

# On the Reconstruction of Block-Sparse Signals With an Optimal Number of Measurements

Mihailo Stojnic, Farzad Parvaresh, and Babak Hassibi

**Abstract**—Let  $A$  be an  $M$  by  $N$  matrix ( $M < N$ ) which is an instance of a real random Gaussian ensemble. In compressed sensing we are interested in finding the sparsest solution to the system of equations  $Ax = y$  for a given  $y$ . In general, whenever the sparsity of  $x$  is smaller than half the dimension of  $y$  then with overwhelming probability over  $A$  the sparsest solution is unique and can be found by an exhaustive search over  $x$  with an exponential time complexity for any  $y$ . The recent work of Candés, Donoho, and Tao shows that minimization of the  $\ell_1$  norm of  $x$  subject to  $Ax = y$  results in the sparsest solution provided the sparsity of  $x$ , say  $K$ , is smaller than a certain threshold for a given number of measurements. Specifically, if the dimension of  $y$  approaches the dimension of  $x$ , the sparsity of  $x$  should be  $K < 0.239 N$ . Here, we consider the case where  $x$  is block sparse, i.e.,  $x$  consists of  $n = N/d$  blocks where each block is of length  $d$  and is either a zero vector or a nonzero vector (under nonzero vector we consider a vector that can have both, zero and nonzero components). Instead of  $\ell_1$ -norm relaxation, we consider the following relaxation:

$$\min_x \|X_1\|_2 + \|X_2\|_2 + \cdots + \|X_n\|_2, \quad \text{subject to } Ax = y \quad (*)$$

where  $X_i = (x_{(i-1)d+1}, x_{(i-1)d+2}, \dots, x_{id})^T$  for  $i = 1, 2, \dots, n$ . Our main result is that as  $n \rightarrow \infty$ , (\*) finds the sparsest solution to  $Ax = y$ , with overwhelming probability in  $A$ , for any  $x$  whose sparsity is  $k/n < (1/2) - O(\varepsilon)$ , provided  $m/n > 1 - 1/d$ , and  $d = \Omega(\log(1/\varepsilon)/\varepsilon^3)$ . The relaxation given in (\*) can be solved in polynomial time using semi-definite programming.

**Index Terms**—Compressed sensing, block-sparse signals, semi-definite programming.

## I. INTRODUCTION

LET  $A$  be an  $M$  by  $N$  instance of the real random Gaussian ensemble and  $x$  be an  $N$  dimensional signal from  $\mathbb{R}^N$  with sparsity  $K$ , i.e., only  $K$  elements of  $x$  are nonzero. Set  $y = Ax$  which is an  $M$  dimensional vector in  $\mathbb{R}^M$ . In compressed sensing  $y$  is called the *measurement vector* and  $A$  the *Gaussian measurement matrix*. Compressed sensing has applications in many different fields such as data mining [36], error-correcting

codes [9], [10], [42], DNA microarrays [25], [39], [49], astronomy, tomography, digital photography, and A/D converters.

In general, when  $K \ll N$  one can hope that the solution of  $y = Ax$  is unique for large enough  $M$  which is much smaller than  $N$ . In other words, instead of sensing an  $N$  dimensional signal  $x$  with sparsity  $K$  we can measure  $M$  random linear functionals of  $x$  where  $M \ll N$  and find  $x$  by solving the under-determined system of equations  $y = Ax$  with the extra condition that  $x$  is  $K$  sparse. The reconstruction can be presented as the following optimization problem:

$$\min_x \|x\|_0 \quad \text{subject to } Ax = y \quad (1)$$

where the  $\ell_0$  norm or the Hamming norm is the number of nonzero elements of  $x$ .

Define  $\alpha \stackrel{\text{def}}{=} M/N$  and  $\beta \stackrel{\text{def}}{=} K/N$ . In [1], the authors show that if  $\beta > (1/2)\alpha$  then for any measurement matrix  $A$  one can construct different  $K$  sparse signals  $x_1$  and  $x_2$  such that  $Ax_1 = Ax_2$ . In addition, if  $\beta \leq (1/2)\alpha$  then there exists an  $A$  such that the  $K$  sparse solution to  $y = Ax$  is unique for any  $y$ ; specifically, for random Gaussian measurements the uniqueness property holds with overwhelming probability in the choice of  $A$ . However, the reconstruction of  $x$  for a given  $y$  can be cumbersome. One of the fundamental questions in compressed sensing is whether one can *efficiently* recover  $x$  using an optimal number of measurements.

### A. Prior Work

A naive exhaustive search can reconstruct the  $K$  sparse solution  $x$  to the systems of equations  $y = Ax$  with  $O(\binom{N}{K}M^3)$  complexity. Recently, Candés, Romberg, Tao and Donoho [7], [8], [16], show that the  $\ell_0$  optimization can be relaxed to  $\ell_1$  minimization if the sparsity of the signal is  $K = O(M/\log(N/M))$ . In this case, the sparse signal is the solution to the following  $\ell_1$  norm optimization with high probability in the choice of  $A$ :

$$\min_x \|x\|_1 \quad \text{subject to } Ax = y \quad (2)$$

This optimization can be solved efficiently using linear programming. Faster algorithms were discovered in [12], [27], [28], [50]. For a comprehensive list of papers and results in compressed sensing please check [41].

Donoho and Tanner [17], [20], [21] determined the region  $(\alpha, \beta)$  for which the  $\ell_1$  and  $\ell_0$  coincide under Gaussian measurements for every (or almost every)  $K$ -sparse vector  $x$ . From a refinement of their result given in [23], when  $\alpha$  approaches 1 the sparsity has to be smaller than  $0.239N$ . Notice that, ideally, one should be able to recover the signal if the sparsity is less than

Manuscript received April 06, 2008; accepted March 11, 2009. First published April 10, 2009; current version published July 15, 2009. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Christine Guillemot.

M. Stojnic is with the School of Industrial Engineering, Purdue University, West Lafayette, IN 47907-2023 USA (e-mail: mstojnic@purdue.edu).

F. Parvaresh is with the Center for Mathematics of Information, California Institute of Technology, Pasadena, CA 91125 USA (e-mail: farzad@systems.caltech.edu).

B. Hassibi is with the Department of Electrical Engineering, California Institute of Technology, Pasadena, CA 91125 USA (e-mail: hassibi@systems.caltech.edu).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TSP.2009.2020754

$1/2N$  (assuming that  $\alpha < 1$ ). We have to mention that with Vandermonde measurements we can recover the sparse signal with optimal number of measurements efficiently [1]. However, it is not clear whether the resulting algorithms (which are variations of recovering a measure from its moments) are robust with respect to imperfections, such as noise [15], [35], [47], [51]. Also, results similar to those valid for Gaussian matrices  $A$  have been established for several different ensembles, e.g., Fourier (see e.g., [8]).

In this paper, we will focus on developing robust efficient algorithms that work for Gaussian measurements.

**B. Our Main Result**

We consider the reconstruction of **block-sparse** signals from their random measurements. A signal of dimension  $N$  which consists of  $n$  blocks (see Fig. 1) of size  $d = N/n$  is  $k$ -block sparse if only  $k$  blocks of the signal out of  $n$  are nonzero (under nonzero block we consider a block that can have both zero and nonzero components). Such signals arise in various applications, e.g., DNA microarrays, equalization of sparse communication channels, magnetoencephalography, etc. (see, e.g., [3], [13], [24], [25], [29], [38], [46], [49], and the references therein). We measure the signal with an  $md \times nd$  random Gaussian matrix  $\mathbf{y} = \mathbf{A}\mathbf{x}$  (for the ease of the exposition we will assume that  $m$  is an integer; however, our results will hold even if  $m$  is not an integer). More on a scenario similar to this one the interested reader can find in e.g., [11], [14], [40], [45], [46]. Using the  $\ell_1$  relaxation for reconstructing  $\mathbf{x}$  does not exploit the fact that the signal is block-sparse, i.e., that the potential nonzero entries can occur only in consecutive positions. Instead, different techniques were used throughout the literature. In [46] the authors adapt standard orthogonal matching pursuit algorithm (used normally in case  $d = 1$ ) to the block-sparse case. In [11], [14], [30], and [45], the authors use certain convex or non-convex relaxations (mostly different from the standard  $\ell_1$ ) and discuss their performances. Generalization of the block-sparse problem to the case when the number of blocks is infinite was considered in the most recent paper [40]. We should also mention that the above mentioned works [11], [14], [40], [45], [46] indeed consider the block structure of the unknown signal  $x$ , i.e., the same signal structure that we are analyzing in this paper. However, what is different in our case is the structure of the measurement matrix  $A$ . Namely, while in other cases the measurement matrix  $A$  is assumed to be block diagonal in our case the matrix  $A$  is assumed to be full.

In this paper, we consider the following convex relaxation for the recovery of  $\mathbf{x}$ :

$$\min_{\mathbf{x}} \|\mathbf{X}_1\|_2 + \|\mathbf{X}_2\|_2 + \dots + \|\mathbf{X}_n\|_2, \quad \text{subject to } \mathbf{A}\mathbf{x} = \mathbf{y} \tag{3}$$

where  $\mathbf{X}_i = (\mathbf{x}_{(i-1)d+1}, \mathbf{x}_{(i-1)d+2}, \dots, \mathbf{x}_{id})^T$ , for  $i = 1, 2, \dots, n$ . If one considers  $\ell_1$  optimization as minimization of the sum of the magnitudes of the components of a vector then (3) can be viewed in the same way. The only difference is that in (3) we have as objective function  $\ell_1$  norm of the vector

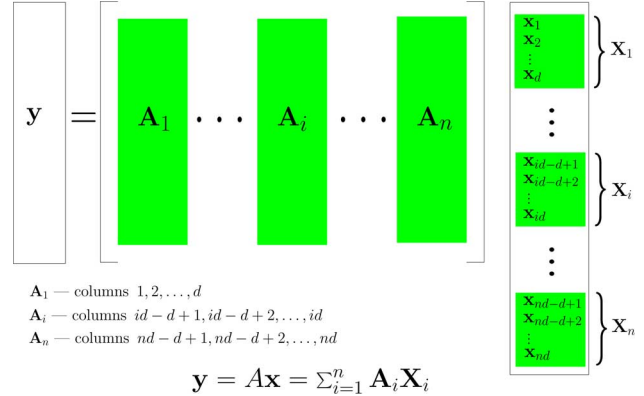


Fig. 1. Block sparse model.

$(\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n)^T$  whose components are now actually vectors  $\mathbf{X}_i = (\mathbf{x}_{(i-1)d+1}, \mathbf{x}_{(i-1)d+2}, \dots, \mathbf{x}_{id})^T, 1 \leq i \leq n$ . We will analyze theoretical performance of (3) and show that for a large enough  $d$ , independent of  $n$ , if  $\alpha$  approaches one,  $\beta$  can approach  $1/2$  and the optimization of (3) will give the unique sparse solution with overwhelming probability over the choice of  $A$  for any  $\mathbf{y}$  (given that we mentioned earlier that  $\beta \geq \alpha/2$  is not possible our result is then near optimal). We will also briefly outline how (3) can be posed as a semi-definite program and therefore solved efficiently in polynomial time by a host of numerical methods. Furthermore, we will demonstrate how (3) can be adapted for practical considerations. Numerical results that we will present indicate that in practice a modified version of (3) (given in Section IV) will even for moderate values of  $d$  be able to recover most of the signals with sparsity fairly close to the number of measurements. Before proceeding further, we state the main result of this work in the following theorem.

