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

Jingbo Liu, Jian Jin, and Yuantao Gu*

Received Dec. 27, 2012.

Abstract

A recent trend in compressed sensing is to consider non-convex optimization techniques for sparse recovery. A general class of such optimizations, called F-minimization, has become of particular interest, since its exact reconstruction condition (ERC) in the noiseless setting can be precisely characterized by null space property (NSP). However, little work has been done concerning its robust reconstruction condition (RRC) in the noisy setting. In this paper we look at the null space of the measurement matrix as a point on the Grassmann manifold, and then study the relation of the ERC and RRC sets on the Grassmannian. It is shown that the RRC set is exactly the topological interior of the ERC set. From this characterization, a previous result of the equivalence of ERC and RRC for l_p-minimization follows easily as a special case. Moreover, when F is non-decreasing, it is shown that the ERC and RRC sets are equivalent up to a set of measure zero. As a consequence, the probabilities of ERC and RRC are the same if the measurement matrix is randomly generated according to a continuous distribution. Finally, we provide several rules for comparing the performances of different cost functions, as applications of the above results.

Keywords: Reconstruction algorithms, compressed sensing, minimization methods, robustness, null space

1 Introduction

Compressed Sensing is a method of recovering a sparse signal from a set of under-determined linear measurements. Ideally, the optimal reconstruction method is the l_0 norm minimization method:

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

*J. Liu is now with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544 USA (e-mail: jingbo@princeton.edu). Part of this work was done when he was with the Department of Electronic Engineering, Tsinghua University, Beijing, 100084 China. J. Jin and Y. Gu are both with the Department of Electronic Engineering, Tsinghua University (e-mail: jinjian620@gmail.com, gyt@tsinghua.edu.cn). The corresponding author of this paper is Yuantao Gu.

where $\mathbf{A}$ is an $m \times n$ measurement matrix, $\mathbf{y} \in \mathbb{R}^m$ is the linear measurements, and we assume that $m < n$. It can be proved that l_0 minimization method requires the least possible number of measurements; however, the l_0 minimization method is computational intractable since it is a hard combinatorial problem. Therefore, many algorithms have been proposed to reduce the computational complexity. Roughly speaking, these algorithms fall into two categories: 1) minimization techniques, where the sparse solution is retrieved by minimizing an appropriate cost function [1, 2], and 2) greedy pursuits, a representative of which is the orthogonal matching pursuit (OMP) [3].

In general, the greedy algorithms often incur less computational complexity, but the minimization techniques are more advantageous in terms of accuracy. The most basic minimization technique is the l_1-minimization, or Basis Pursuit (BP) [1, 4, 5]:

$$\min_{\mathbf{x}\in\mathbb{R}^n} \|\mathbf{x}\|_1 \quad \text{s.t. } \mathbf{y} = \mathbf{A}\mathbf{x}, \tag{2}$$

which is a simple convex optimization and can be recast as a linear program. Recently there is trend to consider minimizing non-convex cost functions. Examples include:

- l_p cost function. The l_p-minimization ($0 < p < 1$) [6–9] considers an optimization problem similar to (2) but the cost function is replaced with $\|\mathbf{x}\|_p^p$.
- Approximate l_0 cost function [2, 10–12].

Although the non-convex nature of these cost functions makes it difficult to exactly solve the corresponding optimization problems, various practical algorithms can be adapted to these non-convex problems, including the iteratively re-weighted least squares minimization (IRLS) [13, 14], iterative thresholding algorithm (IT) [15], which are based on fixed point iteration; and the zero point attracting projection algorithm (ZAP) [2, 16, 17], which is based on Newton's method for solving nonlinear optimization. In general the non-convex 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 remain important.

Most of these non-convex optimizations can be subsumed in a general category called "F-minimization" [18], in which the cost function satisfies some desirable properties, such as subadditivity. The precise definition of the class of cost functions of our interest will be given in the next section.

Two concepts arise naturally in the compressed sensing problem: The *exact recovery condition* (ERC) in the noiseless setting and the *robust recovery condition* (RRC) in the noisy setting. In the literature, ERC typically requires that all sparse signals can be exactly recovered. In addition to this, RRC requires that if the measurement is noisy, the reconstruction error is bounded by the norm of the noise vector multiplied by a constant factor.

While the rigorous definitions of ERC and RRC are deferred to Section 2, we remark here in passing that RRC trivially implies ERC, because ERC can be seen as a special case of RRC where the measurement is free of noise. Conversely, it is not obvious whether ERC also implies RRC, or RRC is *strictly* stronger than ERC. Early work in compressed sensing have provided sufficient conditions for ERC and RRC of the l_1-minimization, based on the so-called restricted isometry property (RIP) [4], and those sufficient conditions appear to be identical. However, analysis based on RIP generally fails to provide exact (necessary and sufficient) condition for ERC and RRC. Another line of research has considered the null space property (NSP), which gives a both necessary and sufficient condition for ERC of the l_p-minimization. In addition, [19] provided a sufficient condition, called NSP', for RRC of l_p-minimization. Later Aldroubi et al proved in [18] that NSP and NSP' are in fact equivalent. Hence, we have that ERC and RRC are actually the same condition for l_p-minimization.

In contrast to the special case of l_p-minimization, the robust recovery condition for the more general case of F-minimization has been recognized as "not easy to establish" [18], merely based on the idea of NSP. The fundamental issue of robustness in F-minimization has remained relatively unexplored.

The purpose of this paper is to give an exact characterization of the relationship between ERC and RRC in the general F-minimization problem. We first show that ERC and RRC depends only on the configuration of the null space of the measurement matrix (the entire entries of the matrix is of course sufficient, but not necessary, information). Moreover, since the null spaces are linear subspaces of the Euclidean space, they can be viewed as points on a Grassmann manifold, which has a natural topological structure, hence concepts such as open sets and interior are well defined for collections/sets of the null spaces. We denote by Ω and Ω^r the sets that consist of the null spaces satisfying ERC and RRC for the F-minimization, respectively. We show that Ω^r is exactly the interior of Ω (Theorem 2). Hence we can give an alternative proof of the equivalence of ERC and RRC in l_p-minimization, by simply showing that Ω is open in this special case. We would like to remark that this analytical framework also gives rise to new ideas and results, including:

• Equivalence of ERC and RRC in probability. Under some mild assumptions we show that Ω and Ω^r differ by a set of measure zero. Building on this, we show that ERC and RRC hold true with the same probability if the measurement matrix is randomly generated according to a continuous distribution.

• Comparison between different sparseness measures. It is interesting and valuable to know how the performances between different sparseness measures compare. Gribonval et al [8, Lemma 7] provided a condition when one spareness measure is better than another in the sense of ERC. Combining this with our result, we show that this condition also provides a comparison in terms of RRC. Moreover, with the concept of measure zero set on the Grassmannian, we are able to provide addition comparison rules which guarantee one sparse measure is better than the other in terms of probability of ERC/RRC.

The organization of the paper is as follows. In Section 2 we present the mathematical formulation of the problem and a brief introduction to null space property and the Grassmann manifold. Section 3 studies the relationship between ERC and RRC: Part A gives an exact characterization of RRC set as the interior of ERC set on the Grassmannian; in Part B we show than the ERC and RRC sets differ by a set of measure zero. In Section 4 we provide some rules for comparing the performance of different sparse measures. Section V compares our approach and definitions with similar ones in the literature. Finally in Section 6 we conclude by reviewing the results and pointing out possible directions for future work.

2 Problem Setup and Key Definitions

This section provides the mathematical formulation of the problem and the definitions of some key concepts. We shall use lower case bold letters for vectors, and upper case bold letters for matrices. Notation $\mathbb{M}(m,n)$ denotes the set of $m \times n$ real matrices. Throughout the paper we suppose the observation matrix is $m \times n$, and set $l := n - m$, unless otherwise indicated. $\|\mathbf{x}\|_0$ refers to the l_0 norm[1] of $\mathbf{x}$, i.e., the number of non-zero elements in the vector, and $\|\mathbf{x}\|_p := (\sum_k |x(k)|^p)^{1/p}$ denotes the l_p norm of $\mathbf{x}$.

