STRONG THRESHOLDS FOR ℓ_2/ℓ_1-OPTIMIZATION IN BLOCK-SPARSE COMPRESSED SENSING

Mihailo Stojnic

Purdue University, West Lafayette, IN
e-mail: mstojnic@purdue.edu

ABSTRACT

It has been known for a while that l_1-norm relaxation can in certain cases solve an under-determined system of linear equations. Recently, [5, 10] proved (in a large dimensional and statistical context) that if the number of equations (measurements in the compressed sensing terminology) in the system is proportional to the length of the unknown vector then there is a sparsity (number of non-zero elements of the unknown vector) also proportional to the length of the unknown vector such that l_1-norm relaxation succeeds in solving the system. In this paper we determine sharp lower bounds on the values of allowable sparsity for any given number (proportional to the length of the unknown vector) of equations for the case of the so-called block-sparse unknown vectors considered in [25].

Index Terms: compressed sensing, l_1-optimization, block-sparse

1. INTRODUCTION

Standard compressed sensing assumes solving an under-determined system of equation (more on compressed sensing the interested reader can find in [13, 5, 23] and references therein)

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

where A is an $M \times N$ measurement matrix, $\mathbf{y}$ is an $M \times 1$ measurement vector, and $\mathbf{x}$ is an $N \times 1$ unknown K-sparse vector (K-sparse will mean that the number of non-zero entries of $\mathbf{x}$ is not greater than K; more on the recovery of the so-called K-approximately sparse unknown vectors can be found in [8], [26]). A particular way of solving (1) which will be the subject of this paper is l_1-norm relaxation [5]. (More on different algorithms the interested reader can find in excellent references [1, 4, 20, 18, 19].) l_1-norm relaxation proposes solving the following problem

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

In a series of works [5, 10, 24, 26] the authors were able to show that if the elements of the matrix A are drawn according to certain probability distributions and if $M = \alpha N$ (where α is a constant) then there is a constant $\beta = \frac{K}{N}$ such that the solutions of (1) and (2) coincide. The best known values of the constant (threshold) β for the l_1-norm relaxation are those obtained in [10] ([11] in the context of the so-called signed vectors $\mathbf{x}$).

What we described above is the standard compressed sensing setup. In this paper we will be interested in the so-called block-sparse compressed sensing problems [25, 20, 14, 2] (more on a related group of problems and various application the interested reader can find in [28, 3, 6, 9, 27, 29] and references therein). To introduce block-sparse signals and facilitate the subsequent exposition we will assume that integers N and d are chosen such that $n = \frac{N}{d}$ is an integer and it represents the total number of blocks that $\mathbf{x}$ consists of. Clearly d is the length of each block. Furthermore, we will assume that $m = \frac{M}{d}$ is an integer as well and that $X_i = \mathbf{x}_{(i-1)d+1:id}, 1 \leq i \leq n$ are the n blocks of $\mathbf{x}$. Then we will call any signal $\mathbf{x}$ k-block-sparse if its at most $k = \frac{K}{d}$ blocks X_i are non-zero. Since k-block-sparse signals are K-sparse one could then use (2) to recover the solution of (1). While this is possible, it clearly uses the block structure of $\mathbf{x}$ in no way. To exploit the block structure of $\mathbf{x}$ in [25] the following combination of l_2 and l_1 optimizations was proposed

$$\begin{array}{ll} \min & \sum_{i=1}^{n} \|\mathbf{x}_{(i-1)d+1:id}\|_2 \\ \text{subject to} & A\mathbf{x} = \mathbf{y}. \end{array} \tag{3}$$

Extensive simulations in [25] demonstrated that as d grows the algorithm in (3) significantly outperforms standard l_1. The following was shown in [25] as well: let A be an $M \times N$ matrix with a basis of null-space comprised of i.i.d. Gaussian elements; if $\alpha = \frac{M}{N} \to 1$ then there is a constant d such that all k-block-sparse signals $\mathbf{x}$ with sparsity $K \leq \beta N, \beta \to \frac{1}{2}$ can be recovered with overwhelming probability by solving (3). The precise relation between d and how fast $\alpha \longrightarrow 1$ and $\beta \longrightarrow \frac{1}{2}$ was quantified in [25] as well. In the present paper we provide results of a similar flavor for the entire range of α, i.e. for $0 \leq \alpha \leq 1$. More precisely, for any given constant $0 \leq \alpha \leq 1$ we will determine a constant $\beta = \frac{K}{N}$ and a constant d such that (3) recovers any k-block-sparse signal with sparsity less then K. Our analysis will be a combination of results from [25], [24], and [16]. In the following section we briefly recall on a key ingredient of the analysis from [25] that we will reuse in this paper.

