Derivation of Compressive Sensing Theorems from the Spherical Section property

Stephen A. Vavasis*

March 5, 2009

Abstract

In this note, we prove several of the main results of compressive sensing from the spherical section property.

1 Compressive sensing

In the past four years, there has been extensive activity on compressive sensing. Recently, Kashin and Temlyakov [7] and Zhang [8] have developed simplified proofs of some of the main theorems of compressive sensing using the spherical section property. This notes summarizes proofs of the main theorems from the spherical section inequality. In contrast, Candès and Tao have proved the results using either the uniform uncertainty principle (UUP) or the restricted isometry principle (RIP), while Donoho has used hypotheses called CS1 to CS3.

Zhang notes the advantage of the KGG inequality over these other properties, namely, the UUP, RIP, and CS1-3 are not invariant under left-multiplication by an arbitrary nonsingular matrix even though the compressive sensing algorithm is invariant.

Definition. Let m, n be two positive integers such that $m < n$. Let V be an $(n-m)$ dimensional subspace of $\mathbf{R}^n$. Say that V has the Δ *spherical section property* if for any nonzero $\mathbf{v} \in V$,

$$\frac{\|\mathbf{v}\|_1}{\|\mathbf{v}\|_2} \geq \sqrt{\frac{m}{\Delta}}. \tag{1}$$

Here, Δ is called the *distortion* of V. Zhang quotes Gluskin and Milman [5], who attribute the following theorem to Kashin [6] and Gluskin and Garnaev [4].

Theorem 1. *There is a universal constant c_0 as follows. Let m, n be two positive integers such that $m < n$. Let V be an $(n-m)$-dimensional subspace of $\mathbf{R}^n$ chosen at random with the usual probability measure on choice of subspace. Then with probability at least $1 - e^{c_0(n-m)}$, V has the Δ-spherical section property for $\Delta = c_1(\log(n/m) + 1)$, where c_1 is a universal constant.*

*Department of Combinatorics and Optimization, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada, vavasis@math.uwaterloo.ca

It is also known that random ± 1 matrices and random Fourier information have the Δ-spherical section property.

2 Main results

Now we state the two main results of compressive sensing in the case of a sparse solution. similar to Donoho [3], which are (a) and (b) in the following theorem. If $\mathbf{x} \in \mathbf{R}^n$, let $\operatorname{supp}(\mathbf{x})$ be the set of indices (a subset of $\{1, \ldots, n\}$) in which x_i is nonzero. Let $|\operatorname{supp}(\mathbf{x})|$ denote the cardinality of $\operatorname{supp}(\mathbf{x})$, which is sometimes also denoted by $\|\mathbf{x}\|_0$. Like Zhang, we can increase the generality of this theorem by assuming that $\bar{\mathbf{x}}$ satisfies some side constraints determined by a set Ω; it is commonplace in the literature, however, to simply take $\Omega = \mathbf{R}^n$

Theorem 2. *Suppose* $\operatorname{null}(A)$ *has the* Δ*-spherical section property. Let* $\Omega \subset \mathbf{R}^n$ *be a nonempty set, let* $\bar{\mathbf{x}} \in \Omega$ *be a nonzero vector, and suppose* $\mathbf{b} = A\bar{\mathbf{x}}$.

(a) Provided that

$$|\operatorname{supp}(\bar{\mathbf{x}})| \leq \frac{c_2 m}{\Delta}, \tag{2}$$

$\bar{\mathbf{x}}$ *is the unique vector satisfying* $\mathbf{x} \in \Omega$, $A\mathbf{x} = \mathbf{b}$ *and* $|\operatorname{supp}(\mathbf{x})| \leq \frac{c_2 m}{\Delta}$.

