Compressed Sensing over the Grassmann Manifold: A Unified Analytical Framework

Weiyu Xu
Department of Electrical Engineering
California Institute of Technology
Pasadena CA 91125, USA
Email: weiyu@systems.caltech.edu

Babak Hassibi
Department of Electrical Engineering
California Institute of Technology
Pasadena CA 91125, USA
Email: hassibi@systems.caltech.edu

***Abstract*—It is well known that compressed sensing problems reduce to finding the sparse solutions for large under-determined systems of equations. Although finding the sparse solutions in general may be computationally difficult, starting with the seminal work of [2], it has been shown that linear programming techniques, obtained from an l_1-norm relaxation of the original non-convex problem, can provably find the unknown vector in certain instances. In particular, using a certain restricted isometry property, [2] shows that for measurement matrices chosen from a random Gaussian ensemble, l_1 optimization can find the correct solution with overwhelming probability even when the support size of the unknown vector is proportional to its dimension. The paper [1] uses results on neighborly polytopes from [6] to give a "sharp" bound on what this proportionality should be in the Gaussian case. In this paper we shall focus on finding sharp bounds on the recovery of "approximately sparse" signals (also possibly for noisy measurements). While the restricted isometry property can be used to study the recovery of approximately sparse signals (and also in the presence of noisy measurements), the obtained bounds can be quite loose. On the other hand, the neighborly polytopes technique which yields sharp bounds in the ideal case cannot be generalized to approximately sparse signals. In this paper, starting from a necessary and sufficient condition for achieving a certain signal recovery accuracy, using high-dimensional geometry, we give a unified *null-space Grassmannian angle*-based analytical framework for compressive sensing. This new framework gives sharp quantitative tradeoffs between the signal sparsity and the recovery robustness of the l_1 optimization for approximately sparse signals. As it will turn out, the neighborly polytopes result of [1] for ideally sparse signals can be viewed as a special case of ours. Our result concerns fundamental properties of linear subspaces and so may be of independent mathematical interest.**

Index Terms: compressed sensing, basis pursuit, l_1-optimization, k-balancedness, Grassmann manifold, Grassmann angle, high-dimensional integral geometry, geometric probability, convex polytopes, random linear subspaces, neighborly polytopes

I. Introduction

In this paper we are interested in compressed sensing problems. Namely, we would like to find $\mathbf{x}$ such that

$$A\mathbf{x} = \mathbf{y} \tag{1}$$

where A is an $m \times n$ measurement matrix and $\mathbf{y}$ is $m \times 1$ measurement vector. In usual compressed sensing context $\mathbf{x}$ is $n \times 1$ unknown k-sparse vector. This assumes that $\mathbf{x}$ has only k nonzero components. In this paper we will consider a more general version of the k-sparse vector $\mathbf{x}$. Namely, we will assume that k components of the vector $\mathbf{x}$ have large magnitude and that the vector comprised of the remaining $n-k$ components has l_1-norm less than δ. We will refer to this type of signals as approximately k-sparse signals, or for brevity only approximately sparse signals. More on similar type of problems the interested can find in [13].

In the rest of the paper we will further assume that the number of the measurements is $m = \alpha n$ and the number of the "large" components of $\mathbf{x}$ is $k = \zeta n$, where $0 < \zeta < 1$ and $0 < \alpha < 1$ are constants independent of n (clearly, $\alpha > \zeta$). This problem setup is more realistic of practical applications than the standard compressed sensing of ideally k-sparse signals (see, e.g., [14], [21] and the references therein).

A particular way of solving (1) which recently generated a large amount of research is called l_1-optimization (basis pursuit) [2]. It proposes solving the following problem

$$\begin{aligned} \min \quad & \|\mathbf{x}\|_1 \\ \text{subject to} \quad & A\mathbf{x} = \mathbf{y}. \end{aligned} \tag{2}$$

Quite remarkably in [2] the authors were able to show that if the number of the measurements is $m = \alpha n$ and if the matrix A satisfies a special property called the restricted isometry property (RIP), then any unknown vector $\mathbf{x}$ with no more than $k = \zeta n$ (where ζ is an absolute constant which is a function of α, but independent of n, and explicitly bounded in [2]) non-zero elements can be recovered by solving (2). As expected, this assumes that $\mathbf{y}$ was in fact generated by that $\mathbf{x}$ and given to us (more on the case when the available measurements are noisy versions of $\mathbf{y}$ can be found in e.g. [11], [12]).

As can be immediately seen, the previous result heavily relies on the assumption that the measurement matrix A satisfies the RIP condition. It turns out that for several specific classes of matrices, such as matrices with independent zero-mean Gaussian entries or independent Bernoulli entries, the RIP holds with overwhelming probability [2], [4], [5]. However, it should be noted that the RIP is only a sufficient condition for l_1-optimization to produce a solution of (1).

Instead of characterizing the $m \times n$ matrix A through the RIP condition, in [1], [3] the authors assume that A constitutes a k-neighborly poly-tope. It turns out (as shown in [1]) that this characterization of the matrix A is in fact a necessary and sufficient condition for (2) to produce the solution of (1). Furthermore, using the results of [6], it can be shown that if the matrix A has i.i.d. zero-mean Gaussian entries with overwhelming probability it also constitutes a k-neighborly poly-tope. The precise relation between m and k in order for this to happen is characterized in [1] as well. It should also be noted that for a given value m i.e. for a given value of the constant α, the value of the constant ζ is significantly better in [1], [3] than in [2]. Furthermore, the values of constants ζ obtained for different values of α in [1] approach the ones obtained by simulation as $n \longrightarrow \infty$.

