows : Let $G_1, \cdots, G_M$ be a partition of $\{1, \cdots, p\}$. We denote $\beta_{G_k} = (\beta_j)_{j \in G_k}$ and, for every $1 \leq p < \infty$, we define the mixed $(2,q)$-norm and $(2,\infty)$-norm of β as follows :

$$\begin{cases} |\beta|_{2,q} = \left(\sum_{k=1}^{M} \left(\sum_{j \in G_k} \beta_j^2 \right)^{q/2} \right)^{1/q} \\ |\beta|_{2,\infty} = \max\limits_{1 \leq k \leq M} |\beta_{G_k}|_2 \end{cases}$$

For any $\beta \in \mathbb{R}^p$, we define the regularization norm $\|.\|$ as follows :

$$\|\beta\| = |\beta|_{2,1} = \sum_{k=1}^{M} |\beta_{G_k}|_2 \tag{27}$$

The Group LASSO estimator is a solution of the convex minimization problem

$$\hat{\beta} \in \arg\min_{\beta \in \mathbb{R}^p} \|\mathbb{X}\beta - y\|_n^2 + \lambda \sum_{k=1}^{M} |\beta_{G_k}|_2 \tag{28}$$

Without loss of generality, we assume, in the following, that the groups G_k have the same cardinality $|G_k| = p/M, k = 1, \cdots, M$. Since we consider mainly sparse vectors, it is convenient to define a generalisation of the support conecept. To any $\beta \in \mathbb{R}^p$, let :

$$\mathscr{K}(\beta) = \{k \in \{1, \cdots, M\} : \beta_{G_k} \neq 0\}.$$

It plays the role of support by block of vector β. As for the LASSO application, we begin by verifying the decomposability assumptions and we bound the compatibility factor for the heavy tailed example introduced above.

Lemma 4.3. *Consider $\mathbb{X}$ is a matrix $\in \mathbb{R}^{n \times p}$ and $\beta \in \mathbb{R}^p$,*

1. *the decomposability assumptions'2.1 are satisfied when $\mathcal{P}_\beta$ is the orthogonal projection operator onto the linear span of $\left\{ e_j : j \in \bigcup_{k \in \mathscr{K}(\beta)} G_k \right\}$.*
2. *The event Ω defined in (6) takes the following form :*

$$\Omega = \left\{ \max_{1 \leq k \leq M} \frac{1}{n} \left| \mathbb{X}_{G_k}^T \xi \right|_2 \leq \tau \right\}$$

where $\mathbb{X}_{G_k}$ is the $n \times |G_k|$ sub-matrix of $\mathbb{X}$ formed by the columns indexed by G_k.

Proof. 1. For $\beta \in \mathbb{R}^p$ and $\mathcal{P}_\beta = \text{Proj}_{\left\{ e_j : j \in \bigcup_{k \in \mathscr{K}(\beta)} G_k \right\}}$, We have

$$\begin{aligned} \mathcal{P}_\beta \beta &= \sum_{j=1}^{p} \beta_j \text{Proj}_{\left\{ e_j : j \in \bigcup_{k \in \mathscr{K}(\beta)} G_k \right\}} e_j \\ &= \sum_{j=1}^{p} \beta_j e_j \mathbb{1}_{\left\{ j \in \bigcup_{k \in \mathscr{K}(\beta)} G_k \right\}} = \sum_{j \in \bigcup_{k \in \mathscr{K}(\beta)} G_k} \beta_j e_j = \beta. \end{aligned}$$

Consider $\acute{\beta} = \sum_{j=1}^{p} \acute{\beta}_j e_j$. We have,

$$\begin{aligned} |\beta|_{2,1} - \left| \acute{\beta} \right|_{2,1} &= \sum_{k \in \mathscr{K}(\beta)} |\beta_{G_k}|_2 - \sum_{k=1}^{M} \left| \acute{\beta}_{G_k} \right|_2 \\ &= \sum_{k \in \mathscr{K}(\beta)} |\beta_{G_k}|_2 - \sum_{k \in \mathscr{K}(\beta)} \left| \acute{\beta}_{G_k} \right|_2 - \sum_{k \in \mathscr{K}^c(\beta)} \left| \acute{\beta}_{G_k} \right|_2. \end{aligned}$$

