A Remark on Compressed Sensing

B. S. Kashin[1*] and V. N. Temlyakov[2**]

[1]Steklov Mathematical Institute, Moscow, Russia
[2]University of South Carolina, Columbia, USA
Received August 15, 2007

Abstract—Recently, a new direction in signal processing — "Compressed Sensing" is being actively developed. A number of authors have pointed out a connection between the Compressed Sensing problem and the problem of estimating the Kolmogorov widths, studied in the seventies and eighties of the last century. In this paper we make the above mentioned connection more precise.

1. INTRODUCTION

Recently, Compressed Sensing (Compressive Sampling) has attracted a lot of attention of both mathematicians and computer scientists. Compressed Sensing refers to the problem of *economical* recovery of an unknown vector $u \in \mathbb{R}^m$ from the information provided by linear measurements $\langle u, \varphi_j \rangle$, $\varphi_j \in \mathbb{R}^m$, $j = 1, \ldots, n$. The goal is to design an algorithm that finds (approximates) u from the information $y = (\langle u, \varphi_1 \rangle, \ldots, \langle u, \varphi_n \rangle) \in \mathbb{R}^n$. We note that the most important case is when the number of measurements n is much smaller than m. The crucial step here is to build a *sensing* set of vectors $\varphi_j \in \mathbb{R}^m$, $j = 1, \ldots, n$, that is *good* for all vectors $u \in \mathbb{R}^m$. Clearly, the terms *economical* and *good* should be clarified in a mathematical setting of the problem. For instance, economical may mean a polynomial time algorithm. The natural variant of such a setting discussed here uses the concept of *sparsity*. Sparse representations of a function are not only a powerful analytic tool but they are utilized in many application areas such as image/signal processing and numerical computation. The backbone of finding sparse representations is the concept of m-term approximation of the target function by elements of a given system of functions (dictionary). Since the elements of the dictionary used in the m-term approximation are allowed to depend on the function being approximated, this type of approximation is very efficient. We call a vector $u \in \mathbb{R}^m$ *k-sparse* if it has at most k nonzero coordinates. Now, for a given pair (m, n), we want to understand what is the biggest sparsity $k(m, n)$ such that there exists a set of vectors $\varphi_j \in \mathbb{R}^m$, $j = 1, \ldots, n$, and an economical algorithm A mapping y into $\mathbb{R}^m$ in such a way that, for any u of sparsity $k(m, n)$, one would have an exact recovery $A(y(u)) = u$. In other words, we want to describe matrices Φ with rows $\varphi_j \in \mathbb{R}^m$, $j = 1, \ldots, n$, such that there exists an economical algorithm of solving the above formulated sparse recovery problem.

In a number of cases the sparse recovery problem is equivalent to the following problem of finding the sparsest vector (column) $u^0 := u^0_\Phi(y) \in \mathbb{R}^m$:

$$\min \|v\|_0 \qquad \text{subject to} \quad \Phi v = y, \qquad\qquad (P_0)$$

where $\|v\|_0 := |\operatorname{supp}(v)|$. Donoho with coauthors (see, for instance, [1], [2] and the history therein) suggested an economical algorithm and begun a systematic study of the following question. For which

---

*[*]E-mail: kashin@mi.ras.ru.
*[**]E-mail: temlyak@math.sc.edu.

measurement matrices Φ the highly non-convex combinatorial optimization problem (P_0) should be equivalent to its convex relaxation problem

$$\min \|v\|_1 \qquad \text{subject to} \quad \Phi v = y, \tag{P_1}$$

where $\|v\|_1$ denotes the ℓ_1-norm of the vector $v \in \mathbb{R}^m$? It is known that the problem (P_1) can be solved by linear programming techniques. The ℓ_1-minimization algorithm A_Φ from (P_1) is the economical algorithm that we consider in this paper. Denote the solution to (P_1) by $A_\Phi(y)$. It is known (see, for instance, [2]) that, for M-coherent matrices Φ, one has $u_\Phi^0(\Phi u) = A_\Phi(\Phi u) = u$ provided u is k-sparse with $k < (1 + 1/M)/2$. This allows us to build rather simple deterministic matrices Φ with $k(m, n) \asymp n^{1/2}$ and recover with the ℓ_1-minimization algorithm A_Φ from (P_1).