However, all these mentioned results rely on the assumption that the unknown vector $\mathbf{x}$ has only k non-zero components. As mentioned earlier, in this paper we will be interested in the case of approximately k-sparse signals. Since in this case the unknown vector $\mathbf{x}$ in general has no zeros it is relatively easy to see that its exact recovery from a reduced number of measurements is not possible. Instead, we will prove that, if the unknown approximately k-sparse vector is $\mathbf{x}$ and $\hat{\mathbf{x}}$ is the solution of (2) then for any given constant $0 \leq \alpha \leq 1$ there will be a constant ζ such that

$$||\hat{\mathbf{x}} - \mathbf{x}||_1 \leq \frac{2(C+1)\delta}{C-1} \tag{3}$$

where $C > 1$ is a given constant determining how close in l_1 norm the recovered vector $\hat{\mathbf{x}}$ should be to the original approximately k-sparse vector $\mathbf{x}$. As expected, ζ will be a function of C and α. However, ζ will be an absolute constant independent of n. A similar problem was considered with different proof technique in [13] based on the restricted isometry property from [2], where no explicit values of ζ were given. Since the RIP condition is only a sufficient condition, it generally gives rather loose bounds on the explicit values of ζ even in the ideally sparse case. In this paper we will provide sharp bounds on the explicit values of the allowable constants ζ for the general cases $C \geq 1$ based on high-dimensional geometry. Certainly there were also discussions of compressive sensing under different definitions of non-ideally sparse signals in the literature, for example, [23] discussed compressive sensing for signals from a l_p ball with $0 < p \leq 1$ using sufficient conditions based on results of the Gel'fand n-widths. However, the results in this paper are dealing directly with approximately sparse signals defined in terms of the concentration of l_1 norm, and furthermore, we give a neat *necessary and sufficient* condition for l_1 optimization to work and we are also able to explicitly give much sharper compressive sensing performance bounds.

To prove the previous statements we will make use of an characterization that constitutes both necessary and sufficient conditions which the matrix A should satisfy in order that the solution of (2) approximates the original signal accurately enough such that (3) holds. This characterization will be equivalent to the neighborly polytope characterization from [1] in the "ideally sparse" case. Furthermore, as we will see later in the paper, in the perfectly sparse signal case (which allows $C \longrightarrow 1$), our result for allowable ζ will match the result of [1]. Our analysis will be directly based on the null-space Grassmannian angle result in high dimensional integral geometry, which gives a unified analytic framework for l_1 minimization.

II. The Null-space characterization

In this section we introduce a useful characterization of the matrix A. The characterization will establish a necessary and sufficient condition on the matrix A so that solution of (2) approximates the solution of (1) such that (3) holds. (See [8], [9], [10], [13], [24], [25], [27] etc. for variations of this result).[1]

Theorem 1: Assume that an $m \times n$ measurement matrix A is given. Further, assume that $\mathbf{y} = A\mathbf{x}$ and that $\mathbf{w}$ is an $n \times 1$ vector. Let K be any subset of $\{1, 2, \ldots, n\}$ such that $|K| = k$ and let K_i denote the i-th element of K. Further, let $\bar{K} = \{1, 2, \ldots, n\}/K$. Then the solution $\hat{\mathbf{x}}$ produced by (2) will satisfy $\|\mathbf{x} - \hat{\mathbf{x}}\|_1 \leq \frac{2(C+1)}{C-1}\|\mathbf{x}_{\bar{K}}\|_1$, with $C > 1$, if and only if

$$(\forall \mathbf{w} \in \mathbf{R}^n | A\mathbf{w} = 0) \quad \text{and} \quad \forall K, \quad C\sum_{i=1}^{k} |\mathbf{w}_{K_i}| \leq \sum_{i=1}^{n-k} |\mathbf{w}_{\bar{K}_i}|. \tag{4}$$

Proof: Sufficiency: Suppose the matrix A has the claimed null-space property. Now the solution $\hat{\mathbf{x}}$ of (2) satisfies $\|\hat{\mathbf{x}}\|_1 \leq \|\mathbf{x}\|_1$, where $\mathbf{x}$ is the original signal. Since $A\hat{\mathbf{x}} = \mathbf{y}$, it easily follows that $\mathbf{w} = \hat{\mathbf{x}} - \mathbf{x}$ is in the null space of A. Therefore we can further write $\|\mathbf{x}\|_1 \geq \|\mathbf{x} + \mathbf{w}\|_1$. Using the triangular inequality for the l_1 norm we obtain

$$\begin{aligned}
\|\mathbf{x}_K\|_1 + \|\mathbf{x}_{\bar{K}}\|_1 &= \|\mathbf{x}\|_1 \\
&\geq \|\hat{\mathbf{x}}\|_1 = \|\mathbf{x} + \mathbf{w}\|_1 \\
&\geq \|\mathbf{x}_K\|_1 - \|\mathbf{w}_K\|_1 + \|\mathbf{w}_{\bar{K}}\|_1 - \|\mathbf{x}_{\bar{K}}\|_1 \\
&\geq \|\mathbf{x}_K\|_1 + (C-1)\|\mathbf{w}_K\|_1 - \|\mathbf{x}_{\bar{K}}\|_1 \\
&\geq \|\mathbf{x}_K\|_1 - \|\mathbf{x}_{\bar{K}}\|_1 + \frac{C-1}{C+1}\|\mathbf{w}\|_1
\end{aligned}$$

where the last two inequalities are from the claimed null-space property. Relating the first equality and the last inequality above, we finally get $2\|\mathbf{x}_{\bar{K}}\|_1 \geq \frac{(C-1)}{C+1}\|\mathbf{w}\|_1$, as desired.