*Theorem 1:* Let  $A$  be an  $md \times nd$  matrix. Further, let  $A$  be an instance of random Gaussian ensemble. Also, assume that  $n$  tends to infinity,  $m = \alpha n$ , and the block-sparsity of  $\mathbf{x}$  is smaller than  $\beta n$ . Assume that  $\epsilon$  is a small positive number, i.e.,  $0 < \epsilon \ll 1$  and  $\beta = 1/2 - O(\epsilon)$ . Then, there are  $d = \Omega(\log(1/\epsilon)/\epsilon^3)$  and  $\alpha > 1 - 1/d$  such that *all* block-sparse signals  $\mathbf{x}$  that have blocks of length  $d$  can be reconstructed with overwhelming probability from  $\mathbf{y} = \mathbf{A}\mathbf{x}$  by solving the optimization problem (3).

*Remark:* Three points from the above theorem deserve clarification: 1) For any two real functions  $q(\epsilon)$  and  $w(\epsilon)$  we have  $q(\epsilon) = O(w(\epsilon))$  if there exist constants  $\kappa > 0$  and  $\epsilon_0$  such that  $q(\epsilon) \leq \kappa w(\epsilon)$  for any  $\epsilon > \epsilon_0$ ; 2) for any two real functions  $q(\epsilon)$  and  $w(\epsilon)$  we have  $q(\epsilon) = \Omega(w(\epsilon))$  if there exist constants  $\kappa > 0$  and  $\epsilon_0$  such that  $q(\epsilon) \geq \kappa w(\epsilon)$  for any  $\epsilon > \epsilon_0$ ; and 3) under overwhelming probability we assume a probability  $\pi$  such that  $1 - \pi < e^{-\ln}$  for some constant  $\ln$ .

We will prove the above Theorem in Section 3. Our proof technique does not use the restricted isometry property of the measurement matrix  $A$ , introduced in the work of Candés and Tao [8] and further discussed in [4], nor does it rely on the  $k$ -neighborliness of the projected polytopes presented in the work of Donoho and Tanner [17], [20], [21], [48]. Instead, we look at the null-space of the measurement matrix  $A$  and use a generalization of a necessary and sufficient condition given in [43] for the equivalence of (1) and (3).

II. NULL-SPACE CHARACTERIZATION

In this section, we introduce a necessary and sufficient condition on the measurement matrix  $A$  so that the optimizations of (1) and (3) are equivalent for all  $k$ -block sparse  $\mathbf{x}$  (see [18], [26], [31]–[33], [37], [43], and [53] for variations of this result). Throughout the paper, we set  $K$  to be the set of all subsets of size  $k$  of  $\{1, 2, \dots, n\}$  and by  $\bar{K}$  we mean the complement of the set  $\mathcal{K} \subset K$  with respect to  $\{1, 2, \dots, n\}$ , i.e.,  $\bar{K} = \{1, 2, \dots, n\} \setminus \mathcal{K}$ .

*Theorem 2:* Assume that  $A$  is a  $dm \times dn$  measurement matrix. Then the solutions of (1) and (3) coincide for any  $k$ -block sparse  $\mathbf{x}$  if and only if for all nonzero  $\mathbf{w} \in \mathbb{R}^{dn}$  where  $A\mathbf{w} = 0$  and all  $\mathcal{K} \in K$

$$\sum_{i \in \mathcal{K}} \|\mathbf{W}_i\|_2 < \sum_{i \in \bar{\mathcal{K}}} \|\mathbf{W}_i\|_2 \tag{4}$$

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

*Proof:* The proof goes along the lines similar to those in the proofs in [18], [26], [31]–[33], [37], [43], and [53]. The only difference is that each component of the vector is now replaced by the two norm of the subvector. First, we prove that if (4) is satisfied then the solution of (3) coincides with the solution (1). Let  $\bar{\mathbf{x}}$  be the solution of (1) and let  $\hat{\mathbf{x}}$  be the solution of (3). Also, let  $\bar{\mathbf{x}} \neq \hat{\mathbf{x}}$ . Further, let  $\bar{\mathbf{X}}_i = (\bar{\mathbf{x}}_{(i-1)d+1}, \bar{\mathbf{x}}_{(i-1)d+2}, \dots, \bar{\mathbf{x}}_{id})^T$ , for  $i = 1, 2, \dots, n$  and  $\hat{\mathbf{X}}_i = (\hat{\mathbf{x}}_{(i-1)d+1}, \hat{\mathbf{x}}_{(i-1)d+2}, \dots, \hat{\mathbf{x}}_{id})^T$ , for  $i = 1, 2, \dots, n$ . Set  $\mathcal{K}$  to be the support of  $\bar{\mathbf{x}}$ , then we can write

$$\begin{aligned} \sum_{i=1}^n \|\hat{\mathbf{X}}_i\|_2 &= \sum_{i=1}^n \|\hat{\mathbf{X}}_i - \bar{\mathbf{X}}_i + \bar{\mathbf{X}}_i\|_2 \\ &= \sum_{i \in \mathcal{K}} \|\hat{\mathbf{X}}_i - \bar{\mathbf{X}}_i + \bar{\mathbf{X}}_i\|_2 + \sum_{i \in \bar{\mathcal{K}}} \|\hat{\mathbf{X}}_i - \bar{\mathbf{X}}_i + \bar{\mathbf{X}}_i\|_2 \\ &= \sum_{i \in \mathcal{K}} \|\hat{\mathbf{X}}_i - \bar{\mathbf{X}}_i + \bar{\mathbf{X}}_i\|_2 + \sum_{i \in \bar{\mathcal{K}}} \|\hat{\mathbf{X}}_i - \bar{\mathbf{X}}_i\|_2 \\ &\geq \sum_{i=1}^n \|\bar{\mathbf{X}}_i\|_2 - \sum_{i \in \mathcal{K}} \|\hat{\mathbf{X}}_i - \bar{\mathbf{X}}_i\|_2 + \sum_{i \in \bar{\mathcal{K}}} \|\hat{\mathbf{X}}_i - \bar{\mathbf{X}}_i\|_2. \end{aligned} \tag{5}$$

Since  $\bar{\mathbf{x}} - \hat{\mathbf{x}}$  lies in the null-space of  $A$ , we have  $\sum_{i \in \mathcal{K}} \|\hat{\mathbf{X}}_i - \bar{\mathbf{X}}_i\|_2 < \sum_{i \in \bar{\mathcal{K}}} \|\hat{\mathbf{X}}_i - \bar{\mathbf{X}}_i\|_2$ . Thus, (5) implies  $\sum_{i=1}^n \|\hat{\mathbf{X}}_i\|_2 > \sum_{i=1}^n \|\bar{\mathbf{X}}_i\|_2$ , which is a contradiction. Therefore,  $\bar{\mathbf{x}} = \hat{\mathbf{x}}$ . Now we prove the converse. Assume (4) does not hold. Then there exists  $\mathbf{w} \in \mathbb{R}^{dn}$ ,  $A\mathbf{w} = 0$ ,  $\mathbf{w} = \begin{pmatrix} \mathbf{w}_1 \\ \mathbf{w}_2 \end{pmatrix}$ ,  $\mathbf{w}_1 \in \mathbb{R}^{kd}$ ,  $\mathbf{w}_2 \in \mathbb{R}^{(n-k)d}$  such that  $\mathbf{w}_1$  is  $k$ -block sparse and  $\sum_{i \in \mathcal{K}} \|\mathbf{W}_i\|_2 \geq \sum_{i \in \bar{\mathcal{K}}} \|\mathbf{W}_i\|_2$ , where  $\mathcal{K}$  is the support of  $\mathbf{w}_1$ . Take  $\mathbf{x} = \begin{pmatrix} \mathbf{w}_1 \\ \mathbf{0} \end{pmatrix}$  and  $\mathbf{y} = A\mathbf{x}$ . Since  $\mathbf{w}$  is in the null-space of  $A$ ,  $\mathbf{y} = A \begin{pmatrix} \mathbf{0} \\ -\mathbf{w}_2 \end{pmatrix}$ . Therefore, we have found a signal  $\begin{pmatrix} \mathbf{0} \\ -\mathbf{w}_2 \end{pmatrix}$  which is not  $k$ -block sparse and has smaller norm than the  $k$ -block sparse  $\begin{pmatrix} \mathbf{w}_1 \\ \mathbf{0} \end{pmatrix}$ .  $\begin{pmatrix} \mathbf{0} \\ -\mathbf{w}_2 \end{pmatrix}$  is  $(n - k)$ -block sparse; since  $k < (n)/(2)$  we have  $n - k > k$ . ■

*Remark 1:* One should note that the results from [31]–[33] relate to the use of  $\ell_p$ ,  $0 < p \leq 1$  norm instead of  $\ell_1$  norm in (2). As pointed out by an anonymous reviewer it is reasonable to believe that the null-space characterization given in Theorem 2 (as well as the entire technique presented in this paper) can easily be generalized to the  $\ell_p$  optimization.

*Remark 2:* We need not to check (4) for all subsets  $\mathcal{K}$ ; checking the subset with the  $k$  largest (in two norm) blocks of  $\mathbf{w}$  is sufficient. However, the form of Theorem 2 will be more convenient for our subsequent analysis.

Let  $Z$  be a basis of the null space of  $A$ , so that any  $dn$  dimensional vector  $\mathbf{w}$  in the null-space of  $A$  can be represented as  $Z\mathbf{v}$  where  $\mathbf{v} \in \mathbb{R}^{d(n-m)}$ . For any  $\mathbf{v} \in \mathbb{R}^{d(n-m)}$  write  $\mathbf{w} = Z\mathbf{v}$ . We split  $\mathbf{w}$  into blocks of size  $d$ ,  $\mathbf{W}_i = (\mathbf{w}_{(i-1)d+1}, \mathbf{w}_{(i-1)d+2}, \dots, \mathbf{w}_{id})^T$ , for  $i = 1, 2, \dots, n$ . Then, the condition (4) of Theorem 2 is equivalent to

$$\sum_{i \in \mathcal{K}} \|\mathbf{W}_i\|_2 < \sum_{i \in \bar{\mathcal{K}}} \|\mathbf{W}_i\|_2, \quad \text{for any } \mathbf{v} \in \mathbb{R}^{d(n-m)} \text{ and } \mathcal{K} \in K, \text{ where } \mathbf{w} = Z\mathbf{v}. \tag{6}$$

We denote by  $I_{\mathbf{v}}$  the event that (6) happens. In the following we find an upper bound on the probability that  $I_{\mathbf{v}}$  fails as  $n$  tends to infinity. We will show that for certain values of  $\alpha, \beta$ , and  $d$  this probability tends to zero.

*Lemma 3:* Let  $A \in \mathbb{R}^{dm \times dn}$  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 left-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  $Z \in \mathbb{R}^{dn \times d(n-m)}$  has i.i.d.  $\mathcal{N}(0, 1)$  entries.

