An Introduction to Compressed Sensing

Toby Sanders
University of South Carolina

December 2011

The purpose of this paper is to give a brief overview of compressed sensing. I will discuss what compressed sensing is, what is the purpose of it, and a few of the major theorems concerning compressed sensing.

Introduction

Compressed sensing, also known as compressed sampling or sparse sampling, is a technique for finding sparse solutions to an underdetermined system of linear equations. In terms of linear algebra, compressed sensing is a technique for finding a sparse solution c to the equation $Ac = f$, where $A \in \mathbb{R}^{M \times N}$, $c \in \mathbb{R}^N$, $f \in \mathbb{R}^M$, and $N \gg M$. As is well known, there are infinitely many solutions to this problem. But we are concerned with only sparse solutions c. This leads to the following problem:

$$\min_{c \in \mathbb{R}^n} \|c\|_{\ell_o} \quad \text{such that} \quad Ac = f, \tag{1}$$

where $\|c\|_{\ell_0} := |\{k : c_k \neq 0, k = 1, 2, \ldots, n\}|$, i.e. $\|c\|_{\ell_o}$ counts the number of nonzero entries in the vector c. Unfortunately this problem is NP-hard, meaning no polynomial time algorithms exist to solve the problem. However, if we are given some special conditions on A, it can be shown [1], [2] that we may instead solve the problem

$$\min_{c \in \mathbb{R}^n} \|c\|_{l_1} \quad \text{such that} \quad Ac = f, \tag{2}$$

where $\|c\|_{l_1} := \sum_{k=1}^{n} |c_k|$, and in solving this problem, we will obtain the same solution as solving (1). And it is well known [3] that (2) can be solved using linear programming. Hence, the equivalence of solving (1) and (2) is a crucial property and becomes our motivation.

Importance of compressed sensing

The work related to solving (1) and (2) has strong connections with the problem of recovering signals and images from highly incomplete measurements. Common practice in recovering signals or images follows the Shannon-Hartley theorem, which essentially states that in order to recover the signal the sampling rate must be at least twice the maximum frequency present in the signal. This principle underlies most all signal acquisition used in consumer audio and visual electronics, medical imaging devices, radio receivers, and so on. Compressed sensing goes against this common wisdom, and asserts that we may instead recover most signals and images from far fewer samples than suggested by the Shannon-Hartley theorem.

For a practical example, consider attempting to recover an image from only a small number of measurements. Typical cameras record measurements of every pixel in the image, and then later compress the number of measurments for memory purposes. Compressed sensing is an attempt to instead take a much smaller set of measurements and construct the entire image from only the small set of measurements. One may think this does not seem possible. However,if we consider only the set of realistic images that may be constructed from the measurements, then the problem seems feasible. Most images in our world are not completely random, and there are many "uniformities." Hence, we rely on the idea that the image has a sparse representation with respect to some basis. Thus with enough samples, we may have hope of reconstructing the image. But how do we properly sample the image in order to reduce the number of overall measurements needed? And even then, how do we estimate how many measurements that we will need for reconstruction? These topics arise in the sections that follow and are strongly related to the "incoherence" of the sampling matrix A.

The above applications are stated here in the context of our problem: suppose we would like to recover a discrete signal $s \in \mathbb{R}^n$, given that $As = f$, where $A \in \mathbb{R}^{M \times N}$ is our sensing or sampling matrix and f is our incomplete set of measurements. If s has a sparse representation with respect to some basis $\{\psi_i\}_{i=1}^N$, so that $\Psi c = s$, where $\Psi = [\psi_1, \psi_2, ..., \psi_N]$ and c is sparse, then we may solve

$$\min_{c \in \mathbb{R}^N} \|c\|_{\ell_1} \quad \text{such that} \quad A\Psi c = f,$$

which is equivalent to solving (2).