Necessity: Since every step in the proof of the sufficiency can be reversed if equality is achieved in the triangular equality, the condition $C\sum_{i=1}^{k} |\mathbf{w}_{K_i}| \leq \sum_{i=1}^{n-k} |\mathbf{w}_{\bar{K}_i}|$ is also a necessary condition for $\|\mathbf{x} - \hat{\mathbf{x}}\|_1 \leq \frac{2(C+1)}{C-1}\|\mathbf{x}_{\bar{K}}\|_1$ to hold for every $\mathbf{x}$. ■

It should be noted that if the condition (4) is satisfied then $2\|\mathbf{x}_{\bar{K}}\|_1 \geq \frac{(C-1)}{C+1}\|\mathbf{w}\|_1 = \frac{(C-1)}{C+1}\|\hat{\mathbf{x}} - \mathbf{x}\|_1$ for any K or $\bar{K}$. Hence it is also true for the set K which corresponds to the k largest components of the vector $\mathbf{x}$. In that case we can write $2\delta \geq \frac{(C-1)}{C+1}\|\hat{\mathbf{x}} - \mathbf{x}\|_1$ which exactly corresponds to (3). In fact,

[1] We should also mention that the result given below is a "strong result" (in the sense of [1]) since it guarantees recovery for *all* approximately k-sparse x. It is also possible to consider signal recovery in a weak sense (again, as done in [1]), where now recovery is guaranteed for *almost all* approximately k-sparse x. However, in the interest of space, we shall not do so here.

the condition (4) is also a sufficient and necessary condition for unique exact recovery of ideally k-sparse signals after we take $C=1$ and let (4) take strict inequality for all $\mathbf{w} \neq 0$ in the null space of $\mathcal{A}$. To see this, suppose the ideally k-sparse signal $\mathbf{x}$ is supported over the set K, namely, $\|\mathbf{x}_{\bar{K}}\|_1 = 0$. Then from the same triangular inequality derivation of Theorem 1, we know that $\|\hat{\mathbf{x}} - \mathbf{x}\|_1 = 0$, namely $\hat{\mathbf{x}} = \mathbf{x}$. Or we can just let C be arbitrarily close to 1 and since $\|\mathbf{x} - \hat{\mathbf{x}}\|_1 \leq \frac{2(C+1)}{C-1}\|\mathbf{x}_{\bar{K}}\|_1 = 0$, we also get $\hat{\mathbf{x}} = \mathbf{x}$. In this sense, when $C=1$, the null-space condition is equivalent to the neighborly polytope condition[1] for unique exact recovery of ideally sparse signals.

Remark: Clearly, we need not to check (4) for all subsets K; checking the subset with the k largest (in absolute value) elements of $\mathbf{w}$ is sufficient. However, Theorem 1 will be more convenient for our subsequent analysis.

In the following section, for a given value $\alpha = \frac{m}{n}$ and any value $C \geq 1$, we will determine the value of feasible $\zeta = \frac{k}{n}$ for which there exists a sequence of A such that (4) is satisfied when n goes to infinity and $\frac{m}{n} = \alpha$. It turns out that for a specific A, it is very hard to check whether the condition (4) is satisfied or not. Instead, we consider randomly choosing A from a certain distribution, and analyze for what ζ, the condition (4) for its null-space is satisfied with overwhelming probability as n goes to infinity.

The standard results on compressed sensing assume that the matrix A has i.i.d. $\mathcal{N}(0,1)$ entries. In this case, the following lemma gives a characterization of the resulting null-space.

Lemma 1: Let $A \in R^{m \times n}$ be a random matrix with i.i.d. $\mathcal{N}(0,1)$ entries. Then the following statements hold:

- The distribution of A is left-rotationally invariant, $P_A(A) = P_A(A\Theta)$, $\Theta\Theta^* = \Theta^*\Theta = I$
- The distribution of Z, any basis of the null-space of A is right-rotationally invariant. $P_Z(Z) = P_Z(\Theta^* Z)$, $\Theta\Theta^* = \Theta^*\Theta = I$
- It is always possible to choose a basis for the null-space such that $Z \in R^{n \times (n-m)}$ has i.i.d. $\mathcal{N}(0,1)$ entries.

In view of Theorem 1 and Lemma 1 what matters is that the null-space of A be rotationally invariantly. Sampling from this rotationally invariant distribution is equivalent to uniformly sampling a random $(n-m)$-dimensional subspace from the Grassmann manifold $\text{Gr}_{(n-m)}(n)$. Here the Grassmannian manifold $\text{Gr}_{(n-m)}(n)$ is the set of $(n-m)$-dimensional subspaces in the n-dimensional Euclidean space R^n [7]. For any such A and ideally sparse signals, the sharp bounds of [1], for example, apply. In this paper, we shall use the unified Grassmannian angle framework to analyze the null-space property with applications to compressive sensing for approximately sparse signals.