*Proof:*

- The first part trivially follows.
- To show the second part we first focus on an orthonormal basis for the null space. Thus, assume that

$$AY = 0 \quad \text{and} \quad Y^*Y = I$$

For this particular basis (i.e., the orthonormal basis), we have

$$p_{Y|A}(Y|A) = \delta(A Y) \delta(Y^* Y - I).$$

Then

$$\begin{aligned} p_Y(Y) &= \int_A dA p_{Y|A}(Y|A) p_A(A) \\ &= \int_A dA \delta(A Y) \delta(Y^* Y - I) p_A(A). \end{aligned}$$

Therefore, for any deterministic unitary  $\Theta$

$$\begin{aligned} p_Z(\Theta^* Z) &= \int_A dA \delta(A \Theta^* Z) \delta(Z^* \Theta \Theta^* Z - I) p_A(A) \\ &= \int_A dA \delta(A \Theta^* Z) \delta(Z^* Z - I) p_A(A). \end{aligned}$$

Now doing the change of variables  $T = A\Theta^*$ , we have  $dA = dT$  (the Jacobian of the transformation is 1), and therefore

$$p_Z(\Theta^*Z) = \int_T dT \delta(TZ) \delta(Z^*Z - I) p_A(T\Theta) = \int_T dT \delta(TZ) \delta(Z^*Z - I) p_A(T) = p_Z(Z)$$

where we have used the fact that  $A$  is right rotationally invariant, i.e.,  $T = A\Theta^*$  has the same distribution as  $A$ . This shows the left rotational invariance of the orthonormal basis. Any other basis can be written as

$$Z = YR$$

where  $R$  is square, nonsingular and independent of the orthonormal basis  $Y$ . It is clear that any such  $Z$  is left rotationally invariant.

- The third part now follows from the second part by carefully selecting the random matrix  $R$ . Namely, we choose  $R$  to be an upper-triangular random matrix in the following way: the  $i$ th entry on the main diagonal of  $R$  is a  $\chi^2$  distributed random variable with  $n - i + 1$  degrees of freedom and any strictly upper-triangular entry of  $R$  is  $\mathcal{N}(0, 1)$ . All diagonal and strictly upper-triangular elements of  $R$  are independent random variables. Furthermore, as mentioned above, the random entries of  $R$  are independent of the entries in left rotationally invariant orthonormal basis  $Y$ .  $Z$  then has i.i.d.  $\mathcal{N}(0, 1)$  entries. (The interested reader can find more on this construction of the Gaussian matrix  $Z$  in Lemma 1 and in Remark 2 after Lemma 1 in [34]. Furthermore the Proof of Lemma 1 is given in Appendix A of [34].) ■

In view of Theorem 2 and Theorem 3, for any  $A$  whose null-space is rotationally invariant the sharp bounds of [19], for example, apply (of course, if  $d = 1$ ). In this paper, we shall analyze the null-space directly.

### III. PROBABILISTIC ANALYSIS OF THE NULL-SPACE CHARACTERIZATION

Assume  $Z$  is an  $dn \times d(n - m)$  matrix whose components are i.i.d. zero-mean unit-variance Gaussian random variables. Define  $Z_i$  to be the matrix which consists of the  $\{(i - 1)d + 1, (i - 1)d + 2, \dots, id\}$  rows of  $Z$  and define  $Z_{ij}$  to be the  $j$ th column of  $Z_i$  (see Fig. 2).

Let  $\alpha = 1 - \gamma, 0 < \gamma \ll 1$  where  $\gamma$  is a constant independent of  $n$ . Then we will find a  $d$  such that  $\beta \rightarrow 1/2$  and

$$\lim_{n \rightarrow \infty} P(I_{\mathbf{v}}) = 1. \tag{7}$$

Proving (7) is enough to show that for all random matrix ensembles which have isotropically distributed null-space, (3) with overwhelming probability solves (1). In order to prove (7), we will actually look at the complement of the event  $I_{\mathbf{v}}$  and we show that

$$\lim_{n \rightarrow \infty} P_f \stackrel{\text{def}}{=} \lim_{n \rightarrow \infty} P(\bar{I}_{\mathbf{v}}) = 0 \tag{8}$$

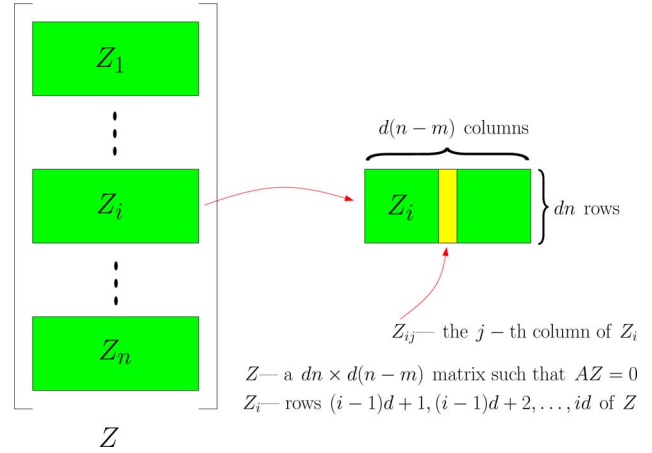


Fig. 2. Block structure of matrix  $Z$ .

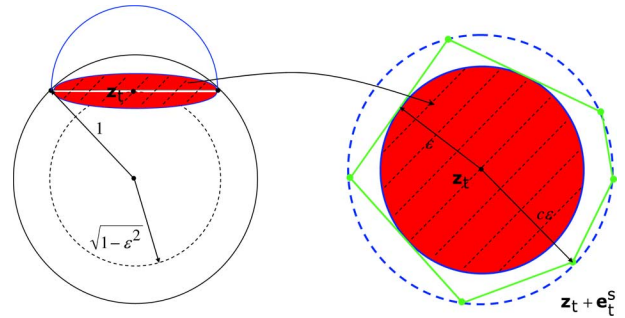


Fig. 3. Covering of the unit sphere.

where  $\bar{I}_{\mathbf{v}}$  denotes the complement of the event  $I_{\mathbf{v}}$ . Using the union bound, we can write

$$P_f \leq \sum_{K \in \mathcal{K}} P \left( \exists \mathbf{v} \in \mathbb{R}^{d(n-m)} : \sum_{i \in K} \|Z_i \mathbf{v}\|_2 \geq \sum_{i \in \bar{K}} \|Z_i \mathbf{v}\|_2 \right). \tag{9}$$

Clearly, the size of  $K$  is  $\binom{n}{k}$ . Since the probability in (9) is insensitive to scaling of  $\mathbf{v}$  by a constant we can restrict  $\mathbf{v}$  to lie on the surface of a shape  $\mathcal{C}$  that encapsulates the origin. Furthermore, since the elements of the matrix  $Z$  are i.i.d. all  $\binom{n}{k}$  terms in the first summation on the right-hand side of (9) will then be equal. Therefore, we can further write

$$P_f \leq \binom{n}{k} \cdot P \left( \exists \mathbf{v} \in \mathcal{C} : \sum_{i=1}^k \|Z_i \mathbf{v}\|_2 \geq \sum_{i=k+1}^n \|Z_i \mathbf{v}\|_2 \right). \tag{10}$$

The main difficulty in computing the probability on the right-hand side of (10) is in the fact that the vector  $\mathbf{v}$  is continuous. Our approach will be based on the discrete covering of

the unit sphere. In order to do that we will use small spheres of radius  $\epsilon$ . It can be shown [22], [42], [52] that  $\epsilon^{-d(n-m)}$  spheres of radius  $\epsilon$  is enough to cover the surface of the  $d(n-m)$ -dimensional unit sphere. Let the coordinates of the centers of these  $\epsilon^{-d(n-m)}$  small spheres be the vectors  $\mathbf{z}_t, t = 1, 2, \dots, \epsilon^{-d(n-m)}$ . Clearly,  $\|\mathbf{z}_t\|_2 = \sqrt{1 - \epsilon^2}$ . Further, let  $S_t, t = 1, 2, \dots, \epsilon^{-d(n-m)}$  be the intersection of the unit sphere and the hyperplane through  $\mathbf{z}_t$  perpendicular on the line that connects  $\mathbf{z}_t$  and the origin. It is not difficult to see that  $\bigcup_{t=1}^{\epsilon^{-d(n-m)}} S_t$  forms a body which completely encapsulates the origin. This effectively means that for any point  $\mathbf{v}$  such that  $\|\mathbf{v}\|_2 > 1$ , the line connecting  $\mathbf{v}$  and the origin will intersect  $\bigcup_{t=1}^{\epsilon^{-d(n-m)}} S_t$ . Hence, we set  $\mathcal{C} = \bigcup_{t=1}^{\epsilon^{-d(n-m)}} S_t$  and apply union bound over  $S_t$  to get

$$P_f \leq \binom{n}{k} \epsilon^{-d(n-m)} \times \max_t \left[ P \left( \exists \mathbf{v} \in S_t : \sum_{i=1}^k \|Z_i \mathbf{v}\|_2 \geq \sum_{i=k+1}^n \|Z_i \mathbf{v}\|_2 \right) \right]. \quad (11)$$

Every vector  $\mathbf{v} \in S_t$  can be represented as  $\mathbf{v} = \mathbf{z}_t + \mathbf{e}$  where  $\|\mathbf{e}\|_2 \leq \epsilon$ . Then we have

$$\begin{aligned} & \max_t \left[ P \left( \exists \mathbf{v} \in S_t : \sum_{i=1}^k \|Z_i \mathbf{v}\|_2 \geq \sum_{i=k+1}^n \|Z_i \mathbf{v}\|_2 \right) \right] \\ &= \max_t \left[ P \left( \exists \mathbf{e} : \|\mathbf{e}\|_2 \leq \epsilon \text{ and } \sum_{i=1}^k \|Z_i(\mathbf{z}_t + \mathbf{e})\|_2 \geq \sum_{i=k+1}^n \|Z_i(\mathbf{z}_t + \mathbf{e})\|_2 \right) \right]. \end{aligned} \quad (12)$$

Given the symmetry of the problem (i.e., the rotational invariance of the  $Z_i$ ) it should be noted that, without loss of generality, we can assume  $\mathbf{z}_t = [\|\mathbf{z}_t\|_2, 0, 0, \dots, 0]$ . Then clearly the maximization over  $t$  in (12) can be omitted. To further simplify notation we set  $\mathbf{z}_t = \mathbf{z}_0 = [\|\mathbf{z}_0\|_2, 0, 0, \dots, 0]$  Further, using the results from [6] we have that  $\eta^{d(n-m)-1}$  points can be located on the sphere of radius  $c\epsilon$  centered at  $\mathbf{z}_0$  such that  $S_0$  (which is centered at  $\mathbf{z}_0$  and lies in a  $(d(n-m) - 1)$ -dimensional space, and whose radius is  $\epsilon$ ) is inside a polytope determined by them and

$$c \leq \begin{cases} \frac{1}{(1 - \ln(\eta)) \sqrt{2 \ln(\eta) - \frac{\ln(d(n-m)-1)}{d(n-m)-1}}}, & \text{if } \eta < \sqrt{2} \\ \frac{1}{1 - (1 + \frac{1}{\eta^2})^{\frac{1}{2\eta^2}}}, & \text{otherwise.} \end{cases} \quad (13)$$

To get a feeling for what values  $\eta$  and  $c$  can take we refer to [5] where it was stated that  $3^{d(n-m)-1}$  points can be located on

the sphere of radius  $\sqrt{(9)/(8)}\epsilon$  centered at  $\mathbf{z}_0$  such that  $S_0$  (the sphere of radius  $\epsilon$  centered at  $\mathbf{z}_0$ ) is inside a polytope determined by them.

