ISIT'2013 1569719445

Relation between Exact and Robust Recovery for F-minimization: A Topological Viewpoint

Jingbo Liu
Dept. of Electrical Engineering
Princeton University
Princeton, NJ 08544, USA
Email: jingbo@princeton.edu

Jian Jin
Dept. of Electronic Engineering
Tsinghua University
Beijing, 100084, China
Email: jinjian620@gmail.com

Yuantao Gu
Dept. of Electronic Engineering
Tsinghua University
Beijing, 100084, China
Email: gyt@tsinghua.edu.cn

***Abstract*—Recent work in compressed sensing has shown the possibility reducing the number of measurements via non-convex optimization methods. Most of these methods can be studied in the general framework called "F-minimization", for which the relation between the noiseless exact recovery condition (ERC) and noisy robust recovery condition (RRC) was not fully understood. In this paper, we associate each set of nulls spaces of the measurement matrices satisfying ERC/RRC as a subset of a Grassmannian, and show that the RRC set is exactly the interior of the ERC set. Then, a previous result of the equivalence of ERC and RRC for l_p-minimization follows easily as a special case. We also show under some mild but necessary additional assumption that the ERC and RRC sets differ by a set of measure zero.**

I. Introduction

Consider the problem of recovering a sparse signal $\bar{\mathbf{x}} \in \mathbb{R}^n$ from a set of linear measurements $\mathbf{y} = \mathbf{A}\bar{\mathbf{x}} \in \mathbb{R}^m$, where $m \leq n$. Ideally, the optimal reconstruction method is the l_0 norm minimization method:

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

Since (1) is a hard combinatorial problem, many algorithms have been proposed to reduce the computational complexity, most notably the l_1-minimization, or Basis Pursuit (BP) [1], [2], which is a convex relaxation of (1). Recently however, there is a trend to consider nonlinear cost functions, such as:

- l_p cost function ($0 < p < 1$) [3], in the form of $\|\mathbf{x}\|_p^p$.
- Approximate l_0 cost functions, examples of which can be found in [4], [5], [6].

Various practical algorithms can be adapted to these nonlinear problems, including the iteratively reweighted least squares minimization (IRLS) [7], iterative thresholding algorithm (IT) [8], and the zero point attracting projection algorithm (ZAP) [5]. In general the nonlinear algorithms have empirically outperformed BP in the various respects, because nonlinear cost functions can better promote sparsity than the l_1 cost function. Thus, a detailed study of the reconstruction properties of these sparse recovery methods remains important. Noting that most of the aforementioned cost functions enjoy some nice properties (subadditivity, vanishing at zero, etc), one can study these minimization techniques in a general framework, called F-minimization [9]. Now two questions naturally arise, as in any other compressed sensing problem: 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.

By definitions RRC trivially implies ERC, while the converse is not obvious. The well known restricted isometry property (RIP) [1] is capable of finding bounds on the number of measurements needed, but not very useful in determining the exact relation between ERC and RRC. On the other hand, previous work [3], [9], [10] showed that the null space property (NSP) is equivalent to ERC for F-minimizations, and is also equivalent to NSP for the special case of l_p-minimization.

In contrast, the relation between ERC and RRC for general F-minimization is more difficult to establish merely based on the previous methods. In this paper we address this problem by studying the sets of the null spaces of the measurement matrices which satisfy ERC and RRC, respectively. For measurement matrices in the general position, these two sets are subsets of the Grassmann manifold $G_{n-m}(\mathbb{R}^n)$, denoted by Ω and Ω^r, respectively. We show that Ω^r is exactly the interior of Ω. In the special case of l_p-minimization, we show that Ω is open, hence ERC and RRC are equivalent, which agrees with the previous result in the literature. This amounts to a new "topological" proof to a "non-topological" result, since this special case can be simply stated pointwise without referring to the topology of the Grassmannian. Moreover, under an additional monotonicity assumption on the cost functions, we show that Ω and Ω^r differ by a set of measure zero. The additional monotonicity assumption does not cause problems for most practical applications, and a counterexample also shows the necessity of an additional assumption. Building on this, we show that ERC and RRC hold with the same probability if the measurement matrix is randomly generated according to a continuous distribution. In a word, ERC and RRC are no longer strictly equivalent when passing to the general case of F-minimization, but are still comparable or "equivalent" in some weak sense.