Recent progress (see the surveys [3], [4]) in Compressed Sensing resulted in proving the existence of matrices Φ with $k(m, n) \asymp n/\log(m/n)$, which is substantially greater than $n^{1/2}$. A number of authors (see, for instance, [5], [6]) have pointed out a connection between the Compressed Sensing problem and the problem of estimating the widths of finite-dimensional sets, studied at the end of seventies and the beginning of eighties of the 20th century. In this paper, we make the above-mentioned connection more precise. We proceed to a detailed discussion of recent results.

We begin with results from [5]. Donoho [5] formulated the following three properties of matrices Φ with normalized in ℓ_2 columns and proved the existence of matrices satisfying these conditions. Let T be a subset of indices from $[1, m]$. Denote by Φ_T a matrix consisting of columns of Φ with indices from T.

CS1. The minimal singular value of Φ_T satisfies $\geq \eta_1 > 0$ uniformly for all T such that $|T| \leq \rho n/\log m$.

CS2. Let W_T denote the range of Φ_T. Assume that, for any T satisfying $|T| \leq \rho n/\log m$, one has

$$\|w\|_1 \geq \eta_2 n^{1/2} \|w\|_2 \qquad \forall\, w \in W_T, \quad \eta_2 > 0.$$

CS3. Denote $T^c := \{j\}_{j=1}^m \setminus T$. For any T, $|T| \leq \rho n/\log m$, and for any $w \in W_T$, one has for any v satisfying $\Phi_{T^c} v = w$

$$\|v\|_{\ell_1(T^c)} \geq \eta_3 \left(\log\left(\frac{m}{n}\right) \right)^{-1/2} \|w\|_1, \qquad \eta_3 > 0.$$

It was proved in [5] that if Φ satisfies CS1−CS3, then there exists a $\rho_0 > 0$ for which we have the relation $u_\Phi^0(\Phi u) = A_\Phi(\Phi u) = u$, provided that $|\operatorname{supp} u| \leq \rho_0 n/\log m$. The analysis in [5] relates the Compressed Sensing problem to the problem of estimating the Kolmogorov widths and their dual, the Gelfand widths.

We give the corresponding definitions. For a compact set $F \subset \mathbb{R}^m$, its *Kolmogorov width* has the form

$$d_n(F, \ell_p) := \inf_{L_n: \dim L_n \leq n} \sup_{f \in F} \inf_{a \in L_n} \|f - a\|_p,$$

where L_n is a linear subspace of $\mathbb{R}^m$ and $\|\cdot\|_p$ denotes the ℓ_p-norm. The *Gelfand width* is defined as follows

$$d^n(F, \ell_p) := \inf_{V_n} \sup_{f \in F \cap V_n} \|f\|_p,$$

where the infimum is taken over linear subspaces V_n of dimension $\geq m - n$. It is well known that the Kolmogorov and the Gelfand widths are related by the duality formula. S. M. Nikol'skii was the first to use the duality idea in approximation theory even before the introduction of the notion of Gelfand width. For instance (see [7]), for the case in which $F = B_p^m$ is a unit ℓ_p-ball in $\mathbb{R}^m$ and $1 \leq q, p \leq \infty$, one has

$$d_n(B_p^m, \ell_q) = d^n(B_{q'}^m, \ell_{p'}), \qquad p' := \frac{p}{p-1}. \tag{1.1}$$

In the particular case $p = 2$, $q = \infty$ of our interest, (1.1) gives

$$d_n(B_2^m, \ell_\infty) = d^n(B_1^m, \ell_2). \tag{1.2}$$

It has been established in approximation theory (see [8] and [9]) that

$$d_n(B_2^m, \ell_\infty) \leq C\left(\frac{1 + \log(m/n)}{n}\right)^{1/2}. \tag{1.3}$$