Let us call the polytope determined by  $\eta^{d(n-m)-1}$  points  $P_0$ . Let  $\mathbf{e}^s, s = 1, 2, \dots, \eta^{d(n-m)-1}$  be its  $\eta^{d(n-m)-1}$  corner points. Since  $\|Z_i \mathbf{z}_0\|_2 - \|Z_i \mathbf{e}\|_2 \leq \|Z_i(\mathbf{z}_0 + \mathbf{e})\|_2$ , and  $S_0 \subset P_0$  we have

$$\begin{aligned} & P \left( \exists \mathbf{e}, \|\mathbf{e}\|_2 \leq \epsilon \text{ s.t. } \sum_{i=1}^k \|Z_i(\mathbf{z}_0 + \mathbf{e})\|_2 \geq \sum_{i=k+1}^n \|Z_i(\mathbf{z}_0 + \mathbf{e})\|_2 \right) \\ & \leq P \left( \exists \mathbf{e}, (\mathbf{z}_0 + \mathbf{e}) \in P_0 \text{ s.t. } \sum_{i=1}^k \|Z_i(\mathbf{z}_0 + \mathbf{e})\|_2 \geq \sum_{i=k+1}^n (\|Z_i \mathbf{z}_0\|_2 - \|Z_i \mathbf{e}\|_2) \right) \\ & \leq P \left( \max_s \left( \sum_{i=k+1}^n \|Z_i \mathbf{e}^s\|_2 + \sum_{i=1}^k \|Z_i(\mathbf{z}_0 + \mathbf{e}^s)\|_2 \geq \sum_{i=k+1}^n \|Z_i \mathbf{z}_0\|_2 \right) \right) \end{aligned} \quad (14)$$

where the second inequality follows from the property that the maximum of a convex function over a polytope is achieved at its corner points and that function inside the  $\max_s$  is convex as it is a sum of convex norms. Connecting (11), (12), and (14), we obtain

$$P_f \leq \frac{\binom{n}{k}}{\epsilon^{d(n-m)}} P \left( \max_s \left( \sum_{i=k+1}^n \|Z_i \mathbf{e}^s\|_2 + \sum_{i=1}^k \|Z_i(\mathbf{z}_0 + \mathbf{e}^s)\|_2 \geq \sum_{i=k+1}^n \|Z_i \mathbf{z}_0\|_2 \right) \right). \quad (15)$$

Using the union bound over  $s$ , we further have

$$\begin{aligned} & P \left( \max_s \left( \sum_{i=k+1}^n \|Z_i \mathbf{e}^s\|_2 + \sum_{i=1}^k \|Z_i(\mathbf{z}_0 + \mathbf{e}^s)\|_2 \geq \sum_{i=k+1}^n \|Z_i \mathbf{z}_0\|_2 \right) \right) \\ & \leq \sum_{s'=1}^{\eta^{d(n-m)-1}} P \left( \left( \sum_{i=k+1}^n \|Z_i \mathbf{e}^{s'}\|_2 + \sum_{i=1}^k \|Z_i(\mathbf{z}_0 + \mathbf{e}^{s'})\|_2 \geq \sum_{i=k+1}^n \|Z_i \mathbf{z}_0\|_2 \right) \right). \end{aligned} \quad (16)$$

Given that only the first component of  $\mathbf{z}_0$  is not equal to zero and the symmetry of the problem we can write

$$\begin{aligned} & \sum_{s'=1}^{\eta^{d(n-m)-1}} P \left( \left( \sum_{i=k+1}^n \|Z_i \mathbf{e}^{s'}\|_2 + \sum_{i=1}^k \|Z_i(\mathbf{z}_0 + \mathbf{e}^{s'})\|_2 \right) \right. \\ & \quad \left. \geq \sum_{i=k+1}^n \|Z_i \mathbf{z}_0\|_2 \right) \\ & \leq \eta^{d(n-m)-1} \max_{s'} P \left( \left( \sum_{i=k+1}^n \left\| \sum_{j=2}^{d(n-m)} Z_{ij}(\mathbf{e}^{s'})_j \right\|_2 \right. \right. \\ & \quad \left. \left. + \sum_{i=1}^k \|Z_i(\mathbf{z}_0 + \mathbf{e}^{s'})\|_2 \right) \right. \\ & \quad \left. \geq \sum_{i=k+1}^n \|Z_{i1}\|_2 (\|\mathbf{z}_0\|_2 - |(\mathbf{e}^{s'})_1|) \right) \end{aligned} \quad (17)$$

where  $(\mathbf{e}^{s'})_j$  denotes  $j$ th components of  $\mathbf{e}^{s'}$ . Let  $B_l = Z_l(\mathbf{z}_0 + \mathbf{e}^{s'})_1$ ,  $1 \leq l \leq k$ ,  $C_i = Z_{i1}(\|\mathbf{z}_0\|_2 - |(\mathbf{e}^{s'})_1|)$ , and  $D_i = \sum_{j=2}^{d(n-m)} Z_{ij}(\mathbf{e}^{s'})_j$ . Clearly,  $B_l$ ,  $C_i$ , and  $D_i$  are independent zero-mean Gaussian random vectors of length  $d$ . Then we can rewrite (17) as

$$\begin{aligned} & \sum_{s'=1}^{\eta^{d(n-m)-1}} P \left( \left( \sum_{i=k+1}^n \|Z_i \mathbf{e}^{s'}\|_2 + \sum_{i=1}^k \|Z_i(\mathbf{z}_0 + \mathbf{e}^{s'})\|_2 \right) \right. \\ & \quad \left. \geq \sum_{i=k+1}^n \|Z_i \mathbf{z}_0\|_2 \right) \\ & \leq (\eta)^{d(n-m)-1} \max_{s'} P \left( \sum_{i=k+1}^n \|D_i\|_2 \right. \\ & \quad \left. + \sum_{l=1}^k \|B_l\|_2 \geq \sum_{i=k+1}^n \|C_i\|_2 \right). \end{aligned} \quad (18)$$

Let  $B_{l_p}$ ,  $C_{i_p}$ , and  $D_{i_p}$  denote the  $p$ th components of the vectors  $B_l$ ,  $C_i$ ,  $D_i$ , respectively. Then for any  $1 \leq p \leq d$  it holds

$$\begin{aligned} \text{var}(B_{l_p}) &= \|\mathbf{z}_0 + \mathbf{e}^{s'}\|_2^2 = 1 - \epsilon^2 + c^2 \epsilon^2, \\ \text{var}(C_{i_p}) &= (\|\mathbf{z}_0\|_2 - |(\mathbf{e}^{s'})_1|)^2, \\ \text{var}(D_{i_p}) &= \|\mathbf{e}^{s'}\|_2^2 - |(\mathbf{e}^{s'})_1|^2. \end{aligned}$$

Let  $G_i$ ,  $k+1 \leq i \leq n$ ,  $F_i$ ,  $k+1 \leq i \leq n$  be independent zero-mean Gaussian random vectors such that for any  $1 \leq p \leq d$

$$\begin{aligned} \text{var}(G_{i_p}) &= (\|\mathbf{z}_0\|_2 - \|\mathbf{e}^{s'}\|_2)^2 \\ &= (\|\mathbf{z}_0\|_2 - c\epsilon)^2 = (\sqrt{1 - \epsilon^2} - c\epsilon)^2 \\ \text{var}(F_{i_p}) &= \|\mathbf{e}^{s'}\|_2^2 = c^2 \epsilon^2 \end{aligned}$$

where we used the fact that  $\|\mathbf{e}^{s'}\|_2 = c\epsilon$ . Since  $\text{var}(G_{i_p}) \leq \text{var}(C_{i_p})$ , and  $\text{var}(F_{i_p}) \geq \text{var}(D_{i_p})$  we have from (18)

$$\begin{aligned} & \eta^{d(n-m)-1} \max_{s'} P \left( \sum_{i=k+1}^n \|D_i\|_2 + \sum_{l=1}^k \|B_l\|_2 \geq \sum_{i=k+1}^n \|C_i\|_2 \right) \\ & \leq \eta^{d(n-m)-1} \max_{s'} P \left( \sum_{i=k+1}^n \|F_i\|_2 + \sum_{l=1}^k \|B_l\|_2 \right. \\ & \quad \left. \geq \sum_{i=k+1}^n \|G_i\|_2 \right). \end{aligned} \quad (19)$$

Since  $\|\mathbf{e}^{s'}\|_2$  does not depend on  $s'$ , the outer maximization can be omitted. Using the Chernoff bound we further have

$$\begin{aligned} & \eta^{d(n-m)-1} P \left( \sum_{l=1}^k \|B_l\|_2 \geq \sum_{i=k+1}^n (\|G_i\|_2 - \|F_i\|_2) \right) \\ &= \eta^{d(n-m)-1} P \left( \sum_{l=1}^k \|B_l\|_2 - \sum_{i=k+1}^n \|G_i\|_2 \right. \\ & \quad \left. + \sum_{i=k+1}^n \|F_i\|_2 \geq 0 \right) \\ &= \eta^{d(n-m)-1} P \left( \mu \sum_{l=1}^k \|B_l\|_2 - \mu \sum_{i=k+1}^n \|G_i\|_2 \right. \\ & \quad \left. + \mu \sum_{i=k+1}^n \|F_i\|_2 \geq 0 \right) \\ & \leq \eta^{d(n-m)-1} \\ & \quad \times E e^{\mu \sum_{l=1}^k \|B_l\|_2 - \mu \sum_{i=k+1}^n \|G_i\|_2 + \mu \sum_{i=k+1}^n \|F_i\|_2} \\ & \leq \eta^{d(n-m)-1} \\ & \quad \times E e^{\mu \sum_{l=1}^k \|B_l\|_2} e^{-\mu \sum_{i=k+1}^n \|G_i\|_2} e^{\mu \sum_{i=k+1}^n \|F_i\|_2} \\ & \leq \eta^{d(n-m)-1} (E e^{\mu \|B_1\|_2})^k \\ & \quad \times (E e^{-\mu \|G_1\|_2})^{n-k} (E e^{\mu \|F_1\|_2})^{n-k} \end{aligned} \quad (20)$$

where  $\mu$  is any positive constant and  $B_1$ ,  $G_1$ , and  $F_1$  are independent zero-mean Gaussian random vectors such that for any  $1 \leq p \leq d$ ,  $\text{var}(B_{1_p}) = \|\mathbf{z}_0 + \mathbf{e}^{s'}\|_2^2 = 1 - \epsilon^2 + c^2 \epsilon^2$ ,  $\text{var}(G_{1_p}) = (\|\mathbf{z}_0\|_2 - c\epsilon)^2 = (\sqrt{1 - \epsilon^2} - c\epsilon)^2$ ,  $\text{var}(F_{1_p}) = c^2 \epsilon^2$ . Connecting (15)–(20), we have

$$P_f \leq \binom{n}{k} \frac{1}{\eta} \left( \frac{\eta}{\epsilon} \right)^{d(n-m)} (E e^{\mu \|B_1\|_2})^k \left( \frac{E e^{-\mu \|G_1\|_2}}{(E e^{\mu \|F_1\|_2})^{-1}} \right)^{n-k}. \quad (21)$$

After setting  $k = \beta n$ ,  $m = \alpha n$ , and using the fact that  $\binom{n}{k} \approx e^{-nH(\beta)}$ , we finally obtain

$$\lim_{n \rightarrow \infty} P_f \leq \lim_{n \rightarrow \infty} \xi^n \quad (22)$$

where

$$\xi = \frac{(\eta/\epsilon)^{d(1-\alpha)}}{e^{H(\beta)}} (E e^{\mu \|B_1\|_2})^\beta \left( \frac{E e^{-\mu \|G_1\|_2}}{(E e^{\mu \|F_1\|_2})^{-1}} \right)^{1-\beta}. \quad (23)$$

and  $H(\beta) = \beta \ln \beta + (1 - \beta) \ln(1 - \beta)$ . We now set  $\mu = \sqrt{2d-1} \delta \sqrt{2}$ ,  $\delta \ll 1$ . In the appendixes, we will determine  $E e^{\sqrt{2d-1} \delta \sqrt{2} \|B_1\|_2}$ ,  $E e^{\sqrt{2d-1} \delta \sqrt{2} \|F_1\|_2}$ , and  $E e^{-\sqrt{2d-1} \delta \sqrt{2} \|G_1\|_2}$ .