2. NULL-SPACE CHARACTERIZATION FOR BLOCK-SPARSE SIGNALS

In this section we introduce a necessary and sufficient condition on the measurement matrix A so that the solutions of (2) and (3) coincide for all k-block-sparse $\mathbf{x}$. (see [25, 12, 17, 15, 30, 26] for variations of this result). Throughout the paper we set $\mathcal{K}$ to be the set of all subsets of size k of $\{1, 2, \ldots, n\}$; also if $K \subset \mathcal{K}$ then $\bar{K} = \{1, 2, \ldots, n\} \setminus K$.

Theorem 1 *([25]) Assume that A is a $dm \times dn$ measurement matrix, $\mathbf{y} = A\mathbf{x}$ and $\mathbf{x}$ is k-block-sparse. Then the solutions of (3) and (1) coinside if and only if for all nonzero $\mathbf{w} \in R^{dn}$ where $A\mathbf{w} = 0$ and all $K \in \mathcal{K}$*

$$\sum_{i\in K}||\mathbf{W}_i||_2 < \sum_{i\in \bar{K}}||\mathbf{W}_i||_2 \quad (4)$$

where $\mathbf{W}_i=(\mathbf{w}_{(i-1)d+1},\mathbf{w}_{(i-1)d+2},\dots,\mathbf{w}_{id})$, $i=1,2,\dots,n$.

3. PROBABILISTIC ANALYSIS OF THE NULL-SPACE CHARACTERIZATION

In this section we probabilistically analyze the validity of (4). We will assume that the matrix A has a basis of the null-space distributed uniformly in the Grassmanian. Under this assumption for any constant $\alpha=\frac{m}{n}$ $(0\leq\alpha\leq 1)$ we will determine a constant $\beta=\frac{k}{m}$ such that (4) holds with overwhelming probability. In the standard case $d=1$ (when there is no effect of block-sparsity and optimizations (2) and (3) are equivalent) results of this type were already computed in [5, 10, 26, 24]. Throughout the analysis we will make use of a few crucial observations of [24]. The first one is in fact a brilliant result of Gordon related to Grassmann manifolds.

Theorem 2 *([16]) Let S be a subset of the unit Euclidean sphere S^{dn-1} in R^{dn}. Let Y be a random $(dn-dk)$-dimensional subspace of R^{dn}, distributed uniformly in the Grassmanian with respect to the Haar measure. Let*

$$w(S)=E\sup_{\mathbf{w}\in S}|\mathbf{g}\mathbf{w}| \quad (5)$$

where $\mathbf{g}$ is a random row vector in R^{dn} with i.i.d. $\mathcal{N}(0,1)$ components. Assume that $w(S)<\sqrt{dm}-\frac{1}{4\sqrt{dm}}$. Then

$$P(Y\cap S=0)>1-2.5e^{-\frac{\left(\sqrt{dm}-\frac{1}{4\sqrt{dm}}-w(S)\right)^2}{18}}. \quad (6)$$

Remark: Gordon's original constant 3.5 was substituted by 2.5 in [24]. Both constants are fine for our subsequent analysis.

As masterly noted in [24] Theorem 2 can be used in probabilistic analysis of (4). Namely, let S in (5) be

$$S=\{\mathbf{w}\in S^{dn-1}|\exists K\in\mathcal{K},\sum_{i\in K}||\mathbf{W}_i||_2\geq\sum_{i\in\bar{K}}||\mathbf{W}_i||_2\} \quad (7)$$

where $\mathbf{W}_i=(\mathbf{w}_{(i-1)d+1},\mathbf{w}_{(i-1)d+2},\dots,\mathbf{w}_{id})^T$, $i=1,2,\dots,n$. Let Y be an $d(n-m)$ dimensional subspace of R^{dn} uniformly distributed in Grassmanian. Furthermore, let Y be the null-space of A. Then as long as $w(S)<\sqrt{dm}-\frac{1}{4\sqrt{dm}}$, Y will miss S (i.e. (4) will be satisfied) with probability no smaller than the one given in (6). More precisely, if $\alpha=\frac{m}{n}$ is a constant (the case of interest in this paper), n,m are large, and $w(S)$ is smaller than but proportional to $\sqrt{dm}$ then $P(Y\cap S=0)\longrightarrow 1$. This in turn is equivalent to having

$$P(\forall\mathbf{w}\in R^{dn}|A\mathbf{w}=0,\forall K\in\mathcal{K}\sum_{i\in K}||\mathbf{W}_i||_2<\sum_{i\in\bar{K}}||\mathbf{W}_i||_2)\longrightarrow 1$$

which according to Theorem 1 means that solutions of (3) and (2) coincide with probability 1. For any given value of $\alpha\in(0,1)$ a threshold value of β can be determined as a maximum β such that $w(S)<\sqrt{dm}-\frac{1}{4\sqrt{dm}}$. Since computing $w(S)$ does not appear easy we introduce a set D in the following way (see [24])

$$D=conv\{\mathbf{w}\in S^{dn-1}||\{i\in\{1,2,\dots,n\}|\mathbf{W}_i\neq\mathbf{0}\}|\leq k\} \quad (8)$$

where $\mathbf{0}$ is a vector of d zeros. Now, if we can compute an upper bound on $w(D)$ and show that $S\subset\xi D$ (for some $\xi>0$) we will effectively be able to establish an upper bound on $w(S)$. Equalling that potential upper bound with $\sqrt{dm}-\frac{1}{4\sqrt{dm}}$ would give us a way to compute thresholds for β. In Subsection 3.1 we compute an upper bound on $w(D)$ while in Subsection 3.2 we determine a ξ such that $S\subset\xi D$.