2.1 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 $T := supp(\bar{\mathbf{x}})$ be the support of $\bar{\mathbf{x}}$. Vector $\bar{\mathbf{x}}$ is called k-sparse if $|T| \leq k$. The linear measurement $\mathbf{y}$ is given by

$$\mathbf{y} = \mathbf{A}\bar{\mathbf{x}} + \mathbf{v}. \tag{3}$$

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

$$J(\mathbf{x}) := \sum_{k=1}^{n} F(|x(k)|). \tag{4}$$

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

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

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 (4) 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 [18].

[1] Strictly speaking, the l_0 norm and $l_p (0 < p < 1)$ norm defined here do not satisfy the definition of norm in mathematics.

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

$$\min_{\mathbf{x}\in\mathbb{R}^n} J(\mathbf{x}) \quad s.t.\ \mathbf{A}\mathbf{x} = \mathbf{y}. \tag{5}$$

We say $\mathbf{A}$, 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 (5).*

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

$$\min_{\mathbf{x}\in\mathbb{R}^n} J(\mathbf{x}) \quad s.t.\ \|\mathbf{A}\mathbf{x} - \mathbf{y}\| < \epsilon, \tag{6}$$

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

$$\|\bar{\mathbf{x}} - \hat{\mathbf{x}}\| < C\epsilon, \tag{7}$$

where C *is a constant.*

2.2 Null Space Property

The null space property [8,20,21] is useful for the analysis of a special class of cost functions, which we introduce as follows:

Definition 3 (sparseness measure) *Function*

$$F : [0, +\infty) \to [0, +\infty) \tag{8}$$

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.*[2]

In this paper we assume that the function F is a sparseness measure as in Definition 3. This is a rather loose assumption, so that the key optimization problems in many of the sparse recovery algorithms can be subsumed in our framework, including l_p-minimization and ZAP algorithm. 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:

[2]For our purpose, the definition of sparseness measure in this paper does not need to require that $F(x)/x$ is non-increasing. A comparison with other definitions of the sparseness measure is given in Section 5, Part B.

Lemma 1 (Null space property [8](Lemma 6)) *If $F \in \mathcal{M}$, then a necessary and sufficient condition for ERC is*

$$J(\mathbf{z}_T) < J(\mathbf{z}_{T^c}), \quad \forall \mathbf{z} \in \mathcal{N}(\mathbf{A}) \setminus \{\mathbf{0}\},\ |T| \leq k. \tag{9}$$

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

It's useful to define the *null space constant [8]*, especially when one wants to study l_p-minimization or to compare it with F-minimization:

Definition 4 (Null space constant, NSC) *Suppose $F \in \mathcal{M},\ q \in (0,1]$. Define the null space constant is defined as:*

$$\theta_J := \sup_{\mathbf{z} \in \mathcal{N}(\mathbf{A}) \setminus \{\mathbf{0}\}} \max_{|T| \leq k} \frac{J(\mathbf{z}_T)}{J(\mathbf{z}_{T^c})}. \tag{10}$$

In the same spirit, we denote by θ_{l_p} the null space constant associated with l_p cost function.

The null space constant is closely associated with NSP, and hence characterizes the performance of F-minimization. We have the following result, which is a direct consequence of Definition 4 and Lemma 1.

Lemma 2
1) $\theta_J \leq 1$ 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. [19]), 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.*

2.3 Preliminaries of the Grassmann Manifold

In this part we briefly review some relevant properties of the Grassmann manifold. More detailed treatment of the subject can be found in many standard texts, such as [22,23]. The main thrust for considering this object is that, the property of exact recovery of a particular measurement matrix is completely determined by its null space, from Lemma 1. Of course, $\mathcal{N}(\mathbf{A})$ is an $l := n - m$ dimensional linear subspace of $\mathbb{R}^n$ when $\mathbf{A}$ is of full rank.

Geometrically, the Grassmann manifold $G_l(\mathbb{R}^n)$ can be conceived as the collection of all the l dimensional subspaces (l-planes) of $\mathbb{R}^n$. One can introduce a topology on $G_l(\mathbb{R}^n)$ by defining a metric on it: for arbitrary $\nu, \nu' \in G_l(\mathbb{R}^n)$, the distance between ν, ν' can be defined as [24]:

$$d(\nu, \nu') := \|\mathbf{P}_\nu - \mathbf{P}_{\nu'}\|, \tag{11}$$

where $\mathbf{P}_\nu$ (resp. $\mathbf{P}_{\nu'}$) is the projection matrix onto ν (resp. ν'), and $\|\cdot\|$ denotes the spectral norm. The Grassmann manifold is then a compact metric space.

We shall next define the coordinates on $G_l(\mathbb{R}^n)$ to introduce its differential manifold structure. Let $F(n,l)$ be the set of all non-degenerate (invertible) $n \times l$ matrices, and let $\sim$ be the following equivalence relation: If $\mathbf{X}, \mathbf{Y} \in F(n,l)$, then $\mathbf{X} \sim \mathbf{Y}$ means $\mathbf{X}, \mathbf{Y}$ are equivalent up to a non-degenerate column transform, i.e. $\mathbf{X}, \mathbf{Y}$ spans the same linear subspace. Hence the Grassmann manifold can be defined as a quotient space $G_l(\mathbb{R}^n) := F(n,l)/\sim$, for which we denote by $\pi : F(n,l) \to G_l(\mathbb{R}^n)$ the associated natural projection. For any arbitrary collection of indices $1 \le i_1 < i_2 < \cdots < i_l \le n$, let $1 \le \bar{i}_1 < \bar{i}_2 < \cdots < \bar{i}_{n-l} \le n$ be the remaining indices. Given an index set $I = \{i_1, i_2, \ldots, i_l\}$, we denote by $\mathbf{X}_I$ the $l \times l$ sub-matrix formed by the rows of $\mathbf{X}$ indexed by I. Define

$$U_I := \{\mathbf{X} \in F(n,l) : \det \mathbf{X}_I \neq 0\}, \quad V_I := \pi(U_I). \tag{12}$$

Then $\{V_I\}$ constitutes an open covering of $G_l(\mathbb{R}^n)$. For any arbitrarily chosen $\mathbf{Y} \in \pi^{-1}(v)$, where $v \in V_I$, the following matrix is $\sim$ equivalent with $\mathbf{Y}$, and is independent with the specific choice of $\mathbf{Y}$ which represents v:

$$\mathbf{X} = \mathbf{Y} \cdot (\mathbf{Y}_I)^{-1} = \left(\begin{array}{ccc} 1 & & \\ & \ddots & \\ & & 1 \\ \hline z_{\bar{i}_1,1} & \cdots & z_{\bar{i}_1,l} \\ \vdots & & \vdots \\ z_{\bar{i}_{n-l},1} & \cdots & z_{\bar{i}_{n-l},l} \end{array}\right) \begin{array}{c} (i_1) \\ \vdots \\ (i_l) \\ (\bar{i}_1) \\ \vdots \\ (\bar{i}_{n-l}) \end{array} \tag{13}$$

Note that in the above, we have performed a row permutation to the last matrix for clarity of display. Define $\phi_I : V_I \to \mathbb{M}(n-l, l), v \mapsto \mathbf{X}_{\bar{I}}$, then $\{(V_I, \phi_I) : 1 \le i_1 < \cdots < i_l \le n\}$ forms an atlas of $G_l(\mathbb{R}^n)$.

Concepts such as open sets and interior are well-defined once a topology on $G_l(\mathbb{R}^n)$ has been unambiguously chosen. One might notice that there are possibly two topologies defined on $G_l(\mathbb{R}^n)$: the metric topology arising from the metric defined in (11), and the manifold topology (which is connected to the standard topology on $\mathbb{R}^{ml}$ by all the homeomorphisms $\{\phi_I\}$). Unsurprisingly these two topologies agree, since standard calculations would show that the metric on U_I induced from the Euclidean metric on $\phi_I(U_I)$ is topologically equivalent to the metric defined in (11).