By C we denote an absolute constant throughout the paper. In other words, it was proved (see (1.3) and (1.2)) that, for any pair (m, n), there exists a subspace V_n, $\dim V_n \geq m - n$ such that, for any $x \in V_n$, one has

$$\|x\|_2 \leq C\left(\frac{1 + \log(m/n)}{n}\right)^{1/2}\|x\|_1. \tag{1.4}$$

It was made clear in [5] that properties of the null space $\mathcal{N}(\Phi) := \{x : \Phi x = 0\}$ of a measurement matrix Φ play an important role in the Compressed Sensing problem. In [5], Donoho introduced the following two characteristics associated with Φ that are formulated in terms of $\mathcal{N}(\Phi)$ as follows:

$$w(\Phi, F) := \sup_{x \in F \cap \mathcal{N}(\Phi)} \|x\|_2, \qquad \nu(\Phi, T) := \sup_{x \in \mathcal{N}(\Phi)} \frac{\|x_T\|_1}{\|x\|_1},$$

where x_T is a restriction of x onto T: $(x_T)_j = x_j$ for $j \in T$ and $(x_T)_j = 0$ otherwise. He proved that if Φ obeys the following two conditions

$$\nu(\Phi, T) \leq \eta_1, \qquad |T| \leq \frac{\rho_1 n}{\log m}, \tag{A1}$$

$$w(\Phi, B_1^m) \leq \eta_2 \left(\frac{\log m}{n}\right)^{1/2}, \tag{A2}$$

then, for any $u \in B_1^m$, we have

$$\|u - A_\Phi(\Phi u)\|_2 \leq C\left(\frac{\log m}{n}\right)^{1/2}.$$

We now proceed to the contribution of E. Candes, J. Romberg, and T. Tao published in a series of papers. They (see [10]) introduced the following Restricted Isometry Property (RIP) of a sensing matrix Φ: $\delta_S < 1$ is the *S-restricted isometry constant* of Φ if it is the smallest quantity such that

$$(1 - \delta_S)\|c\|_2^2 \leq \|\Phi_T c\|_2^2 \leq (1 + \delta_S)\|c\|_2^2 \tag{1.5}$$

for all subsets T with $|T| \leq S$ and all coefficient sequences $\{c_j\}_{j \in T}$. Candes and Tao (see [10]) proved that if $\delta_{2S} + \delta_{3S} < 1$, then, for an S-sparse u, one has $A_\Phi(\Phi u) = u$ (recovery by ℓ_1-minimization is exact). They also proved the existence of sensing matrices Φ obeying the condition $\delta_{2S} + \delta_{3S} < 1$ for large values of sparsity $S \asymp n/\log(m/n)$. For a positive number a, denote

$$\sigma_a(v)_1 := \min_{w \in \mathbb{R}^m : |\operatorname{supp}(w)| \leq a} \|v - w\|_1.$$

In [11], the authors proved that if

$$\delta_{3S} + 3\delta_{4S} < 2,$$

then

$$\|u - A_\Phi(\Phi u)\|_2 \leq C S^{-1/2} \sigma_S(u)_1. \tag{1.6}$$

We note that properties of RIP-type matrices have already been employed in [8] for the estimation of widths. The inequality (1.3) with an extra factor $(1 + \log m/n)$ was established in [8]. The proof in [8] is based on properties of a random matrix Φ with elements $\pm 1/\sqrt{n}$. It was proved in [8] that a random matrix with elements $\pm 1/\sqrt{n}$ satisfies (with probability close to 1) the left-hand inequality in (1.5) for $S \asymp n/(1 + \log m/n)$ (see (13) and (30) in [8]). It was also proved in [8] that this matrix satisfies the inequality

$$\|\Phi_T c\|_2^2 \leq C\left(1 + \log \frac{m}{n}\right)\|c\|_2^2 \tag{1.7}$$

for any subset T with $|T| \leq n$ and any set of coefficients $\{c_j\}_{j \in T}$ (see (29) in [8]). We note that the proof of the right-hand inequality in (1.5) with $S \asymp n/(1 + \log m/n)$ for a random $n \times m$ matrix with elements $\pm 1/\sqrt{n}$ could be done in a way similar to the proof of (1.7).

