Robustness of Sparse Recovery via F-minimization: A Topological Viewpoint

Center for Science of Information
NSF Science and Technology Center

Jingbo Liu (Princeton Univ.), Jian Jin, Yuantao Gu (Tsinghua Univ.)

Introduction

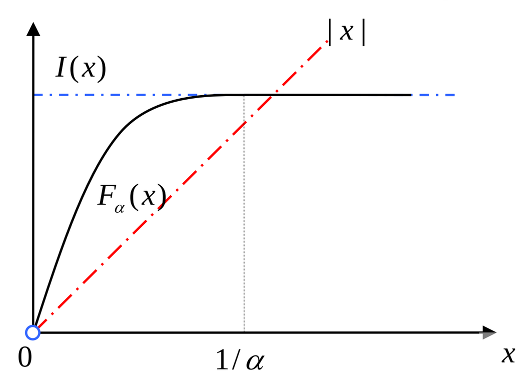
- In fields such as compressed sensing or statistical learning, a basic model problem is to recover a sparse signal $\bar{\mathbf{x}} \in \mathbb{R}^n$ from a set of noisy linear measurements $\mathbf{y} = \mathbf{A}\bar{\mathbf{x}} + \mathbf{v} \in \mathbb{R}^m$, where $m \le n$. Ideally, the optimal reconstruction method is the l_0 norm minimization method:

$$\min_{\mathbf{x}\in\mathbb{R}^n} \|\mathbf{x}\|_0 \quad \text{s.t. } \mathbf{Ax} = \mathbf{y}. \tag{1}$$

Since (1) is a hard combinatorial problem, l_1-minimization (BP) is usually adopted as a computationally tractable alternative to (1).

- Recently however, there is a trend to consider nonlinear functions in place of the l_1 cost functions. In other words, a general cost function $J(\mathbf{x}) := \sum_{k=1}^{n} F(|x(k)|)$ is employed. Examples of such an F include:
  - l_p cost function ($0 < p < 1$) [4], in the form of $\|\mathbf{x}\|_p^p$.
  - Approximate l_0 cost functions, c.f. [5, 2].

  Various practical algorithms can be adapted to these nonlinear problems, e.g. IRLS, iterative thresholding, and zero point attracting projection. In general the nonlinear algorithms empirically outperforms BP, because nonlinear cost functions can better promote sparsity than the l_1 cost function.



- Now two questions naturally arise: the *exact recovery condition* (ERC), which requires that all sparse signals can be exactly recovered in the noiseless case, and the *robust recovery condition* (RRC), which requires that if the measurement is noisy, then the reconstruction error must be bounded by the norm of the noise vector multiplied by a constant factor.
  Previous work [4, 1] showed that ERC is equivalent to RRC for the special case of l_p-minimization. In contrast, the relation between ERC and RRC for general F-minimization cannot be established without new ideas.

Problem Setup and Key Definitions

We consider the linear observation setting described in the Introduction.

- In the noiseless case, the sparse signal is retrieved by solving:

$$\min_{\mathbf{x}\in\mathbb{R}^n} J(\mathbf{x}) \quad \text{s.t. } \mathbf{Ax} = \mathbf{y}. \tag{2}$$

We say $(\mathbf{A}, J)$ satisfies the *exact recovery condition* (ERC) if for any k-sparse $\bar{\mathbf{x}}$ and measurement $\mathbf{y} = \mathbf{A}\bar{\mathbf{x}}$, $\bar{\mathbf{x}}$ is also the unique solution to (2).

- In the noisy measurement ($\mathbf{v} \neq \mathbf{0}$) case, the sparse signal is retrieved from:

$$\min_{\mathbf{x}\in\mathbb{R}^n} J(\mathbf{x}) \quad \text{s.t. } \|\mathbf{Ax} - \mathbf{y}\| < \epsilon, \tag{3}$$

where $\epsilon \in \mathbb{R}^+$ is a constant chosen to tolerate the noise. We say that the *robust recovery condition* (RRC) is satisfied if for any k-sparse signal $\bar{\mathbf{x}}$, any noise $\mathbf{v}$ satisfying $\|\mathbf{v}\| \le \epsilon$, and any feasible solution $\hat{\mathbf{x}}$ satisfying $J(\hat{\mathbf{x}}) \le J(\bar{\mathbf{x}})$, we have

$$\|\bar{\mathbf{x}} - \hat{\mathbf{x}}\| < C\epsilon, \quad \text{for some } C \text{ fixed.} \tag{4}$$

- we shall next consider a class of functions wide enough such that most practical sparsity inducing cost functions are subsumed: a function

$$F : [0, +\infty) \to [0, +\infty) \tag{5}$$

is called a *sparseness measure* if the following two conditions are satisfied:
  - $F(|\cdot|)$ is sub-additive on $\mathbb{R}$;
  - $F(x) = 0$ if and only if $x = 0$.