We now return to the analysis of (23). Replacing the results from (37), (38), and (44) in (23), we finally have

$$\xi \approx \frac{(\eta/\epsilon)^{d(1-\alpha)}}{e^{H(\beta)}} \left( e^{d((\delta b)^2 + 2\delta b)} \right)^\beta \times \left( e^{d((\delta f)^2 + 2\delta f)} \right)^{1-\beta} \left( e^{d((\delta g)^2 - 2\delta g)} \right)^{1-\beta} \quad (24)$$

where we recall that  $b = \sqrt{1 - \epsilon^2 + c^2 \epsilon^2}$ ,  $f = c\epsilon$ , and  $g = \sqrt{1 - \epsilon^2} - c\epsilon$ . Our goal is to find  $d$  such that for  $\alpha = 1 - \gamma$ ,  $0 < \gamma \ll 1$  and  $\beta = (1/2) - \sigma$ ,  $0 < \sigma \ll (1/2)$ ,  $\xi < 1$ . That means we need

$$\ln(\xi) < 0 \quad (25)$$

which implies

$$d(1 - \alpha) \ln\left(\frac{\eta}{\epsilon}\right) + 2d\delta(\beta b + (1 - \beta)f - (1 - \beta)g) + d\delta^2(\beta b^2 + (1 - \beta)f^2 + (1 - \beta)g^2) < H(\beta). \quad (26)$$

To simplify the exposition below, let

$$\beta_{\text{opt}} = \frac{g - f}{g + b - f} \approx \frac{1 - 2c\epsilon}{2 - 2c\epsilon}. \quad (27)$$

Combining the previous results, the following theorem then can be proved.

**Theorem 4:** Assume that the matrix  $A$  has an isotropically distributed null-space and that the number of rows of the matrix  $A$  is  $dm = \alpha dn$ . Fix constants  $c$  and  $\eta$  according to (13) and arbitrarily small numbers  $\epsilon$  and  $\delta$ , say  $\epsilon < 10^{-5}$  and  $\delta = \epsilon^2$ . Let  $b = \sqrt{1 - \epsilon^2 + c^2 \epsilon^2}$ ,  $f = c\epsilon$ , and  $g = \sqrt{1 - \epsilon^2} - c\epsilon$ . Choose  $\beta < \beta_{\text{opt}}$  where  $\beta_{\text{opt}}$  is given by (27). If [see (28) shown at the bottom of the page], then with overwhelming probability for all block-sparse  $\mathbf{x}$  that have blocks of length  $d$  and sparsity  $k < \beta n$ , the solutions to the optimizations (1) and (3) coincide.

*Proof:* Follows from the previous discussion combining (8), (22), (23), (24), (25), and (26). ■

Before moving on to the numerical study of the performance of the algorithm (3), we should also mention that the results from Theorem 4 as well as the theoretical results from [16], [19], and [23] are related to what is often called the *strong threshold* (the interested reader can find more on the definition of the strong threshold in [16], [19], and [21]) for sparsity. As we have said earlier, if the number of the measurements is  $M = \alpha N$  then the strong threshold for sparsity is ideally  $K = \alpha/(2)N$ . Also, the definition of the strong threshold assumes that the reconstructing algorithm [(2), (3), or any other] succeeds for *any* sparse signal

with sparsity below the strong threshold. However, since this can not be numerically verified (we simply can not generate all possible  $k$ -block sparse signals from  $\mathbb{R}^{dn}$ ), a weaker notion of the threshold (called the *weak threshold*) is usually considered in numerical experiments (the interested reader can also find more on the definition of the weak threshold in [16]). The main feature of the weak threshold definition is that it allows failure in reconstruction of a certain small fraction of signals with sparsity below it. However, as expected, the ideal performance in the sense of the weak threshold assumes that if the number of the measurements is  $M = \alpha N$  and the sparsity is  $K = \beta N$ , then  $\beta$  should approach  $\alpha$ . As the numerical experiments in the following sections hint, increasing the block length  $d$  leads to almost ideal performance of the reconstructing technique introduced in this paper.

#### IV. NUMERICAL STUDY OF THE BLOCK SPARSE RECONSTRUCTION

In this section we recall the basics of the algorithm, show how it can efficiently be solved in polynomial time, and demonstrate its performance through numerical simulations.

In order to recover a  $k$ -block sparse signal  $\mathbf{x}$  from the linear measurements  $\mathbf{y} = A\mathbf{x}$  we consider the following optimization problem:

$$\begin{aligned} \min_{\mathbf{x}} \quad & \|\mathbf{X}_1\|_2 + \|\mathbf{X}_2\|_2 + \dots + \|\mathbf{X}_n\|_2 \\ \text{subject to} \quad & A\mathbf{x} = \mathbf{y} \end{aligned} \quad (29)$$

where  $\mathbf{X}_i = (\mathbf{x}_{(i-1)d+1}, \mathbf{x}_{(i-1)d+2}, \dots, \mathbf{x}_{id})^T$ , for  $i = 1, 2, \dots, n$ . Since the objective function is convex this is clearly a convex optimization problem. Since this problem is convex it is solvable in polynomial time. Furthermore, we can transform it to a bit more convenient form in the following way:

$$\begin{aligned} \min_{\mathbf{x}, t_1, t_2, \dots, t_n} \quad & \sum_{i=1}^n t_i \\ \text{subject to} \quad & \|\mathbf{X}_i\|_2^2 \leq t_i^2, \quad t_i \geq 0, \quad 1 \leq i \leq n \\ & A\mathbf{x} = \mathbf{y} \end{aligned} \quad (30)$$

where as earlier  $\mathbf{X}_i = (\mathbf{x}_{(i-1)d+1}, \mathbf{x}_{(i-1)d+2}, \dots, \mathbf{x}_{id})^T$ , for  $i = 1, 2, \dots, n$ . This is a second-order cone programming (SOCP) optimization problem which can be solved by a host of numerical methods [2]. Finally, it is not that difficult to see that (30) can be transformed to

$$\begin{aligned} \min_{\mathbf{x}, t_1, t_2, \dots, t_n} \quad & \sum_{i=1}^n t_i \\ \text{subject to} \quad & \begin{bmatrix} t_i I & \mathbf{X}_i^* \\ \mathbf{X}_i & t_i \end{bmatrix} \geq 0, \quad t_i \geq 0, \quad 1 \leq i \leq n \\ & A\mathbf{x} = \mathbf{y} \end{aligned} \quad (31)$$

---


$$\begin{aligned} d &> \frac{H(\beta) - \ln\left(\frac{\eta}{\epsilon}\right)}{2\delta(\beta b + (1 - \beta)f - (1 - \beta)g) + \delta^2(\beta b^2 + (1 - \beta)f^2 + (1 - \beta)g^2)} \text{ and} \\ \alpha &> 1 - \frac{1}{d} \end{aligned} \quad (28)$$

with  $\mathbf{X}_i = (\mathbf{x}_{(i-1)d+1}, \mathbf{x}_{(i-1)d+2}, \dots, \mathbf{x}_{id})^T$ , for  $i = 1, 2, \dots, n$ . Clearly, (31) is a semi-definite program and can be solved by a host of numerical methods in polynomial time. From the numerical point of view, it is more efficient to solve (30) than (31).

To further improve the reconstruction performance we introduce an additional modification of (31). Assume that  $\tilde{\mathbf{X}}_i$ ,  $1 \leq i \leq n$  is the solution of (31). Further, sort  $\|\tilde{\mathbf{X}}_i\|_2$  and assume that  $\hat{\mathcal{K}}$  is the set of  $k$  indexes which correspond to the  $k$  vectors  $\mathbf{X}_i$  with the largest norm. Let these indexes determine the positions of the nonzero blocks. Then let  $A_{\hat{\mathcal{K}}}$  be the submatrix of  $A$  obtained by selecting the columns with the indexes  $\hat{\mathcal{K}}$  from the first  $kd$  rows of  $A$ . Also let  $\mathbf{y}_{\hat{\mathcal{K}}}$  be the first  $kd$  components of  $\mathbf{y}$ . Then we generate the nonzero part of the reconstructed signal  $\hat{\mathbf{x}}$  as  $\hat{\mathbf{x}}_{\hat{\mathcal{K}}} = A_{\hat{\mathcal{K}}}^{-1} \mathbf{y}_{\hat{\mathcal{K}}}$ . We refer to this procedure of reconstructing the sparse signal  $\mathbf{x}$  as  $\ell_2/\ell_1$  algorithm and in the following subsection we show its performance. Also, we mention that to fully implement the above described algorithm one needs to know the sparsity value  $K$  (this is in a sharp contrast with solving (31) which does not require the knowledge of  $K$ ). The necessity of using the knowledge of  $K$  can be avoided if one implements a bit more sophisticated interior-point methods for solving (31). However, a practical implementation of an interior-point optimization algorithm goes beyond the scope of this paper, and we therefore do not pursue it further here.

## Recovery of block-sparse signals

**Input:** Measured vector  $\mathbf{y} \in \mathbb{R}^m$ , size of blocks  $d$ , and measurement matrix  $A$ .

**Output:** Block-sparse signal  $\mathbf{x} \in \mathbb{R}^n$ .

- 1: Solve the following optimization problem

$$\begin{aligned} \min_{\mathbf{x}} \quad & \|\mathbf{X}_1\|_2 + \|\mathbf{X}_2\|_2 + \dots + \|\mathbf{X}_n\|_2 \\ \text{subject to} \quad & A\mathbf{x} = \mathbf{y} \end{aligned}$$

using semi-definite programming.

- 2: Sort  $\|\mathbf{X}_i\|_2$  for  $i = 1, 2, \dots, n$ , such that  $\|\mathbf{X}_{j_1}\|_2 \geq \|\mathbf{X}_{j_2}\|_2 \geq \dots \geq \|\mathbf{X}_{j_n}\|_2$ .
- 3: Set  $\bar{A}$  to be the submatrix of  $A$  containing columns of first  $kd$  rows of  $A$  that correspond to the blocks  $j_1, j_2, \dots, j_k$ .
- 4: Let  $\bar{\mathbf{x}}$  represent the nonzero blocks of  $\mathbf{x}$ . Set  $\bar{\mathbf{x}} = \bar{A}^{-1} \mathbf{y}$  and the rest of blocks of  $\mathbf{x}$  to zero.
- 5:  $\mathbf{x}$ .