Conditions on the sampling matrix A

For convenience, we will introduce some notation that will be used throughout the remainder of the paper. Suppose we have a S-sparse vector $c \in \mathbb{R}^N$, meaning that c is supported on some set $T \subset \{1, 2, \dots, N\}$ with $|T| \leq S$. Also, suppose that we are given $A \in \mathbb{R}^{M \times N}$ as before. The set $\{1, 2, \dots, N\}$ will be denoted by N. A_T will denote the submatrix of A with only the column indices $j \in T$. The notation a_j will be used to denote the jth column of A. Likewise c_j will denote the jth entry of any vector c, and c_T denotes the vector containing only the entries in which c is supported (those entries of indices $j \in T$). Since c is supported on T, we then have

$$Ac = A_T x_T = \sum_{j \in T} c_j a_j,$$

As will be shown, assuming that c is at most S-sparse, in order for (2) to recover c that we desire, we need A to be such that A_T obeys some "restricted isometry properties" for all possible sets T, where $|T| \leq S$. The formal definition of the "restricted isometry properties" are given below.

Definition 1 (Restricted isometry constants)**.** Let $A \in \mathbb{R}^{M \times N}$ be the matrix with a finite collection of vectors $(a_j)_{j \in N} \in \mathbb{R}^M$ as columns. For every integer $1 \leq S \leq |N|$, we define the S-restricted isometry constant δ_S to be the smallest quantity such that A_T obeys

$$(1 - \delta_S)\|c\|^2 \leq \|A_T c\|^2 \leq (1 + \delta_S)\|c\|^2 \tag{3}$$

for all subsets $T \subset N$ of cardinality of at most S, and all real coefficients $(c_j)_{j \in T}$. Similarly, we define the S, S'-restricted orthogonality constants $\theta_{S,S'}$ for $S + S' \leq |N|$ to be the smallest quantity such that

$$|\langle A_T c, A_{T'} c' \rangle| \leq \theta_{S,S'} \|c\| \|c'\| \tag{4}$$

holds for all disjoint sets $T, T' \subset N$ of cardinality $|T| \leq S$ and $|T'| \leq S'$.

We will see that the smaller these constants are, the more likely it is that solving (2) is the same solution as solving (1), which is exactly what we would like.

Informally, the restricted isometry contants δ_S tells us that the mapping of any vector under A_T must preserve the length of the vector with respect to some tolerance. For some insight to why this is important, suppose for example that there exists a set T, $|T| \leq S$ and a vector c, with $c_T \neq 0$ such that $A_T c_T = 0$. Hence, $\delta_{|T|} = 1$, and A_T does not preserve the length of vectors. Then $\sum_{j \in T} c_j a_j = 0$, and since $c_T \neq 0$ there exists disjoint sets T_1 and T_2, so that $T_1 \cup T_2 = T$ and

$$f = \sum_{j \in T_1} c_j a_j = -\sum_{j \in T_2} c_j a_j \neq 0.$$

Thus the nonzero vector f, which is supported on T, has two distinct sparse representations. Thus given $Ac = f$, where f is the same f described above, we have two unique sparse solutions $c = (c_j)_{j\in T_1}$ and $c' = (c_j)_{j\in T_2}$. Then how do we know which is the sparse solution that we are looking for? Thus we need A to have the property that $\delta_S < 1$ for all subsets T with $|T| \le S$.

The importance of the restricted orthogonality constants are less obvious though. Suppose we are given two vectors $c_1, c_2 \in \mathbb{R}^N$ such that c_1 is supported on a set T_1 disjoint from T_2, the set support of c_2, with $|T_1| \le S$ and $|T_2| \le S'$. If we then have $\theta_{S,S'} \ll 1$, then we may conclude that $Ac_1 = f_1$ and $Ac_2 = f_2$ are "approximately orthogonal." Essentially, having two vectors with disjoint support will give us very samples f_1 and f_2, thus we can hope that our sparse solution c to $Ac = f$ is unique. Of course if the columns of A are orthogonal then it is easy to see that $\theta_{S,S'} = 0$, but this isn't possible since there are many more columns than rows. For the remainder of the paper we will use θ_S to denote $\theta_{S,S}$.

Lemma 2. *We have* $\theta_{S,S'} \le \delta_{S+S'} \le \theta_{S,S'} + max(\delta_S, \delta'_S)$ *for all* S, S'.