We denote by $\mathcal{M}$ the set of all sparseness measures.

Equivalence Lost: A Topological Characterization of RRC

We show that the equivalence of ERC and RRC no longer holds when passing from l_p-minimization to F-minimization.

- For Lebesgue-almost all measurement matrix $\mathbf{A}$, the null space of $\mathbf{A}$ corresponds to an element in $G_l(\mathbb{R}^n)$; and we can show that this element is sufficient to determine whether ERC and RRC are satisfied. Hence we shall examine $\Omega_J, \Omega_J^r \subset G_l(\mathbb{R}^n)$, which denote the collection of the null spaces satisfying ERC and RRC, respectively.
- For example, if two cost functions induced from the sparseness measures $F, G \in \mathcal{M}$ satisfy the following condition

$$\Omega_{J_G} \subseteq \Omega_{J_F}, \tag{6}$$

then ERC for G-minimization implies ERC for F-minimization, i.e., F is better a sparseness than G in the sense of ERC. In the light of this we can describe and compare the performances of different sparseness measures in terms of ERC by a simple set inclusion relation like (6).

- **Theorem 1**: *Consider the minimization problem in (3). The RRC holds if and only if there exists a $d > 0$, such that for each $\mathbf{z} \in \mathcal{N}(\mathbf{A}) \setminus \{0\}$, $\mathbf{n} \in \mathbb{R}^n$, $T \subseteq \{1, \dots, n\}$ satisfying $\|\mathbf{n}\| < d\|\mathbf{z}\|$, and $|T| \le k$, we have the following:*

$$J(\mathbf{z}_T + \mathbf{n}_T) < J(\mathbf{z}_{T^c} + \mathbf{n}_{T^c}). \tag{7}$$

- There is a nice interpretation of Theorem 1 in terms of point set topology:
  Theorem 2: *With the standard topology on $G_l(\mathbb{R}^n)$, the following relation holds.*

$$\Omega_J^r = \text{int}(\Omega_J). \tag{8}$$

  Proof idea: Theorem 1 and canonical coordinates of $G_l(\mathbb{R}^n)$.

- Define the null space constant

$$\theta_J := \sup_{\mathbf{z}\in\mathcal{N}(\mathbf{A})\setminus\{0\}} \max_{|T|\le k} \frac{J(\mathbf{z}_T)}{J(\mathbf{z}_{T^c})}. \tag{9}$$

  Corollary 1: *If F is continuous, then $\theta_J : G_l(\mathbb{R}^n) \to [0, +\infty)$ is a lower semi-continuous function. Further, $\theta_{l_p} : G_l(\mathbb{R}^n) \to [0, +\infty)$ is a continuous function.*

- For the special case of l_p-minimization, our interior set characterization leads to a topological proof to a previous result:
  Corollary 2: *If $0 < p \le 1$, then Ω_{l_p} is open, hence $\Omega_{l_p}^r = \Omega_{l_p}$.*
  Proof idea: Show that Ω_{l_p} is the pre-image of the open set $(-\infty, 1)$ under the continuous map $\theta_{l_p} : G_l(\mathbb{R}^n) \to \mathbb{R}$.

Equivalence Regained: the Probabilistic Equivalence

In practice the measurement matrix $\mathbf{A}$ is usually generated by using i.i.d. entries drawn from some continuous distribution. In this case, with some mild (but not redundant) monotonicity assumption of F, we show that RRC and ERC still occur with the same probability. (Note that this does not immediately follows from Theorem 2).