3.1. Upper-bounding of $w(D)$

By definition given in (5) we have

$$w(D)=E\sup_{\mathbf{w}\in D}|\mathbf{g}\mathbf{w}|. \quad (9)$$

Let $\mathbf{G}_i=(\mathbf{g}_{(i-1)d+1},\mathbf{g}_{(i-1)d+2},\dots,\mathbf{g}_{id}), i=1,2,\dots,n$. For a given $\mathbf{g}$ the function $|\mathbf{g}\mathbf{w}|$ is convex in $\mathbf{w}$ and hence achieves the maximum at the extreme points of D (by the definition D is a convex set). Also, by the definition of the set D given in (8) its extreme points can have at most k non-zero vectors (blocks) $\mathbf{W}_i$. Therefore we have

$$w(D)=E\sup_{\mathbf{w}\in D}|\mathbf{g}\mathbf{w}|=E\sup_{\mathbf{w}\in D}|\sum_{i=1}^n\mathbf{G}_i\mathbf{W}_i|=E\sup_{|J|=k}\left(\sum_{i\in J}\|\mathbf{G}_i\|_2^2\right)^{\frac{1}{2}}$$

where $J\subset\{1,2,\dots,n\}$. Using the Holder's inequality trick from [24] one for $p>1$ obtains

$$w(D)=E\sup_{|J|=k}\left(\sum_{i\in J}\|\mathbf{G}_i\|_2^2\right)^{\frac{1}{2}}\leq E(\sum_{|J|=k}(\sum_{i\in J}\|\mathbf{G}_i\|_2^2)^{\frac{p}{2}})^{\frac{1}{p}}$$

$$\leq\binom{n}{k}^{\frac{1}{p}}(E(\sum_{i\in J}\|\mathbf{G}_i\|_2^2)^{\frac{p}{2}})^{\frac{1}{p}}\leq\left(\frac{en}{k}\right)^{\frac{k}{p}}(E(\sum_{i\in J}\|\mathbf{G}_i\|_2^2)^{\frac{p}{2}})^{\frac{1}{p}}.$$

The second inequality follows by the concavity of the function $()^{\frac{1}{p}}$. As it will soon be clear this inequality is the key point for the success of our analysis. Namely, one should note that the combinatorial term $\binom{n}{k}^{\frac{1}{p}}$ does not change as d grows. The reason is exactly the structure of the sparse signal, i.e. the fact that the outer sum on the right side of the first inequality runs over all subsets of k blocks $\mathbf{W}_i$ rather than over all subsets od dk elements of $\mathbf{w}$. Since $\mathbf{G}_i$ is d-dimensional vector with $\mathcal{N}(0,1)$ i.i.d. components $E(\sum_{i\in J}\|\mathbf{G}_i\|_2^2)^{\frac{p}{2}}$ is the p-th central moment of a chi-distributed random variable with dk degrees of freedom. Hence $E(\sum_{i\in J}\|\mathbf{G}_i\|_2^2)^{\frac{p}{2}}=(2^{\frac{p}{2}}\frac{\Gamma(\frac{p+dk}{2})}{\Gamma(\frac{dk}{2})})^{\frac{1}{p}}$. The Stirling's formula and bounding the gamma functions similarly as in [24] give

$$w(D)\leq\left(\frac{en}{k}\right)^{\frac{k}{p}}\left(\frac{p+dk}{e}\right)^{\frac{1}{2}}\left(1+\frac{p}{dk}\right)^{\frac{dk-1}{2p}}e^{\frac{1}{12pdk}}. \quad (10)$$

Let $p=d^{\epsilon}k$, where $\epsilon>0$ is arbitrarily small. Then from (10) we have

$$\begin{aligned} w(D) &\leq (d^{\epsilon}k+dk)^{\frac{1}{2}}\left[\frac{e^{-\frac{1}{2}}(1+d^{(\epsilon-1)})^{\frac{d^{(1-\epsilon)}}{2}}e^{\frac{1}{12pdk}}}{\left(\frac{\beta}{e}\right)^{d^{-\epsilon}}(1+d^{(\epsilon-1)})^{\frac{1}{2d^{\epsilon}k}}}\right] \\ &= (dk)^{\frac{1}{2}}\left[\frac{e^{-\frac{1}{2}}(1+d^{(\epsilon-1)})^{\frac{d^{(1-\epsilon)}}{2}}e^{\frac{1}{12pdk}}}{\left(\frac{\beta}{e}\right)^{d^{-\epsilon}}(1+d^{(\epsilon-1)})^{\left(\frac{1}{2d^{\epsilon}k}-\frac{1}{2}\right)}}\right]. \end{aligned} \quad (11)$$