(b) Provided that

$$|\operatorname{supp}(\bar{\mathbf{x}})| \leq \frac{c_3 m}{\Delta}, \tag{3}$$

and that the right-hand side of (3) *is less than* $n/2$, $\bar{\mathbf{x}}$ *is the unique solution to the optimization problem of minimizing* $\|\mathbf{x}\|_1$ *subject to* $A\mathbf{x} = \mathbf{b}$, $\mathbf{x} \in \Omega$.

Proof. We start the proof with following simple observation that holds for all $\mathbf{x}$: $\|\mathbf{x}\|_1 \leq \sqrt{|\operatorname{supp}(\mathbf{x})|}\|\mathbf{x}\|_2$. To see this, let $\mathbf{u}$ be the vector $\operatorname{sgn}(\mathbf{x})$, i.e., $u_i = 1$ if $x_i > 0$, $u_i = -1$ if $x_i < 0$, and $u_i = 0$ else. Then apply the Cauchy-Schwarz inequality to $\mathbf{x}^T\mathbf{u}$.

For part (a), let $c_2 = 1/3$. We prove the contrapositive of (a), i.e., we assume that $\mathbf{y}$ is a second solution to $A\mathbf{x} = \mathbf{b}$, $\mathbf{x} \in \Omega$ and conclude that $\mathbf{y}$ is not sufficiently sparse. Suppose $\mathbf{y} \in \Omega$ also satisfies $A\mathbf{y} = \mathbf{b}$, assume $\mathbf{y} \neq \bar{\mathbf{x}}$, and let $k = |\operatorname{supp}(\mathbf{y})|$. Let $\mathbf{z} = \bar{\mathbf{x}} - \mathbf{y}$, so $\mathbf{z} \neq \mathbf{0}$ and $\mathbf{z} \in \operatorname{null}(A)$. Furthermore, $\operatorname{supp}(\mathbf{z}) \subset \operatorname{supp}(\bar{\mathbf{x}}) \cup \operatorname{supp}(\mathbf{y})$. This implies that

$$|\operatorname{supp}(\mathbf{z})| \leq |\operatorname{supp}(\bar{\mathbf{x}})| + |\operatorname{supp}(\mathbf{y})| \leq k + c_2 m/\Delta,$$

which implies by the argument in the preceding paragraph that

$$\frac{\|\mathbf{z}\|_1}{\|\mathbf{z}\|_2} \leq \sqrt{k + c_2 m/\Delta}.$$

Since $\operatorname{null}(A)$ has the spherical section property, this implies that

$$\sqrt{k + c_2 m/\Delta} \geq \sqrt{m/\Delta},$$

i.e.,

$$k + c_2 m/\Delta \geq m/\Delta.$$

Subtracting $c_2 m/\Delta$ from both sides and using the fact that $c_2 = 1/3$ yields $k \geq (2/3)m/\Delta$, which proves that $\mathbf{y}$ is not sufficiently sparse to satisfy $|\operatorname{supp}(\mathbf{y})| \leq c_2 m/\Delta$.

For part (b), let $c_3 = 1/2$. Again, we prove the contrapositive. Suppose $\bar{\mathbf{x}} \in \Omega$ and $A\bar{\mathbf{x}} = \mathbf{b}$. Let us suppose that $\bar{\mathbf{x}}$ is not the unique minimizer of $\|\mathbf{x}\|_1$ subject to to these constraints; we will conclude that $\bar{\mathbf{x}}$ violates (3). In particular, suppose that $\bar{\mathbf{x}} + \mathbf{z}$ is a minimizer not equal to $\bar{\mathbf{x}}$. Then $A(\bar{\mathbf{x}} + \mathbf{z}) = \mathbf{b}$, $\mathbf{z} \neq \mathbf{0}$, $\|\bar{\mathbf{x}} + \mathbf{z}\|_1 \leq \|\bar{\mathbf{x}}\|_1$, and $\mathbf{z} \in \operatorname{null}(A)$. Let $S = \operatorname{supp}(\bar{\mathbf{x}})$ and $\bar{S} = \{1, \ldots, n\} - S$. Then