In Sec. 3, we give an elaboration of the argument in [8] that allows us to get rid of the extra log-factor in the estimate of $d_n(B_2^m, \ell_\infty^m)$. We note that this argument does not use the duality formula, contrary to the first proof of the sharp result from [9].

Further investigation of the Compressed Sensing problem was conducted by A. Cohen, W. Dahmen, and R. DeVore ([6]). They proved that if Φ satisfies the RIP of order $2k$ with $\delta_{2k} < \delta < 1/3$, then one has $\delta_{2k} < \delta < 1/3$,

$$\|u - A_\Phi(\Phi u)\|_1 \leq \frac{2 + 2\delta}{1 - 3\delta} \, \sigma_k(u)_1. \tag{1.8}$$

In the proof of (1.8), the authors used the following property (*null space property*) of matrices Φ satisfying RIP of order $3k/2$: for any $x \in \mathcal{N}(\Phi)$ and any T with $|T| \leq k$, we have

$$\|x\|_1 \leq C\|x_{T^c}\|_1. \tag{1.9}$$

The null space property (1.9) is closely related to the property (A1) from [5]. The proof of (1.8) from [6] gives an inequality similar to (1.8) under the assumption that Φ has the null space property (1.9) with $C < 2$.

We now discuss the results of the present paper. We say that a measurement matrix Φ has a *Strong Compressed Sensing Property* (SCSP) if, for any $u \in \mathbb{R}^m$, we have

$$\|u - A_\Phi(\Phi u)\|_2 \leq Ck^{-1/2}\sigma_k(u)_1 \tag{1.10}$$

for $k \asymp n/\log(em/n)$. We define the *Weak Compressed Sensing Property* (WCSP) by replacing (1.10) by the weaker inequality

$$\|u - A_\Phi(\Phi u)\|_2 \leq Ck^{-1/2}\|u\|_1. \tag{1.11}$$

We say that Φ satisfies the *Width Property* (WP) if (1.4) holds for the null space $\mathcal{N}(\Phi)$. The main result of our paper states that the above three properties of Φ are equivalent. The equivalence is understood in the following way. For example, we say that the WCSP implies the SCSP if (1.11) with a constant C implies (1.10) with other constant C'. We stress that we are interested here in the asymptotic behavior of the quantities as m and n go to infinity.

2. NEW RESULTS

We mentioned in the Introduction that it is known that for any pair (m, n), $n < m$, there exists a subspace $\Gamma \subset \mathbb{R}^m$ with $\dim \Gamma \geq m - n$ such that

$$\|x\|_2 \leq Cn^{-1/2}\left(\ln \frac{em}{n}\right)^{1/2}\|x\|_1 \qquad \forall \, x \in \Gamma. \tag{2.1}$$

We will study some properties of subspaces Γ satisfying (2.1) that are useful in compressed sensing. Denote

$$S := S(m, n) := C^{-2}n\left(\ln \frac{em}{n}\right)^{-1}.$$

For $x = (x_1, \ldots, x_m) \in \mathbb{R}^m$, denote $\operatorname{supp}(x) := \{j : x_j \neq 0\}$.

Lemma 2.1. *Let Γ satisfy* (2.1)*, and let $x \in \Gamma$. Then either $x = 0$ or $|\operatorname{supp}(x)| \geq S(m, n)$.*

Proof. Assume $x \neq 0$. Then $\|x\|_1 > 0$. Denote $\Lambda := \operatorname{supp}(x)$. We have

$$\|x\|_1 = \sum_{j \in \Lambda} |x_j| \leq |\Lambda|^{1/2} \left(\sum_{j \in \Lambda} |x_j|^2 \right)^{1/2} \leq |\Lambda|^{1/2} \|x\|_2. \tag{2.2}$$