For any arbitrarily small $\epsilon>0$ one can always find a large constant d such that

$$\left[\frac{e^{-\frac{1}{2}}(1+d^{(\epsilon-1)})^{\frac{d^{(1-\epsilon)}}{2}}e^{\frac{1}{12pdk}}}{\left(\frac{\beta}{e}\right)^{d^{-\epsilon}}(1+d^{(\epsilon-1)})^{\left(\frac{1}{2d^{\epsilon}k}-\frac{1}{2}\right)}}\right]^2<1+\delta$$

and $\delta>0$ is arbitrarily small. Term $\frac{(1+d^{(\epsilon-1)})^{\frac{d^{(1-\epsilon)}}{2}}}{e^{\frac{1}{2}}}\rightarrow 1$ for a fixed small ϵ and sufficiently large d and the terms $e^{\frac{1}{12pdk}}$, $\left(\frac{\beta}{e}\right)^{d^{-\epsilon}}$, and $(1+d^{(\epsilon-1)})^{\left(\frac{1}{2d^{\epsilon}k}-\frac{1}{2}\right)}$ clearly go to 1 for a fixed small ϵ and sufficiently large d. Therefore from (11) we have that for any arbitrarily small $\delta>0$ there is a sufficiently large d such that

$$w(D)^2\leq dk(1+\delta). \quad (12)$$

3.2. Computing ξ

In this subsection we compute a ξ such that $S \subset \xi D$. As earlier

$$D = conv\{\mathbf{w} \in S^{dn-1} || \{i \in \{1,2,\dots,n\} | \mathbf{W}_i \neq \mathbf{0}\}| \leq k\}$$

$$S = \{\mathbf{w} \in S^{dn-1} | \exists K \in \mathcal{K}, \sum_{i \in K} ||\mathbf{W}_i||_2 \geq \sum_{i \in \bar{K}} ||\mathbf{W}_i||_2\} \quad (13)$$

and $\mathbf{W}_i = (\mathbf{w}_{(i-1)d+1}, \mathbf{w}_{(i-1)d+2}, \dots, \mathbf{w}_{id})$, $i = 1,2,\dots,n$. Clearly, $\mathbf{w} = (\mathbf{W}_1, \mathbf{W}_2, \dots, \mathbf{W}_n)^T$. Let $\mathbf{w}^*$ be a vector obtained as a permutation of the blocks of $\mathbf{w}$ so that their norms-2 form a non-increasing sequence. Then $\mathbf{w}^* = (\mathbf{W}_1^*, \mathbf{W}_2^*, \dots, \mathbf{W}_n^*)$,

$$S = \{\mathbf{w} \in S^{n-1} | \sum_{i=1}^{k} \|\mathbf{w}^*_{(i-1)d+1:id}\|_2 \geq \sum_{i=k+1}^{n} \|\mathbf{w}^*_{(i-1)d+1:id}\|_2\}. \quad (14)$$

(Notation: $\mathbf{w}^*_{(i-1)d+1:id} = (\mathbf{w}^*_{(i-1)d+1}, \mathbf{w}^*_{(i-1)d+2}, \dots, \mathbf{w}^*_{id})$.) Since D is a convex set there will be a norm denoted $\|.\|_D$ whose unit ball in R^n is D. It easily follows that ξ can be computed as

$$\xi = \max_{\mathbf{w} \in S} \|\mathbf{w}\|_D. \quad (15)$$

To simplify the analysis we now assume that $\gamma = \beta^{-1} = \frac{n}{k}$ is an integer and $\gamma \geq 3$. We also break the analysis into two cases.

3.2.1. *Case* 1*:* $\mathbf{w} \in S, \|\mathbf{w}^*_{1:dk}\|_2 \leq \sqrt{\frac{\gamma-1}{\gamma}}$

Using (14) and the quadratic-arithmetic mean inequality

$$\sqrt{\frac{\gamma-1}{\gamma}} \geq \|\mathbf{w}^*_{1:dk}\|_2 \geq \frac{\sum_{i=1}^{k} \|\mathbf{W}_i^*\|_2}{\sqrt{k}} \geq \frac{\sum_{i=k+1}^{n} \|\mathbf{W}_i^*\|_2}{\sqrt{k}}. \quad (16)$$