### A. Simulation Results

In this section, we discuss the performance of the  $\ell_2/\ell_1$  algorithm. We conducted four numerical experiments for four different values of the block length  $d$ . In cases when  $d = 1, 4, 8$ , we set the length of the sparse vector to be  $N = 800$  and in the case  $d = 16$  we set  $N = 1600$ . For fixed values of  $d$  and  $N$  we then generated a random Gaussian measurement matrix  $A$  for  $\alpha \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$ . For each of these matrices we randomly generate 100 different signals of a given sparsity  $\beta$ , form a measurement vector  $\mathbf{y}$ , and run the  $\ell_2/\ell_1$  algorithm. The percentage of success (perfect recovery of the sparse signal) is shown in Fig. 4. The threshold values  $\beta/\alpha$  are shown in Table I.

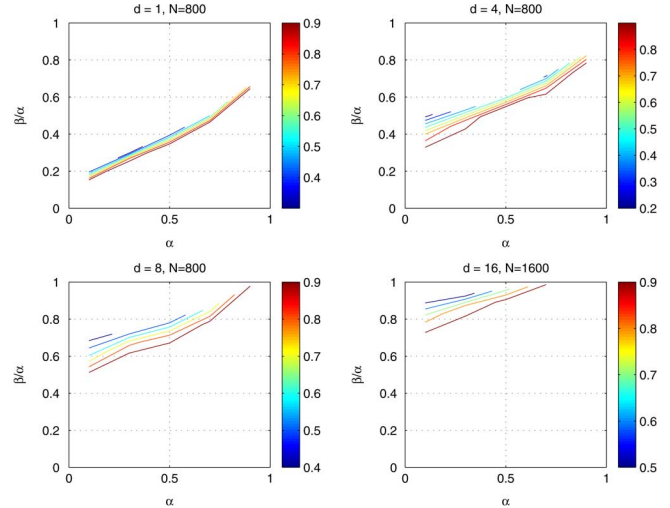


Fig. 4. Threshold for  $\beta$  for a given  $\alpha$  (the colors of the curves indicate the probability of success of  $\ell_2/\ell_1$  algorithm calculated over 100 independent instances of block sparse signals  $\mathbf{x}$  and a fixed Gaussian measurement matrix  $A$ );  $d$  is the length of each block of  $\mathbf{x}$ .

TABLE I  
THEORETICAL AND SIMULATION RESULTS FOR RECOVERY OF BLOCK-SPARSE SIGNALS WITH DIFFERENT BLOCK SIZE.  $\rho_S$  IS THE STRONG THRESHOLD FOR  $\ell_1$  OPTIMIZATION AND  $\rho_W$  IS THE WEAK THRESHOLD FOR  $\ell_1$  OPTIMIZATION [19], [21].  $d$  REPRESENTS THE BLOCK SIZE IN VARIOUS SIMULATIONS. THE DATA FOR  $\beta/\alpha$  ARE TAKEN FROM THE CURVES WITH PROBABILITY OF SUCCESS MORE THAN %95

	$\alpha = 0.1$	$\alpha = 0.3$	$\alpha = 0.5$	$\alpha = 0.7$	$\alpha = 0.9$
$\rho_S$	0.049	0.070	0.089	0.111	0.140
$\rho_W$	0.188	0.292	0.385	0.501	0.677
$d = 1, \beta/\alpha$	0.10	0.23	0.30	0.41	0.62
$d = 4, \beta/\alpha$	0.30	0.33	0.50	0.57	0.72
$d = 8, \beta/\alpha$	0.50	0.60	0.60	0.71	0.89
$d = 16, \beta/\alpha$	0.70	0.80	0.80	0.91	0.94

These values are taken from Fig. 4 as the values for which the probability of perfect recovery is at least 95%. The case  $d = 1$  corresponds to the basic  $\ell_1$  relaxation. As can be seen from Fig. 4, increasing the block length significantly improves the threshold for allowable sparsity.

## V. CONCLUSION

In this paper, we studied the efficient recovery of block sparse signals using an under-determined system of equations generated from random Gaussian matrices. Such problems arise in different applications, such as DNA microarrays, equalization of sparse communication channels, magnetoencephalography, etc. We analyzed the minimization of a mixed  $\ell_2/\ell_1$  type norm, which can be reduced to solving a semi-definite program (in fact it can be reduced to solving an SOCP—a special type of general semi-definite programs). We showed that, as the number of measurements approaches the number of unknowns, the  $\ell_2/\ell_1$  algorithm can uniquely recover all block-sparse signals whose sparsity is up to half the number of measurements with overwhelming probability over the measurement matrix. This coincides with the best that can be achieved via exhaustive search. Extensive numerical simulations demonstrated that  $\ell_2/\ell_1$  optimization (or its a variant) can be very successful when applied in recovery of block-sparse compressed signals. We believe that the proving technique presented in this paper is powerful and



potentially could be used to obtain similar thresholds for any number of measurements. However, due to inability to deal numerically with certain parabolic cylinder functions, our proof technique (which involves a certain union bound) appears to give a loose bound when the number of measurements is a fixed fraction of the number of unknowns.

For future work, it would be interesting to see if one could obtain “sharp” bounds on when signal recovery is possible (similar to the sharp bounds in [17]) for  $\ell_2/\ell_1$  method. In this paper, we were able to use some probabilistic arguments to show that, for a random Gaussian measurement matrix, (4) holds with overwhelming probability. In our proof we used a union bound to upper bound the probability that (4) fails; this makes our bound loose for  $\alpha$  less than one. We expect to get sharp thresholds for other values of  $\alpha$  by generalizing the idea of looking at the neighborliness of randomly projected simplices presented in [17], [20], and [21]. However, for relaxation in (3) instead of simplices we would have to work with the convex hull  $\mathcal{B}$  of  $nd$ -dimensional spheres. Specifically, one would need to compute the probability that a random  $h$ -dimensional affine plane that passes through a point on the boundary of  $\mathcal{B}$  will be inside the tangent cone of that given point. Solving this problem seems to be rather difficult.

APPENDIX I

COMPUTING  $E \left[ e^{\sqrt{2d-1}\delta\sqrt{2}\|B_1\|_2} \right]$  AND  $E \left[ e^{\sqrt{2d-1}\delta\sqrt{2}\|F_1\|_2} \right]$

Now we turn to computing  $E e^{\sqrt{2d-1}\delta\sqrt{2}\|B_1\|_2}$  and  $E e^{\sqrt{2d-1}\delta\sqrt{2}\|F_1\|_2}$ . Let us first consider  $E e^{\mu\|B_1\|_2}$ . Since  $B_1$  is a  $d$  dimensional vector let  $B_1 = [B_{1_1}, B_{1_2}, \dots, B_{1_d}]$ . As we have stated earlier  $B_{1_p}, 1 \leq p \leq d$  are i.i.d. zero-mean Gaussian random variables with variance  $\text{var}(B_{1_p}) = 1 - \epsilon^2 + \epsilon^2 \epsilon^2 = b^2, 1 \leq p \leq d$ . Then we can write

$$E e^{\sqrt{2d-1}\delta\sqrt{2}\|B_1\|_2} = \frac{1}{\sqrt{2\pi}^d} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \exp \left( \sqrt{2d-1}\delta\sqrt{2} b \sqrt{\sum_{p=1}^d B_{1_p}^2 - \frac{\sum_{p=1}^d B_{1_p}^2}{2}} \right) dB_1.$$

Using the spherical coordinates, it is not that difficult to show that the previous integral can be transformed to

$$\begin{aligned} & E e^{\sqrt{2d-1}\delta\sqrt{2}\|B_1\|_2} \\ &= \frac{1}{\sqrt{2\pi}^d} \frac{2\sqrt{\pi}^d}{\Gamma(\frac{d}{2})} \int_0^\infty r^{d-1} e^{\sqrt{2d-1}\delta\sqrt{2}br - \frac{r^2}{2}} dr \\ &= \frac{\Gamma(d) e^{\frac{(2d-1)(\delta b)^2}{2}} e^{-\frac{(2d-1)(\delta b)^2}{2}}}{\Gamma(\frac{d}{2}) 2^{\frac{d}{2}-1} \Gamma(d)} \\ &\quad \times \int_0^\infty r^{d-1} e^{\sqrt{2d-1}\delta\sqrt{2}br - \frac{r^2}{2}} dr \\ &= \frac{\Gamma(d) e^{\frac{(2d-1)(\delta b)^2}{2}}}{\Gamma(\frac{d}{2}) 2^{\frac{d}{2}-1}} U \left( \frac{2d-1}{2}, -\sqrt{2d-1}\delta\sqrt{2}b \right) \end{aligned} \quad (32)$$

where  $U$  is *parabolic cylinder function* (see, e.g., [44]). Before proceeding further we recall the asymptotic results for  $U$  from [44]. Namely, from [44] we have that if  $\zeta \gg 0$  and  $t \geq 0$

$$U \left( \frac{\zeta^2}{2}, -\zeta t\sqrt{2} \right) \approx \frac{h(\zeta) e^{\zeta^2 \tilde{\rho}} \sqrt{2\pi}}{\Gamma \left( \frac{\zeta^2+1}{2} \right) (t^2+1)^{\frac{1}{4}}} \quad (33)$$

where

$$\begin{aligned} h(\zeta) &= 2^{-\frac{\zeta^2}{4} - \frac{1}{4}} e^{-\frac{\zeta^2}{4}} \zeta^{\frac{\zeta^2}{2} - \frac{1}{2}}, \\ \tilde{\rho} &= \frac{1}{2} (t\sqrt{1+t^2} + \ln(t + \sqrt{1+t^2})). \end{aligned} \quad (34)$$

From (33) and (34) we have

$$\begin{aligned} & U \left( \frac{2d-1}{2}, -\sqrt{2d-1}\delta b\sqrt{2} \right) \\ &\approx \frac{1}{\Gamma(d)} \left( (2e)^{-\frac{2d-1}{4}} \sqrt{2d-1}^{\frac{2d-1}{2} - \frac{1}{2}} 2^{-\frac{1}{4}} \right) \\ &\quad \times \frac{e^{\frac{2d-1}{2} 2 \left( \frac{1}{2} (\delta b \sqrt{1+(\delta b)^2} + \ln(\delta b + \sqrt{1+(\delta b)^2})) \right)}}{(1 + (\delta b)^2)^{\frac{1}{4}}}. \end{aligned} \quad (35)$$

Connecting (32) and (35), we finally obtain for  $d \gg 0$  and  $\delta \ll 1$  ( $\delta$  is a constant independent of  $d$ )

$$\begin{aligned} & E e^{\sqrt{2d-1}\delta\sqrt{2}\|B_1\|_2} \\ &\approx \frac{e^{\frac{2d-1}{2} (\delta b)^2}}{\Gamma \left( \frac{d}{2} \right) 2^{\frac{d}{2}-1}} \left( (2e)^{-\frac{2d-1}{4}} \sqrt{2d-1}^{\frac{2d-1}{2} - \frac{1}{2}} 2^{-\frac{1}{4}} \right) \\ &\quad \times \frac{e^{\frac{2d-1}{2} 2 \left( \frac{1}{2} (\delta b \sqrt{1+(\delta b)^2} + \ln(\delta b + \sqrt{1+(\delta b)^2})) \right)}}{(1 + (\delta b)^2)^{\frac{1}{4}}}. \end{aligned} \quad (36)$$