Applying the triangular inequality $\sum_{k\in\mathscr{K}(\beta)}|\beta_{G_k}|_2-\sum_{k\in\mathscr{K}(\beta)}\left|\acute{\beta_{G_k}}\right|_2\leq\sum_{k\in\mathscr{K}(\beta)}\left|(\beta-\acute{\beta})_{G_k}\right|_2=\left|\mathcal{P}_\beta(\beta-\acute{\beta})\right|_{2,1}$, we have

$$|\beta|_{2,1}-\left|\acute{\beta}\right|_{2,1}\leq\left|\mathcal{P}_\beta(\beta-\acute{\beta})\right|_{2,1}-\left|\mathcal{P}_\beta^\perp\acute{\beta}\right|_{2,1}.$$

2. The event $\Omega=\left\{\sup_{v\in\mathbb{R}^p:|v|_{2,1}\leq1}\frac{1}{n}\varepsilon^T\mathbb{X}v\leq\tau\right\}$ become

$$\Omega=\left\{\sup_{v\in\mathbb{R}^p:|v|_{2,1}\leq1}\frac{1}{n}\sum_{k=1}^{M}\varepsilon^T\mathbb{X}_{G_k}v_{G_k}\leq\tau\right\}$$
$$=\left\{\sup_{v\in\mathbb{R}^p:|v|_{2,1}\leq1}\frac{1}{n}\left(\left(\mathbb{X}_{G_1}\dots\mathbb{X}_{G_M}\right)^T\xi\right)^T\begin{pmatrix}v_{G_1}\\\vdots\\v_{G_M}\end{pmatrix}\leq\tau\right\}.$$

From the duality between ℓ_1 and l_∞ norms and the mixed $(2,\infty)$-norm, we have,

$$\sup_{v\in\mathbb{R}^p:|v|_{2,1}\leq1}\frac{1}{n}\left(\left(\mathbb{X}_{G_1}\dots\mathbb{X}_{G_M}\right)^T\xi\right)^T\begin{pmatrix}v_{G_1}\\\vdots\\v_{G_M}\end{pmatrix}=\left|\frac{1}{n}\left(\mathbb{X}_{G_1}\dots\mathbb{X}_{G_M}\right)^T\xi\right|_{2,\infty}=\max_{1\leq k\leq M}\frac{1}{n}\left|\mathbb{X}_{G_k}^T\xi\right|_2.$$

□

We consider now to control the compatibility $\mu_{c_0}(\beta)$ for any group sparse vector $\beta\in\mathbb{R}^p$ which means that $|\mathscr{K}(\beta)|$ is much smaller than M the number of groups. For this purpose, we will need the following RE(s) condition introduced in [11, Assumption 3.1].

Assumption 4.1 (RE(s) condition)**.** *For any $S\subset\{1,\dots,M\}$, let $c_0>0$ be a constant and $1\leq s\leq M$ be an integer that gives an upper bound on the group sparsity of a vector δ, the following condition holds.*

$$\kappa_G^2(s,c_0)=\min\left\{\frac{\|\mathbb{X}\delta\|_n^2}{|\delta_S|_2^2}:|S|\leq s,\delta\in\mathbb{R}^p,\sum_{k\in S^c}|\delta_{G_k}|_2\leq c_0\sum_{k\in S}|\delta_{G_k}|_2\right\}>0,\tag{29}$$

where S^c denotes the complement of the set of indices S.

In order to apply theorem 3.1, we need to find τ such that $\mathbb{P}(\Omega)>1/2$. The lower bound of τ is determined in the following lemma.