Since $\|\mathbf{W}_1^*\|_2 \geq \|\mathbf{W}_2^*\|_2 \geq, \dots, \geq \|\mathbf{W}_k^*\|_2$ it follows that $\|\mathbf{W}_k^*\|_2 \leq \frac{1}{\sqrt{k}}\sqrt{\frac{\gamma-1}{\gamma}}$. $\|\mathbf{W}_k^*\|_2 \geq \|\mathbf{W}_{k+1}^*\|_2 \geq, \dots, \geq \|\mathbf{W}_n^*\|_2$ implies

$$\frac{1}{\sqrt{k}}\sqrt{\frac{\gamma-1}{\gamma}} \geq \|\mathbf{W}_{k+1}^*\|_2 \geq \|\mathbf{W}_{k+2}^*\|_2 \geq, \dots, \geq \|\mathbf{W}_n^*\|_2. \quad (17)$$

Similarly to [24] the extreme points of the intersection of the regions given in (16) and (17) (we view the intersection only for $\mathbf{w}^*_{dk+1:n}$ part of $\mathbf{w}^*$, i.e. for this case we assume below that $\mathbf{w}^*_{dk+1:n}$ is padded with zeros to full length dn) are points that have k non-zero blocks-each with norm-2 equal to $\frac{1}{\sqrt{k}}\sqrt{\frac{\gamma-1}{\gamma}}$. To see this let $\mathbf{w}^*$ be a point such that (16) and (17) are satisfied. Then, $(n-k)$-dimensional point $(\|\mathbf{W}_{k+1}^*\|_2, \|\mathbf{W}_2^*\|_2, \dots, \|\mathbf{W}_n^*\|_2)$ is a convex combination, say *Conv*, of vectors $\pi^{(1)}, \pi^{(2)}, \dots, \pi^{(\binom{n-k}{k})}$, where each $\pi^{(j)}$ is either zero-vector or an $(n-k)$-dimensional vector with k-non-zero components equal to $\frac{1}{\sqrt{k}}\sqrt{\frac{\gamma-1}{\gamma}}$ (see [24]). Let $\Pi^{(j)}$ be a $d(n-k)$-dimensional vector obtained by expanding $\pi^{(j)}$ in the following way: $\Pi^{(j)}_{(i-k-1)d+1:(i-k)d} = \pi^{(j)}_{i-k} \frac{\mathbf{w}^*_{(i-1)d+1:id}}{\|\mathbf{w}^*_{(i-1)d+1:id}\|_2}$, $(k+1) \leq i \leq n$ ($\pi^{(j)}_{i-k}$ is the $(i-k)$-th component of $\pi^{(j)}$ and $\Pi^{(j)}_{(i-k-1)d+1:(i-k)d}$ is the $(i-k)$-th block of $\Pi^{(j)}$). Then it easily follows that $\mathbf{w}^*_{dk+1:n}$ is the convex combination *Conv* of $\Pi^{(j)}, 1 \leq j \leq \binom{n-k}{k}$ (each $\Pi^{(j)}$ is either zero-vector or has k non-zero blocks-each of norm-2 equal to $\frac{1}{\sqrt{k}}\sqrt{\frac{\gamma-1}{\gamma}}$). Ignoring zero-vector, one therefore concludes that the extreme points indeed have k non-zero blocks-each of norm-2 equal to $\frac{1}{\sqrt{k}}\sqrt{\frac{\gamma-1}{\gamma}}$ (note that if $\mathbf{w}^*_{dk+1:n}$ is in the region then $-\mathbf{w}^*_{dk+1:n}$ is as well, hence zero-vector is not an extreme point).

D is block symmetric and $\mathbf{w}^*$ is a block permutated $\mathbf{w}$, hence $\|\mathbf{w}\|_D = \|\mathbf{w}^*\|_D \leq \|(\mathbf{w}^*_{1:dk}, 0_{dk+1:dn})\|_D + \|(0_{1:dk}, \mathbf{w}^*_{dk+1:n})\|_D$. The maximum of $\|(0_{1:dk}, \mathbf{w}^*_{dk+1:n})\|_D$ is achieved at the extreme points of the region defined by (16) and (17). Let $\mathbf{z}$ be the extreme point where $\|(0_{1:dk}, \mathbf{w}^*_{dk+1:n})\|_D$ is maximized. Since $\mathbf{z}$ has no more than k non-zero blocks $\|\mathbf{z}\|_D = \|\mathbf{z}\|_2$ we then have

$$\|(0_{1:dk}, \mathbf{w}^*_{dk+1:n})\|_D \leq \|\mathbf{z}\|_D = \|\mathbf{z}\|_2 = \sqrt{\frac{\gamma-1}{\gamma}}$$

$$\|(\mathbf{w}^*_{1:dk}, 0_{dk+1:dn})\|_D = \|(\mathbf{w}^*_{1:dk}, 0_{dk+1:dn})\|_2 \leq \sqrt{\frac{\gamma-1}{\gamma}}. \quad (18)$$