Further, since $G_l(\mathbb{R}^n)$ is a C^∞ (therefore differentiable) manifold, the concept of measure zero set can be defined as follows:

Definition 5 *[23, Definition 1.16] A subset A of a differentiable manifold has* measure zero *if $\phi(A \cap U)$ has Lebesgue measure zero for every chart (U, ϕ).*

There is a unique (up to a scalar factor) rotational invariant measure on $G_l(\mathbb{R}^n)$, i.e., the Haar measure. The requirement that a set A has zero Haar measure agrees with Definition

5. [3] 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. We use μ to denote the normalized Haar measure on $G_l(\mathbb{R}^n)$, which can also be understood as a probability.

3 The Relationship between ERC and RRC

3.1 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:

$$\Omega_J := \{\nu \in G_l(\mathbb{R}^n) : J(\mathbf{z}_T) < J(\mathbf{z}_{T^c}), \forall \mathbf{z} \in \nu \setminus \{\mathbf{0}\}, |T| \leq k\}. \tag{16}$$

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

$$\Omega_{J_G} \subseteq \Omega_{J_F}, \tag{17}$$

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 (17).

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 (6). 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*

[3] This is because the Haar measure on $G_l(\mathbb{R}^n)$ can be associated with an exterior differential form $\mathrm{d}X$ and an orientation (c.f. [25, Section 1.4]) such that $\mu(A) = \int_A \mathrm{d}X$. The integral is define as follows: given an arbitrary oriented atlas $\{(U_\alpha, \phi_\alpha) | \alpha \in \mathcal{A}\}$ and a partition of unity $\{\eta_\alpha | \alpha \in \mathcal{A}\}$ obeying:

$$\eta_\alpha \geq 0, \quad \mathrm{supp}\eta_\alpha \subseteq U_\alpha, \quad \sum_{\alpha \in \mathcal{A}} \eta_\alpha \equiv 1,$$

then the probability of A is given by

$$\int_A \mathrm{d}X = \sum_{\alpha \in \mathcal{A}} \int_{\phi(A \cap U_\alpha)} f_\alpha(\phi_\alpha^{-1}(x)) \eta_\alpha \mathrm{d}x_1 \ldots \mathrm{d}x_{ml},$$

where $f_\alpha \in C^\infty(G_l(\mathbb{R}^n))$ is a positive function such that $\mathrm{d}X|_{U_\alpha} = f_\alpha \wedge_{i=1}^{ml} \mathrm{d}x_i$. Since the last integral is the Lebesgue integral of a positive function on $\mathbb{R}^{ml}$, we deduce that $\lambda(\phi(A \cap U_\alpha)) = 0$ if and only if the integral vanishes.

$\|\mathbf{n}\| < d\|\mathbf{z}\|$, *and* $|T| \leq k$, *we have the following:*

$$J(\mathbf{z}_T + \mathbf{n}_T) < J(\mathbf{z}_{T^c} + \mathbf{n}_{T^c}). \tag{18}$$

Proof See Appendix A. ■

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 the 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:

$$\begin{aligned}\Omega_J^r :=&\{\nu \in G_l(\mathbb{R}^n) : \exists d > 0, \text{s.t. } J(\mathbf{z}_T + \mathbf{n}_T) < J(\mathbf{z}_{T^c} + \mathbf{n}_{T^c}),\\ &\forall \mathbf{z} \in \nu \setminus \{\mathbf{0}\}, \|\mathbf{n}\| < d\|\mathbf{z}\|, |T| \leq k\}.\end{aligned} \tag{19}$$

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:

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

$$\Omega_J^r = int(\Omega_J). \tag{20}$$

Proof See Appendix B. ■

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 [18], 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.*

Proof See Appendix C. ■

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 $0 < p \leq 1$, 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. A comparison of different proof methods can be found in Section 5, Part A.*

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 $\Omega^r_{l_p} = int(\Omega_{l_p}) = \Omega_{l_p}$. ■

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*

$$F(x) := x + 1 - e^{-x}. \tag{21}$$

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 $\mathbb{R}^3$

$$\mathcal{N} := [x, y, z]^T, \tag{22}$$

where the homogenous coordinates $[x, y, z]^T$ denotes the subspace 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:

$$J(\mathbf{z}_T) < J(\mathbf{z}_{T^c}). \tag{23}$$

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

$$F((1-d)xt) + F(yt) < F(zt). \tag{24}$$

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 $\|\mathbf{n}\|/\|\mathbf{z}\| \leq d$; on the other hand (18) doesn't hold because of (24). Therefore RRC is not fulfilled as a result of Theorem 1. ■

3.2 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 Example 1. Using the subadditivity property and the Taylor expansion of F at the origin, one can explicitly write out:

$$\Omega_J = \left\{ [x_1, x_2, x_3] : 2 \max_{i=1,2,3} |x_i| \leq \sum_{i=1,2,3} |x_i| \right\}, \tag{25}$$

and

$$\Omega_J^r = \left\{ [x_1, x_2, x_3] : 2 \max_{i=1,2,3} |x_i| < \sum_{i=1,2,3} |x_i| \right\}. \tag{26}$$

We recall that μ denotes the Haar measure on $G_l(\mathbb{R}^n)$. From (25) and (26) it is intuitively clear in this simple case that $\mu(\Omega_J) = \mu(\Omega_J^r)$, 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 2, 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 $\mu(\Omega^r) = \mu(int(\Omega^r))$, as will be shown in the remark at the end of this section. However we can show the following result:

Theorem 3 *Suppose $F \in \mathcal{M}$ is a non-decreasing function, then $\Omega_J - int(\Omega_J)$ is a measure zero set, that is, $\mu(\Omega_J \setminus \Omega_J^r) = 0$.*[4]

PROOF See Appendix D. ■

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*

$$F(x) = \begin{cases} 0 & x = 0; \\ 0.1 & x > 0 \text{ and } x \text{ is rational}; \\ 1 & x > 0 \text{ and } x \text{ is irrational}, \end{cases} \tag{27}$$

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

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[5]. 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 $\mu(\Omega_J) = 1$. One 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 $\mu(int(\Omega_J)) = 0 \neq \mu(\Omega_J)$.

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,[6] then $P(\Omega_J \setminus \Omega_J^r) = 0$. Then it is not counter-intuitive that this should be true if the entries of $\mathbf{A}$ are i.i.d. generated from a certain continuous distribution, which is a common practice used in generating the observation matrix. Nevertheless, the above speculation requires a formal justification. We formulate this result as a corollary, the proof of which is deferred to Appendix E.

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 [26, 27]. 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 $G_l(\mathbb{R}^n)$.*

4 Comparison of Different Sparseness Measures

In this section we provide some methods to compare the performance between two sparseness measures in terms of ERC or RRC, as an application of the results from the previous sections. It turns out that both the topological characterization of RRC and the probabilistic (measure-theoretic) viewpoint become particularly useful when passing from the l_p cost functions to general sparseness measures.

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

[6] The measure μ_1 is said to be absolutely continuous with respect to the measure μ_2 if $\mu_2(E) = 0$ implies $\mu_1(E) = 0$, for arbitrary measurable set E.

The following lemma comes from the corresponding result for ERC in [8] and our interior point characterization of RRC:

Lemma 5 *Suppose $F, G \in \mathcal{M}$. If F, G are non-decreasing and F/G is non-increasing on $\mathbb{R}^+$, then we have $\Omega_{J_G} \subseteq \Omega_{J_F}$ and $\Omega^r_{J_G} \subseteq \Omega^r_{J_F}$.*

Proof The fact that $\Omega_{J_G} \subseteq \Omega_{J_F}$ comes from [8, Lemma 7]. It then follows that $\Omega^r_{J_G} \subseteq \Omega^r_{J_F}$ from Theorem 2. ■

The set inclusions formulas in Lemma 5 means that the sparseness measure F is better than G, in the sense that whenever the cost function J_G guarantees ERC/RRC, so does the J_F. By letting $G(x) := x^q$ in this lemma we can obtain the following result:

Corollary 4 *Suppose $F \in \mathcal{M}$, $p \in (0, 1]$. If F is non-decreasing and $F(x)/x^p$ is non-increasing on $\mathbb{R}^+$, then we have $\Omega_{l_p} \subseteq \Omega_{J_F}$ and $\Omega^r_{l_p} \subseteq \Omega^r_{J_F}$.*

Corollary 4 gives a condition such that J_F is better than l_p in the sense of ERC and RRC. Conversely, we shall show that the asymptotic of F around 0^+ and $+\infty$ gives a sufficient condition that l_p is better than J_F in terms of probability.