III. The Grassmannian Angle Framework for the Null-space Characterization

In this section we give the Grassmannian angle-based framework for analyzing the bounds on $\zeta = \frac{k}{n}$ such that the condition (4) holds for the null-space of the measurement matrix A. From the definition of the condition (4), there is a tradeoff between the sparsity level k and the parameter C, which in turn is related to the allowable signal recovery imperfection. Before proceeding further, let us make clear the problem that we are trying to solve: Let Z be the null-space of the randomly sampled measurement matrix A. Given a certain constant $C > 1$ (or $C \geq 1$), which corresponds to a certain level of recovery accuracy for the approximately sparse signals, we are interested in how large the sparsity level k can be while satisfying the following condition on Z (here $|K| = k$)

$$\forall K \subset \{1,2,...,n\}, C\|\mathbf{w}_K\|_1 \leq \|\mathbf{w}_{\bar{K}}\|_1, \forall \mathbf{w} \in Z, \quad (5)$$

with overwhelming probability.

We note that for a certain subset $K \subset \{1,2,...,n\}$ with $|K| = k$, the event that the null-space Z satisfies

$$C\|\mathbf{w}_K\|_1 \leq \|\mathbf{w}_{\bar{K}}\|_1, \forall \mathbf{w} \in Z \quad (6)$$

is equivalent to the event that $\forall \mathbf{x}$ supported on the k-set K (or supported on a subset of K):

$$\|\mathbf{x}_K + \mathbf{w}_K\|_1 + \|\frac{\mathbf{w}_{\bar{K}}}{C}\|_1 \geq \|\mathbf{x}_K\|_1, \forall \mathbf{w} \in Z \quad (7)$$

This claim can be derived using similar arguments as in the proof of Theorem 1, based on triangular inequalities for the l_1 norm. We will omit the detailed proof here.

Then the event that the condition (5) on the null-space Z holds if and only if $\forall K \subset \{1,2,...,n\}$ with $|K| = k$, and $\forall \mathbf{x}$ supported on the set K (or on a subset of K),

$$\|\mathbf{x}_K + \mathbf{w}_K\|_1 + \|\frac{\mathbf{w}_{\bar{K}}}{C}\|_1 \geq \|\mathbf{x}_K\|_1, \forall \mathbf{w} \in Z \quad (8)$$

We are now in a position to derive the probability that condition (5) holds for the sparsity $|K| = k$ if we uniformly sample a random $(n-m)$-dimensional subspace Z from the Grassmann manifold $\text{Gr}_{(n-m)}(n)$. From the previous discussions, we can equivalently consider its complementary probability P, namely the probability there exists a subset $K \subset \{1,2,...,n\}$ with $|K| = k$, and a vector $\mathbf{x} \in R^n$ supported on the set K (or a subset of K) failing the condition (7). Due to the vector proportionality in the linear subspace Z, we can restrict our attention to those vectors $\mathbf{x}$ from the cross-polytope $\{\mathbf{x} \in R^n \mid \|\mathbf{x}\|_1 = 1\}$ that are only supported on the set K (or a subset of K).

First, we upper bound the probability P by a union bound over all the possible support sets $K \subset \{1,2,...,n\}$ and all the sign patterns of the k-sparse vector $\mathbf{x}$. Since the k-sparse vector $\mathbf{x}$ has $\binom{n}{k}$ possible support sets and 2^k possible sign patterns (nonnegative or non-positive), we have

$$P \leq \binom{n}{k} \times 2^k \times P_{K,-} \quad (9)$$

,where $P_{K,-}$ is the probability that for a specific *support set*, there exist a k-sparse vector $\mathbf{x}$ of a specific *sign pattern* which fails the condition (7). By symmetry, without loss of generality, we assume the signs of the elements of $\mathbf{x}$ to be non-positive.

Now we can focus on deriving the probability $P_{K,-}$. Since $\mathbf{x}$ is a non-positive k-sparse vector supported on the set K (or a subset of K) and can be restricted to the crosspolytope

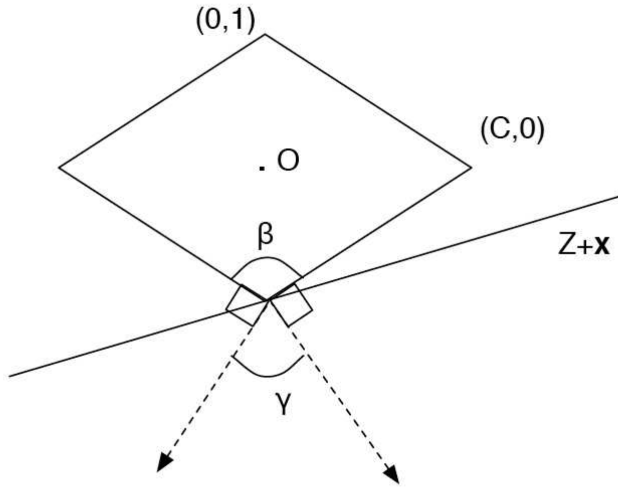

Fig. 1. The Grassmannian Angle for a Skewed Cross-polytope

$\{\mathbf{x} \in R^n \mid \|\mathbf{x}\|_1 = 1\}$, $\mathbf{x}$ is also on a $(k-1)$-dimensional face, denoted by F, of the skewed cross-polytope SP:

$$\text{SP} = \{\mathbf{y} \in R^n \mid \|\mathbf{y}_K\|_1 + \|\frac{\mathbf{y}_{\bar{K}}}{C}\|_1 \leq 1\} \tag{10}$$