Using the facts that  $\delta b \ll 1$  and  $\Gamma((d/2)) \approx ((d/2)e)^{(d/2)}$  when  $d$  is large, (36) can be rewritten as

$$E e^{\sqrt{2d-1}\delta\sqrt{2}\|B_1\|_2} \approx e^{\frac{2d-1}{2} 2 (\delta b)^2} \times e^{\frac{2d-1}{2} 2 \left( \frac{1}{2} (\delta b \sqrt{1+(\delta b)^2} + \ln(\delta b + \sqrt{1+(\delta b)^2})) \right)}.$$

Since  $\delta b \ll 1$  it further follows

$$\begin{aligned} & E e^{\sqrt{2d-1}\delta\sqrt{2}\|B_1\|_2} \\ &\approx e^{\frac{2d-1}{2} ((\delta b)^2 + \frac{2}{3} (\delta b \sqrt{1+(\delta b)^2} + \ln(\delta b + \sqrt{1+(\delta b)^2})))} \\ &\approx e^{\frac{2d-1}{2} ((\delta b)^2 + \frac{2}{3} (\delta b (1 + \frac{(\delta b)^2}{2}) + \ln(\delta b + (1 + \frac{(\delta b)^2}{2})))} \\ &\approx e^{\frac{2d-1}{2} ((\delta b)^2 + \frac{2}{3} (\delta b (1 + \frac{(\delta b)^2}{2}) + \delta b)} \\ &\approx e^{\frac{2d-1}{2} ((\delta b)^2 + 2\delta b)} \\ &\approx e^{d((\delta b)^2 + 2\delta b)}. \end{aligned} \quad (37)$$

To compute  $E e^{\mu\|F_1\|_2}$  we first note that  $F_1$  is a  $d$  dimensional vector. Let  $F_1 = [F_{1_1}, F_{1_2}, \dots, F_{1_d}]$ . As we have stated earlier  $F_{1_p}, 1 \leq p \leq d$  are i.i.d. zero-mean Gaussian random variables with variance  $\text{var}(F_{1_p}) = \epsilon^2 \epsilon^2 = f^2, 1 \leq p \leq d$ . Then the rest of the derivation for computing  $E e^{\sqrt{2d-1}\delta\sqrt{2}\|F_1\|_2}$  follows directly as in the case of  $E e^{\sqrt{2d-1}\delta\sqrt{2}\|B_1\|_2}$ . Hence, we can write similarly to (37)

$$E e^{\sqrt{2d-1}\delta\sqrt{2}\|F_1\|_2} \approx e^{d((\delta f)^2 + 2\delta f)}. \quad (38)$$

APPENDIX II

COMPUTING  $E \left[ e^{-\sqrt{2d-1}\delta\sqrt{2}\|G_1\|_2} \right]$

Now we turn to computing  $E e^{-\sqrt{2d-1}\delta\sqrt{2}\|G_1\|_2}$ . Since  $G_1$  is a  $d$  dimensional vector let  $G_1 = [G_{1_1}, G_{1_2}, \dots, G_{1_d}]$ . As we have stated earlier  $G_{1_p}, 1 \leq p \leq d$  are i.i.d. zero-mean Gaussian random variables with variance  $\text{var}(G_{1_p}) = (\sqrt{1-\epsilon^2} - \epsilon\epsilon)^2 = g^2, 1 \leq p \leq d$ . Then we can write

$$\begin{aligned} & E e^{-\sqrt{2d-1}\delta\sqrt{2}\|G_1\|_2} = \frac{1}{\sqrt{2\pi}^d} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \\ &\quad \times \exp \left( -\sqrt{2d-1}\delta\sqrt{2}g \sqrt{\sum_{p=1}^d G_{1_p}^2 - \frac{\sum_{p=1}^d G_{1_p}^2}{2}} \right) dG_1. \end{aligned}$$

Similarly as in the previous subsection using the spherical coordinates it is not that difficult to show that the previous integral can be transformed to

$$\begin{aligned}
& Ee^{-\sqrt{2d-1}\delta\sqrt{2}\|G_1\|_2} \\
&= \frac{1}{\sqrt{2\pi}^d} \frac{2\sqrt{\pi}^d}{\Gamma(\frac{d}{2})} \int_0^\infty r^{d-1} e^{-\sqrt{2d-1}\delta\sqrt{2}gr - \frac{r^2}{2}} dr \\
&= \frac{\Gamma(d)e^{\frac{(2d-1)(\delta g)^2}{2}}}{\Gamma(\frac{d}{2})2^{\frac{d}{2}-1}} \frac{e^{-\frac{(2d-1)(\delta g)^2}{2}}}{\Gamma(d)} \\
&\quad \times \int_0^\infty r^{d-1} e^{-\sqrt{2d-1}\delta\sqrt{2}gr - \frac{r^2}{2}} dr \\
&= \frac{\Gamma(d)e^{\frac{(2d-1)(\delta g)^2}{2}}}{\Gamma(\frac{d}{2})2^{\frac{d}{2}-1}} U\left(\frac{2d-1}{2}, \sqrt{2d-1}\delta\sqrt{2}g\right) \quad (39)
\end{aligned}$$

where as earlier  $U$  is *parabolic cylinder function*. Before proceeding further we again recall another set of the asymptotic results for  $U$  from [44]. Namely, from [44] we have that if  $\zeta \gg 0$  and  $t \geq 0$

$$U\left(\frac{\zeta^2}{2}, \zeta t\sqrt{2}\right) \approx \frac{\tilde{h}(\zeta)e^{-\zeta^2\tilde{p}}}{(t^2+1)^{\frac{1}{4}}} \quad (40)$$

where

$$\begin{aligned}
\tilde{h}(\zeta) &= 2^{\frac{\zeta^2}{4}-\frac{1}{4}} e^{\frac{\zeta^2}{4}} \zeta^{-\frac{\zeta^2}{2}-\frac{1}{2}}, \\
\tilde{p} &= \frac{1}{2}(t\sqrt{1+t^2} + \ln(t + \sqrt{1+t^2})). \quad (41)
\end{aligned}$$

From (40) and (41) we have

$$\begin{aligned}
& U\left(\frac{2d-1}{2}, \sqrt{2d-1}\delta g\sqrt{2}\right) \\
&\approx \left((2e)^{\frac{2d-1}{4}} \sqrt{2d-1}^{-\frac{2d-1}{2}-\frac{1}{2}} 2^{-\frac{1}{4}}\right) \\
&\quad \times \frac{e^{-\frac{2d-1}{2}2\left(\frac{1}{2}(\delta g\sqrt{1+(\delta g)^2} + \ln(\delta g + \sqrt{1+(\delta g)^2}))\right)}}{(1+(\delta g)^2)^{\frac{1}{4}}}. \quad (42)
\end{aligned}$$

Connecting (39) and (42) we finally obtain for  $d \gg 0$  and  $\delta \ll 1$  (as earlier  $\delta$  is a constant independent of  $d$ )

$$\begin{aligned}
& Ee^{-\sqrt{2d-1}\delta\sqrt{2}\|G_1\|_2} \\
&\approx \frac{\Gamma(d)e^{\frac{2d-1}{2}(\delta g)^2}}{\Gamma(\frac{d}{2})2^{\frac{d}{2}-1}} \left((2e)^{\frac{2d-1}{4}} \sqrt{2d-1}^{-\frac{2d-1}{2}-\frac{1}{2}} 2^{-\frac{1}{4}}\right) \\
&\quad \times \frac{e^{-\frac{2d-1}{2}2\left(\frac{1}{2}(\delta g\sqrt{1+(\delta g)^2} + \ln(\delta g + \sqrt{1+(\delta g)^2}))\right)}}{(1+(\delta g)^2)^{\frac{1}{4}}}. \quad (43)
\end{aligned}$$

Using the facts that  $\Gamma((d/2)) \approx ((d/2e))^{(d/2)}$  and  $\Gamma(d) \approx ((d/e))^d$  when  $d$  is large, (43) can be rewritten as

$$Ee^{-\sqrt{2d-1}\delta\sqrt{2}\|G_1\|_2} \approx e^{-\frac{2d-1}{2}(\delta g)^2} \frac{e^{\frac{2d-1}{2}2\left(\frac{1}{2}(\delta g\sqrt{1+(\delta g)^2} + \ln(\delta g + \sqrt{1+(\delta g)^2}))\right)}}{(1+(\delta g)^2)^{\frac{1}{4}}}.$$

Since  $\delta g \ll 1$  it further follows

$$\begin{aligned}
& Ee^{-\sqrt{2d-1}\delta\sqrt{2}\|G_1\|_2} \\
&\approx e^{\frac{2d-1}{2}((\delta g)^2 - \frac{2}{2}(\delta g\sqrt{1+(\delta g)^2} + \ln(\delta g + \sqrt{1+(\delta g)^2})))} \\
&\approx e^{\frac{2d-1}{2}((\delta g)^2 - \frac{2}{2}(\delta g(1 + \frac{(\delta g)^2}{2}) + \ln(\delta g + (1 + \frac{(\delta g)^2}{2})))} \\
&\approx e^{\frac{2d-1}{2}((\delta g)^2 - \frac{2}{2}(\delta g(1 + \frac{(\delta g)^2}{2}) + \delta g)} \\
&\approx e^{\frac{2d-1}{2}((\delta g)^2 - 2\delta g)} \\
&\approx e^{d((\delta g)^2 - 2\delta g)}. \quad (44)
\end{aligned}$$