Theorem 4 *Suppose $F \in \mathcal{M}$, $p \in (0, 1]$. If $\lim_{x \to 0^+} F(x)/x^p$ or $\lim_{x\to\infty} F(x)/x^p$ exist and is positive, then $\Omega_{J_F} \subseteq \overline{\Omega}_{l_p}$, and $\mu(\Omega_{J_F}) \le \mu(\Omega_{l_p})$.*

Proof See Appendix F. ■

We Remark that $\mu(\Omega_{J_F}) \le \mu(\Omega_{l_p})$ in Theorem 4 cannot be replaced by the stronger set inclusion relation $\Omega_{J_F} \subseteq \Omega_{l_p}$, which holds for l_p cost functions but fails for general sparseness measures. Thus the measure-theoretic viewpoint allows us to restore a comparison criteria when extending l_p-minimization to the F-minimization.

From the above result, we can immediately derive the relation between ZAP [2] and l_1-minimization. The typical form of sparseness measure used in the ZAP algorithm is the following:

$$F(x) = \begin{cases} \alpha x - \alpha^2 x^2 & x < 1/\alpha; \\ 1 & \text{otherwise,} \end{cases} \tag{28}$$

where the tuning parameter α is usually chosen as the inverse of the standard deviation of the non-zero entries in $\bar{\mathbf{x}}$. Our following result says that, while ZAP performs far better than l_1-minimization in the average case, as shown in the numerical experiments [2], the worst case performance (requiring all sparse vectors can be constructed) of the two cost functions are the same:

Corollary 5

$$\mu(\Omega_{ZAP}) = \mu(\Omega_{l_1}). \tag{29}$$

PROOF Using Corollary 4 and Theorem 4 with $p = 1$ one can obtain both the lower and upper bound on $\mu(\Omega_{ZAP})$ respectively. ■

We end this section by summarizing the relationship between the various requirements on F appeared in this section:

Proposition 2 *Assuming that* $0 \le p \le 1$, $F : [0, +\infty) \to [0, +\infty)$, *and* $F(0) = 0$, *we have*
(1) F *is concave* $\Longrightarrow$ $F(t)/t$ *is non-increasing;*
(2) $F(t)/t^p$ *is non-increasing* $\Longrightarrow$ $F(t)/t$ *is non-increasing;*
(3) $F(t)/t$ *is non-increasing* $\Longrightarrow$ F *is sub-additive.*

5 Comparison with Other Works

5.1 The ERC/RRC Equivalence for l_p-minimization

To the best of our knowledge, the *exact* characterization of robustness of l_p-minimization first appeared in [19], where the definition of robustness is the same as in our paper. In [19] a variant of the null space property, called NSP', was proposed as a sufficient condition for the robustness of l_p minimization. The NSP' is obviously stronger than NSP, but the reverse situation is not immediately clear. Later Aldroubi et al adopted the same approach in [18], and proved that NSP and NSP' are in fact equivalent (see also [28]). The proof method in [18] requires a lemma from matrix analysis [18, Lemma 2.1]. We remark that this lemma, from a slightly more general viewpoint, can be seen as a classical application of the open mapping theorem in functional analysis [29, Chapter 4, Corollary 3.2]. Thus it is established that NSP, NSP', ERC and RRC are all equivalent for l_p-minimization.

While the NSP' approach is nice for the l_p case, it is hard to be extended to the general F-minimization problem. This is because NSP' consists of a homogeneous inequality, which appears to work well only for homogeneous cost functions such as the l_p norm. In contrast, the heart of our approach is the interior point characterization of RRC (Theorem 2) for the general F-minimization problem. Then our proof of the ERC/RRC equivalence for l_p-minimization, although involves some basic facts about topological spaces, follows almost immediately as a corollary. Note this application is particularly interesting since the statement of ERC/RRC equivalence does not involve topology at all. Nevertheless, we emphasize that the significance of Theorem 2 is to provide a simple, accurate, and general characterization of the robustness of F-minimization; and the proof of ERC/RRC equivalence for l_p is one of its applications in a special setting.

5.2 The Notion of Sparseness Measure

The sparseness measure defines the class of cost functions of our interest, and is therefore of great importance. In general we want to consider a class wide enough to cover most applications, but also small enough to possess important recovery properties. Intuitively, the cost function should penalize non-zero coefficients, and not penalize the zero coefficients. However there are additional reasonable requirements, the precise definitions of which differ in the literature. For clarifications we compare these different requirements on F as follows (Recall that $\mathcal{M}$ denotes the set of sparseness measures defined in Definition 3):

- $F \in \mathcal{M}$. This is the class of functions mainly considered in our paper as well as [18]. This seems to be most general class of functions that can be studied by the null space property.
- $F \in \mathcal{M}$ and F is non-decreasing. This requirement appears in Theorem 3. As shown in the counter example in the remark following the theorem, the assumption that F being non-decreasing cannot be dropped.
- $F \in \mathcal{M}$, F is non-decreasing, and $F(t)/t$ is non-increasing[7]. This requirement is considered in [8, 30], and it guarantees 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 [8, Lemma 7]. Finally, l_1 norm is the only convex cost function whose corresponding F satisfies this definition of sparseness measure [30, Proposition 2.1].

6 Conclusion

F-minimization refers to a broad family of non-convex optimizations for sparse recovery which has outperformed conventional l_1 minimization experimentally. However because of some technical difficulties, the robustness of F-minimization was not fully understood before, even though its exact recovery property has been studied by using the null space property. The novel approach of this paper is to view the collection of null spaces as a topological manifold, called the Grassmann manifold, and provide an exact characterization of the relationship between robust recovery condition (RRC) and exact recovery condition (ERC): the set of null spaces satisfying RRC is the interior of the one satisfying ERC (Theorem 2). Building on this characterization, the previous result of the equivalence of exact recovery and robust recovery in the l_p-minimization follows as an easy consequence. Besides some rather direct applications of the our interior characterization of RRC, such as the comparison of different sparseness measures, we showed another main result that if F is non-decreasing then the sets associated with ERC and RRC differ by a set of measure zero (Theorem 3). The practical significance of this result is that ERC and RRC will occur with equal probability when the measurement matrix is randomly generated according to a continuous distribution.

Further improvements may include finding more general conditions on F than non-decreasing in order that Theorem 3 still holds. Studies of the robustness under perturbation in the measurement matrix may also be of interest.

[7] The assumption of $F(t)/t$ being non-increasing guarantees that F is sub-additive, as shown in Proposition 1.