Main part of this work was done when the author was with the Dept. of Electronic Engineering, Tsinghua University, China. The corresponding author of this paper is Yuantao Gu. This work was partially supported by National Natural Science Foundation of China (NSFC 60872087 and NSFC U0835003).

II. Problem Setup and Key Definitions

A. Basic Model

Let $\bar{\mathbf{x}} \in \mathbb{R}^n$, $\mathbf{A} \in \mathbb{M}(m,n)$, $\mathbf{v} \in \mathbb{R}^m$ be the sparse signal, the measurement matrix, and the additive noise, respectively. Let $T := \mathrm{supp}(\bar{\mathbf{x}})$ be the support of $\bar{\mathbf{x}}$. A vector $\bar{\mathbf{x}}$ is called k-sparse if $|T| \le k$. The linear measurement $\mathbf{y}$ is given by

$$\mathbf{y} = \mathbf{A}\bar{\mathbf{x}} + \mathbf{v}. \tag{2}$$

We consider the problem of recovering $\bar{\mathbf{x}}$ through an optimization. Supposing $F : [0,+\infty) \to [0,+\infty)$ is a given function, we define the cost function

$$J(\mathbf{x}) := \sum_{k=1}^{n} F(|x(k)|). \tag{3}$$

With a slight abuse of the notation, we shall also use the notations:

$$J(\mathbf{x}_T) := \sum_{k\in T} F(|x(k)|), \quad J(\mathbf{x}_{T^c}) := \sum_{k\in T^c} F(|x(k)|),$$

where $\mathbf{x}_T \in \mathbb{R}^{|T|}$, $\mathbf{x}_{T^c} \in \mathbb{R}^{n-|T|}$ denote the restriction of $\mathbf{x}$ on the set T, T^c, respectively. Clearly (3) is a very general model: For example, if one chooses $F(x) = 1_{x>0}$ then $J(\mathbf{x}) = \|\mathbf{x}\|_0$; if $F(x) = x^p$ then $J(\mathbf{x}) = \|\mathbf{x}\|_p^p$.

The conditions ERC and RRC are commonly formulated as follows, see for example [9].

Definition 1 (Exact recovery condition): In the noiseless case, the sparse signal is retrieved via the following optimization:

$$\min_{\mathbf{x}\in\mathbb{R}^n} J(\mathbf{x}) \quad \text{s.t. } \mathbf{A}\mathbf{x} = \mathbf{y}. \tag{4}$$

We say a pair of $\mathbf{A}$, J satisfy the *exact recovery condition* (ERC) if for any measurement $\mathbf{y} = \mathbf{A}\bar{\mathbf{x}}$, where $\bar{\mathbf{x}}$ is k-sparse, the vector $\bar{\mathbf{x}}$ is also the unique solution to (4).

Definition 2 (Robust recovery condition): In the noisy measurement ($\mathbf{v} \neq \mathbf{0}$) case, the sparse signal is retrieved via the following optimization:

$$\min_{\mathbf{x}\in\mathbb{R}^n} J(\mathbf{x}) \quad \text{s.t. } \|\mathbf{A}\mathbf{x} - \mathbf{y}\| < \epsilon, \tag{5}$$

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

$$\|\bar{\mathbf{x}} - \hat{\mathbf{x}}\| < C\epsilon, \tag{6}$$

where C is a constant.

B. Null Space Property

The null space property [3], [11] is useful for the analysis of a special class of cost functions, which we introduce as follows:

Definition 3 (sparseness measure): Function

$$F : [0,+\infty) \to [0,+\infty) \tag{7}$$

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.

One can verify that the key optimization problems in many of the sparse recovery algorithms can be subsumed in our framework, including l_p-minimization; see also the cost functions used in [5], [6]. The definition is also quite natural, since it can be checked that F is a sparseness measure if and only if its corresponding cost function J induces a metric on $\mathbb{R}^n$ via $d(\mathbf{x},\mathbf{y}) := J(\mathbf{x}-\mathbf{y})$.

When $F \in \mathcal{M}$, the *null space property* (NSP) turns out to be equivalent with ERC:

Lemma 1 (Null space property [3](Lemma 6)): If $F \in \mathcal{M}$, then a necessary and sufficient condition for ERC is[1]

$$J(\mathbf{z}_T) < J(\mathbf{z}_{T^c}), \quad \forall \mathbf{z} \in \mathcal{N}(\mathbf{A}) \setminus \{\mathbf{0}\},\ |T| \le k. \tag{8}$$

where $\mathcal{N}(\mathbf{A})$ denotes the null space of $\mathbf{A}$.