- **Theorem 3:** *Suppose $F \in \mathcal{M}$ is a non-decreasing function, then $\mu(\Omega_J \setminus \Omega_J^r) = 0$, where μ denotes the Haar measure on the Grassmannian $G_l(\mathbb{R}^n)$.*
  Proof idea: Use Lebesgue density theorem in measure theory and the canonical coordinates of $G_l(\mathbb{R}^n)$.
- **Remark:** If the assumption that $F \in \mathcal{M}$ is a non-decreasing is dropped, then one can cook up counterexample in which $\mu(\Omega_J \setminus \Omega_J^r) > 0$.
- A trivial observation from Theorem 2 is that the probability of ERC and RRC are the same if the observation matrix $\mathbf{A}$ has i.i.d. Gaussian entries, since in this case the probability agrees with the measure μ. More generally, suppose P is the probability measure corresponding to the distribution of the null space of $\mathbf{A}$, and P is absolutely continuous with respect to μ, then $P(\Omega_J \setminus \Omega_J^r) = 0$. One can show that this absolute continuity holds if the entries of $\mathbf{A}$ are i.i.d. generated from a certain continuous distribution, therefore we have:
  Corollary 3: *Suppose $F \in \mathcal{M}$ is a non-decreasing function, and the distribution of the matrix $\mathbf{A}$ is absolutely continuous with respect to the Lebesgue measure on $\mathbb{M}(m, n)$. Then the probability of ERC and RRC are the same. This holds true in particular when $\mathbf{A}$ has i.i.d. entries drawn from a continuous distribution.*
  Proof idea: absolute continuity of a measure on $\mathbf{R}^{m\times n} \Rightarrow$ absolute continuity of the induced measure on $G_l(\mathbb{R}^n)$.

A Toy Example

Consider the function

$$F(x) := x + 1 - e^{-x} \tag{10}$$

defined on $[0, +\infty)$ which is a spareness measure. Suppose that $x, y > 0$, $z = x + y$, $k = 1$, and that the null space of the measurement matrix is a one dimensional subspace of $\mathbb{R}^3$:

$$\mathcal{N} := [x, y, z]^T, \tag{11}$$

where the homogenous coordinates $[x, y, z]^T$ denotes the subspace spanned by $(x, y, z)^T$. Conclusion: ERC is satisfied, but not RRC.

- In fact, using the null space property [4, 3](a famous property in the study of compressed sensing, which we do not explain in detail here) and Theorem 1, one can show that

$$\Omega_J = \left\{ [x_1, x_2, x_3] : 2 \max_{i=1,2,3} |x_i| \le \sum_{i=1,2,3} |x_i| \right\}, \tag{12}$$

$$\Omega_J^r = \left\{ [x_1, x_2, x_3] : 2 \max_{i=1,2,3} |x_i| < \sum_{i=1,2,3} |x_i| \right\}. \tag{13}$$

- In the above it is clear that Ω_J is an closed set while Ω_J^r is an open set, since they are preimages of a closed (resp. an open) set under a continuous mapping. Without bothering with rigorous proofs, it's also intuitively clear that $\nu(\Omega_J \setminus \Omega_J^r) = 0$ (think of a low dimensional analogy of sets $(-\infty, 1]$ and $(-\infty, 1)$).

References

- Aldroubi, A., Chen, X., Powell, A.: Stability and robustness of l_q minimization using null space property.
  Proceedings of SampTA 2011 (2011)
- Gasso, G., Rakotomamonjy, A., Canu, S.: Recovering sparse signals with non-convex penalties and dc programming.
  IEEE Trans. Signal Processing **57**(12)
- Gribonval, R., Nielsen, M.: Sparse representations in unions of bases.
  IEEE Trans. Inf. Theory **49**(12), 3320–3325 (2003)
- Gribonval, R., Nielsen, M.: Highly sparse representations from dictionaries are unique and independent of the sparseness measure.
  Applied and Computational Harmonic Analysis **22**(3), 335–355 (2007).
  DOI 10.1016/j.acha.2006.09.003
- Jin, J., Gu, Y., Mei, S.: A stochastic gradient approach on compressive sensing signal reconstruction based on adaptive filtering framework.
  IEEE Journal of Selected Topics in Signal Processing **4**(2), 409–420 (2010)

Acknowledgments

This work was partially supported by National Natural Science Foundation of China (NSFC 60872087 and NSFC U0835003)
A draft of this work will is available on arxiv: http://arxiv.org/abs/1301.0093. Part of this work will also be presented on ISIT 2013.