$$
\begin{aligned}
\|\bar{\mathbf{x}}\|_1 &\geq \|\bar{\mathbf{x}} + \mathbf{z}\|_1 \\
&= \|\bar{\mathbf{x}}_S + \mathbf{z}_S\|_1 + \|\bar{\mathbf{x}}_{\bar{S}} + \mathbf{z}_{\bar{S}}\|_1 \\
&= \|\bar{\mathbf{x}}_S + \mathbf{z}_S\|_1 + \|\mathbf{z}_{\bar{S}}\|_1 \\
&\geq \|\bar{\mathbf{x}}_S\|_1 - \|\mathbf{z}_S\|_1 + \|\mathbf{z}_{\bar{S}}\|_1. \\
&= \|\bar{\mathbf{x}}\|_1 - \|\mathbf{z}_S\|_1 + \|\mathbf{z}_{\bar{S}}\|_1.
\end{aligned}
$$

The third line follows from the fact that $\bar{\mathbf{x}}_{\bar{S}} = \mathbf{0}$ by definition of "supp($\mathbf{x}$)". The fourth line is the triangle inequality. Comparing the first and last lines, we conclude that $\|\mathbf{z}_{\bar{S}}\|_1 \leq \|\mathbf{z}_S\|_1$.

Now we consider the maximum possible value of $\|\mathbf{z}\|_1/\|\mathbf{z}\|_2$ assuming the inequality derived in the previous paragraph, namely, $\|\mathbf{z}_{\bar{S}}\|_1 \leq \|\mathbf{z}_S\|_1$. This defines an optimization problem for $\mathbf{z}$ that we now proceed to solve. Observe that the signs are irrelevant, so we can assume all entries of $\mathbf{z}$ are nonnegative. Since this optimization problem is invariant under rescaling, we can fix a scaling by assuming that $z_1^2 + \cdots + z_n^2 = 1$. This changes the problem to finding the maximum of $z_1 + \cdots + z_n$ subject to the inequalities $z_i \geq 0$, $\sum_{i \in \bar{S}} z_i \leq \sum_{i \in S} z_i$, and $z_1^2 + \cdots + z_n^2 \leq 1$. (It is obvious that the maximizer will be attained only when there is equality in the last constraint.) This is a convex optimization problem, and therefore we can exhibit its maximizer in closed form if we can exhibit a solution to the KKT conditions. Let us assume $|S| \leq n/2$ since otherwise we have already shown that $\bar{\mathbf{x}}$ violates (3) and hence the proof is done. It is easy to check that the solution is $z_i = x$ for $i \in S$ and $z_i = y$ for $i \in \bar{S}$, where $x = a/|S|$ and $a = \left[|S| \cdot |\bar{S}|/(|S| + |\bar{S}|)\right]^{1/2} = \left[|S| \cdot |\bar{S}|/n\right]^{1/2}$ while $y = a/|\bar{S}|$. The two inequalities $\sum_{i \in \bar{S}} z_i \leq \sum_{i \in S} z_i$ and $z_1^2 + \cdots + z_n^2 \leq 1$ are both active at this point, and their KKT multipliers are the solutions to the linear equations $\lambda_1 + 2\lambda_2 y = 1$ and $-\lambda_1 + 2\lambda_2 x = 1$, which are seen to be nonnegative under the assumption that $y \leq x$, i.e., $|S| \leq |\bar{S}|$. The objective value of this optimum solution is a.

Thus, the maximum possible value of $\|\mathbf{z}\|_1/\|\mathbf{z}\|_2$ is $\left[|S| \cdot |\bar{S}|/n\right]^{1/2}$. Since $\mathbf{z} \in \operatorname{null}(A)$ and $\operatorname{null}(A)$ has the Δ-spherical section property, we conclude that

$$|S| \cdot |\bar{S}|/n \geq m/\Delta.$$

Since $|\bar{S}| \leq n$,

$$|S| \geq m/\Delta.$$

Since $c_3 = 1/2$, this shows that $\bar{\mathbf{x}}$ does not satisfy (3). This concludes the proof of the theorem. □

We now extend this result to "stable" reconstruction when the data is approximately sparse.