Now the probability $P_{K,-}$ is the probability that there exists an $\mathbf{x} \in F$, and there exists a $\mathbf{w} \in Z$ ($\mathbf{w} \neq 0$) such that

$$\|\mathbf{x}_K + \mathbf{w}_K\|_1 + \|\frac{\mathbf{w}_{\bar{K}}}{C}\|_1 \leq \|\mathbf{x}_K\|_1 = 1. \tag{11}$$

We start by studying the case for a specific point $\mathbf{x} \in F$ and, without loss of generality, we assume $\mathbf{x}$ is in the relative interior of this $(k-1)$ dimensional face F. For this particular $\mathbf{x}$ on F, the probability, denoted by $P'_{\mathbf{x}}$, that $\exists \mathbf{w} \in Z$ ($\mathbf{w} \neq 0$) such that

$$\|\mathbf{x}_K + \mathbf{w}_K\|_1 + \|\frac{\mathbf{w}_{\bar{K}}}{C}\|_1 \leq \|\mathbf{x}_K\|_1 = 1, \tag{12}$$

is essentially the probability that a uniformly chosen $(n-m)$ dimensional subspace Z shifted by the point $\mathbf{x}$, namely $(Z + \mathbf{x})$, intersects the skewed crosspolytope

$$\text{SP} = \{\mathbf{y} \in R^n \mid \|\mathbf{y}_K\|_1 + \|\frac{\mathbf{y}_{\bar{K}}}{C}\|_1 \leq 1\} \tag{13}$$

non-trivially, namely, at some other point besides $\mathbf{x}$.

From the linear property of the subspace Z, the event that $(Z + \mathbf{x})$ intersects the skewed crosspolytope SP is equivalent to the event that Z intersects nontrivially with the cone SP-Cone($\mathbf{x}$) obtained by observing the skewed polytope SP from the point $\mathbf{x}$. (Namely, SP-Cone($\mathbf{x}$) is conic hull of the point set $(\text{SP} - \mathbf{x})$ and of course SP-Cone($\mathbf{x}$) has the origin of the coordinate system as its apex.) However, as noticed in the geometry for convex polytopes [15][16], the SP-Cone($\mathbf{x}$) are identical for any $\mathbf{x}$ lying in the relative interior of the face F. This means that the probability $P_{K,-}$ is equal to $P'_{\mathbf{x}}$, regardless of the fact $\mathbf{x}$ is only a single point in the relative interior of the face F. (The acute reader may have noticed some singularities here because $\mathbf{x} \in F$ may not be in the relative interior of F, but it turns out that the SP-Cone($\mathbf{x}$) in this case is only a subset of the cone we get when $\mathbf{x}$ is in the relative interior of F. So we do not lose anything if we restrict $\mathbf{x}$ to be in the relative interior of the face F).

Since $P_{K,-} = P'_{\mathbf{x}}$, we only need to determine $P'_{\mathbf{x}}$. From its definition, $P'_{\mathbf{x}}$ is exactly the ***complementary Grassmann angle*** [15] for the face F with respect to the polytope SP under the Grassmann manifold $\text{Gr}_{(n-m)}(n)$:[2] the probability of a uniformly distributed $(n-m)$-dimensional subspace Z from the Grassmannian manifold $\text{Gr}_{(n-m)}(n)$ intersecting non-trivially with the cone SP-Cone($\mathbf{x}$) formed by observing the skewed cross-polytope SP from the relative interior point $\mathbf{x} \in F$.

Built on the works by L.A.Santalö[19] and P.McMullen[20] etc. in high dimensional integral geometry and convex polytopes, the complementary Grassmann angle for the $(k-1)$-dimensional face F can be explicitly expressed as the sum of products of internal angles and external angles[16]:

$$2 \times \sum_{s \geq 0} \sum_{G \in \Im_{m+1+2s}(\text{SP})} \beta(F, G)\gamma(G, \text{SP}), \tag{14}$$

where s is any nonnegative integer, G is any $(m+1+2s)$-dimensional face of the skewed crosspolytope ($\Im_{m+1+2s}(\text{SP})$ is the set of all such faces), $\beta(.,.)$ stands for the internal angle and $\gamma(.,.)$ stands for the external angle. According to [16][20], the internal angles and external angles are basically defined as:

- An internal angle $\beta(F_1, F_2)$ is the fraction of the hypersphere S covered by the cone obtained by observing the face F_2 from the face F_1. [3] The internal angle $\beta(F_1, F_2)$ is defined to be zero when $F_1 \nsubseteq F_2$ and is defined to be one if $F_1 = F_2$.
- An external angle $\gamma(F_3, F_4)$ is the fraction of the hypersphere S covered by the cone of outward normals to the hyperplanes supporting the face F_4 at the face F_3. The external angle $\gamma(F_3, F_4)$ is defined to be zero when $F_3 \nsubseteq F_4$ and is defined to be one if $F_3 = F_4$.