Appendix A Proof of Theorem 1

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

$$\|\mathbf{A}(\hat{\mathbf{x}} - \bar{\mathbf{x}})\| \le \|\mathbf{A}\hat{\mathbf{x}} - \mathbf{y}\| + \|\mathbf{A}\bar{\mathbf{x}} - \mathbf{y}\| \le 2\epsilon. \tag{30}$$

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

$$J(\mathbf{u}_T) \ge J(\mathbf{u}_{T^c}). \tag{31}$$

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 (18), hence from the assumption we must have:

$$\|\mathbf{n}\| \ge d\|\mathbf{z}\|. \tag{32}$$

Therefore

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

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

Necessity: We will show by contradiction. Assuming that

$$\forall d > 0, \exists \|\mathbf{n}\| < d\|\mathbf{z}\|, \mathbf{z} \in \mathcal{N}(\mathbf{A}), \text{such that } J(\mathbf{z}_T + \mathbf{n}_T) \ge J(\mathbf{z}_{T^c} + \mathbf{n}_{T^c}), \tag{33}$$

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

$$\|\mathbf{x}_1 - \mathbf{x}_2\| \ge \frac{2(1-d)\epsilon}{d\|\mathbf{A}\|}. \tag{34}$$

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 (33). Define[8] $\mathbf{u} := \mathbf{z} + \mathbf{n}$, $\mathbf{x}_1 = (\mathbf{u}_T)^T$, $\mathbf{x}_2 = -(\mathbf{u}_{T^c})^{T^c}$, $\mathbf{v} = \mathbf{A}(\mathbf{x}_1 - \mathbf{x}_2)/2$, $\|\mathbf{v}\| = \epsilon$. Then $\|\mathbf{A}\mathbf{x}_1 - (\mathbf{A}\mathbf{x}_2 + \mathbf{v})\| = \epsilon$. Hence

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

Thus the relation (34) holds, as desired.

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

Appendix B Proof of Theorem 2

Lemma 6 *Suppose $\nu \in G_l(\mathbb{R}^n)$. For all $\mathbf{z} \in \nu \setminus \{\mathbf{0}\}, \|\mathbf{n}\| < d\|\mathbf{z}\|$, there exists $\nu' \in G_l(\mathbb{R}^n)$ such that $\mathbf{z} + \mathbf{n} \in \nu'$ and $d(\nu, \nu') < d$.*

Proof Since distances are preserved under a rotation, we can assume without generality that $\nu = span(\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_l)$,[9] and $span(\mathbf{z}) = span(\mathbf{e}_l)$. We then define $\nu' = span(\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_{l-1}, \mathbf{z}+\mathbf{n})$. There is a column transformation that transforms $(\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_{l-1}, \mathbf{z} + \mathbf{n})$ into

$$\mathbf{M} := \left(\begin{array}{c|c} \mathbf{I}_{l-1} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{w} \end{array}\right), \tag{35}$$

where the vector $\mathbf{w} \in \mathbb{R}^{n-l+1}$ satisfies $\|\mathbf{w}\| = 1$. Define $\tilde{\mathbf{w}} := (w_2, w_3, \ldots, w_{l+1})^T \in \mathbb{R}^l$. With some basic algebra we get $\|\tilde{\mathbf{w}}\| < d$. Now the column vectors of $\mathbf{M}$ still spans ν', and one has that

$$\mathbf{P}_\nu = \left(\begin{array}{c|c} \mathbf{I}_l & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} \end{array}\right); \tag{36}$$

$$\mathbf{M}^T\mathbf{M} = \mathbf{I}_l; \tag{37}$$

$$\begin{aligned} \mathbf{P}_{\nu'} &= \mathbf{M}(\mathbf{M}^T\mathbf{M})^{-1}\mathbf{M}^T \\ &= \left(\begin{array}{c|c} \mathbf{I}_{l-1} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{w}\mathbf{w}^T \end{array}\right). \end{aligned} \tag{38}$$

Then with some linear algebra we obtain

$$(\mathbf{P}_\nu - \mathbf{P}_{\nu'})^2 = \left(\begin{array}{c|c|c} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \hline \mathbf{0} & \tilde{\mathbf{w}}^T\tilde{\mathbf{w}} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} & \tilde{\mathbf{w}}\tilde{\mathbf{w}}^T \end{array}\right). \tag{39}$$

Since $\|\tilde{\mathbf{w}}\| < d$, (39) implies that $\|\mathbf{P}_\nu - \mathbf{P}_{\nu'}\| < d$. ■

The proof of Theorem 2 is now in reach. We shall first show that $\Omega^r \subseteq int(\Omega)$. Suppose $\nu \in \Omega^r$, and let $d < 1$ be one feasible parameter that appears in the definition (19). It suffices to prove that the neighbourhood $U = \{\nu' : d(\nu, \nu') < d/(1+d)\}$ is a subset of Ω, i.e. any ν' in this neighbourhood satisfies the condition in (16). This is because for any $\nu' \in U$ and $\mathbf{z}' \in \nu' \setminus \{\mathbf{0}\}$, one can find $\mathbf{z} := \mathbf{P}_\nu\mathbf{z}' \in \nu$ such that $\|\mathbf{z} - \mathbf{z}'\|/\|\mathbf{z}'\| < d/(1+d)$, which implies that $\|\mathbf{z} - \mathbf{z}'\|/\|\mathbf{z}\| < d$. This combined with the fact that $\mathbf{z} \in \Omega^r$ shows $J(\mathbf{z}'_T) < J(\mathbf{z}'_{T^c})$ for every $|T| \leq k$. Hence $\nu' \subseteq \Omega$ follows from the arbitrariness of $\mathbf{z}'$, as desired.

Next we have to show the converse. If $\nu \in int(\Omega)$, then by definition there exists $d > 0$ such that

$$\nu' \in \Omega, \quad \forall d(\nu, \nu') < d. \tag{40}$$

[9] $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ here denotes the standard basis of $\mathbb{R}^n$.

Now $\forall \mathbf{z} \in \nu \setminus \{\mathbf{0}\}, \|\mathbf{n}\| < d\|\mathbf{z}\|$, there exist ν' such that $\mathbf{z} + \mathbf{n} \in \nu'$ and $d(\nu, \nu') < d$. Hence $\nu' \in \Omega$, meaning that $J(\mathbf{z}_T + \mathbf{n}_T) < J(\mathbf{z}_{T^c} + \mathbf{n}_{T^c})$ for every $|T| \leq k$. This implies that $\nu \in \Omega^r$. By the arbitrariness of ν we conclude that $int(\Omega) \subseteq \Omega^r$.

Appendix C Proof of Lemma 4

1) The lower semi-continuity of g is obvious; indeed, it follows from the fact that g is defined as the supremum of a collection of continuous functions [31, P38 (c)].

2) To show that g is also upper semi-continuous when M is compact: We will prove that g is upper semi-continuous at an arbitrary $x_0 \in X$: let y_0 be a point in M such that $g(x_0) = f(x_0, y_0)$ (Here we used the compactness of M). Suppose otherwise, that g is not supper semi-continuous at x_0, then there exists $\epsilon > 0$ such that:

$$\limsup_{x \to x_0} g(x) > g(x_0) + \epsilon. \tag{41}$$

This implies that we can find sequences $x_n, y_n (n \geq 1)$ such that $\lim_{n\to\infty} x_n = x_0$ and the following holds:

$$f(x_n, y_n) > g(x_0) + \epsilon. \tag{42}$$

Since M is compact, we can find a subsequence $y_{n_k}, (k \geq 1)$ converging to some point $y^* \in M$. Hence

$$\begin{aligned} g(x_0) &= f(x_0, y_0) \\ &\geq f(x_0, y_0) \\ &= \lim_{k\to\infty} f(x_{n_k}, y_{n_k}) \\ &\geq g(x_0) + \epsilon, \end{aligned}$$

which is an apparent contradiction.

Remark 4 *In the above proof, the assumption that X, M are metrical spaces rather than topological spaces is useful only when showing the existence of the sequences $x_n, y_n, (n \geq 1)$. Therefore, for full generality we may just assume that X, M are topological spaces satisfying the first countable theorem [32].*

Appendix D Proof of Theorem 3

Notation 1 *In this section we define Ω_T as follows. (This is not to be confused with the definition of Ω_J or Ω_{l_p}.)*

$$\begin{aligned} \Omega_T := \{&\nu \in G_l(\mathbb{R}^n) : \\ &J(\mathbf{z}_T) < J(\mathbf{z}_{T^c}), \forall \mathbf{z} \in \nu \setminus \{\mathbf{0}\}\}, \end{aligned} \tag{43}$$

hence $\Omega = \bigcap_{|T|=k} \Omega_T$.