A related concept is the null space constant, defined as follows:

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

The null space constant associated with l_p cost function, can be defined analogously, which we denote by θ_{l_p}. We have the following result, which is a direct consequence of the definition of the null space constant and Lemma 1.

Lemma 2:
1) $\theta_J \le 1$ is a necessary condition for ERC;
2) $\theta_J < 1$ is a sufficient condition for ERC.

In the case of l_p-minimization, one can obtain the following characterization (c.f.[10]), which is more exact than the case of F-minimization as described in Lemma 2:

Lemma 3: For l_p cost functions, $\theta_{l_p} < 1$ is a both necessary and sufficient condition for ERC.

C. Preliminaries of the Grassmann Manifold

From Lemma 1, the property of exact recovery of a particular measurement matrix is completely determined by its null space. When $\mathbf{A}$ is of full rank, $\mathcal{N}(\mathbf{A})$ is an $l := n - m$ dimensional linear subspace of $\mathbb{R}^n$, which can be conceived as a point on the Grassmann manifold $G_l(\mathbb{R}^n)$ [12]. Concepts such as open sets and interior make sense after specifying a topological structure on $G_l(\mathbb{R}^n)$.

The smooth manifold $G_l(\mathbb{R}^n)$ can be described by a collection of C^∞ compatible charts; and then measure sets can be defined based on this differential structure: in general, a subset A of a differentiable manifold is said to have *measure zero* if $\phi(A \cap U)$ has Lebesgue measure zero for every chart (U,ϕ) [12, Definition 1.16]. A related concept is the Haar measure, which is a rotational invariant measure on $G_l(\mathbb{R}^n)$. The requirement that a set A has zero Haar measure agrees with the first definition of measure zero set. The Haar measure is of practical importance, since it coincides with the distribution of the null space of $\mathbf{A}$ when $\mathbf{A}$ is a Gaussian random matrix.

[1]Here the notation '\' denotes the set minus.

III. The Relationship between ERC and RRC

A. A Topological Characterization of RRC

We have mentioned earlier that NSP is a necessary and sufficient condition for ERC. If $\mathbf{A} \in \mathbb{M}(m,n)$ is in a general position (i.e., the rows of $\mathbf{A} \in \mathbb{M}(m,n)$ are linearly independent), then $\mathbf{A}$ is of full rank, and $\mathcal{N}(\mathbf{A})$ is a l-dimensional subspace in $\mathbb{R}^n$ (recall that $l = n-m$). Therefore almost every measurement matrix (except for the set of $\mathbf{A}$'s not in a general position, which is of Lebesgue measure zero) corresponds to an element in $G_l(\mathbb{R}^n)$; and this element is sufficient to determine whether NSC, and therefore ERC, is satisfied. By Lemma 1, the set of null spaces such that ERC is satisfied is as follows:

$$\begin{aligned}\Omega_J := &\{\nu \in G_l(\mathbb{R}^n) : J(\mathbf{z}_T) < J(\mathbf{z}_{T^c}),\\ &\forall \mathbf{z} \in \nu \setminus \{\mathbf{0}\}, |T| \leq k\}.\end{aligned} \tag{10}$$

If two cost functions induced from the sparseness measures $F, G \in \mathcal{M}$ satisfy the following condition

$$\Omega_{J_G} \subseteq \Omega_{J_F}, \tag{11}$$

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 (11).

In Lemma 1, the necessary and sufficient condition for exact recovery is fully characterized by the structure of the null space. Inspired by this fact we now provide a necessary and sufficient condition for robust recovery:

Theorem 1: Consider the minimization problem in (5). The RRC holds if and only if there exists a $d > 0$, such that for each $\mathbf{z} \in \mathcal{N}(\mathbf{A}) \setminus \{\mathbf{0}\}$, $\mathbf{n} \in \mathbb{R}^n$, $T \subseteq \{1, ..., n\}$ satisfying $\|\mathbf{n}\| < d\|\mathbf{z}\|$, and $|T| \leq k$, we have the following:

$$J(\mathbf{z}_T + \mathbf{n}_T) < J(\mathbf{z}_{T^c} + \mathbf{n}_{T^c}). \tag{12}$$