Using (2.1), we get from (2.2)

$$\|x\|_1 \leq |\Lambda|^{1/2} S(m,n)^{-1/2} \|x\|_1.$$

Thus,

$$|\Lambda| \geq S(m,n). \quad \square$$

Lemma 2.2. *Let* Γ *satisfy* (2.1)*, and let* $x \neq 0$*,* $x \in \Gamma$*. Then, for any* Λ *such that* $|\Lambda| < S(m,n)/4$*, one has*

$$\sum_{j \in \Lambda} |x_j| < \frac{\|x\|_1}{2}.$$

Proof. Similarly to (2.2),

$$\sum_{j \in \Lambda} |x_j| \leq |\Lambda|^{1/2} S(m,n)^{-1/2} \|x\|_1 < \frac{\|x\|_1}{2}. \quad \square$$

Lemma 2.3. *Let* Γ *satisfy* (2.1)*. Suppose that* $u \in \mathbb{R}^m$ *is sparse with* $|\operatorname{supp}(u)| < S(m,n)/4$*. Then, for any* $v = u + x$*,* $x \in \Gamma$*,* $x \neq 0$*, one has*

$$\|v\|_1 > \|u\|_1.$$

Proof. Let $\Lambda := \operatorname{supp}(u)$. Then

$$\|v\|_1 = \sum_{j \in [1,m]} |v_j| = \sum_{j \in \Lambda} |u_j + x_j| + \sum_{j \notin \Lambda} |x_j|$$

$$\geq \sum_{j \in \Lambda} |u_j| - \sum_{j \in \Lambda} |x_j| + \sum_{j \notin \Lambda} |x_j| = \|u\|_1 + \|x\|_1 - 2\sum_{j \in \Lambda} |x_j|.$$

By Lemma 2.2,

$$\|x\|_1 - 2\sum_{j \in \Lambda} |x_j| > 0. \quad \square$$

Lemma 2.3 guarantees that the following algorithm, known as the Basis Pursuit (see A_Φ from the Introduction), will find a sparse u exactly, provided $|\operatorname{supp}(u)| < S(m,n)/4$:

$$u_\Gamma := u + \arg\min_{x \in \Gamma} \|u + x\|_1.$$

Theorem 2.1. *Let* Γ *satisfy* (2.1)*. Then, for any* $u \in \mathbb{R}^m$ *and* u' *such that* $\|u'\|_1 \leq \|u\|_1$*,* $u - u' \in \Gamma$*, one has*

$$\|u - u'\|_1 \leq 4\sigma_{S/16}(u)_1, \tag{2.3}$$

$$\|u - u'\|_2 \leq \left(\frac{S}{16} \right)^{-1/2} \sigma_{S/16}(u)_1. \tag{2.4}$$

Proof. It is given that $u - u' \in \Gamma$. Thus, (2.4) follows from (2.3) and (2.1). We now prove (2.3). Let Λ, $|\Lambda| = [S/16]$, be the set of indices of the biggest (in absolute value) coordinates of u. Denote by u_Λ the restriction of u onto this set, i.e.,

$$(u_\Lambda)_j = u_j \quad \text{for } j \in \Lambda \qquad \text{and} \qquad (u_\Lambda)_j = 0 \quad \text{for } j \notin \Lambda.$$

Also denote $u^\Lambda := u - u_\Lambda$. Then

$$\sigma_{S/16}(u)_1 = \sigma_{|\Lambda|}(u)_1 = \|u - u_\Lambda\|_1 = \|u^\Lambda\|_1. \tag{2.5}$$

We have

$$\|u - u'\|_1 \leq \|(u - u')_\Lambda\|_1 + \|(u - u')^\Lambda\|_1.$$

Next,

$$\|(u - u')^\Lambda\|_1 \leq \|u^\Lambda\|_1 + \|(u')^\Lambda\|_1.$$

Using $\|u'\|_1 \leq \|u\|_1$, we obtain

$$\|(u')^\Lambda\|_1 - \|u^\Lambda\|_1 = \|u'\|_1 - \|u\|_1 - \|u'_\Lambda\|_1 + \|u_\Lambda\|_1 \leq \|(u - u')_\Lambda\|_1.$$