Theorem 3. *Suppose that A is an $m \times n$ matrix such that* null(A) *has the* Δ*-spherical section property. Then for every* $\bar{\mathbf{x}} \in \mathbf{R}^n$ *and every* $k < \min(1/16)m/\Delta, n/4)$

$$\|\mathbf{x}^* - \bar{\mathbf{x}}\|_1 \le 4\|\bar{\mathbf{x}}(k) - \bar{\mathbf{x}}\|_1 \tag{4}$$

where $\bar{\mathbf{x}}(k)$ *is the best* k*-term approximation to* $\mathbf{x}$ *(i.e., the vector that agrees with* $\mathbf{x}$ *in its* k *components of largest magnitude and is zero in the remaining* $n-k$ *components), and where* $\mathbf{x}^*$ *is the solution to the optimization problem* $\min \|\mathbf{x}\|_1$ *subject to* $A\mathbf{x} = A\bar{\mathbf{x}}$*.*

Proof. Choose $k \in \{1, \ldots, n-1\}$ and let $S \subset \{1, \ldots, n\}$ index the k largest entries of $\mathbf{x}$. Let $\mathbf{z} = \mathbf{x}^* - \bar{\mathbf{x}}$, so that $\mathbf{z} \in$ null(A). Then

$$\begin{aligned} \|\bar{\mathbf{x}}\|_1 &\ge \|\mathbf{x}^*\|_1 \\ &= \|\bar{\mathbf{x}} + \mathbf{z}\|_1 \\ &= \|\bar{\mathbf{x}}_S + \mathbf{z}_S\|_1 + \|\bar{\mathbf{x}}_{\bar{S}} + \mathbf{z}_{\bar{S}}\|_1 \\ &\ge \|\bar{\mathbf{x}}_S\|_1 - \|\mathbf{z}_S\|_1 + \|\mathbf{z}_{\bar{S}}\|_1 - \|\mathbf{x}_{\bar{S}}\|_1 \\ &\ge \|\bar{\mathbf{x}}\|_1 - \|\mathbf{z}_S\|_1 + \|\mathbf{z}_{\bar{S}}\|_1 - 2\|\mathbf{x}_{\bar{S}}\|_1. \end{aligned}$$

Thus,

$$\begin{aligned} \|\mathbf{z}_{\bar{S}}\|_1 &\le \|\mathbf{z}_S\|_1 + 2\|\mathbf{x}_{\bar{S}}\|_1 &\quad (5) \\ &= \|\mathbf{z}_S\|_1 + 2\|\bar{\mathbf{x}} - \bar{\mathbf{x}}(k)\|_1. &\quad (6) \end{aligned}$$

Let

$$R = \frac{\|\mathbf{z}\|_1}{\|\bar{\mathbf{x}} - \bar{\mathbf{x}}(k)\|_1}. \tag{7}$$

The goal of the remainder of the proof is to establish a constant upper bound on R ($R \le 4$ in particular). Substitute (7) into (6) to obtain

$$\|\mathbf{z}_{\bar{S}}\|_1 \le \|\mathbf{z}_S\|_1 + 2\|\mathbf{z}\|_1/R.$$

Since $\|\mathbf{z}\|_1 = \|\mathbf{z}_S\|_1 + \|\mathbf{z}_{\bar{S}}\|_1$, the preceding inequality may be rewritten

$$(1 - 2/R)\|\mathbf{z}_{\bar{S}}\|_1/2 \le (1 + 2/R)\|\mathbf{z}_S\|_1.$$

If $1 - 2/R \le 0$, i.e., $R \le 2$, then we are done, since we have shown established an upper bound on R.

