

Sparse Support Recovery with Non-smooth Loss Functions

Kévin Degraux

ISPGROUP/ICTEAM, FNRS
UCLouvain, Belgium

Gabriel Peyré

CNRS, École Normale Supérieure,
DMA, Paris, France

Jalal M. Fadili

Normandie Univ, ENSICAEN,
CNRS, GREYC, Caen, France

Laurent Jacques

ISPGROUP/ICTEAM, FNRS
UCLouvain, Belgium

Abstract—In this work, we study the support recovery guarantees of underdetermined sparse regression using the ℓ_1 -norm as a regularizer and a non-smooth loss function for data fidelity. More precisely, we focus on the ℓ_1 and ℓ_∞ losses, and contrast them with the usual ℓ_2 smooth loss. We identify an “extended support” for the vector to recover and derive a sharp condition which ensures that it is stable to small additive noise in the observations. We give a detailed numerical analysis of the support stability of compressed sensing recovery with these different losses. This highlights different parameter regimes, ranging from support stability to increasing support instability.

I. INTRODUCTION

This work studies sparse linear inverse problems of the form

$$y = \Phi x_0 + w,$$

where $x_0 \in \mathbb{R}^n$ is the vector to estimate, assumed non-zero and sparse with support $I \stackrel{\text{def}}{=} \text{supp}(x_0)$, $w \in \mathbb{R}^m$ is some additive noise and the design matrix $\Phi^{m \times n}$ is in general rank deficient, *i.e.*, typically in the high-dimensional regime where $m \ll n$. In order to recover x_0 , we consider the following sparsity-promoting optimization problem

$$x_\tau \in \underset{x \in \mathbb{R}^n}{\text{Argmin}} \{ \|x\|_1 \text{ s.t. } \|\Phi x - y\|_\alpha \leq \tau \}, \quad (\mathcal{P}_\alpha^\tau(y))$$

where the constraint size $\tau \geq 0$ should be adapted to the noise level. The usual “smooth” ℓ_2 loss function has been studied in depth in the literature. In contrast, the ℓ_1 and ℓ_∞ loss functions which are the focus of this paper, are polyhedral and non-smooth. They lead to significantly different estimation results. The ℓ_1 case corresponds to a “robust” loss function, and is important to cope with impulse noise or outliers contaminating the data (see [9], [7]). At the extreme opposite, the ℓ_∞ loss is typically used to handle uniform noise such as in quantization (see [8]). This paper studies the stability of the support $\text{supp}(x_\tau)$. In particular, we provide a sharp analysis for the polyhedral ℓ_1 and ℓ_∞ cases that allows one to control the deviation of $\text{supp}(x_\tau)$ from $\text{supp}(x_0)$ if $\|w\|_\alpha$ is not too large and τ is chosen proportionally to $\|w\|_\alpha$. The general case is studied numerically in a compressed sensing experiment where we compare $\text{supp}(x_\tau)$ and $\text{supp}(x_0)$ for $\alpha \in [1, +\infty]$.

The main assumption of this work is that x_0 is *identifiable*, *i.e.*, it is solution of the Basis-Pursuit problem [1],

$$\min_x \{ \|x\|_1 \text{ s.t. } \Phi x = \Phi x_0 \}. \quad (\mathcal{P}^0(\Phi x_0))$$

Assuming x_0 is identifiable, $\text{supp}(x_\tau)$ will be characterized by an element in the set of *minimum norm certificates* (see [10])

$$p_\beta \in \underset{p \in \mathbb{R}^m}{\text{Argmin}} \{ \|p\|_\beta \text{ s.t. } \Phi_{\cdot, I}^* p = \text{sign}(x_{0, I}), \|\Phi^* p\|_\infty \leq 1 \}, \quad (1)$$

where β is defined by $\frac{1}{\alpha} + \frac{1}{\beta} = 1$ and for sets of indices S and I , $\Phi_{S, I}$ is the submatrix of Φ restricted to the rows in S and the columns in I . When all are kept, a dot replaces the index set (*e.g.*, $\Phi_{\cdot, I}$). In turn, we define the *extended support* as

$$J \stackrel{\text{def}}{=} \text{sat}(\Phi^* p_\beta) = \{ i \in \{1, \dots, n\} \mid |(\Phi^* p_\beta)_i| = 1 \}. \quad (2)$$

