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 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 $I \overset{\text{def.}}{=} \operatorname{supp}(x_0)$, $w \in \mathbb{R}^m$ 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

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

where the constraint size $\tau \geqslant 0$ 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 $\operatorname{supp}(x_\tau)$. In particular, we provide a sharp analysis for the polyhedral ℓ_1 and ℓ_∞ cases that allows one to control the deviation of $\operatorname{supp}(x_\tau)$ from $\operatorname{supp}(x_0)$ 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 $\operatorname{supp}(x_\tau)$ and $\operatorname{supp}(x_0)$ for $\alpha \in [1, +\infty]$.

Under the assumption that x_0 is solution of Basis-Pursuit [1],

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

we prove that $\operatorname{supp}(x_\tau)$ will be characterized by an element in the set of *minimum norm certificates* (see [10])

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

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. In turn, we define the *extended support* as

$$J \overset{\text{def.}}{=} \operatorname{sat}(\Phi^* p_\beta) = \left\{i \in \{1,\dots,n\} \mid |(\Phi^* p_\beta)_i| = 1\right\}. \qquad (2)$$

Finally, let us define the *model tangent subspace* $T_\beta \overset{\text{def.}}{=} \operatorname{par}(\partial\|p_\beta\|_\beta)^\perp$ [11], where $\operatorname{par}(\mathcal{C})$ is the subspace parallel to the convex set $\mathcal{C}$, ∂ 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

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

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 non-smooth losses $\alpha \in \{1, +\infty\}$.

Theorem 1. *Let $\alpha \in \{1, 2, +\infty\}$. Suppose that x_0 is solution to $(\mathcal{P}^0(\Phi x_0))$ 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 $v_{\beta,J} \overset{\text{def.}}{=} (P_{T_\beta}\Phi_{\cdot,J})^+ P_{T_\beta}\partial\|p_\beta\|_\beta$. Then there exist constants $c_1, c_2 > 0$ depending only on Φ and p_β such that, for any (w, τ) satisfying*

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

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

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

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 $\operatorname{sign}(x_{\tau,I}) = \operatorname{sign}(x_{0,I})$, but might exhibit outliers if $\tilde{J} \overset{\text{def.}}{=} J\backslash I \neq \emptyset$. The special case $I = J$ characterizes the exact support stability ("sparsistency"). The hypotheses as well as the contants c_1 and c_2 depend on Φ and, even though it goes beyond our scope, one could study the influence of m on these assumptions.

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 $\operatorname{supp}(x_\tau)$ and $\operatorname{supp}(x_0)$ for a sweep over $\alpha \in [1, \infty]$. It shows that under the assumptions of Theorem 1, the ℓ_2 data fidelity constraint provides the highest support stability and the ℓ_1 loss function has a small advantage over the ℓ_∞ loss.

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, Barcelone, Spain. See https://arxiv.org/abs/1611.01030 for supplementary material.

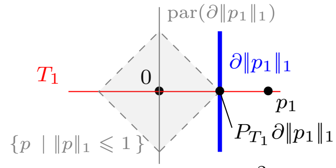

Fig. 1: *Model tangent subspace T_β in $\mathbb{R}^2$ for $(\alpha, \beta) = (\infty, 1)$.*

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

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

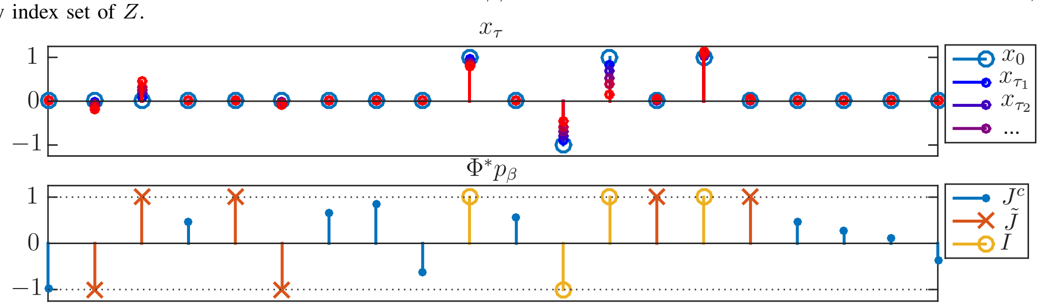

Fig. 2: *(best observed in color) Simulated compressed sensing example showing x_τ (above) for increasing values of τ and random noise w respecting the hypothesis of Theorem 1 and $\Phi^* p_\beta$ (below) with $\tilde{J} \overset{\text{def.}}{=} J \backslash I$ and J^c is the complementary index set of $J(= \mathrm{sat}(\Phi^* p_\beta))$ 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 $\Phi_{i,j} \sim_{i.i.d.} \mathcal{N}(0,1)$ and we use CVX/MOSEK [6], [5] at best precision. The noise w is uniformly distributed with $w_i \sim_{i.i.d.} \mathcal{U}(-\delta, \delta)$ 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 $\mathrm{supp}(x_\tau) = J$.*

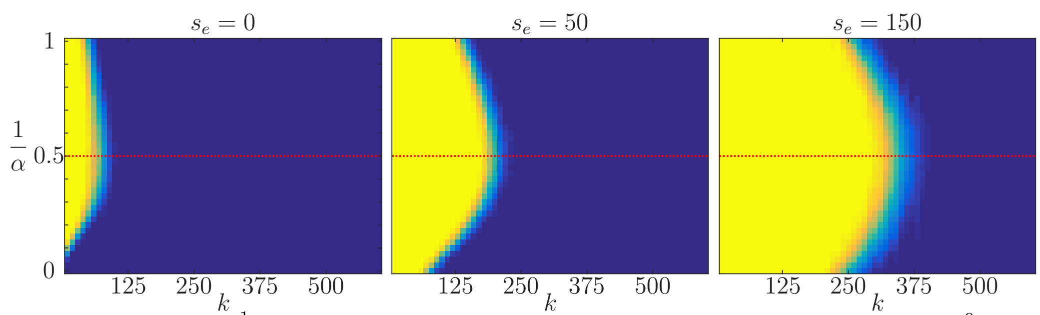

Fig. 3: *(best observed in color) Sweep over $\frac{1}{\alpha} \in [0,1]$ of the empirical probability as a function of k that x_0 is solution to $(\mathcal{P}^0(\Phi x_0))$ and $|J_{p_\beta} \backslash I| \leqslant s_e$ for three values of the* support excess threshold *$s_e \in \{0, 50, 150\}$. The 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 $\Phi_{ij} \sim_{\mathrm{i.i.d}} \mathcal{N}(0,1)$. We generate 60 different k-sparse vectors x_0 with support I where $k \overset{\text{def.}}{=} |I|$ 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 $|J_{p_\beta} \backslash I| \leqslant s_e$. 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.