Otherwise, assume $1 - 2/R > 0$. Let $\gamma = (1 + 2/R)/(1 - 2/R)$ so that the previous equation is written $\|\mathbf{z}_{\bar{S}}\|_1 \le \gamma\|\mathbf{z}_S\|_1$. We can impose the further assumption that $\gamma \le 3$; if not, then $(1 + 2/R)/(1 - 2/R) \ge 3$, which can be simplified to $R \le 4$, and if $R \le 4$ then again we are done since the bound is established.

As in the earlier analysis, we can consider the optimization problem of maximizing $\|\mathbf{z}\|_1/\|\mathbf{z}\|_2$ subject to $\|\mathbf{z}_{\bar{S}}\|_1 \le \gamma\|\mathbf{z}_S\|_1$. Again using the same arguments, we can conclude that the optimal solution has the form $z_i = a/(n-k)$ for $i \in \bar{S}$ and $z_i = a/(\gamma k)$ for $i \in S$, where $a = \gamma\sqrt{k(n-k)/(\gamma^2 k + (n-k))}$. Optimality requires that the multipliers satisfy $\lambda_1 + 2\lambda_2 y = 1$, $-\gamma\lambda_1 + 2\lambda_2 x = 1$ and $\lambda_1, \lambda_2 \ge 0$. This is satisfied if $x \ge y$, which requires the assumption that $k \le n/(\gamma + 1)$; since $\gamma \le 3$, the hypothesis that $k \le n/4$ assures this.

The optimal value is $(\gamma+1)a/3\gamma$. Thus,

$$\begin{aligned}
\max_{\|\mathbf{z}_{\bar{S}}\|_1 \le \gamma \|\mathbf{z}_S\|_1} \frac{\|\mathbf{z}\|_1}{\|\mathbf{z}\|_2} &= \frac{(\gamma+1)a}{\gamma} \\
&= \gamma + 1\sqrt{\frac{k(n-k)}{k+9(n-k)}} \\
&\ge \sqrt{m/\Delta},
\end{aligned}$$

where the last line follows from the assumption about null(A). Since $n-k \le (k+9(n-k))/9$, the preceding inequality implies

$$(\gamma+1)\sqrt{k} \ge \sqrt{m/\Delta}.$$

On the other hand, the hypothesis of the theorem is

$$k \le (1/16)m/\Delta.$$

Comparing the two preceding equations that $\gamma \ge 3$ must hold. On the other hand, we have already assumed that $\gamma \le 3$, so the only way to end up in this case is $\gamma = 3$. If $\gamma = 3$ then $R = 2$, so again an upper bound is established on R. □

This theorem is comparable to Zhang's Theorem 4, although some of the details are different. It represents a partial strengthening of Zhang because his hypothesis (34) is not needed; Zhang himself identifies hypothesis (34) as an undesirable feature of his result.

We can also deduce Candès and Tao's [2] equation (3.34) from this result. In particular, we consider the case of the weak ℓ_p ball denoted $w\ell_p(r)$ by Candès and Tao. This set is defined as

$$w\ell_p(r) = \{\mathbf{x} \in \mathbf{R}^n : |x_{(k)}| \le rk^{-1/p}\},$$

where $x_{(k)}$ is notation for the entry of $\mathbf{x}$ whose magnitude is the kth largest. They prove (eq. (3.34)) that if $\bar{\mathbf{x}} \in w\ell_p(r)$, then, using the same definition of $\mathbf{x}^*$,

$$\|\mathbf{x} - \mathbf{x}^*\|_1 \le c_{p,\alpha} r(\alpha m/\Delta)^{1-1/p}$$

for any α sufficiently small. (Note that [2] has $\Delta = c_1 \log n$ in place of $\Delta = c_1(1+\log(n/m))$; however, they strengthen their result to $1+\log(n/m)$ in the followup paper [1].)