Let $P^{n-1}(\mathbb{R}) = G_1(\mathbb{R}^n)$ be the $n-1$ dimensional projective space (i.e., the set of one dimensional linear subspaces of $\mathbb{R}^n$), and $\rho^n : \mathbb{R}^n \setminus \{\mathbf{0}\} \to P^{n-1}(\mathbb{R})$ be the natural projection from $\mathbb{R}^n \setminus \{\mathbf{0}\}$ to $P^{n-1}(\mathbb{R})$. Then $X \in P^{n-1}(\mathbb{R})$ can be expressed in homogenous coordinates as $X = [x_1, \ldots, x_n]$ if $\rho^n((x_1, \ldots, x_n)) = X$. We note that the function d defined in (11) can be naturally extended to the case where the linear subspaces ν, ν' are of different dimensions, in particular the case where one of the two subspace belongs to $G_l(\mathbb{R})$ and the other belongs to $P^{n-1}(\mathbb{R})$. The following two observations about d, although simple in natural, are useful when dealing with the distance between elements in $G_l(\mathbb{R})$ and $P^{n-1}(\mathbb{R})$:

- If $\nu_i, i = 1, 2, 3$ are linear subspaces of $\mathbb{R}^n$ (with possibly different dimensions), then

$$d(\nu_1, \nu_2) \le d(\nu_1, \nu_3) + d(\nu_2, \nu_3). \tag{44}$$

- If $\nu_i, i = 1, 2, 3$ are linear subspaces of $\mathbb{R}^n$ (with possibly different dimensions) and $\nu_1 \subseteq \nu_2$, then

$$d(\nu_2, \nu_3) \le d(\nu_1, \nu_3). \tag{45}$$

Notation 2 *There is a partial relation $\succcurlyeq_T$ on $P^{n-1}(\mathbb{R})$, defined as follows: If $X, Y \in P^{n-1}(\mathbb{R})$ and X, Y can be expressed in homogenous coordinates as $[x_1, \ldots, x_n], [y_1, \ldots, y_n]$, where $|x_i| \ge |y_i|,\ i \in T$ and $|x_i| \le |y_i|,\ i \in T$, then $X \succcurlyeq_T Y$.*

Since F is non-decreasing, we observe the following property:

Property 1 *Suppose an l-plane $\nu \in \Omega_T^c$ passes through a line X, and an l-plane ν' passes through a line X', with the condition $X' \succcurlyeq_T X$. Then ν' must also belong to Ω_T^c.*

Definition 6 *[33] Suppose E is a measurable set in $\mathbb{R}^L$, the* Lebesgue density *of E at a point $\mathbf{x} \in \mathbb{R}^L$ is defined as $\lim_{r\to 0} \frac{\lambda(B(\mathbf{x},r)\cap E)}{\lambda(B(\mathbf{x},r))}$ where λ denotes the Lebesgue measure. If the density exists and is equal to 1, $\mathbf{x}$ is said to have the Lebesgue density of E.*

The Lebesgue density theorem [33, Chapter 3, Corollary 1.5] claims that almost all $\mathbf{x} \in E$ (except for a set of Lebesgue measure zero) has the Lebesgue density of E. Thus this theorem essentially says that the set E is "robust" when taking those points that are of the Lebesgue density of E. While a point in $int(E)$ must has the Lebesgue density of E, the converse is not always true. However, the idea in the proof of Theorem 3 is to show a converse statement as such by using the monotonicity of F.

PROOF (PROOF OF THEOREM 3) Note that in the Definition 5 of measure zero set, the "every chart" can be replaced by "a collection of charts where the coordinate neighbourhoods cover the manifold", which is an easy consequence of the fact that C^∞ functions on the Euclidean space maps measure zero sets into measure zero sets. With this in mind we only have to show that for any chart (U_I, ϕ_I) and T it holds that $\lambda(\phi_I([\Omega - int(\Omega)] \cap U_I)) = 0$.

Since

$$\begin{aligned} int(\Omega) - \Omega &= int(\bigcap_{|T|=k} \Omega_T) - \bigcap_{|T|=k} \Omega_T \\ &= \bigcap_{|T|=k} int(\Omega_T) - \bigcap_{|T|=k} \Omega_T \\ &\subseteq \bigcup_{|T|=k} (int(\Omega_T) - \Omega_T), \end{aligned} \tag{46}$$

we in turn only need to prove

$$\lambda(\phi_I([\Omega_T - int(\Omega_T)] \cap U_I)) = 0 \tag{47}$$

for every $|T| \leq k$. Moreover, define

$$\begin{aligned} S &= \{\nu \in G_l(\mathbb{R}^n) : \forall \mathbf{x} \in \nu \text{ has at most } l-1 \text{ zero entries}\} \\ &= \bigcap_{\substack{I \subseteq \{1,\ldots,n\} \\ |I|=l}} U_I. \end{aligned} \tag{48}$$

Then obviously S^c is a zero measure set. Hence we only have to prove

$$\lambda(\phi_I([\Omega_T - int(\Omega_T)] \cap S)) = 0 \tag{49}$$

for every $|T| \leq k$.

For any $\nu \in S - int(\Omega_T)$, there exists a sequence $\nu^l \in S - \Omega_T, l = 1, 2, \ldots, \infty$ such that $\nu^l \to \nu$ as $l \to \infty$. Then for each l we can find $\mathbf{z}^l \in \nu^l - \{0\}$ such that $J(\mathbf{z}^l_T) \geq J(\mathbf{z}^l_{T^c})$. Because $P^{n-1}(\mathbb{R})$ is compact, $\rho^n(\mathbf{z}^l)$ has a convergent subsequence converging to an $X \in P^{n-1}(\mathbb{R})$. Moreover from the observations in (44) and (45) we have

$$\begin{aligned} d(X, \nu) &\leq d(X, \rho^n(\mathbf{z}^l)) + d(\rho^n(\mathbf{z}^l), \nu) \\ &\leq d(X, \rho^n(\mathbf{z}^l)) + d(\nu^l, \nu) \\ &\to 0, \end{aligned} \tag{50}$$

therefore $X \in \nu$.

Suppose $\mathbf{x} = (x_1, \ldots, x_n)^T$ and $X = \rho^n(\mathbf{x})$. Since $\nu \notin S$, at most $l-1$ of the entries of $\mathbf{x}$ can be zero. Hence there exists an l-element index set $I' \subseteq \{1, \ldots, n\}$ such that $\{x_i : i \in I'\}$ has at least one non-zero element and $\{x_i : i \notin I'\}$ has no zero element, and therefore we can choose (with much liberty) an invertible matrix $\mathbf{B} \in \mathbb{M}(l, l)$ such that the first column of $\mathbf{B}$ is $\mathbf{x}_{I'}$. With this construction and in view of Property 1, it is obvious that the Lebesgue density of $\mathbf{B} \circ \phi_{I'}(S - \Omega)$ at $\mathbf{B} \circ \phi_{I'}(\nu)$ is at least $2^{-m} > 0$. Since $(\mathbf{B} \circ \phi_{I'}, U_{I'})$ and (ϕ_I, U_I) are C^∞-compatible, the Lebesgue density of $\phi_I(S - \Omega)$ at $\phi_I(\nu)$ is also positive, therefore the Lebesgue density of $\phi_I(S \cap \Omega)$ at $\phi_I(\nu)$ cannot be 1. Recalling that ν is arbitrarily chosen from $S - int(\Omega)$, by the Lebesgue density theorem $\phi_I(S \cap [\Omega - int(\Omega)])$ must be a zero measure set. ∎

Appendix E Proof of Corollary 3

Informally, the idea of the proof can be described as follows. Suppose $\mu(\Omega_J \setminus \Omega_J^r) = 0$. Let $H \subseteq G_m(\mathbb{R}^n)$ be the set of orthogonal complements of the l-dimensional subspaces in $\Omega_J \setminus \Omega_J^r$. Since $G_m(\mathbb{R}^n)$ is isomorphic to $G_l(\mathbb{R}^n)$ (recall that $l := n - m$), we have $\mu(H) = 0$ as well[10]. Notice that an $m \times n$ matrix can be described as an m-dimensional linear space together with m^2 coordinates, we can think of $\mathbb{M}(m,n)$ as having a similar structure as $G_m(\mathbb{R}^n) \times \mathbb{R}^{m^2}$. However $E \times \mathbb{R}^{m^2}$ has measure 0 by the property of the product measure, therefore the set of $m \times n$ matrices whose row spaces lie in E has Lebesgue measure zero. Then the probabilistic equivalence of ERC and RRC follows from the definition of absolute continuity.