Finally, let us define the *model tangent subspace* $T_\beta \stackrel{\text{def}}{=} \text{par}(\partial \|\Phi p_\beta\|_\beta)^\perp$ [11], where $\text{par}(C)$ is the subspace parallel to the convex set C , ∂ is the subdifferential operator, P_{T_β} is the orthogonal projection onto T_β ; see Figure 1 for illustration. The *restricted injectivity condition* then reads

$$\text{Ker}(P_{T_\beta} \Phi_{\cdot, J}) = \{0\}. \quad (\text{INJ}_\alpha)$$

It is possible to show [2] that this condition actually holds almost surely, *e.g.*, when the entries of the design matrix follow a continuous distribution. Table I summarizes for the three specific cases $\alpha \in \{1, 2, +\infty\}$ the quantities introduced here and in the next section.

II. MAIN RESULT

Our main contribution is Theorem 1 below. A similar result is known to hold in the case of the smooth ℓ_2 loss [4], [3]. Our paper extends it to the more challenging case of non-smooth losses $\alpha \in \{1, +\infty\}$. Though we focus here on the case of ℓ_α data loss functions, for $\alpha \in \{1, 2, \infty\}$, our analysis can be extended to more general non-smooth losses, including coercive gauges.

Theorem 1. *Let $\alpha \in \{1, 2, +\infty\}$. Suppose that x_0 is identifiable, and let p_β be a minimal norm certificate (see (1)) with associated extended support J (see (2)). Suppose that the restricted injectivity condition (INJ_α) is satisfied and let $v_{\beta, J} \stackrel{\text{def}}{=} (P_{T_\beta} \Phi_{\cdot, J})^+ P_{T_\beta} \partial \|\Phi p_\beta\|_\beta$. Then there exist constants $c_1, c_2 > 0$ depending only on Φ and p_β such that, for any (w, τ) satisfying*

$$\|w\|_\alpha < c_1 \tau \quad \text{and} \quad \tau \leq c_2 \underline{x} \quad \text{where} \quad \underline{x} \stackrel{\text{def}}{=} \min_{i \in I} |x_{0, I}|, \quad (3)$$

a solution x_τ of $(\mathcal{P}_\alpha^\tau(\Phi x_0 + w))$ with support equal to J is given by

$$x_{\tau, J} \stackrel{\text{def}}{=} x_{0, J} + (P_{T_\beta} \Phi_{\cdot, J})^+ w - \tau v_{\beta, J}. \quad (4)$$

This theorem shows that if the noise level $\|w\|_\alpha$ is small and τ is chosen in proportion to the minimal signal-to-noise ratio, then there is a solution supported exactly in the extended support J . Note in particular that this solution (4) has the correct sign pattern $\text{sign}(x_{\tau, I}) = \text{sign}(x_{0, I})$, but might exhibit outliers if $\tilde{J} \stackrel{\text{def}}{=} J \setminus I \neq \emptyset$. The special case $I = J$ characterizes the exact support stability (“sparsistency”).

III. NUMERICAL EXPERIMENTS

To shed light on this result, we show on Figure 2, a small simulated CS example for $(\alpha, \beta) = (\infty, 1)$. On Figure 3 we address numerically the problem of comparing $\text{supp}(x_\tau)$ and $\text{supp}(x_0)$ for a sweep over $\alpha \in [1, \infty]$.

KD and LJ are funded by the Belgian F.R.S.-FNRS. JF is partly supported by Institut Universitaire de France. GP is supported by the European Research Council (ERC project SIGMA-Vision). The paper corresponding to this extended abstract is [2], accepted to NIPS 2016, Barcelona, Spain. See <https://arxiv.org/abs/1611.01030> for supplementary material.

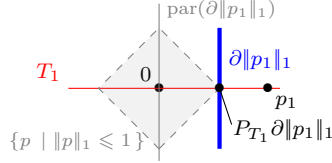


Fig. 1: Model tangent subspace T_β in $\mathbb{R}^2$ for $(\alpha, \beta) = (\infty, 1)$.