Proof: Sufficiency: Suppose $\hat{\mathbf{x}}$ is the recovered signal. From the constraint of the optimization we have

$$\|\mathbf{A}(\hat{\mathbf{x}} - \bar{\mathbf{x}})\| \leq \|\mathbf{A}\hat{\mathbf{x}} - \mathbf{y}\| + \|\mathbf{A}\bar{\mathbf{x}} - \mathbf{y}\| \leq 2\epsilon. \tag{13}$$

Define $\mathbf{u} := \bar{\mathbf{x}} - \hat{\mathbf{x}}$; from the optimality of $\hat{\mathbf{x}}$ we have

$$J(\mathbf{u}_T) \geq J(\mathbf{u}_{T^c}). \tag{14}$$

Decompose $\mathbf{u} = \mathbf{z}+\mathbf{n}$, such that $\mathbf{z}$ belongs to the null space of $\mathbf{A}$. The above inequality is in contradiction with (12), hence from the assumption we must have:

$$\|\mathbf{n}\| \geq d\|\mathbf{z}\|. \tag{15}$$

Therefore

$$\begin{aligned}2\epsilon &\geq \|\mathbf{A}(\hat{\mathbf{x}} - \bar{\mathbf{x}})\|\\ &= \|\mathbf{A}\mathbf{n}\|\\ &\geq \sigma_{min}\|\mathbf{n}\|\\ &\geq \sigma_{min}\frac{d}{1+d}\|\mathbf{u}\|\\ &= \sigma_{min}\frac{d}{1+d}\|\hat{\mathbf{x}} - \bar{\mathbf{x}}\|,\end{aligned}$$

where σ_{min} is the smallest singular value of $\mathbf{A}$. Thus RRC holds.

Necessity: We will show by contradiction. Assuming that

$$\begin{aligned}&\forall d > 0, \exists \|\mathbf{n}\| < d\|\mathbf{z}\|, \mathbf{z} \in \mathcal{N}(\mathbf{A}),\\ &\text{such that } J(\mathbf{z}_T + \mathbf{n}_T) \geq J(\mathbf{z}_{T^c} + \mathbf{n}_{T^c}),\end{aligned} \tag{16}$$

we will show that the recovery is not robust. To do this, we will construct $\mathbf{x}_1$, $\mathbf{x}_2$ with $J(\mathbf{x}_2) \geq J(\mathbf{x}_1)$, and $\mathbf{v}$ with $\|\mathbf{v}\| = \epsilon$, $\|\mathbf{A}\mathbf{x}_1 - (\mathbf{A}\mathbf{x}_2 + \mathbf{v})\| = \epsilon$; but

$$\|\mathbf{x}_1 - \mathbf{x}_2\| \geq \frac{2(1-d)\epsilon}{d\|\mathbf{A}\|}. \tag{17}$$

Since d is arbitrary and hence unbounded from below, the constant $\frac{2(1-d)\epsilon}{d\|\mathbf{A}\|}$ will be unbounded from above.

For any d, choose $\mathbf{n}, \mathbf{z}$ satisfying (16). Define[2] $\mathbf{u} := \mathbf{z} + \mathbf{n}$, $\mathbf{x}_1 = (\mathbf{u}_T)^T$, $\mathbf{x}_2 = -(\mathbf{u}_{T^c})^{T^c}$, $\mathbf{v} = \mathbf{A}(\mathbf{x}_1 - \mathbf{x}_2)/2$, $\|\mathbf{v}\| = \epsilon$. Then $\|\mathbf{A}\mathbf{x}_1 - (\mathbf{A}\mathbf{x}_2 + \mathbf{v})\| = \epsilon$. Hence

$$\begin{aligned}2\epsilon &= \|\mathbf{A}(\mathbf{x}_1 - \mathbf{x}_2)\|\\ &= \|\mathbf{A}\mathbf{n}\|\\ &= \|\mathbf{A}\|\|\mathbf{n}\|\\ &\leq \|\mathbf{A}\|\frac{d}{1-d}\|\mathbf{u}\|.\end{aligned}$$

Thus the relation (17) holds, as desired. ■

We remark that RRC trivially implies ERC, as can be seen in their definitions (Letting $\mathbf{v} = \mathbf{0}$ in the definition of RRC would result in the definition of ERC), as well as in Theorem 1 (Letting $\mathbf{n} = \mathbf{0}$).