Proof. We first show that $\theta_{S,S'} \le \delta_{S,S'}$. We may normalize (3) and (4) to obtain A_T obeying

$$(1-\delta_S) \le \|A_T u\| \le (1+\delta_S) \quad \text{for all } \|u\| = 1$$

and

$$|\langle A_T u, A_{T'} u'\rangle| \le \theta_{S,S'} \quad \text{for all } \|u\| = \|u'\| = 1.$$

Thus, showing $\theta_{S,S'} \le \delta_{S,S'}$ is equivalent to showing that

$$|\langle \sum_{j\in T} c_j a_j, \sum_{j'\in T'} c_{j'} a_{j'}\rangle| \le \delta_{S+S'}$$

whenever $|T| \le S$, $|T'| \le S'$, T, T' are disjoint, and $\sum_{j\in T} |c_j|^2 = \sum_{j'\in T'} |c_{j'}|^2 = 1$. Here $\|\cdot\| := \|\cdot\|_{l_2}$. Now (3) gives

$$2(1-\delta_{S+S'}) \le \|\sum_{j\in T} c_j a_j + \sum_{j'\in T'} c_{j'} a_{j'}\|^2 \le 2(1+\delta_{S+S'})$$

as well as

$$2(1-\delta_{S+S'}) \le \|\sum_{j\in T} c_j a_j - \sum_{j'\in T'} c_{j'} a_{j'}\|^2 \le 2(1+\delta_{S+S'}).$$

The claim then follows from the parallelogram identity

$$|\langle \sum_{j\in T} c_j a_j, \sum_{j'\in T'} c_{j'} a_{j'}\rangle| = \frac{\|\sum_{j\in T} c_j a_j - \sum_{j'\in T'} c_{j'} a_{j'}\|^2 + \|\sum_{j\in T} c_j a_j + \sum_{j'\in T'} c_{j'} a_{j'}\|^2}{4}$$

It remains to show that $\delta_{S+S'} \le \theta_S + \max(\delta_S, \delta_{S'})$. Again by normalizing (3), we see that $\|A_{T_0} u\|^2 \le (1+\delta_{S+S'})$ for all $\|u\|_{l_2} = 1$ and sets T_0 where $|T_0| \le S$. From here it suffices to show that

$$|\langle \sum_{j\in T_0} u_j a_j, \sum_{j\in T_0} u_j a_j\rangle - 1| \le \theta_S + \max(\delta_S, \delta_{S'}).$$

To prove this property, partition T_0 as $T_0 = T \cup T'$ where $|T| \le S$ and $|T'| \le S'$ and write $\sum_{j\in T} u_j^2 = \alpha$ and $\sum_{j\in T'} u_j^2 = 1-\alpha$. Then (3) and (4) give the following

$$\begin{aligned}
(1-\delta_S)\alpha \le &\langle \sum_{j\in T} u_j a_j, \sum_{j\in T} u_j a_j\rangle \le (1+\delta_S)\alpha \\
(1-\delta_{S'})(1-\alpha) \le &\langle \sum_{j\in T'} u_j a_j, \sum_{j\in T'} u_j a_j\rangle \le (1+\delta_{S'})(1-\alpha) \\
&|\langle \sum_{j\in T} u_j a_j, \sum_{j'\in T'} u_{j'} a_{j'}\rangle| \le \theta_{S,S'} \alpha^{1/2}(1-\alpha)^{1/2}.
\end{aligned}$$

Now, adding the equations above we obtain

$$\begin{aligned}&\langle\sum_{j\in T}u_ja_j,\sum_{j\in T}u_ja_j\rangle+\langle\sum_{j\in T'}u_ja_j,\sum_{j\in T'}u_ja_j\rangle+|\langle\sum_{j\in T}u_ja_j,\sum_{j'\in T'}u_{j'}a_{j'}\rangle|\\ \leq&(1+\delta_S)\alpha+(1+\delta_{S'})(1-\alpha)+\theta_{S,S'}\alpha^{1/2}(1-\alpha)^{1/2}\end{aligned}$$