Now the major deficiency of the above intuition is that $\mathbb{M}(m,n)$ is not indeed identical to the product space $G_m(\mathbb{R}^n) \times \mathbb{R}^{m^2}$. Thus we need the following concept:

Definition 7 *[23, 34]: A* vector bundle *is a 4-tuple $(E, X, \pi, \mathbb{R}^k)$ in which the* base space *X is a C^r manifold, and the following conditions are satisfied:*

(a) Each $x \in X$ is associated with a fiber*, which is a k-dimensional vector space of E_x, and we can write the* total space *E as a disjoint union:*

$$E = \coprod_{x \in X} E_x, \tag{51}$$

and the projection *map $\pi : E \to X$ satisfies*

$$\pi(\xi) = x, \quad \forall \xi \in E_x. \tag{52}$$

(b) There exists an open cover $\mathcal{U}$ of X, such that for each $U \in \mathcal{U}$, there is a corresponding family of frames

$$e_U(x) = (e_{U,1}(x), \dots, e_{U,k}(x)).$$

Here, $e_U(x)$ is a frame of E_x for each $x \in U$. If $U, V \in \mathcal{U}$ and $U \cap V \neq \emptyset$, both $e_U(x)$ and $e_V(x)$ are defined on $U \cap V$, therefore there is a $k \times k$ non-degenerate matrix $A_{UV}(x)$ such that

$$e_U(x) = e_V(x) A_{UV}(x). \tag{54}$$

Moreover, the map $A_{UV} : U \cap V \to GL(\mathbb{R}^k)$ is C^r.

Remark 5 *If the above conditions are met, then the space E can be parameterized to be a C^r manifold. The C^∞ structure imposed on E is such that for each $U \in \mathcal{U}$, the map $h_U : U \times \mathbb{R}^k \to \pi^{-1}(U), (x, \alpha_1, \dots, \alpha_k) \mapsto \sum_{i=1}^k \alpha_i e_{U,i}(x)$ is a C^∞ homeomorphism. See for example [34, Theorem 2.4].*

[10] Here we abused the notation by denoting μ the Haar measure both on $G_l(\mathbb{R}^n)$ and on $G_m(\mathbb{R}^n)$, since the two manifolds are isomorphic.

We now construct a particular vector bundle in which the base space is $G_m(\mathbb{R}^n)$. The total space is denoted as $E_m(\mathbb{R}^n)$ which consists of all pairs $(\nu, \mathbf{p}_1, \ldots, \mathbf{p}_m)$ where $\mathbf{p}_i \in \mathbb{R}^m,\ 1 \leq i \leq m$ are vectors in the plane ν. The projection π maps $(\nu, \mathbf{p}_1, \ldots, \mathbf{p}_m)$ into ν. The linear space structure on the fiber attached to ν is the linear structure of the linear transforms on ν. We can choose the open cover $\mathcal{U}$ as the family of $\{U_I\}_{|I|=m}$ in a similar manner as described in Section 2, Part C. For an $n \times m$ matrix $\mathbf{Y}$ whose columns vectors spans ν, we define the frame at ν by

$$e_{U_I}(\nu) = \mathbf{Y} \cdot (\mathbf{Y}_I)^{-1}. \tag{55}$$

It is obvious that this definition of the frame is independent on the specific choice of $\mathbf{Y}$, and it is routine to check the compatibility condition (54) for intersecting sets $U_I, U_{I'}$ and that $A(U_I, U_{I'})$ is C^∞. In this way we have constructed a vector bundle according to Definition 7.[11] We summarize as follows:

Proposition 3 $(E_m(\mathbb{R}^n), G_m(\mathbb{R}^n), \pi, \mathbb{R}^n)$ *is a vector bundle, thus in particular* $E_m(\mathbb{R}^n)$ *is an* mn*-dimensional* C^∞ *manifold.*

Recall that measure zero sets are well defined on C^∞ manifolds. From Remark 5 it's easy to see the following fact:

Proposition 4 *The natural projection* $\rho : E_m(\mathbb{R}^n) \to \mathbb{M}(m, n),\ (\nu, \mathbf{p}_1, \ldots, \mathbf{p}_m) \mapsto [\mathbf{p}_1^t, \ldots, \mathbf{p}_m^t]^t$ *is* C^∞*, therefore maps measure zero subsets of* $E_m(\mathbb{R}^n)$ *into measure zero subsets of* $\mathbf{M}(m, n)$*.*

Returning to the proof of Corollary 3, we first obtain that $\mu(H) = 0$, where $H \subseteq G_m(\mathbb{R}^n)$ is the set of orthogonal complements of the l-dimensional subspaces in $\Omega_J \setminus \Omega_J^r$, as before. Then for each U_I, $H \cap U_I$ is a measure zero subset of U_I. By the property of product measure we have that $(H \cap U_I) \times \mathbb{R}^{m^2}$ is a measure zero subset of $U_I \times \mathbb{R}^{m^2}$. Since h in Remark 5 is a C^∞ homeomorphism we obtain that $\pi^{-1}(H \cap U_I)$ is a measure zero subset of $\pi^{-1}(U_I)$. Finally $\pi^{-1}(H) = \bigcup_I \pi^{-1}(H \cap U_I)$ is a measure zero subset of $\pi^{-1}(G_m(\mathbb{R}^n)) = E_m(\mathbb{R}^n)$. This combined with Proposition 4 shows that $\mathcal{N}(\mathbf{A}) \in (\Omega_J \setminus \Omega_J^r))$ only if $\mathbf{A}$ falls into a Lebesgue measure zero set on $\mathcal{M}(m, n)$, which is of probability zero if the probability distribution of $\mathbf{A}$ is absolutely continuous (with respect to the Lebesgue measure).

Appendix F Proof of Theorem 4

Since the value of θ_J depends on the null space of the measurement matrix, in the following it is considered as a function of $\nu \in G_l(\mathbb{R}^n)$.

Lemma 7 *Let* $F \in \mathcal{M}, q \in (0, 1]$. *If* $\lim_{x \downarrow 0} F(x)/x^q$ *or* $\lim_{x \to \infty} F(x)/x^q$ *exists and is positive, then* $\theta_{l_q} \leq \theta_J$ *for any* $\nu \in G_l(\mathbb{R}^n)$*.*

[11] Alternatively, we can construct the same bundle as the Whitney sum of m *universal bundles* [23, Definition 2.9, Definition 2.26].

PROOF We only prove for the case where $\lim_{t\downarrow 0} F(t)/t^q$ exists and is positive, because the case where $\lim_{t\to\infty} F(t)/t^q$ exists and is positive is essentially similar. By definition we only have to prove the following for any $\mathbf{z} \in \nu \setminus \{\mathbf{0}\}$ and T satisfying $|T| \leq k$:

$$\frac{\|\mathbf{z}_T\|_q^q}{\|\mathbf{z}_{T^c}\|_q^q} \leq \theta_J. \tag{56}$$

Notice that for any $t \in \mathbb{R}$, vector $t\mathbf{z}$ still belongs to $\mathcal{N}(\mathbf{A})$, hence

$$\text{left side of(56)} = \lim_{t\downarrow 0} \frac{J(t\mathbf{z}_T)}{J(t\mathbf{z}_{T^c})} \tag{57}$$

$$\leq \theta_J, \tag{58}$$

where (57) is because $\lim_{t\downarrow 0} F(t)/t^q$ exists and is positive, and (58) is from the definition of supremum. ∎

Lemma 8 $\overline{\Omega_{l_q}} = \{\nu : \theta_{l_q} \leq 1\}$.