Lemma 4.4. *Let $\mathbb{X}$ be deterministic and ξ having a distribution of the form (18) with $\alpha>2$ and denote by $\|\mathbb{X}_{G_k}\|_{sp}=\sup_{\substack{v\in\mathbb{R}^p:\\|v|_2\leq1}}|\mathbb{X}_{G_k}v|_2$ the spectral norm of matrix $\mathbb{X}_{G_k}$ and $\|\mathbb{X}_{G_k}\|_{Fr}=\sqrt{\mathbb{X}_{G_k}^T\mathbb{X}_{G_k}}$ its Frobenius norm. Then, set $\psi_{sp}^*=\max_{k=1,\dots,M}\frac{\|\mathbb{X}_{G_k}\|_{sp}}{\sqrt{n}}$ and $\psi_{Fr}^*=\max_{k=1,\dots,M}\frac{\|\mathbb{X}_{G_k}\|_{Fr}}{\sqrt{n}}$. If*

$$\tau\geq\frac{(Mc_\alpha n)^{\frac{1}{\alpha}}\psi_{sp}^*}{\sqrt{n}}+\sqrt{\frac{2}{(\alpha-1)(\alpha-2)}}\frac{\psi_{Fr}^*}{\sqrt{n}}+\bar{c}$$

then $\mathbb{P}(\Omega)\geq1/2$.

Proof. Define $f:\mathbb{R}^n\to\mathbb{R}$ by :

$$f(z)=\left|\mathbb{X}_{G_k}^Tz\right|_2.$$

For every $z_1,z_2\in\mathbb{R}^n$, $|f(z_1)-f(z_2)|\leq\sqrt{n}\psi_{sp}^*|z_1-z_2|_2$. Therefore, f is a Lipschitz function with Lipschitz constant $\sqrt{n}\psi_{sp}^*$. By the concentration inequality [2, equation 3], there exists constants $t_0(\alpha)>e$ and $C(\alpha)$ such that for all $t>t_0(\alpha)$,

$$\mathbb{P}\left(\frac{1}{n}\left|\mathbb{X}_{G_k}^T\xi\right|_2\geq\frac{tn^{\frac{1}{\alpha}}\psi_{sp}^*}{\sqrt{n}}+Med\left(\frac{1}{n}\left|\mathbb{X}_{G_k}^T\xi\right|_2\right)\right)\leq\frac{1}{2}\varphi(t)$$

where $\varphi(t) = C(\alpha)\left(\frac{\log(t)}{t}\right)^{\alpha}$. Replacing median by mean, we have,

$$\mathbb{P}\left(\frac{1}{n}\left|\mathbb{X}_{G_k}^T\xi\right|_2 \geq \frac{tn^{\frac{1}{\alpha}}\psi_{sp}^*}{\sqrt{n}} + \mathbb{E}\left(\frac{1}{n}\left|\mathbb{X}_{G_k}^T\xi\right|_2\right) + \bar{c}\right) \leq \frac{1}{2}\varphi(t)$$

where $\bar{c} = t_0(\alpha) + \int_{t_0(\alpha)}^{\infty} C(\alpha)\left(\frac{\log(t)}{t}\right)^{\alpha} \mathrm{d}t$. As $\left(\mathbb{E}\left(\left|\mathbb{X}_{G_k}^T\xi\right|_2^2\right)\right)^{1/2} \geq \mathbb{E}\left(\left|\mathbb{X}_{G_k}^T\xi\right|_2\right)$ then, for all $t > t_0(\alpha)$, we get,

$$\mathbb{P}\left(\frac{1}{n}\left|\mathbb{X}_{G_k}^T\xi\right|_2 \geq \frac{tn^{\frac{1}{\alpha}}\psi_{sp}^*}{\sqrt{n}} + \left(\mathbb{E}\left(\frac{1}{n^2}\left|\mathbb{X}_{G_k}^T\xi\right|_2^2\right)\right)^{1/2} + \bar{c}\right) \leq C(\alpha)\left(\frac{\log(t)}{t}\right)^{\alpha}.$$