α	T_β	(INJ_α)	$(P_{T_\beta} \Phi_{\cdot, J})^+$	$v_{\beta, J}$
2	$\mathbb{R}^m$	$\text{Ker}(\Phi_{\cdot, J}) = \{0\}$	$\Phi_{\cdot, J}^+$	$\Phi_{\cdot, J}^+ \frac{p_2}{\ p_2\ _2}$
∞	$\{u \mid \text{supp}(u) = S\}$	$\text{Ker}(\Phi_{S, J}) = \{0\}$	$\Phi_{S, J}^{-1} \text{Id}_{S, \cdot}$	$\Phi_{S, J}^{-1} \text{sign}(p_1, S)$
1	$\{u \mid u_Z = \rho \text{sign}(p_{\infty, Z}), \rho \in \mathbb{R}\}$	$\text{Ker}\left(\begin{bmatrix} \Phi_{Z^c, J} \\ q_Z^* \Phi_{Z, J} \end{bmatrix}\right) = \{0\}$	$\begin{bmatrix} \Phi_{Z^c, J} \\ q_Z^* \Phi_{Z, J} \end{bmatrix}^{-1} \begin{bmatrix} \text{Id}_{Z^c, \cdot} \\ q_Z^* \text{Id}_{Z, \cdot} \end{bmatrix}$	$\begin{bmatrix} \Phi_{Z^c, J} \\ q_Z^* \Phi_{Z, J} \end{bmatrix}^{-1} \begin{bmatrix} 0_{ Z^c } \\ 1 \end{bmatrix}$

TABLE I: Model tangent subspace, restricted injectivity condition and $v_{\beta, J}$ with $S \stackrel{\text{def}}{=} \text{supp}(p_1)$, $Z \stackrel{\text{def}}{=} \text{sat}(p_\infty)$ and $q_Z \stackrel{\text{def}}{=} \text{sign}(p_{\infty, Z})$. Note that Z^c is the complementary index set of Z .

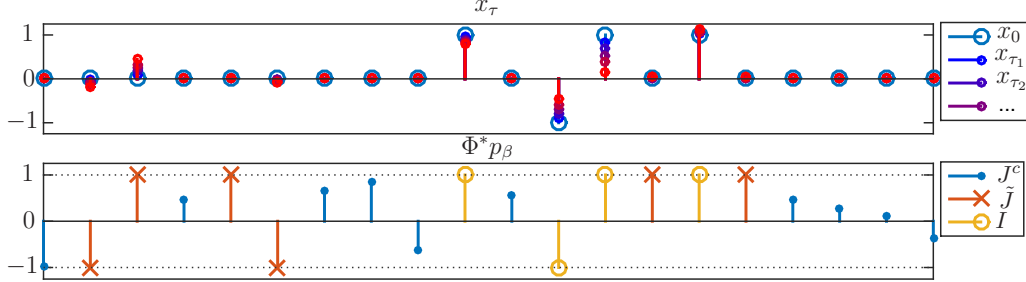


Fig. 2: (best observed in color) Simulated compressed sensing example showing x_τ (above) for increasing values of τ and random noise w respecting the hypothesis of Theorem 1 and $\Phi^* p_\beta$ (below) with $\tilde{J} \stackrel{\text{def}}{=} J \setminus I$ and J^c is the complementary index set of $J (= \text{sat}(\Phi^* p_\beta))$ which predicts the support of x_τ when $\tau > 0$. The parameters are $n = 20$, $m = 10$, $|I| = 4$, $x_{0, I} \in \{\pm 1\}^{|I|}$ and $\Phi \in \mathbb{R}^{m \times n}$ with $\Phi_{i, j} \sim \text{i.i.d. } \mathcal{N}(0, 1)$ and we use CVX/MOSEK [6], [5] at best precision. The noise w is uniformly distributed with $w_i \sim \text{i.i.d. } \mathcal{U}(-\delta, \delta)$ and δ chosen appropriately to ensure that the hypotheses hold. Observe that as we increase τ , new non-zero entries appear in x_τ but because w and τ are small enough, as predicted, we have $\text{supp}(x_\tau) = J$.