Finally using (18) we obtain

$$\max_{\mathbf{w} \in S, \|\mathbf{w}^*_{1:dk}\|_2 \leq \sqrt{\frac{\gamma-1}{\gamma}}} \|\mathbf{w}\|_D \leq 2\sqrt{\frac{\gamma-1}{\gamma}}. \quad (19)$$

From (19) we easily have

$$\left(\mathbf{w} \in S, \|\mathbf{w}^*_{1:dk}\|_2 \leq \sqrt{\frac{\gamma-1}{\gamma}}\right) \Longrightarrow \mathbf{w} \in 2\sqrt{\frac{\gamma-1}{\gamma}}D. \quad (20)$$

3.2.2. *Case* 2*:* $\mathbf{w} \in S, \|\mathbf{w}^*_{1:dk}\|_2 \geq \sqrt{\frac{\gamma-1}{\gamma}}$

Let $D_1 = \left\{\mathbf{w} \in S | \sum_{i=1}^{\gamma} \|\mathbf{w}^*_{(i-1)dk:idk}\|_2 \leq 2\sqrt{\frac{\gamma-1}{\gamma}}\right\}$. Also, let $\phi^{(i)}, 1 \leq i \leq \gamma$ be a vector of length n such that only its coordinates $(i-1)dk+1, (i-1)dk+2, \dots, idk$ are non-zero. Furthermore, let $\phi^{(i)}_{(i-1)dk+1:idk} = \frac{\mathbf{w}^*_{(i-1)dk+1:idk}}{\|\mathbf{w}^*_{(i-1)dk+1:idk}\|_2}, 1 \leq i \leq \gamma$. Then we have that each element of D_1 can be represented as a linear combination $\frac{1}{2\sqrt{\frac{\gamma-1}{\gamma}}}(\|\mathbf{w}^*_{1:dk}\|_2, \|\mathbf{w}^*_{dk+1:2dk}\|_2, \dots, \|\mathbf{w}^*_{n-dk+1:n}\|_2)$ of γ vertices $2\sqrt{\frac{\gamma-1}{\gamma}}(\phi^{(1)}, \phi^{(2)}, \dots, \phi^{(\gamma)})$ of the convex set $2\sqrt{\frac{\gamma-1}{\gamma}}D$. Since $\sum_{i=1}^{\gamma} \frac{1}{2\sqrt{\frac{\gamma-1}{\gamma}}}(\|\mathbf{w}^*_{(i-1)dk+1:idk}\|_2) \leq 1$ by the definition of D_1, it then follows that $D_1 \subset 2\sqrt{\frac{\gamma-1}{\gamma}}D$.

We will now show that if $\mathbf{w} \in S$ and $\|\mathbf{w}^*_{1:dk}\|_2 \geq \sqrt{\frac{\gamma-1}{\gamma}}$ then $\mathbf{w} \in D_1$. The quadratic-arithmetic mean inequality gives

$$\sum_{i=1}^{\gamma} \|\mathbf{w}^*_{(i-1)dk:idk}\|_2 = \|\mathbf{w}^*_{1:dk}\|_2 + \sum_{i=2}^{\gamma} \|\mathbf{w}^*_{(i-1)dk:idk}\|_2$$

$$\leq \|\mathbf{w}^*_{1:dk}\|_2 + \sqrt{\gamma-1} \sum_{i=dk+1}^{n} (\mathbf{w}^*_i)^2. \quad (21)$$

Since $\mathbf{w} \in S$, by definition $\|\mathbf{w}\|_2 = \|\mathbf{w}^*\|_2 = 1$. Therefore $\sum_{i=dk+1}^{n} (\mathbf{w}^*_i)^2 = \sqrt{1 - \|\mathbf{w}^*_{1:dk}\|_2}$. Then from (21) we obtain

$$\sum_{i=1}^{\gamma} \|\mathbf{w}^*_{(i-1)dk:idk}\|_2 \leq \|\mathbf{w}^*_{1:dk}\|_2 + \sqrt{\gamma-1}\sqrt{1 - \|\mathbf{w}^*_{1:dk}\|_2}. \quad (22)$$

