Sparse Support Recovery with Non-smooth Loss Functions

Kévin Degraux  
ISG Group/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  
ISG Group/ICTEAM, FNRS  
UCLouvain, Belgium

Abstract—In this work, we study the support recovery guarantees of underdetermined sparse regression using the ℓ₂-norm as a regularizer and a non-smooth loss function for data fidelity. More precisely, we focus on the ℓ₁ and ℓ∞ losses, and contrast them with the usual ℓ₂ 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 \( J_0 \equiv \text{supp}(x_0), w \in \mathbb{R}^m \) 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

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

where the constraint size \( \tau \geq 0 \) should be adapted to the noise level. The usual “smooth” ℓ₂ loss function has been studied in depth in the literature. In contrast, the ℓ₁ and ℓ∞ loss functions which are the focus of this paper, are polyhedral and non-smooth. They lead to significantly different estimation results. The ℓ₁ case corresponds to a “robust” loss function, and is important to cope with impulse noise or outliers contaminating the data (see [10], [11]). At the extreme opposite, the ℓ∞ loss is typically used to handle uniform noise such as in quantization (see [3]). This paper studies the stability of the support \( \text{supp}(x_\tau) \). In particular, we provide a sharp analysis for the polyhedral \( \ell_1 \) and \( \ell_\infty \) cases that allows one to control the deviation of support \( \text{supp}(x_\tau) \) from \( \text{supp}(x_0) \); if \( |w|_\alpha \), is not too large and \( \tau \) 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 \( \alpha \in [1, +\infty) \).

II. MAIN RESULT

Our main contribution is Theorem 1 below. A similar result is known to hold in the case of the smooth ℓ₂ 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 \( (P_\alpha^\tau(\Phi x_0)) \) and let \( p_\beta \) be a minimal norm certificate (see [2]) with associated extended support \( J \) (see [1]). Suppose that the restricted injectivity condition \( (\text{INJs}) \) is satisfied and let \( u_{\beta,J} \equiv (P_{\beta,J} \Phi_\cdot,\cdot)^+ P_{\beta,J} \|p_\beta\|_\beta \). Then there exist constants \( c_1, c_2 > 0 \) depending only on \( \Phi \) and \( p_\beta \) such that, for any \( (w, \tau) \) satisfying

\[ |w|_\alpha < c_1 \tau \quad \text{and} \quad \tau \leq c_2 \|x_0\|_I, \quad (3) \]

a solution \( x_\tau \) of \( (P_\alpha^\tau(\Phi x_0 + w)) \) with support equal to \( J \) reads

\[ x_{\tau,J} \equiv x_0,J + (P_{\beta,J} \Phi_{\cdot,J})^+ w - \tau v_{\beta,J}. \quad (4) \]

This theorem shows that if the noise level \( |w|_\alpha \) is small and \( \tau \) 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 \( \text{sign}(x_{\tau,J}) = \text{sign}(x_{0,J}) \), but might exhibit outliers if \( J \subseteq \bar{J} \neq \emptyset \). The special case \( I = J \) characterizes the exact support stability (“sparsistency”). The hypotheses as well as the constants \( c_1 \) and \( c_2 \) depend on \( \Phi \) 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 \((\alpha, \beta) = (2, \infty)\). On Figure 3 we address numerically the problem of comparing \( \text{supp}(x_\tau) \) and \( \text{supp}(x_0) \) for a sweep over \( \alpha \in [1, \infty] \). It shows that under the assumptions of Theorem 1 the ℓ₂ data fidelity constraint provides the highest support stability and the ℓ₁ loss function has a small advantage over the ℓ∞ loss.
TABLE I: Model tangent subspace, restricted injectivity condition and \( v_{\beta,J} \) with \( S^d = \text{supp}(p_1) \), \( Z^d = \text{sat}(p_{\infty}) \) and \( q_Z^d = \text{sign}(p_{\infty,Z}) \). Note that \( Z^c \) is the complementary index set of \( Z \).

<table>
<thead>
<tr>
<th>( \alpha )</th>
<th>( T_{\beta} )</th>
<th>( (\text{INJ}_\alpha) )</th>
<th>( (P_{\beta,J} \Phi,J)^{+} )</th>
<th>( v_{\beta,J} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>( \mathbb{R}^m )</td>
<td>( \text{Ker}(\Phi,J) = {0} )</td>
<td>( \Phi^+_{J} )</td>
<td>( \Phi^+<em>{J} \mid</em>{\text{p}_1^2} )</td>
</tr>
<tr>
<td>( \infty )</td>
<td>( {u \mid \text{supp}(u) = S} )</td>
<td>( \text{Ker}(\Phi_{S,J}) = {0} )</td>
<td>( \Phi_{S,J}^{-1} \text{Id}_{S,J} )</td>
<td>( \Phi_{S,J}^{-1} \text{sign}(p_{1,S}) )</td>
</tr>
<tr>
<td>1</td>
<td>( {u \mid u_Z = \rho \text{sign}(p_{\infty,Z}), \rho \in \mathbb{R}} )</td>
<td>( \text{Ker}(\Phi_{Z,J} \Phi,J) = {0} )</td>
<td>( \Phi_{Z,J}^{-1} \text{Id}_{Z,J} )</td>
<td>( \Phi_{Z,J}^{-1} \text{sign}(p_{1,Z}) )</td>
</tr>
</tbody>
</table>

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

Fig. 2: (best observed in color) Simulated compressed sensing example showing \( x_{\tau} \) (above) for increasing values of \( \tau \) and random noise \( w \) respecting the hypothesis of Theorem 1 and \( \Phi^p_{J} \) (below) with \( J \equiv J' \). \( J \) and \( J' \) is the complementary index set of \( J = \text{sat}(\Phi^p_{\beta,J}) \) which predicts the support of \( x_{\tau} \) when \( \tau > 0 \). The parameters are \( n = 20, m = 10, |I| = 4, x_0 \in \{\pm 1\}^{|I|} \) and \( \Phi \in \mathbb{R}^{m \times n} \) with \( \Phi_{0,J} \sim_{\text{i.i.d.}} N(0, 1) \) and we use CVX/MOSEK \[ 6 \], \[ 5 \] at best precision. The noise \( w \) is uniformly distributed with \( v_0 \sim_{\text{i.i.d.}} \mathcal{U}(-\delta, \delta) \) and \( \delta \) chosen appropriately to ensure that the hypotheses hold. Observe that as we increase \( \tau \), new non-zero entries appear in \( x_{\tau} \) but because \( w \) and \( \tau \) 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 solution to \( \text{p}_1^2(\Phi x) \) and \( |J_{\beta,J}| \leq s_e \) 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 \( \Phi_{0,J} \sim_{\text{i.i.d.}} N(0, 1) \). We generate 60 different \( k \)-sparse vectors \( x_0 \) with support \( I \) where \( k \leq |I| \) 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 \( |J_{\beta,J}| \leq s_e \). It is always (for all \( s_e \)) further to the right at \( \alpha = 2 \) which means that the \( \ell_2 \) data fidelity constraint provides the highest support stability. This maximal \( k \) decreases gracefully as \( \alpha \) moves away from 2 in one way or the other. The \( \ell_1 \) loss function has a small advantage over the \( \ell_{s_e} \) loss.

REFERENCES