The term $\left|\mathbb{X}_{G_k}^T\xi\right|_2^2$ is a quadratic form which can be writted as :

$$\begin{aligned}\left|\mathbb{X}_{G_k}^T\xi\right|_2^2 &= \xi^T\mathbb{X}_{G_k}\mathbb{X}_{G_k}^T\xi \\ &= \sum_{j=1}^{n}(\mathbb{X}_{G_k}\mathbb{X}_{G_k}^T)_{jj}\xi_j^2 + \sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}(\mathbb{X}_{G_k}\mathbb{X}_{G_k}^T)_{ij}\xi_i\xi_j.\end{aligned}$$

Since $\mathbb{X}_{G_k}$ is deterministic and $\mathbb{E}(\xi_j^2) = \frac{2}{(\alpha-1)(\alpha-2)}$ and $\mathbb{E}(\xi_i\xi_j) = 0$, we have:

$$\begin{aligned}\mathbb{E}\left(\left|\mathbb{X}_{G_k}^T\xi\right|_2^2\right) &= \frac{2}{(\alpha-1)(\alpha-2)}\sum_{j=1}^{n}(\mathbb{X}_{G_k}\mathbb{X}_{G_k}^T)_{jj} \\ &= \frac{2}{(\alpha-1)(\alpha-2)}trace(\mathbb{X}_{G_k}\mathbb{X}_{G_k}^T) \\ &= \frac{2}{(\alpha-1)(\alpha-2)}\left\|\mathbb{X}_{G_k}\right\|_{Fr}^2.\end{aligned}$$

Thus, for all $t > t_0(\alpha)$, we obtain,

$$\mathbb{P}\left(\frac{1}{n}\left|\mathbb{X}_{G_k}^T\xi\right|_2 \geq \left(\frac{tn^{\frac{1}{\alpha}}\psi_{sp}^*}{\sqrt{n}} + \sqrt{\frac{2}{(\alpha-1)(\alpha-2)}}\frac{\psi_{Fr}^*}{\sqrt{n}} + \bar{c}\right)\right) \leq \frac{1}{2}C(\alpha)\left(\frac{\log(t)}{t}\right)^{\alpha}.$$

If we denote $\mathcal{U} \triangleq \mathbb{P}\left(\max_{k=1,\ldots,M}\frac{1}{n}\left|\mathbb{X}_{G_k}^T\xi\right|_2 \geq \left(\frac{tn^{\frac{1}{\alpha}}\psi_{sp}^*}{\sqrt{n}} + \sqrt{\frac{2}{(\alpha-1)(\alpha-2)}}\frac{\psi_{Fr}^*}{\sqrt{n}} + \bar{c}\right)\right)$, we will have,

$$\begin{aligned}\mathcal{U} &\leq \sum_{k=1}^{M}\mathbb{P}\left(\frac{1}{n}\left|\mathbb{X}_{G_k}^T\xi\right|_2 \geq \left(\frac{tn^{\frac{1}{\alpha}}\psi_{sp}^*}{\sqrt{n}} + \sqrt{\frac{2}{(\alpha-1)(\alpha-2)}}\frac{\psi_{Fr}^*}{\sqrt{n}} + \bar{c}\right)\right) \\ &\leq \frac{1}{2}MC(\alpha)\left(\frac{\log(t)}{t}\right)^{\alpha} \leq \frac{1}{2}\frac{Mc_\alpha}{t^\alpha}.\end{aligned}$$

Using the union bound, we get,

$$\mathbb{P}\left(\max_{k=1,\ldots,M}\frac{1}{n}\left|\mathbb{X}_{G_k}^T\xi\right|_2 \leq \left(\frac{tn^{\frac{1}{\alpha}}\psi_{sp}^*}{\sqrt{n}} + \sqrt{\frac{2}{(\alpha-1)(\alpha-2)}}\frac{\psi_{Fr}^*}{\sqrt{n}} + \bar{c}\right)\right) \geq 1 - \frac{1}{2}\frac{Mc_\alpha}{t^\alpha}$$

This probability is greater than $1/2$ if $t^\alpha \geq Mc_\alpha$. We deduce then the desired condition,

$$\tau \geq \frac{(Mc_\alpha n)^{\frac{1}{\alpha}}\psi_{sp}^*}{\sqrt{n}} + \sqrt{\frac{2}{(\alpha-1)(\alpha-2)}}\frac{\psi_{Fr}^*}{\sqrt{n}} + \bar{c}$$

□

Combining theorem 4.1 and lemma 4.4, we have an oracle inequality for the Group Lasso estimator with explicit bounds.