Let us take for example the 2-dimensional skewed cross-polytope $\text{SP} = \{(y_1, y_2) \in R^2 \mid \|\mathbf{y}_2\|_1 + \|\frac{\mathbf{y}_1}{C}\|_1 \leq 1\}$ (namely the diamond) in Figure 1, where n=2, $(n-m) = 1$ and $k = 1$. Then the point $\mathbf{x} = (0, -1)$ is a 0-dimensional face (namely a vertex) of the skewed polytope SP. Now from their definitions, the internal angle $\beta(\mathbf{x}, \text{SP}) = \beta$ and the external angle $\gamma(\mathbf{x}, \text{SP}) = \gamma$, $\gamma(\text{SP}, \text{SP}) = 1$. The complementary Grassmann angle for the vertex $\mathbf{x}$ with respect to the polytope SP is the probability that a uniformly sampled 1-dimensional subspace (namely a line, we denote it by Z) shifted by $\mathbf{x}$ intersects non-trivially with $\text{SP} = \{(y_1, y_2) \in R^2 \mid \|\mathbf{y}_2\|_1 + \|\frac{\mathbf{y}_1}{C}\|_1 \leq 1\}$

[2] An Grassman angle and its corresponding complementary Grassmann angle always sum up to 1. There is apparently inconsistency in terms of the definition of which is “Grassmann angle” and which is “complementary Grassmann angle” between [15],[17] and [6] etc. But we will stick to the earliest definition in [15] for Grassmann angle: the measure of the subspaces that intersect trivially with a cone.

[3] Note the dimension of the hypersphere S here matches the dimension of the corresponding cone discussed. Also, the center of the hypersphere is the apex of the corresponding cone. All these defaults also apply to the definition of the external angles.

(or equivalently the probability that Z intersects non-trivially with the cone obtained by observing SP from the point $\mathbf{x}$). It is obvious that this probability is 2β. The readers can also verify the correctness of the formula (14) very easily for this toy example.

For a general polytope, it might be hard to give explicit formula for the external and internal angles. Fortunately in the skewed cross-polytope case, both the internal angles and the external angles can be explicitly computed.

First, let us look at the internal angle $\beta(F, G)$ between the $(k-1)$-dimensional face F and a $(l-1)$-dimensional face G. As we can see, the cone formed by observing G from F is a convex polyhedral cone formed by $(l-k)$ unit vectors with inner product $\frac{1}{1+C^2k}$ between each other. In this case, the internal angle is given by [6], [18]

$$\beta(F, G) = \frac{V_{l-k-1}(\frac{1}{1+C^2k}, l-k-1)}{V_{l-k-1}(S^{l-k-1})}, \quad (15)$$

where $V_i(S^i)$ denotes the i-th dimensional surface measure on the unit sphere S^i, while $V_i(\alpha', i)$ denotes the surface measure for regular spherical simplex with $(i+1)$ vertices on the unit sphere S^i and with inner product as α' between these $(i+1)$ vertices. Thus the equation (15) is given by $B(\frac{1}{1+C^2k}, l-k)$, where

$$B(\alpha', m') = \theta^{\frac{m'-1}{2}} \sqrt{(m'-1)\alpha'+1}\pi^{-m'/2}\alpha'^{-1/2} J(m', \theta) \quad (16)$$

with $\theta = (1-\alpha')/\alpha'$ and

$$J(m', \theta) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} (\int_{0}^{\infty} e^{-\theta v^2 + 2iv\lambda} \, dv)^{m'} e^{-\lambda^2} \, d\lambda \quad (17)$$

Also, we can derivation the external angle $\gamma(G, \text{SP})$ between the $(l-1)$-dimensional face G and the skewed cross-polytope SP as:

$$\gamma(G, SP) = \frac{2^{n-l}}{\sqrt{\pi}^{n-l+1}} \int_{0}^{\infty} e^{-x^2} (\int_{0}^{\frac{x}{C\sqrt{k+\frac{l-k}{C^2}}}} e^{-y^2} \, dy)^{n-l} \, dx. \quad (18)$$

These derivations involve the computations of the volumes of cones in high dimensional geometry. Due to space limitations, we have omitted the details of these calculations. The interested reader is referred to the more extensive version of this paper under preparation [26].

IV. Numerical Computations on the Bounds of ζ

In summary, we finally get the probability

$$P \le \binom{n}{k} \times 2^k \times 2 \times \sum_{s \ge 0} \sum_{G \in \Im_{m+1+2s}(\text{SP})} \beta(F, G)\gamma(G, \text{SP}), \quad (19)$$

Recall that we assume $\frac{m}{n} = \alpha$. In order to see how large k can be to make P overwhelmingly converge to zero as $n \to \infty$, we need to analyze the decaying exponents of the right handside of (19). First we define $l = (m+1+2s)+1$ and $\mu = \frac{l}{n}$. In the skewed crosspolytope SP , we notice that there are in total $\binom{n-k}{l-k}2^{l-k}$ faces G of dimension $(l-1)$ such that $F \subset G$ and $\beta(F, G) \neq 0$. Because of the symmetry in the skewed crosspolytope SP, it follows from (19) that

$$P \le \binom{n}{k} 2^k \times 2 \sum_{s \ge 0} \binom{n-k}{l-k} 2^{l-k} \beta(F, G)\gamma(G, \text{SP}), \quad (20)$$

where $l = (m+1+2s)+1$ and G is any single face of dimension $(l-1)$ such that $F \subset G$.