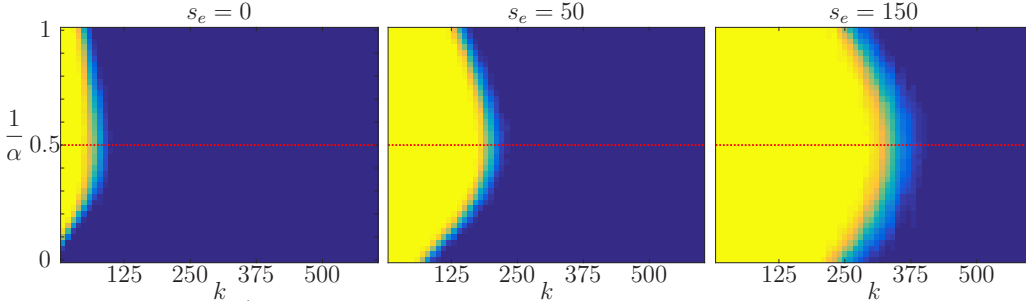


Fig. 3: (best observed in color) Sweep over $\frac{1}{\alpha} \in [0, 1]$ of the empirical probability as a function of k that x_0 is identifiable and $|J_{p_\beta} \setminus I| \leq s_e$ for three values of the support excess threshold $s_e \in \{0, 50, 150\}$. The dotted red line indicates $\alpha = 2$. All computations use CVX/MOSEK [6], [5] at best precision. We set $n = 1000$, $m = 900$ and generate 200 times the sensing matrix $\Phi \in \mathbb{R}^{m \times n}$ with $\Phi_{i, j} \sim \text{i.i.d. } \mathcal{N}(0, 1)$. We generate 60 different k -sparse vectors x_0 with support I where $k \stackrel{\text{def}}{=} |I|$ varies from 10 to 600. The non-zero entries of x_0 are randomly picked in $\{\pm 1\}$ with equal probability. The yellow to blue transition can be interpreted as the maximal k to ensure, with high probability, that $|J_{p_\beta} \setminus I| \leq s_e$. It is always (for all s_e) further to the right at $\alpha = 2$ which means that the ℓ_2 data fidelity constraint provides the highest support stability. This maximal k decreases gracefully as α moves away from 2 in one way or the other. The ℓ_1 loss function has a small advantage over the ℓ_∞ loss.

REFERENCES

- [1] S. S. Chen, D. L. Donoho, and M. A. Saunders. Atomic Decomposition by Basis Pursuit. *SIAM Journal on Scientific Computing*, 20(1):33–61, jan 1998.
- [2] Kévin Degraux, Gabriel Peyré, Jalal M. Fadili, and Laurent Jacques. Sparse support recovery with non-smooth loss functions. In *Advances In Neural Information Processing Systems 29*, pages 4269–4277. Curran Associates, Inc., 2016.
- [3] V. Duval and G. Peyré. Sparse spikes deconvolution on thin grids. Preprint 01135200, HAL, 2015.
- [4] J.-J. Fuchs. On sparse representations in arbitrary redundant bases. *IEEE Transactions on Information Theory*, 50(6):1341–1344, 2004.
- [5] M. Grant and S. Boyd. Graph implementations for nonsmooth convex programs. In V. Blondel, S. Boyd, and H. Kimura, editors, *Recent Advances in Learning and Control*, Lecture Notes in Control and Information Sciences, pages 95–110. Springer-Verlag Limited, 2008.
- [6] M. Grant and S. Boyd. CVX: Matlab software for disciplined convex programming, version 2.1. <http://cvxr.com/cvx>, March 2014.
- [7] L. Jacques. On the optimality of a L1/L1 solver for sparse signal recovery from sparsely corrupted compressive measurements. *Technical Report, TR-LJ-2013.01*, arXiv preprint arXiv:1303.5097, 2013.
- [8] L. Jacques, D. K. Hammond, and Jalal M. Fadili. Dequantizing Compressed Sensing: When Oversampling and Non-Gaussian Constraints Combine. *IEEE Transactions on Information Theory*, 57(1):559–571, jan 2011.
- [9] M. Nikolova. A variational approach to remove outliers and impulse noise. *Journal of Mathematical Imaging and Vision*, 20(1), 2004.
- [10] R. T. Rockafellar. *Conjugate duality and optimization*, volume 16. Siam, 1974.
- [11] S. Vaiter, G. Peyré, and J. Fadili. Model consistency of partly smooth regularizers. Preprint 00987293, HAL, 2014.