Theorem 4.3. *Let ξ and X satisfying conditions of theorem 4.1. Assume the assumption 4.1 on the RE(s) condition holds for any group s-sparse vectors β (i.e.,$\beta \in \mathbb{R}^p$ such that $|\mathscr{K}(\beta)| \leq s$). The Group Lasso estimator defined by (28) with tuning parameter*

$$\lambda \geq \frac{4\left((Mc_\alpha n)^{\frac{1}{\alpha}}\psi_{sp}^* + \sqrt{\frac{2}{(\alpha-1)(\alpha-2)}}\psi_{Fr}^* + \sqrt{n}\bar{c}\right)}{\sqrt{n}}$$

satisfies, with probability at least $1 - \dfrac{c_\alpha\left(\kappa_G(s,c_0)\right)^\alpha n}{2\left(\left((Mc_\alpha n)^{\frac{1}{\alpha}}\psi_{sp}^* + \sqrt{\frac{2}{(\alpha-1)(\alpha-2)}}\psi_{Fr}^* + \sqrt{n}\bar{c}\right)\sqrt{s}\right)^\alpha}$,

$$\left\|\mathbb{X}\hat{\beta} - \boldsymbol{f}\right\|_n^2 - \|\mathbb{X}\beta - \boldsymbol{f}\|_n^2 \leq \frac{32\left((Mc_\alpha n)^{\frac{1}{\alpha}}\psi_{sp}^* + \sqrt{\frac{2}{(\alpha-1)(\alpha-2)}}\psi_{Fr}^* + \sqrt{n}\bar{c}\right)^2 s}{n\kappa_G^2(s,c_0)}.$$

Proof. Let matrix $\mathbb{X} \in \mathbb{R}^{n\times p}$ be deterministic and β any group s-sparse vectors (i.e., $|\mathscr{K}(\beta)| \leq s$). Assume that $RE(s)$ condition holds, we have

$$\left|\beta_{\mathscr{K}(\beta)}\right|_2 = |\mathcal{P}_\beta\beta|_2 \leq \frac{\|\mathbb{X}\beta\|_n}{\kappa_G(\mathscr{K}(\beta),c_0)}$$

where $\mathcal{P}_\beta = \mathrm{Proj}_{\{e_j : j\in\bigcup_{k\in\mathscr{K}(\beta)}G_k\}}$. Using equivalence between mixed $(2,1)$-norm and ℓ_2 norm, $|\delta_S|_{2,1} \leq \sqrt{|S|}\,|\delta_S|_2$, we have:

$$\left|\beta_{\mathscr{K}(\beta)}\right|_{2,1}^2 = |\mathcal{P}_\beta\beta|_{2,1}^2 \leq |\mathscr{K}(\beta)|\,|\mathcal{P}_\beta\beta|_2^2 \leq \frac{|\mathscr{K}(\beta)|}{\kappa_G^2(\mathscr{K}(\beta),c_0)}\|\mathbb{X}\beta\|_n^2.$$

Hence,

$$\mu_{c_0}^2(\beta) \leq \frac{|\mathscr{K}(\beta)|}{\kappa_G^2(\mathscr{K}(\beta),c_0)} \leq \frac{s}{\kappa_G^2(s,c_0)}$$