Since we are considering the case $n \to \infty$, in order to let P go to zero, one sufficient condition is that over $\alpha \le \mu = \frac{l}{n} \le 1$, the exponent for the combinatorial factors

$$\psi_{com} = \lim_{n \to \infty} \frac{\log\left(\binom{n}{k}2^k 2\binom{n-k}{l-k}2^{l-k}\right)}{n} \quad (21)$$

and the negative exponent for the angle factors

$$\psi_{angle} = -\lim_{n \to \infty} \frac{\log\left(\beta(F, G)\gamma(G, \text{SP})\right)}{n} \quad (22)$$

satisfy $\psi_{com} - \psi_{angle} < 0$ for every μ.

From this observation and formula (20), we can actually show that for any $C \ge 1$ and any $\alpha > 0$, there always exists a $\zeta > 0$ such that the probability P goes to zero exponentially as $n \to \infty$. Due to space limitations, in this paper, we will focus on presenting the achievable numerical results. By analyzing the decaying exponents of the external angles and internal angles through the Laplace method, we can give the following numerical results shown in Figure 2 and Figure 3. Following [1] we take $m = 0.5555n$. Then k/n, the largest sparsity level $\zeta = \frac{k}{n}$ which makes the computed failure probability of the event (7) approach zero asymptotically as n goes large, is shown in Figure 2. As we can see, when $C = 1$, we get the same bound of $\zeta = 0.095 \times 0.5555 = 0.0528$ as obtained in the ideally sparse case in [1]. As expected, as the constant C grows, the l_1 minimization achieves higher signal recovery accuracy, and then also requires a smaller sparsity level ζ.

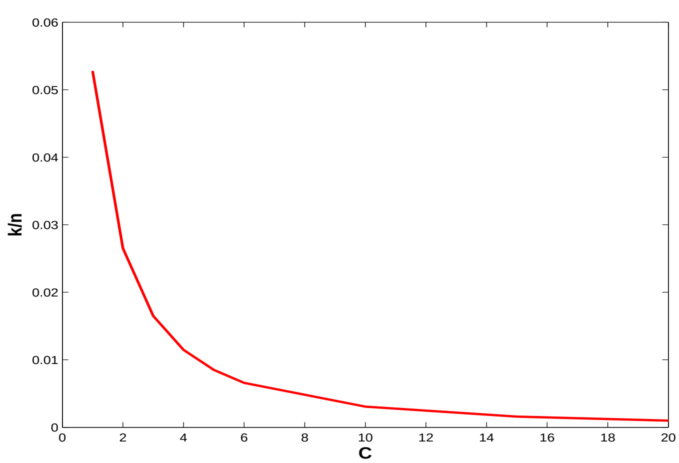

Fig. 2. Allowable sparsity as a function of C (allowable imperfection of the recovered signal is $\frac{2(C+1)\delta}{C-1}$)

In Figure 3 we show the exponents ψ_{com} and ψ_{angle} for $C = 2$, $\alpha = 0.5555$ and $\zeta = 0.0265$. The red solid curve denotes

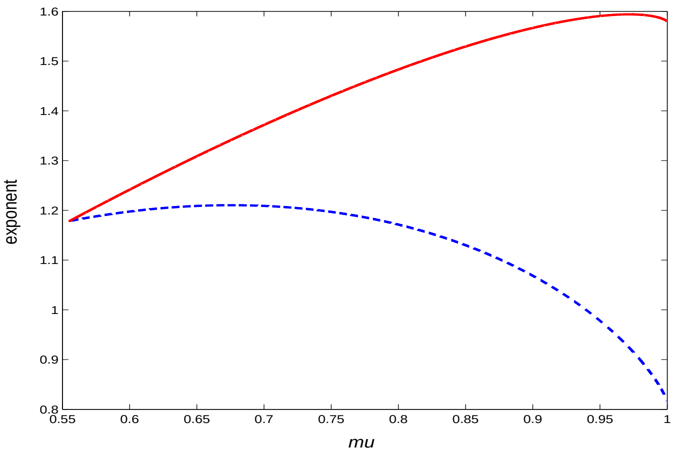

Fig. 3. The Combinatorial and Angle Exponents

ψ_{angle} and the blue dashed line denotes ψ_{com}. It just satisfies $\psi_{com} - \psi_{angle} < 0$ over $\alpha \leq \mu \leq 1$. Indeed, $\zeta = 0.0265$ is the bound shown in Figure 2.

V. Conclusion

It is well known that l_1 optimization can be used to recover ideally sparse and approximately sparse signals in compressed sensing, if the underlying signal is sparse enough. While in the ideally sparse case the recent results of [1] have given us sharp bounds on how sparse the signal can be, sharp bounds for the recovery of approximately sparse signals were not available.

In this paper we developed and analyzed a null-space characterization of the necessary and sufficient conditions for the success of l_1-norm optimization in compressed sensing of the approximately sparse signals. Using high-dimensional geometry, we give a unified *null-space Grassmannian angle*-based analytical framework for compressive sensing. This new framework gives sharp quantitative tradeoffs between the signal sparsity and the recovery robustness of the l_1 optimization for approximately sparse signals. It can therefore be of practical use in applications where the underlying signal is not ideally sparse and where we are interested in the quality of the recovered signal. As expected, the neighborly polytopes result of [1] for ideally sparse signals follows as a special case of our results. This work investigates the fundamental "balancedness" property of linear subspaces, and may be of independent mathematical interest. It is also possible to generalize these results to the analysis of compressive sensing noisy measurements in future work.