From Theorem 1 it is clear that the property of robust recovery of a particular matrix is also completely determined by its null space. Moreover, it implies that the subset of $G_l(\mathbb{R}^n)$ that guarantees RRC is the following:

$$\begin{aligned}\Omega_J^r := &\{\nu \in G_l(\mathbb{R}^n) : \exists d > 0,\\ &\text{s.t. } J(\mathbf{z}_T + \mathbf{n}_T) < J(\mathbf{z}_{T^c} + \mathbf{n}_{T^c}),\\ &\forall \mathbf{z} \in \nu \setminus \{\mathbf{0}\}, \|\mathbf{n}\| < d\|\mathbf{z}\|, |T| \leq k\}.\end{aligned} \tag{18}$$

It is not immediately clear from Lemma 1 and Theorem 1 the connection between ERC and RRC. However there is a nice relation between these two conditions once taking a perspective from the point set topology[3]:

Theorem 2: With the standard topology on $G_l(\mathbb{R}^n)$, the following relation holds.

$$\Omega_J^r = \text{int}(\Omega_J). \tag{19}$$

Two questions then arise: are the conditions ERC and RRC equivalent for generic cost functions? If not, how much do they differ from each other? We shall first address the former question in the remainder of this part, while the second question will be discussed in Part B. In the special case of l_p-minimization, these two conditions are indeed equivalent [9],

[2]For $\mathbf{x} \in \mathbb{R}^{|T|}$, we denote by $\mathbf{x}^T \in \mathbb{R}^n$ the n-vector supported on T satisfying $(\mathbf{x}^T)_T = \mathbf{x}$.

[3]The proofs of some of the results in this section, which can be found in [13], are omitted here due to the limitation of space.

as discussed in the introductory section. In view of Theorem 1, we can show this result by simply proving that the Ω is an open set in the case of l_p-minimization. We first note the following basic fact about generic continuous functions. (It is stated in a slightly stronger and more complete manner than needed for obtaining our final result).

Lemma 4: Suppose X, M are metric spaces. If $f : X \times M \to \mathbb{R}$ is continuous, then $g : X \to \mathbb{R},\ x \mapsto \max_{y \in M} f(x, y)$ is lower semi-continuous on X. Further, if M is compact, then g is also continuous.

It then follows the following result about the null space constant θ, now conceived as a map from $G_l(\mathbb{R}^n)$ to the real numbers:

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.

The openness of Ω_{l_p} then follows easily, from the very definition of continuous functions: that the pre-images of open sets are open.

Corollary 2: If $0 < p \le 1$, then Ω_{l_p} is open, hence $\Omega^r_{l_p} = \Omega_{l_p}$.

Remark 1: The equivalence result of $\Omega^r_{l_p} = \Omega_{l_p}$ in the above is essentially 'non-topological', since it does not involve the concept of open sets on the Grassmann manifold. Different proof methods in the literature are discussed in Section IV.

Proof: By Corollary 1, function θ_{l_p} is continuous with respect to ν. Since Ω_{l_p} is the pre-image of $(-\infty, 1)$ under the continuous mapping of θ_{l_p} (Lemma 3), we conclude that Ω_{l_p} is open, hence $\Omega^r_{l_p} = \mathrm{int}(\Omega_{l_p}) = \Omega_{l_p}$. ■

Next we shall show an example in which RRC is strictly stronger than ERC, i.e., $\Omega^r_J \subsetneqq \Omega_J$.

Proposition 1: The function

$$F(x) := x + 1 - \mathrm{e}^{-x} \tag{20}$$

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

$$\mathcal{N} := [x, y, z]^T, \tag{21}$$

where the homogenous coordinates $[x, y, z]^T$ denotes the sub-space spanned by $(x, y, z)^T$. Conclusion: in this setting ERC is satisfied, but not RRC.

Proof: Let $\mathbf{z} = (x, y, z)$. Since $|z| > |x|, |y|$ and $F(x) + F(y) > F(z)$, for any T such that $|T| = 1$ we have:

$$J(\mathbf{z}_T) < J(\mathbf{z}_{T^c}). \tag{22}$$

Hence NSP is satisfied, and ERC must hold. On the other hand, for any $0 < d < 1$ there exists $t > 0$ such that

$$F((1-d)xt) + F(yt) < F(zt). \tag{23}$$

Now in Theorem 1, take $\mathbf{z} = (xt, yt, zt)^T$, $T = \{3\}$, and $\mathbf{n} = (-dxt, 0, 0)$. On the one hand we have $\|\mathbf{n}\|/\|\mathbf{z}\| \le d$; on the other hand (12) doesn't hold because of (23). Therefore RRC is not fulfilled as a result of Theorem 1. ■