It is not that difficult to check that for $\|\mathbf{w}^*_{1:dk}\|_2 \geq \sqrt{\frac{\gamma-1}{\gamma}}$ the function on the right side of (22) is non-increasing in $\|\mathbf{w}^*_{1:dk}\|_2$. Hence we finally have

$$\sum_{i=1}^{\gamma} \|\mathbf{w}^*_{(i-1)dk:idk}\|_2 \leq \|\mathbf{w}^*_{1:dk}\|_2 + \sqrt{\gamma-1}\sqrt{1 - \|\mathbf{w}^*_{1:dk}\|_2}$$

$$\leq \sqrt{\frac{\gamma-1}{\gamma}} + \sqrt{\gamma-1}\sqrt{1 - \frac{\gamma-1}{\gamma}} = 2\sqrt{\frac{\gamma-1}{\gamma}} \quad (23)$$

which guarantees that if $\mathbf{w} \in S$ and $\|\mathbf{w}^*_{1:dk}\|_2 \geq \sqrt{\frac{\gamma-1}{\gamma}}$ then $\mathbf{w} \in D_1$ as well. Since $D_1 \in 2\sqrt{\frac{\gamma-1}{\gamma}}D$ we have

$$\left(\mathbf{w} \in S, \|\mathbf{w}^*_{1:dk}\|_2 \geq \sqrt{\frac{\gamma-1}{\gamma}}\right) \Longrightarrow \mathbf{w} \in 2\sqrt{\frac{\gamma-1}{\gamma}}D. \quad (24)$$

We can then formulate the following lemma.

Lemma 1 *Let S and D be as defined in (13) and (14). Let $\beta^{-1} = \frac{n}{k}$ be an integer greater than 2. Then*

$$S \subset 2\sqrt{1-\beta}D. \quad (25)$$

Proof 1 *Follows by combing (20), (24), and recalling on the definition of γ introduced below (15).* ■

Theorem 3 *Let A be a $dm \times dn$ measurement matrix in (1) with the null-space uniformly distributed in the Grassmanian. Let the unknown $\mathbf{x}$ in (1) be k-block-sparse with block lengths d as defined above (3). Let k, m, n be large and let $\alpha = \frac{m}{n}$ and $\beta = \frac{k}{n}$ be constants independent of m and n. If α and β satisfy (26) and β^{-1} is an integer then there will always be a sufficiently large* ***constant*** *d such that the solutions of (1) and (3) coincide with overwhelming probability.*

$$4(1-\beta)\beta \leq \alpha \tag{26}$$

Proof 2 *Follows by combining (25), (12), and the results of Theorem 2.* ■

The values of thresholds for β obtained based on (26) are shown on Figure 1. For $\beta^{-1} = 3, 4, 5, \dots$ an α is obtained from (26). Then the remaining points are interpolated by the virtue that the threshold function is non-decreasing. The staircase style of the plot comes from the assumption that β^{-1} is an integer. The entire analysis can be done without this assumption but it becomes significantly more cumbersome and the final results would almost be no different. The bounds on the value of $w(S)$ obtained by using the augmented set D are tight. Therefore, we believe that if one is to get better results for thresholds a modified version of the set D should be used. Also, one should note that our results determine what is called strong threshold for β (i.e. (3) succeeds with overwhelming probability in solving (1) for *all* βn-block-sparse $\mathbf{x}$). Figure 1 also contains a straight line $\frac{1}{2}$ which is theoretically the best value (for strong threshold) for $\frac{\beta}{\alpha}$ that one can hope for no matter what algorithm is used.

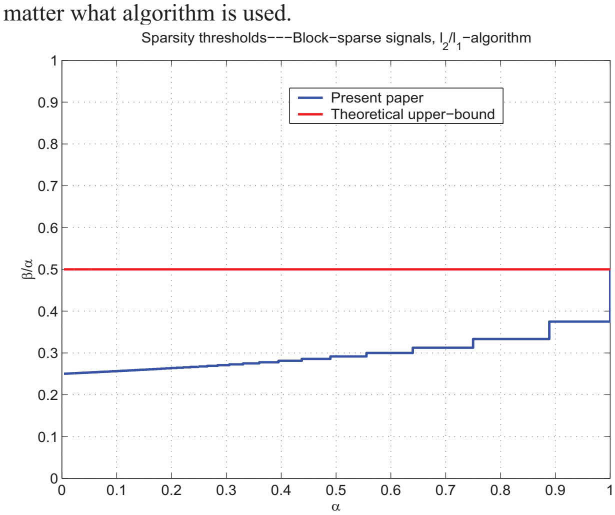

Fig. 1. Block-sparsity thresholds

We should finally make an important observation. Carefully following our derivation, one could note that our result made critical use of an excellent work [16] which on the other hand massively relied on phenomenal results [21, 7] related to the estimates of the normal tail distributions of Lipshitz functions. In a very recent work related to the matrix-rank optimization the authors in [22] successfully applied results of [21, 7] directly without relying on the Gordon's escape through a mesh theorem. It will be interesting to see if our work can be improved as well by directly applying the results of [21, 7].

4. SUMMARY

In this paper we analyzed recovery of the block-sparse signals. We explicitly evaluated lower bounds on the values of the sparsity of the block-sparse signals that a polynomial l_2/l_1 algorithm can recover with overwhelming probability.

5. REFERENCES

[1] M. Akcakaya and V. Tarokh. Performance bounds on sparse representations using redundant frames. *2007*. available online at arXiv:cs/0703045.