For our proof, apply the above theorem, taking $k = \alpha m/\Delta$ for $\alpha < (1/16)$. With this choice of k, we conclude from our theorem that

$$\begin{aligned}
\|\bar{\mathbf{x}} - \mathbf{x}^*\|_1 &\le 4\|\bar{\mathbf{x}} - \bar{\mathbf{x}}(k)\|_1 \\
&\le 4\sum_{i=k+1}^{n} ri^{-1/p} \\
&\le 4r\int_k^\infty x^{-1/p}\,dx \\
&= 4rk^{1-1/p}/(1-1/p) \\
&= 4c_p r(\alpha m/\Delta)^{1-1/p}. \qquad (8)
\end{aligned}$$

In the case of $p = 1$, Candés and Tao instead assume that $\|\bar{\mathbf{x}}\|_1 \leq r$. For such a vector, $|x_{(k)}| \leq r/k$, so the same result holds.

Finally, we can establish Candès and Tao's main result (Theorem 1.2) as follows. Their Theorem 1.2 is like (8), except that the left-hand side is the 2-norm instead of 1-norm, and the right-hand side has an exponent of $1/2 - 1/p$ instead of $1 - 1/p$. Assume again that $\bar{\mathbf{x}} \in w\ell_p(r)$. Following the notation of (4), let $\mathbf{z} = \mathbf{x}^* - \mathbf{x}$. Then since $\mathbf{z} \in \text{null}(A)$ and null(A) has the Δ-spherical section property,

$$\|\mathbf{z}\|_2 \leq \|\mathbf{z}\|_1 \cdot \sqrt{\Delta/m}.$$

Combining this with (4) yields

$$\|\mathbf{x}^* - \bar{\mathbf{x}}\|_2 \leq 4\sqrt{\frac{\Delta}{m}}\|\bar{\mathbf{x}}(k) - \bar{\mathbf{x}}\|_1.$$

Thus, we follow the same chain of inequalities as in (8) except with an additional factor of $\sqrt{\Delta/m}$ on the right-hand side, and thus we obtain

$$\|\bar{\mathbf{x}} - \mathbf{x}^*\|_2 \leq 4c_p r(\alpha m/\Delta)^{1/2-1/p}$$

3 Remark

Note that Zhang obtains a result like (4) for 2-norm, although under his undesirable hypothesis (34). It would be interesting if his result were true without this hypothesis; we eliminated the need for (34) for the 1-norm case, but have not been able to establish the claim for the 2-norm case.

References

[1] E. Candes and T. Tao. Decoding by linear programming. *IEEE Trans. Information Theory*, 51(12):4203–4215, 2005.

[2] E. Candès and T. Tao. Near-optimal signal recovery from random projections: Universal encoding strategies? *IEEE Trans. Information Theory*, 52(12):5406–5425, 2006.

[3] D. Donoho. For most large underdetermined systems of linear equations the minimal ℓ_1-norm solution is also the sparsest solution. *Commun. Pure and Appl. Math.*, 59(6):797–829, 2006.

[4] A. Y. Garnaev and E. D. Gluskin. On the widths of the Euclidean ball. *Soviet Math. Dokl.*, 30:200–203, 1984.

[5] E. Gluskin and V. Milman. Note on the geometric-arithmetic mean inequality. In *Geometric aspects of Functional analysis*, pages 131–135. Springer, Berlin, 2003.

[6] B. Kashin. Diameters of certain finite-dimensional sets in classes of smooth function. *Izv. Akad. Nauk SSSR, Ser. Mat.*, 41(2):334–351, 1977.

[7] B. Kashin and V. Temlyakov. A remark on compressed sensing. *Mathematical notes*, 82:748–755, 2007.

[8] Y. Zhang. On theory of compressive sensing via ℓ_1-minimization: Simple derivations and extensions. Technical Report TR08-11, Department of Computational and Applied Mathematics, Rice University, Houston, Texas, 2008.