B. Equivalence Regained: the Probabilistic Equivalence

While strict equivalence of ERC and RRC is lost when passing from l_p cost functions to generic sparseness measures, as demonstrated in Proposition 1, we will show in this part that the difference is only a set of measure zero on the Grassmann manifold, at least for non-decreasing sparseness measures. First we take a closer look at Proposition 1. Using the subadditivity property and the Taylor expansion of F at the origin, one can explicitly write out:

$$\Omega_J = \left\{ [x_1, x_2, x_3] : 2 \max_{i=1,2,3} |x_i| \le \sum_{i=1,2,3} |x_i| \right\}, \tag{24}$$

and

$$\Omega^r_J = \left\{ [x_1, x_2, x_3] : 2 \max_{i=1,2,3} |x_i| < \sum_{i=1,2,3} |x_i| \right\}. \tag{25}$$

We recall that μ denotes the Haar measure on $G_l(\mathbb{R}^n)$. From (24) and (25) it is intuitively clear in this simple case that $\mu(\Omega_J) = \mu(\Omega^r_J)$, i.e. the set of null spaces satisfying ERC and the set of null spaces satisfying RRC differ at most by a set of measure zero. Recall that the Haar measure agrees with the probability measure in the case of i.i.d. Gaussian random entries, as described in Section II, Part C. This means that if $\mathbf{A}$ is a Gaussian random matrix, then the probability of ERC and RRC are the same, even though the former is implied by the latter.

The general case tends to be much more complicated. Indeed, taking an arbitrary topological measurable space, it is very well possible that the measure of a set is strictly greater than the measure of its interior. (Consider for example the set of all irrational numbers, whose Lebesgue measure is ∞, but whose interior is empty.) In fact merely $F \in \mathcal{M}$ does not guarantee $\mu(\Omega^r) = \mu(\mathrm{int}(\Omega^r))$, as will be shown in the remark at the end of this section. However we have shown the following result, the proof of which is based on Lebesgue density theorem from measure theory.

Theorem 3: Suppose $F \in \mathcal{M}$ is a non-decreasing function, then $\Omega_J \setminus \mathrm{int}(\Omega_J)$ is a measure zero set, that is, $\mu(\Omega_J \setminus \Omega^r_J) = 0$.

Remark 2: Almost all commonly used F-minimizations (e.g. l_p-minimization, ZAP) satisfy the requirement of F being non-decreasing, hence the non-increasing assumption is a very mild one. On the other hand, we remark that the non-decreasing requirement is also essential for the validity of Theorem 3. To see this, consider the following example: Define

$$F(x) = \begin{cases} 0 & x = 0; \\ 0.1 & x > 0 \text{ and } x \text{ is rational;} \\ 1 & x > 0 \text{ and } x \text{ is irrational,} \end{cases} \tag{26}$$

and set $m = 2, n = 3, k = 1$. It can then be verified that F satisfies the definition of sparseness measure in Definition 3. Moreover, for arbitrary $x_1, x_2 \in \mathbb{R}$, denote by $x_1 \simeq x_2$ the equivalence relation that either $x_1/x_2 \in \mathbb{Q} - \{0\}$ or $x_1 = x_2 =$

0 holds[4]. Then for any $\nu \in G_1(\mathbb{R}^3)$, the three homogenous coordinates of $\mathbf{z} \in \nu$ can be grouped into equivalent classes according to $\simeq$, and whether $\nu \in \Omega_J$ is completely determined by how these coordinates are grouped. Now we say ν is of type (say) $(1,1,2)$ if the first two homogenous coordinates of ν are of a same equivalence class and the third homogenous coordinate is of another equivalence class. From the null space property we can check that the type $(1,2,3)$ is in Ω_J, while $(1,1,2)$ is not. Since the null spaces of the type $(1,2,3)$ is of measure 1, we have that $\mu(\Omega_J) = 1$. On the other hand, since the set of one dimensional subspaces corresponding to the type $(1,1,2)$ is dense in $G_1(\mathbb{R}^3)$ and also does not intersect Ω_J, the interior of Ω_J must be vacuous, hence $\mu(\mathrm{int}(\Omega_J)) = 0 \neq \mu(\Omega_J)$.

A trivial observation from Theorem 3 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.