[2] R. Baraniuk, V. Cevher, M. Duarte, and C. Hegde. Model-based compressive sensing. available online at http://www.dsp.ece.rice.edu/cs/.

[3] D. Baron, M. Wakin, M. Duarte, S. Sarvotham, and Richard Baraniuk. Distributed compressed sensing. *Allerton*, 2005.

[4] R. Berinde, A. C. Gilbert, P. Indyk, H. Karloff, and M. J. Strauss. Combining geometry and combinatorics: A unified approach to sparse signal recovery. 2008. available online at http://www.dsp.ece.rice.edu/cs/.

[5] E. Candes and T. Tao. Decoding by linear programming. *IEEE Trans. on Information Theory*, 51:4203–4215, Dec. 2005.

[6] J. Chen and X. Huo. Theoretical results on sparse representations of multiple-measurement vectors. *IEEE Trans. on Signal Processing*, Dec 2006.

[7] B. S. Cirelson, I. A. Ibragimov, and V. N. Sudakov. Norms of gaussian sample functions. *Lect. Notes Math.*, 50, 1976.

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

[9] S. Cotter, B. Rao, K. Engan, and K. Kreutz-Delgado. Sparse solutions to linear inverse problems with multiple measurement vectors. *IEEE Trans. on Signal Porcessing*, July 2005.

[10] D. Donoho. High-dimensional centrally symmetric polytopes with neighborlines proportional to dimension. *Disc. Comput. Geometry*, 35:617–652, 2006.

[11] D. Donoho and J. Tanner. Sparse nonnegative solutions of underdetermined linear equations by linear programming. *Proc. National Academy of Sciences*, 102(27):9446–9451, 2005.

[12] D. L. Donoho and X. Huo. Uncertainty principles and ideal atomic decompositions. *IEEE Trans. Inform. Theory*, 47(7):2845–2862, November 2001.

[13] M. Duarte, M. Davenport, D. Takhar, J. Laska, T. Sun, K. Kelly, and R. Baraniuk. Single-pixel imaging via compressive sampling. *IEEE Signal Processing Magazine*, 25(2):83–91, 2008.

[14] Y. C. Eldar and M. Mishali. Robust recovery of signals from a union of subspaces. 2008. available online at http://www.dsp.ece.rice.edu/cs/.

[15] A. Feuer and A. Nemirovski. On sparse representation in pairs of bases. *IEEE Trans. on Information Theory*, 49:1579–1581, June 2003.

[16] Y. Gordon. On Milman's inequality and random subspaces which escape through a mesh in R^n. *Geometric Aspect of of functional analysis, Isr. Semin. 1986-87, Lect. Notes Math*, 1317, 1988.

[17] R. Gribonval and M. Nielsen. Sparse representations in unions of bases. *IEEE Trans. Inform. Theory*, 49(12):3320–3325, December 2003.

[18] S. Jafarpour, W. Xu, B. Hassibi, and R. Calderbank. Efficient compressed sensing using high-quality expander graphs. available online at http://www.dsp.ece.rice.edu/cs/.

[19] D. Needell and J. A. Tropp. CoSaMP: Iterative signal recovery from incomplete and inaccurate samples. 2008. available online at http://www.dsp.ece.rice.edu/cs/.

[20] F. Parvaresh and B. Hassibi. Explicit measurements with almost optimal thresholds for compressed sensing. *IEEE ICASSP*, Mar-Apr 2008.

[21] G. Pisier. Probabilistic methods in the geometry of banach spaces. *Springer Lecture Notes*, 1206, 1986.

[22] B. Recht, W. Xu, and B. Hassibi. Necessary and sufficient conditions for success of the nuclear norm heuristic for rank minimization. 2008. avialable at arxiv.

[23] J. Romberg. Imaging via compressive sampling. *IEEE Signal Processing Magazine*, 25(2):14–20, 2008.

[24] M. Rudelson and R. Vershynin. On sparse reconstruction from Fourier and Gaussian measurements. *Comm. on Pure and Applied Math.*, 61(8), 2007.

[25] M. Stojnic, F. Parvaresh, and B. Hassibi. On the reconstruction of block-sparse signals with an optimal number of measurements. *submitted to IEEE Trans. on Signal Processing*, 2008. avaiable on arxiv.

[26] M. Stojnic, W. Xu, and B. Hassibi. Compressed sensing of approximately sparse signals. *ISIT, International symposium on information theory*, July 2008.

[27] V.N. Temlyakov. A remark on simultaneous greedy approximation. *East J. Approx.*, 100, 2004.

[28] J. Tropp, A. C. Gilbert, and M. Strauss. Algorithms for simultaneous sparse approximation. part i: Greedy pursuit. *Signal Processing*, Aug 2005.

[29] H. Vikalo, F. Parvaresh, and B. Hassibi. On sparse recovery of compressed DNA microarrays. *Asilomor conference*, November 2007.

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