PROOF First we prove that $\overline{\Omega_{l_q}} \subseteq \{\nu : \theta_{l_q} \leq 1\}$. This is because Lemma 1 shows that $\{\nu : \theta_{l_q} \leq 1\}$ is closed. $\Omega_{l_q} \subseteq \{\nu : \theta_{l_q} \leq 1\}$. On the other hand, it is obvious that $\{\nu : \theta_{l_q} \leq 1\} \subseteq \overline{\Omega_{l_q}}$. The proof is complete. ∎

Lemma 9 *Given $\nu \in G_l(\mathbb{R}^n)$, if $\theta_{l_q} \leq \theta_J$, then $\Omega_J \subseteq \overline{\Omega_{l_q}}$.*

PROOF

$$\Omega_J \subseteq \{\nu : \theta_J \leq 1\} \subseteq \{\nu : \theta_{l_q} \leq 1\} = \overline{\Omega_{l_q}}. \tag{59}$$

∎

Theorem 4 then follows easily from the following lemma:

Lemma 10

$$\mu(\{\nu : \theta_{l_q} < 1\}) = \mu(\{\nu : \theta_{l_q} \leq 1\}). \tag{60}$$

PROOF Define

$$\theta_{l_q,T} := \sup_{\mathbf{z}\in\mathcal{N}(\mathbf{A})\setminus\{\mathbf{0}\}} \frac{\|\mathbf{z}_T\|^q}{\|\mathbf{z}_{T^c}\|^q}. \tag{61}$$

Obviously $\theta_{l_q} = \max_{|T|\leq k} \theta_{l_q,T}$, therefore we only have to prove that for any $|T| \leq k$ we have $\mu(\{\nu : \theta_{l_q,T} < 1\}) = \mu(\{\nu : \theta_{l_q,T} \leq 1\})$. Choose an arbitrary I which does not intersect with T (Since $k + l \leq n$, one can always find such I). Then we only have to prove:

$$\lambda(\{\mathbf{X}_I \in \mathbb{M}(m,l) : \theta_{l_q,T}(\phi_I^{-1}(\mathbf{X}_I)) = 1\}) = 0, \tag{62}$$

In the following we will write $\theta_{l_q,T}$ instead of $\theta_{l_q,T}(\phi_I^{-1}(\mathbf{X}_I))$ for simplicity. From the continuity of the measure, it then suffices to prove:

$$\lambda(M \cap \{\mathbf{X}_I \in \mathbb{M}(m,l) : \theta_{l_q,T} < 1\}) = \lambda(M \cap \{\mathbf{X}_I \in \mathbb{M}(m,l) : \theta_{l_q,T} \leq 1\}) \tag{63}$$

for any bounded set $M \subseteq \mathbb{M}(m,l)$.

From the dilation property of the Lebesgue measure we have

$$\lambda(M^{T,a} \cap \{\mathbf{X}_I \in \mathbb{M}(m,l) : \theta_{l_q,T} \leq a\}) = a^{lk} \lambda(M \cap \{\mathbf{X}_I \in \mathbb{M}(m,l) : \theta_{l_q,T} \leq 1\}), \tag{64}$$

where $M^{T,a}$ is the dilation of the set M in the rows associated with T by the factor of a. Hence

$$\begin{aligned}
\lambda(M \cap \{\mathbf{X}_I \in \mathbb{M}(m,l) : \theta_{l_q,T} < 1\}) =& \lambda(\cup_{l\geq 1}[M \cap \{\mathbf{X}_I \in \mathbb{M}(m,l) : \theta_{l_q,T} \leq 1 - 1/l\}]) \\
\geq& \lambda(\cup_{l\geq 1}[M^{T,1-1/l} \cap \{\mathbf{X}_I \in \mathbb{M}(m,l) : \theta_{l_q,T} \leq 1 - 1/l\}]) \\
=& \lim_{l\to\infty} \lambda(M^{T,1-1/l} \cap \{\mathbf{X}_I \in \mathbb{M}(m,l) : \theta_{l_q,T} \leq 1 - 1/l\}) \\
=& \lim_{l\to\infty} (1 - 1/l)^{kl} \lambda(M \cap \{\mathbf{X}_I \in \mathbb{M}(m,l) : \theta_{l_q,T} \leq 1\}) \\
=& \lambda(M \cap \{\mathbf{X}_I \in \mathbb{M}(m,l) : \theta_{l_q,T} \leq 1\}),
\end{aligned}$$

which proves the validity of (63). ■

References

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

[2] 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.

[3] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, 2007.

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

[5] S. Chen, D. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing*, vol. 20, pp. 33–61, August 1998.

[6] R. Chartrand and V. Staneva, "Restricted isometry properties and nonconvex compressive sensing," *Inverse Problems*, vol. 24, no. 3, pp. 20–35, 2008.

[7] S. Foucart and M. Lai, "Sparsest solutions of underdetermined linear systems via l_q-minimization for $0 < q \leq 1$," *Applied and Computational Harmonic Analysis*, vol. 26, no. 3, pp. 395–407, 2009.

[8] 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.

[9] S. Foucart and M. Lai, "Sparsest solutions of underdetermined linear systems via l_q-minimization for $0 < q \leq 1$," *Applied and Computational Harmonic Analysis.*

[10] 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.

[11] M. Hyder and K. Mahata, "An improved smoothed l_0 approximation algorithm for sparse representation," *IEEE Trans. Signal Processing*, vol. 58, no. 4, pp. 2194–2205, 2010.

[12] H. Mohimani, M. B. Zadeh, and C. Jutten, "A fast approach for overcomplete sparse decomposition based on smoothed l_0 norm," *IEEE Trans. Signal Processing*, vol. 57, no. 1, 2009.

[13] R. Chartrand and W. Yin, "Iteratively reweighted algorithms for compressive sensing," *ICASSP 2008*, pp. 3869–3872, April.

[14] 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.

[15] 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.

[16] X. Wang, Y. Gu, and L. Chen, "Proof of convergence and performance analysis for sparse recovery via zero-point attracting projection," *IEEE Trans. Signal Processing*, vol. 60, no. 8, pp. 4081–4093, 2012.

[17] L. Chen and Y. Gu, "Approximate projected generalized gradient methods with sparsity-inducing penalties," *Manuscript Submitted to IEEE Trans. Signal Processing.*

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

[19] S. Foucart, "Notes on compressed sensing," *available online*, 2009.

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

[21] R. Gribonval and M. Nielsen, *On the Strong Uniqueness of Highly Sparse Representations from Redundant Dictionaries*, vol. 3195 of *Lecture Notes in Computer Science*. Springer, 2004.

[22] W. M. Boothby, *An Introduction to Differentiable Manifolds and Riemannian Geometry.* Elsevier, 2 ed., 2003.

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

[24] P. Mattila, *Geometry of Sets and Measures in Euclidean Spaces.* Cambridge University Press, 1995.

[25] Y. Chikuse, *Statistics on Special Manifolds.* New York : Springer, c2003.

[26] E. Candès and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?," *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5406–5425, 2006.

[27] 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.

[28] A. Aldroubi, X. Chen, and A. M. Powell, "Perturbations of measurement matrices and dictionaries in compressed sensing," *Applied and Computational Harmonic Analysis*, vol. 33, no. 2, pp. 282 – 291, 2012.

[29] R. S. E. M. Stein, *Functional Analysis: Introduction to Further Topics in Analysis.* Princeton University Press, 2011.

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

[31] W. Rudin, *Real and Complex Analysis.* McGraw-Hill, 3 ed., 1987.

[32] J. R. Munkres, *Topology.* Upper Saddle River, NJ Prentice Hall, Inc. c2000, 2nd ed ed.

[33] E. M. Stein and R. Shakarchi, *Real Analysis: Measure Theory, Integration, and Hilbert Spaces.* Princeton University Press, 2005.

[34] Z. Zhang, *New Lectures on Differential Topology [M] (In Chinese).* Peking University Press, July 2002.