Acknowledgment

This work was supported in part by the National Science Foundation under grant no. CCF-0729203, by the David and Lucille Packard Foundation, and by Caltech's Lee Center for Advanced Networking.

References

[1] David Donoho, "High-Dimensional Centrally Symmetric Polytopes with Neighborliness Proportional to Dimension ", *Discrete and Computational Geometry* , 102(27), pp. 617-652, 2006, Springer .

[2] Emmanuel Candes and Terence Tao, "Decoding by linear programming", *IEEE Trans. on Information Theory*, 51(12), pp. 4203 - 4215, December 2005.

[3] David Donoho and Jared Tanner, "Neighborliness of randomly-projected simplices in high dimensions", *Proc. National Academy of Sciences*, 102(27), pp. 9452-9457, 2005.

[4] Richard Baraniuk, Mark Davenport, Ronald DeVore, and Michael Wakin, "A simple proof of the restricted isometry property for random matrices", *To appear in Constructive Approximation*, available online at http://www.dsp.ece.rice.edu/cs/.

[5] Mark Rudelson and Roman Vershynin, "Geometric approach to error correcting codes and reconstruction of signals", *International Mathematical Research Notices*, 64, pp. 4019 - 4041, 2005.

[6] A.M. Vershik and P.V. Sporyshev, "Asymptotic Behavior of the Number of Faces of Random Polyhedra and the Neighborliness Problem", *Selecta Mathematica Sovietica* vol. 11, No. 2 (1992).

[7] W. M. Boothby, An Introduction to Differential Manifolds and Riemannian Geometry, 2nd ed. San Diego, CA: Academic, 1986.

[8] A. Feuer and A. Nemirovski, "On sparse representation in pairs of bases", *IEEE Transactions on Information Theory* 49(6): 1579-1581 (2003).

[9] N. Linial and I. Novik, "How neighborly can a centrally symmetric polytope be?", *Discrete and Computational Geometry*, 36(2006) 273-281.

[10] Y. Zhang, "When is missing data recoverable", available online at http://www.dsp.ece.rice.edu/cs/.

[11] J. Haupt and R. Nowak, "Signal reconstruction from noisy random projections", *IEEE Trans. on Information Theory*, 52(9), pp. 4036-4048, September 2006.

[12] M. J. Wainwright, "Sharp thresholds for high-dimensional and noisy recovery of sparsity", *Proc. Allerton Conference on Communication, Control, and Computing*, Monticello, IL, September 2006.

[13] A. Cohen, W. Dahmen, and R. DeVore, "Compressed sensing and best k-term approximation". (Preprint, 2006), available online at http://www.dsp.ece.rice.edu/cs/.

[14] J. Tropp, M. Wakin, M. Duarte, D. Baron, and R. Baraniuk, "Random filters for compressive sampling and reconstruction". *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, Toulouse, France, May 2006.

[15] Branko Grünbaum, Grassmann angles of convex polytopes. *Acta Math.*, 121:pp.293-302, 1968.

[16] Branko Grünbaum, *Convex polytopes*, volume 221 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 2003. Prepared and with a preface by Volker Kaibel, Victor Klee and GnterüM. Ziegler.

[17] Fernando Affentranger and Rolf Schneider. "Random projections of regular simplices". *Discrete Comput. Geom.*, 7(3):pp.219-226, 1992.

[18] Károly Böröczky, Jr. and Martin Henk."Random projections of regular polytopes". *Arch. Math. (Basel)*, 73(6):pp.465-473, 1999.

[19] L.A.Santaló, Geometría integral en espacios de curvatura constante, Rep.Argetina Publ.Com.Nac.Energí Atómica, Ser.Mat 1,No.1(1952)

[20] Peter McMullen. "Non-linear angle-sum relations for polyhedral cones and polytopes". *Math. Proc. Cambridge Philos. Soc.*, 78(2):pp.247-261, 1975.

[21] Emmanuel Candes, "Compressive sampling", *Proc. International Congress of Mathematics*, 3, pp. 1433-1452, Madrid, Spain, 2006.

[22] Emmanuel Cands, Justin Romberg, and Terence Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information." , *IEEE Trans. on Information Theory*, 52(2) pp. 489-509, February 2006.

[23] David Donoho, "Compressed sensing", *IEEE Trans. on Information Theory*, 52(4), pp. 1289-1306, April 2006

[24] M. Stojnic, W. Xu, and B. Hassibi, "Compressed sensing - probabilistic analysis of a null-space characterization", *IEEE International Conference on Acoustics, Speech and Signal Processing*, Pages:3377-3380, March 31 2008-April 4 2008.

[25] M. Stojnic, W. Xu, and B. Hassibi, "Compressed sensing of approximately sparse signals", *IEEE International Symposium on Information Theory*, July 2008.

[26] W. Xu, and B. Hassibi, "Compressed sensing of approximately sparse signals - A Grassman angle approach", *under preparation*.

[27] Boris S. Kashin and Vladimir N. Temlyakov, "A remark on compressed sensing," *Mathematical Notes*, 82(5-6), pp. 748-755, Nov. 2007.