Therefore,

$$\|(u')^\Lambda\|_1 \leq \|u^\Lambda\|_1 + \|(u - u')_\Lambda\|_1$$

and

$$\|u - u'\|_1 \leq 2\|(u - u')_\Lambda\|_1 + 2\|u^\Lambda\|_1. \tag{2.6}$$

Using the fact that $u - u' \in \Gamma$, we estimate

$$\|(u - u')_\Lambda\|_1 \leq |\Lambda|^{1/2}\|(u - u')_\Lambda\|_2 \leq |\Lambda|^{1/2}\|u - u'\|_2 \leq |\Lambda|^{1/2}S^{-1/2}\|u - u'\|_1. \tag{2.7}$$

Our assumption on $|\Lambda|$ guarantees that $|\Lambda|^{1/2}S^{-1/2} \leq 1/4$. Using this and substituting (2.7) into (2.6), we obtain

$$\|u - u'\|_1 \leq \frac{\|u - u'\|_1}{2} + 2\|u^\Lambda\|_1,$$

which gives (2.3):

$$\|u - u'\|_1 \leq 4\|u^\Lambda\|_1. \quad \square$$

Corollary 2.1. *Let Γ satisfy* (2.1)*. Then, for any $u \in \mathbb{R}^m$, one has*

$$\|u - u_\Gamma\|_1 \leq 4\sigma_{S/16}(u)_1, \tag{2.8}$$

$$\|u - u_\Gamma\|_2 \leq \left(\frac{S}{16}\right)^{-1/2}\sigma_{S/16}(u)_1. \tag{2.9}$$

Proposition 2.1. *Let Γ be such that* (1.11) *holds with u_Γ instead of $A_\Phi(\Phi u)$ and $k = n/\ln(em/n)$. Then Γ satisfies* (2.1)*.*

Proof. Let $u \in \Gamma$. Then $u_\Gamma = 0$ and, from (1.11), we obtain

$$\|u\|_2 \leq C\left(\frac{\ln(em/n)}{n}\right)^{1/2}\|u\|_1. \quad \square$$

Theorem 2.2. *The following three properties of Φ are equivalent*:

- *Strong Compressed Sensing Property;*

- *Weak Compressed Sensing Property;*

- *Width Property.*

Proof. It is obvious that SCSP $\Rightarrow$ WCSP. Corollary 2.1 with $\Gamma = \mathcal{N}(\Phi)$ implies that WP $\Rightarrow$ SCSP. Proposition 2.1 with $\Gamma = \mathcal{N}(\Phi)$ implies that WCSP $\Rightarrow$ WP. Thus the three properties are equivalent. $\square$

The result (1.8) from [11] states that RIP with $S \asymp n/\log(em/n)$ implies the SCSP. Therefore, by Theorem 2.2 it implies the WP. In Sec. 3, we give a direct proof of the above statement.

3. A DIRECT PROOF THAT RIP IMPLIES WP

We will show that any subspace $L \subset \mathbb{R}^m$ generated by a matrix Φ of rank n (L is spanned by the rows of Φ) that satisfies (1.5) with $S \asymp n/(1 + \log m/n)$, approximates in the ℓ_∞^m-metric the Euclidean ball with the optimal error:

$$d(B_2^m, L)_{\ell_\infty^m} \le Cn^{-1/2}\left(1 + \log\frac{m}{n}\right)^{1/2}. \tag{3.1}$$

We can assume that any n columns of the matrix Φ are linearly independent (we always can achieve this by an arbitrarily small change in the elements of matrix Φ). Let $e_1, \dots, e_m$ be the columns of the matrix Φ. Then it is sufficient to prove (see [8]) that, for any decomposition

$$e_{i_{n+1}} = \sum_{s=1}^{n} \lambda_s e_{i_s}, \qquad i_\nu \ne i_\mu \text{ if } \nu \ne \mu, \quad 1 \le \nu, \mu \le n+1, \tag{3.2}$$

the inequality

$$\frac{\|\lambda\|_2}{\|\lambda\|_1}\left(1 + \frac{1}{\|\lambda\|_2}\right) \le Cn^{-1/2}\left(1 + \log\frac{m}{n}\right)^{1/2}, \qquad \lambda = (\lambda_1, \dots, \lambda_n) \tag{3.3}$$

holds. Let us write (3.2) as

$$\sum_{s=1}^{n+1} \widetilde{\lambda}_s e_{i_{\sigma(s)}} = 0, \qquad \text{where} \quad |\widetilde{\lambda}_1| \ge \cdots \ge |\widetilde{\lambda}_{n+1}|$$

and, among the coordinates of the vector $\widetilde{\lambda} = \{\widetilde{\lambda}_s\}_{s=1}^{n+1}$, there is a 1 and all the λ_i, $1 \le i \le n$.