## REFERENCES

- [1] M. Akçakaya and V. Tarokh, "A frame construction and a universal distortion bound for sparse representations," *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2443–2450, Jun. 2008.
- [2] F. Alizadeh and D. Goldfarb, "Second-order cone programming," *Math. Program.*, no. 95, pp. 3–51, 2003.
- [3] R. Baraniuk, V. Cevher, M. Duarte, and C. Hegde, Model-Based Compressive Sensing [Online]. Available: <http://www.dsp.ece.rice.edu/cs/>
- [4] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Construct. Approx.* [Online]. Available: <http://www.dsp.ece.rice.edu/cs, to be published>
- [5] I. Barany and N. Simanyi, "A note on the size of the largest ball inside a convex polytope," *Periodica Mathematica Hungarica*, vol. 51, no. 2, pp. 15–18, 2005.
- [6] K. Boroczky, Jr. and G. Wintsche, S. Basu, Ed., "Covering the sphere by equal spherical balls," *Discrete and Computational Geometry: The Goodman-Pollach Festschrift*, pp. 237–253, 2003.
- [7] E. Candes, "Compressive sampling," in *Proc. Int. Congr. Mathematics*, Madrid, Spain, 2006, vol. 3, pp. 1433–1452.
- [8] E. J. Candés, 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, Feb. 2006.
- [9] E. J. Candés, M. Rudelson, T. Tao, and R. Vershynin, "Error correction via linear programming," in *Proc. 46th Annu. IEEE Symp. Foundations Computer Science (FOCS)*, pp. 295–308.
- [10] E. Candes and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.
- [11] J. Chen and X. Huo, "Theoretical results on sparse representations of multiple-measurement vectors," *IEEE Trans. Signal Process.*, vol. 54, no. 12, pp. 4634–4643, Dec. 2006.
- [12] G. Cormode and S. Muthukrishnan, "Combinatorial algorithms for compressed sensing," in *Proc. 13th Colloquium on Structural Information and Communication Complexity (SIROCCO)*, Chester, U.K., Jul. 3–5, 2006, pp. 280–294.
- [13] S. Cotter and B. Rao, "Sparse channel estimation via matching pursuit with application to equalization," *IEEE Trans. Comm.*, vol. 50, no. 3, pp. 374–377, Mar. 2002.
- [14] S. Cotter, B. Rao, K. Engan, and K. Kreutz-Delgado, "Sparse solutions to linear inverse problems with multiple measurement vectors," *IEEE Trans. Signal Process.*, vol. 53, no. 7, pp. 2477–2488, Jul. 2005.
- [15] R. A. DeVore, "Optimal computation," in *Proc. Int. Congr. Mathematics*, Madrid, Spain, 2006, vol. I.
- [16] D. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [17] D. L. Donoho, "High-dimensional centrally-symmetric polytopes with neighborliness proportional to dimension," *Disc. Comput. Geometry* Dec. 2005 [Online]. Available: <http://www.springerlink.com/content/7253x6471m380t02/>
- [18] D. Donoho and X. Huo, "Uncertainty principles and ideal atomic decompositions," *IEEE Trans. Inf. Theory*, vol. 47, no. 7, pp. 2845–2862, Nov. 2001.
- [19] D. Donoho and J. Tanner, "Neighborliness of randomly-projected simplices in high dimensions," *Proc. Nat. Acad. Sci.*, vol. 102, no. 27, pp. 9452–9457, 2005.
- [20] D. Donoho and J. Tanner, "Sparse nonnegative solutions of underdetermined linear equations by linear programming," *Proc. Nat. Acad. Sci.*, vol. 102, no. 27, pp. 9446–9451, 2005.
- [21] D. Donoho and J. Tanner, "Thresholds for the recovery of sparse solutions via L1 minimization," in *Proc. Conf. Information Sciences Systems*, Mar. 2006.
- [22] I. Dumer, M. S. Pinsker, and V. V. Prelov, "On the thinnest coverings of the spheres and ellipsoids," in *General Theory of Information Transfer and Combinatorics*, ser. Lecture Notes in Computer Science. Berlin, Germany: Springer, 2006, vol. 4123, pp. 883–910.
- [23] C. Dwork, F. McSherry, and K. Talwar, "The price of privacy and the limits of LP decoding," in *Proc. 39th Annu. ACM Symp. Theory Computing (STOC)*, New York, 2007.
- [24] Y. C. Eldar and H. Bolcskei, "Block-sparsity: Coherence and efficient recovery," [Online]. Available: [arxiv:0812.0329](http://arxiv.org/abs/0812.0329)
- [25] S. Erickson and C. Sabatti, "Empirical Bayes estimation of a sparse vector of gene expression," *Statistical Applications in Genetics and Molecular Biology* 2005 [Online]. Available: <http://repositories.cdlib.org/uclastat/papers/2005020102/>
- [26] A. Feuer and A. Nemirovski, "On sparse representation in pairs of bases," *IEEE Trans. Inf. Theory*, vol. 49, no. 6, pp. 1579–1581, 2003.

- [27] A. Gilbert, M. Strauss, J. Tropp, and R. Vershynin, "Algorithmic linear dimension reduction in the  $l_1$  norm for sparse vectors," in *Proc. 44th Annu. Allerton Conf. Communication, Control, Computing*, 2006.
- [28] A. Gilbert, M. Strauss, J. Tropp, and R. Vershynin, "One sketch for all: Fast algorithms for compressed sensing," in *Proc. Symp. Theory Computing*, 2007.
- [29] I. F. Gorodnitsky, J. S. George, and B. Rao, "Neuromagnetic source imaging with FOCUSS: A recursive weighted minimum norm algorithm," *J. Electroencephalogr. Clin. Neurophysiol.*, Oct. 1995.
- [30] I. Gorodnitsky and B. Rao, "Sparse signal reconstruction from limited data using FOCUSS: A re-weighted minimum norm algorithm," *IEEE Trans. Signal Process.*, vol. , no. 3, pp. 600–616, Mar. 1997.
- [31] R. Gribonval and M. Nielsen, "Highly sparse representations from dictionaries are unique and independent of the sparseness measure," *Appl. Comput. Harmon. Anal.*, May 2007.
- [32] R. Gribonval and M. Nielsen, "On the strong uniqueness of highly sparse expansions from redundant dictionaries," in *Proc. Int Conf. Independent Component Analysis (ICA)*, Sep. 2004.
- [33] R. Gribonval and M. Nielsen, "Sparse representations in unions of bases," *IEEE Trans. Inf. Theory*, vol. 49, no. 12, pp. 3320–3325, Dec. 2003.
- [34] B. Hassibi and H. Vikalo, "On the sphere decoding algorithm I. Expected complexity," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 2806–2818, Aug. 2005.
- [35] J. Haupt and R. Nowak, "Signal reconstruction from noisy random projections," *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 4036–4048, Sep. 2006.
- [36] P. Indyk, "Explicit constructions for compressed sensing of sparse signals," Jul. 2007, manuscript.
- [37] N. Linial and I. Novik, "How neighborly can a centrally symmetric polytope be?," *Discrete Comput. Geom.*, vol. 36, pp. 273–281, 2006.
- [38] D. M. Malioutov, M. Cetin, and A. S. Willsky, "Source localization by enforcing sparsity through a Laplacian prior: An SVD-based approach," in *Proc. IEEE Workshop on Statistical Signal Processing*, Sep. 28–Oct. 1, 2003, pp. 573–576.
- [39] O. Milenkovic, R. Baraniuk, and T. Simunic-Rosing, "Compressed sensing meets bionformatics: A new DNA microarray architecture," presented at the Information Theory and Applications Workshop, San Diego, CA, 2007.
- [40] M. Mishali and Y. Eldar, "Reduce and boost: Recovering arbitrary sets of jointly sparse vectors," [Online]. Available: <http://www.dsp.ece.rice.edu/cs>
- [41] Compressed Sensing Resources, Rice DSP Group, Rice University, Houston, TX [Online]. Available: <http://www.dsp.ece.rice.edu/cs>
- [42] M. Rudelson and R. Vershynin, "Geometric approach to error correcting codes and reconstruction of signals," *Int. Math. Res. Notices*, vol. 64, pp. 4019–4041, 2005.
- [43] M. Stojnic, W. Xu, and B. Hassibi, "Compressed sensing—Probabilistic analysis of a null space characterization," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, 2008.
- [44] N. M. Temme, "Numerical and asymptotic aspects of parabolic cylinder functions," *J. Comput. Appl. Math.*, vol. 121, no. 1–2, pp. 221–246, Sep. 1, 2000.
- [45] J. Tropp, "Algorithms for simultaneous sparse approximation. Part II: Convex relaxation," *Signal Process.*, Aug. 2005.
- [46] J. Tropp, A. C. Gilbert, and M. Strauss, "Algorithms for simultaneous sparse approximation. Part I: Greedy pursuit," *Signal Process.*, Aug. 2005.
- [47] J. Tropp, M. Wakin, M. Duarte, D. Baron, and R. Baraniuk, "Random filters for compressive sampling and reconstruction," presented at the IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP), Toulouse, France, May 2006.
- [48] A. M. Vershik and P. V. Sporyshev, "Asymptotic behavior of the number of faces of random polyhedra and the neighborliness problem," *Selecta Math. Sovietica*, vol. 11, no. 2, 1992.
- [49] H. Vikalo, F. Parvaresh, S. Misra, and B. Hassibi, "Recovering sparse signals using sparse measurement matrices in compressed DNA microarrays," in *Proc. Asilomar Conf.*, Nov. 2007.
- [50] W. Xu and B. Hassibi, "Efficient compressive sensing with deterministic guarantees using expander graphs," presented at the IEEE Information Theory Workshop, Lake Tahoe, Sep. 2007.
- [51] M. J. Wainwright, "Sharp thresholds for high-dimensional and noisy recovery of sparsity," presented at the Allerton Conf. Communication, Control, Computing, Monticello, IL, Sep. 2006.

- [52] A. D. Wyner, "Random packings and coverings of the unit  $N$ -sphere," *Bell Syst. Tech. J.*, vol. 46, pp. 2111–2118, 1967.
- [53] Y. Zhang, "When Is Missing Data Recoverable [Online]. Available: <http://www.dsp.ece.rice.edu/cs>



**Mihailo Stojnic** received the M.S. and Ph.D. degrees in electrical engineering from the California Institute of Technology, Pasadena, in 2003 and 2007, respectively.

Currently, he is an Assistant Professor in the School of Industrial Engineering at Purdue University, West Lafayette, IN. His research interests are in optimization, algorithms, and generally in mathematics applied in engineering problems.



**Farzad Parvaresh** was born in Isfahan, Iran, in 1978. He received the B.S. degree in electrical engineering from the Sharif University of Technology, Tehran, Iran, in 2001, and the M.S. and Ph.D. degrees in electrical and computer engineering from the University of California, San Diego, in 2003 and 2007, respectively.

He is currently a Postdoctoral Scholar at the Center for Mathematics of Information, California Institute of Technology, Pasadena. His research interests include error-correcting codes, algebraic decoding algorithms, information theory, networks, and fun math problems.

Dr. Parvaresh was a recipient of the Best Paper Award from the Forty-Sixth Annual IEEE Symposium on Foundations of Computer Science (FOCS) 2005.



**Babak Hassibi** was born in Tehran, Iran, in 1967. He received the B.S. degree from the University of Tehran, Iran, in 1989, and the M.S. and Ph.D. degrees from Stanford University, Stanford, CA, in 1993 and 1996, respectively, all in electrical engineering.

Since January 2001, he has been with the California Institute of Technology, Pasadena, where he is currently Professor and Executive Officer of Electrical Engineering. From October 1996 to October 1998, he was a Research Associate at the Information Systems Laboratory, Stanford University, and from

November 1998 to December 2000 he was a Member of the Technical Staff in the Mathematical Sciences Research Center at Bell Laboratories, Murray Hill, NJ. He has also held short-term appointments at Ricoh California Research Center, the Indian Institute of Science, and Linköping University, Sweden. His research interests include wireless communications and networks, robust estimation and control, adaptive signal processing and linear algebra. He is the coauthor (with A. H. Sayed and T. Kailath) of the books *Indefinite Quadratic Estimation and Control: A Unified Approach to  $H^2$  and  $H^\infty$  Theories* (New York: SIAM, 1999) and *Linear Estimation* (Englewood Cliffs, NJ: Prentice-Hall, 2000).

Dr. Hassibi is a recipient of an Alborz Foundation Fellowship, the 1999 O. Hugo Schuck Best Paper Award of the American Automatic Control Council (with H. Hindi and S. P. Boyd), the 2002 National Science Foundation Career Award, the 2002 Okawa Foundation Research Grant for Information and Telecommunications, the 2003 David and Lucille Packard Fellowship for Science and Engineering, and the 2003 Presidential Early Career Award for Scientists and Engineers (PECASE), and was a participant in the 2004 National Academy of Engineering "Frontiers in Engineering" program. He has been a Guest Editor for the IEEE TRANSACTIONS ON INFORMATION THEORY Special Issue on Space-Time Transmission, Reception, Coding and Signal Processing, was an Associate Editor for Communications of the IEEE TRANSACTIONS ON INFORMATION THEORY from 2004 to 2006, and is currently an Editor for the journal *Foundations and Trends in Information and Communication*.