Remark 3: Apart from the one described in Corollary 3, another popular method for the generation of $\mathbf{A}$ is by randomly selecting m rows in the $n \times n$ Fourier transform matrix [14]. However in this scheme the probability of ERC and RRC may not agree, since the probability distribution of the null space is not continuous on $G_l(\mathbb{R}^n)$.

IV. Comparison with Other Works

For the special case of l_p-minimization, the equivalence of ERC and RRC was first proved in [9], based on a condition called NSP' and a lemma from matrix analysis [9, Lemma 2.1]. However, this approach is hard to be extended to the general F-minimization problem, because it consists of a homogeneous inequality, which appears to work well only for homogeneous cost functions such as the l_p norm. In contrast, our approach applies for general F-minimizations; and for the special case of l_p-minimization, it is particularly interesting to note that our approach amounts to a "topological" proof of a "non-topological" result.

It's also worth mentioning that a slightly stronger definition of sparseness measure also appeared in the literature [15], [3], in which the function F must be non-increasing and $F(t)/t$ be non-increasing[5], in addition to the other requirements in Definition 3. The additional assumptions guarantee that the cost function J_F is better than l_1 norm in the sense of ERC. There is also another nice property relating to the composition of two functions in this class [3, Lemma 7]. Finally, l_1 norm is the only convex cost function whose corresponding F satisfies this definition of sparseness measure [15, Proposition 2.1].

[4] $\mathbb{Q}$ denotes the set of rational numbers.

[5] The assumption of $F(t)/t$ being non-increasing implies that F is sub-additive.

V. Conclusion

The new analytic approach in this paper gives an exact characterization of the relationship between robust recovery condition (RRC) and exact recovery condition (ERC). Building on this characterization, the previous result of the equivalence of exact recovery and robust recovery in the l_p-minimization follows easily as a special case of our result. Under some mild (but essential) additional assumption on the cost functions, we have also shown that the ERC set and RRC set differ by set of measure zero. The practical significance of the last result is that ERC and RRC will occur with the same probability when the measurement matrix is randomly generated according to a continuous distribution.

However our proof method has just shown the existence of the constant C in the definition of RRC without actually deriving it, as opposed to the method in [9]. Finding an explicit bound for such a constant for general F-minimization may serve as a direction for future work.

References

[1] E. J. Candès and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, 2005.

[2] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[3] R. Gribonval and M. Nielsen, "Highly sparse representations from dictionaries are unique and independent of the sparseness measure," *Applied and Computational Harmonic Analysis*, vol. 22, no. 3, pp. 335–355, 2007.

[4] M. B. Zadeh and C. Jutten, "On the stable recovery of the sparsest overcomplete representations in presence of noise," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 6–18, 2010.

[5] J. Jin, Y. Gu, and S. Mei, "A stochastic gradient approach on compressive sensing signal reconstruction based on adaptive filtering framework," *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 409–420, 2010.

[6] G. Gasso, A. Rakotomamonjy, and S. Canu, "Recovering sparse signals with non-convex penalties and dc programming," *IEEE Trans. Signal Processing*, vol. 57, no. 12.

[7] I. Daubechies, R. DeVore, M. Fornasier, and C. S. Gunturk, "Iteratively re-weighted least squares minimization for sparse recovery," *Princeton Univ NJ Program in Applied and Computational Mathematics*, vol. 63, no. 1, p. 35.

[8] I. Daubechies, M. Defrise, and C. DeMol, "An iterative thresholding algorithm for linear inverse problems," *Princeton Univ NJ Program in Applied and Computational Mathematics*, vol. 63, no. 1, p. 35.

[9] A. Aldroubi, X. Chen, and A. Powell, "Stability and robustness of l_q minimization using null space property," *Proceedings of SampTA 2011*, 2011.

[10] S. Foucart, "Notes on compressed sensing," 2009.

[11] R. Gribonval and M. Nielsen, "Sparse representations in unions of bases," *IEEE Trans. Inf. Theory*, vol. 49, no. 12, pp. 3320–3325, 2003.

[12] J. Milnor, *Differential Topology*. 1958 Lecture Notes.

[13] J. Liu, J. Jin, and Y. Gu, "Robustness of sparse recovery via F-minimization: A topological viewpoint," *draft, available at http://arxiv.org/abs/1301.0093*, 2012.

[14] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, 2006.

[15] J. A. Tropp, "Topics in sparse approximation, ph.d. dissertation," August 2004.