Let $c^2 \leq \dfrac{s}{\kappa_G^2(s,c_0)}$ and $\tau = \frac{(Mc_\alpha n)^{\frac{1}{\alpha}}\psi_{sp}^*}{\sqrt{n}} + \sqrt{\frac{2}{(\alpha-1)(\alpha-2)}}\frac{\psi_{Fr}^*}{\sqrt{n}} + \bar{c}$, by lemma 4.4 and theorem 4.1, we have,

$$\begin{aligned}
\left\|\mathbb{X}\hat{\beta} - \mathbf{f}\right\|_n^2 - \|\mathbb{X}\beta - \mathbf{f}\|_n^2 &\leq \frac{(\lambda + 2(\tau+t))^2}{4}\mu_{c_0}^2(\beta) + \frac{(\lambda+2(\tau+t))^2}{4}c^2 \\
&\leq 2\lambda^2\frac{s}{\kappa_G^2(s,c_0)} \\
&\leq \frac{2\times 16\left((Mc_\alpha n)^{\frac{1}{\alpha}}\psi_{sp}^* + \sqrt{\frac{2}{(\alpha-1)(\alpha-2)}}\psi_{Fr}^* + \sqrt{n}\bar{c}\right)^2 s}{n\kappa_G^2(s,c_0)}
\end{aligned}$$

□

Conclusion

In this paper we discussed penalized least squares estimators with convex penalty or regularisation norms. We focused on regression model where the observations noise are independent and follow a probability measure which satisfies a weak spectral gap (or weak Poincaré) inequality. We established oracle inequalities in probability for the prediction error for the Lasso and Group Lasso estimators. Our results have been applied to tow example of non gaussian cases; namely a heavy tailed and a sub-exponential examples. For these cases we explicit the oracle inequalities bounds in a close form.

References

[1] A.Tsybakov A.Bellec G.Lecué. "Towards the study of least squares estimators with convex penalty". In: *Arxiv* (2017).

[2] Franck Barthe, Patrick Cattiaux, and Cyril Roberto. “Concentration for independent random variables with heavy tails”. In: *Applied Mathematics Research eXpress* 2005.2 (2005), pp. 39–60.

[3] Pierre C Bellec, Guillaume Lecué, Alexandre B Tsybakov, et al. “Slope meets lasso: improved oracle bounds and optimality”. In: *Annals of Statistics* 46.6B (2018), pp. 3603–3642.

[4] Pierre Bellec and Alexandre Tsybakov. “Bounds on the prediction error of penalized least squares estimators with convex penalty”. In: *International Conference on Modern Problems of Stochastic Analysis and Statistics*. Springer. 2016, pp. 315–333.

[5] Peter J Bickel, Ya’acov Ritov, Alexandre B Tsybakov, et al. “Simultaneous analysis of Lasso and Dantzig selector”. In: *The Annals of statistics* 37.4 (2009), pp. 1705–1732.

[6] Peter Bühlmann and Sara Van De Geer. *Statistics for high-dimensional data: methods, theory and applications*. Springer Science & Business Media, 2011.

[7] Patrick Cattiaux and Arnaud Guillin. “On the Poincaré constant of log-concave measures”. In: *Geometric Aspects of Functional Analysis*. Springer, 2020, pp. 171–217.

[8] Sara A van de Geer. *Estimation and testing under sparsity*. Springer, 2016.

[9] Christophe Giraud. “Graphical Models”. In: *Introduction to High-Dimensional Statistics*. Chapman and Hall/CRC, 2014, pp. 157–180.

[10] Vladimir Koltchinskii and Shahar Mendelson. “Bounding the smallest singular value of a random matrix without concentration”. In: *International Mathematics Research Notices* 2015.23 (2015), pp. 12991–13008.

[11] Karim Lounici et al. “Oracle inequalities and optimal inference under group sparsity”. In: *Annals of statistics* 39.4 (2011), pp. 2164–2204.

[12] Shahar Mendelson. “Learning without concentration”. In: *Conference on Learning Theory*. PMLR. 2014, pp. 25–39.

[13] Sahand N Negahban et al. “A unified framework for high-dimensional analysis of M-estimators with decomposable regularizers”. In: *Statistical science* 27.4 (2012), pp. 538–557.

[14] Ivan Selesnick. “Sparse regularization via convex analysis”. In: *IEEE Transactions on Signal Processing* 65.17 (2017), pp. 4481–4494.

[15] Jonathan Taylor et al. “The geometry of least squares in the 21st century”. In: *Bernoulli* 19.4 (2013), pp. 1449–1464.