By repeating the reasoning of Lemma 4.1 from [6] (see also Lemma 3 in [9] and [11]) and using (1.5) for $S \asymp n/(1 + \log m/n)$, we obtain

$$\left(\sum_{s=1}^{4S} \widetilde{\lambda}_s^2\right)^{1/2} \le C'S^{-1/2} \sum_{4S+1}^{n+1} |\widetilde{\lambda}_s|, \qquad \sum_{s=1}^{n+1} |\widetilde{\lambda}_s| \le C \sum_{s=S+1}^{n} |\widetilde{\lambda}_s|. \tag{3.4}$$

It follows from (3.4) that

$$\|\widetilde{\lambda}\|_2 \le CS^{-1/2}\|\widetilde{\lambda}\|_1. \tag{3.5}$$

If (3.5) were not true, then the "positive share" of the ℓ^2-norm of the vector $\widetilde{\lambda}$ would be located on the first $4S$ coordinates (see Lemma 4 in [8]), which contradicts (3.4).

Besides, because $\|\widetilde{\lambda}\|_2 > 1$, from (3.5) we obtain $\|\widetilde{\lambda}\|_1 \ge cS^{1/2}$ and, therefore, $\|\lambda\|_1 \ge cS^{1/2}$. Finally, we have

$$\frac{\|\lambda\|_2}{\|\lambda\|_1} \le \frac{\|\widetilde{\lambda}\|_2}{\|\lambda\|_1} \le \frac{2\|\widetilde{\lambda}\|_2}{\|\widetilde{\lambda}\|_1} \le 2CS^{-1/2},$$

which is what we needed to prove.

ACKNOWLEDGMENTS

REFERENCES

1. S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," SIAM Rev. **43** (1), 129–159 (2001).
2. D. L. Donoho, M. Elad, and V. N. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," IEEE Trans. Inform. Theory **52** (1), 6–18 (2006).
3. E. J. Candes, "Compressive sampling," in *International Congress of Mathematicians, Madrid, Spain, August 22–30, 2006* (Eur. Math. Soc., Zürich, 2006), Vol. 3, pp. 1433–1452.
4. R. DeVore, "Optimal computation," in *International Congress of Mathematicians, Madrid, Spain, August 22–30, 2006* (Eur. Math. Soc., Zürich, 2006), Vol. 1, pp. 187–215.
5. D. L. Donoho, "Compressed sensing," IEEE Trans. Inform. Theory **52** (4), 1289–1306 (2006).
6. A. Cohen, W. Dahmen, and R. DeVore, *Compressed Sensing and k-Term Approximation* (Manuscript, 2007).
7. R. S. Ismagilov, "Diameters of sets in normed linear spaces, and the approximation of functions by trigonometric polynomials," Uspekhi Mat. Nauk **3** (177), 161–178 (1974) [Russian Math. Surveys **29** (3), 169–186 (1974)].
8. B. S. Kashin, "Diameters of some finite-dimensional sets and classes of smooth functions," Izv. Akad. Nauk SSSR Ser. Mat. **41**, 334–351 (1977) [Math. USSR-Izv. **11**, 317–333 (1978)].
9. A. Yu. Garnaev and E. D. Gluskin, "On widths of the Euclidean ball," Dokl. Akad. Nauk SSSR **277**, 1048–1053 (1984) [Soviet Math., Dokl. **30** (5), 200–204 (1984)].
10. E. J. Candes and T. Tao, "Decoding by linear programming," IEEE Trans. Inform. Theory **51** (12), 4203–4215 (2005).
11. E. J. Candes, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," Comm. Pure Appl. Math. **59** (8), 1207–1223 (2006).