Hence

$$\begin{aligned}|\langle\sum_{j\in T_0}u_ja_j,\sum_{j\in T_0}u_ja_j\rangle-1|&\leq\delta_S\alpha+\delta_{S'}(1-\alpha)+2\theta_{S,S'}\alpha^{1/2}(1-\alpha)^{1/2}\\ &\leq\max(\delta_S,\delta_{S'})+\theta_{S,S'},\end{aligned}$$

and the proof is complete. Here we use the fact that $\alpha^{1/2}(1-\alpha)^{1/2}\leq 1/2 \quad \text{for } 0\leq\alpha\leq 1.$ □

Lemma 3. *Suppose that $S\geq 1$ is such that $\delta_{2S}<1$, and let $T\subset N$ be such that $|T|\leq S$. Let $f:=A_Tc$ for some arbitrary $|T|$-dimensional vector c. Then the set T and the coefficients $(c_j)_{j\in T}$ can be reconstructed uniquely from knowledge of the vector f and the a_j's.*

Proof. Suppose instead that f has two distinct sparse representation so that $f=A_Tc=A_{T'}c'$ where $|T|,|T'|\leq S$. Then define a vector d so that

$$d_j=\begin{cases}c_j-c_j' & \text{for } j\in T\cap T'\\ c_j & \text{for } T-T'\\ c_j' & \text{for } T'-T.\end{cases}$$

Then it is easy to see $A_{T\cup T'}d=0$. Using (3), the fact that $\delta_{2S}<1$ and $|T\cup T'|\leq 2S$, we may conclude that $\|d\|=0$, contradicting that the representations were unique. □

Main results

The previous lemma is only an abstract existence argument, and gives us no way to go about recovering T and c_j from f. The theorem that follows below was given by Emmanuel Candés and Terence Tao in 2004, [4]. The theorem imposes stronger conditions on the restricted isometry and orthogonality constants of the sensing matrix than that of the previous lemma. With these conditions we are then able to show the equivalence of solving (1) and (2).

Theorem 4. *Suppose that $S\geq 1$ is such that*

$$\delta_S+\theta_S+\theta_{S,2S}<1, \tag{5}$$

and let c be a real vector supported on a set $T\subset N$ obeying $|T|\leq S$. Put $f:=Ac(=A_Tc_T)$. Then c is the unique minimizer to

$$\min\|d\|_{l_1} \quad \textit{such that} \quad Ad=f. \tag{6}$$

In order to prove Theorem 4 we will need a Lemma 5 Lemma 6. We will first summarize the main results of these lemmas so that we may prove Theorem4 directly. As will be seen d is the unique minimizer to (6) if A_T has full rank and if one can find a vector ω with the two properties
$(i)\langle\omega,a_j\rangle=sign(d_j)$ for all $j\in T$,
$(ii)|\langle\omega,a_j\rangle|<1$ for all $j\notin T$
We will now use these two facts to prove Theorem 4.

Proof. We know there must exist at least one unique minimizer d to (6), and we need to show that $d = c$. Since d is the minimizer, we have

$$\|d\|_{l_1} \leq \|c\|_{l_1} = \sum_{j \in T} |c_j|. \tag{7}$$

Now take an ω obeying (i) and (ii), we then have

$$\begin{aligned}
\|d\|_{l_1} &= \sum_{j \in T} |c_j + (d_j - c_j)| + \sum_{j \notin T} |d_j| \\
&\geq \sum_{j \in T} sign(c_j)(c_j + (d_j - c_j)) + \sum_{j \notin T} d_j \langle \omega, a_j \rangle \\
&= \sum_{j \in T} |c_j| + \sum_{j \in T} (d_j - c_j) \langle \omega, a_j \rangle + \sum_{j \notin T} d_j \langle \omega, a_j \rangle \\
&= \sum_{j \in T} |c_j| + \langle \omega, \sum_{j \in N} d_j a_j - \sum_{j \in T} c_j a_j \rangle \\
&= \sum_{j \in T} |c_j| + \langle \omega, f - f \rangle \\
&= \sum_{j \in T} |c_j|
\end{aligned}$$

Comparing with (7) we see that all of the above inequalities are actually equalities. Now using (ii), we can see that the second line in the set of inequalities above would be a strict inequality if there exists some $j \notin T$ so that $d_j \neq 0$. Thus $d_j = 0$ for all $j \notin T$ and $A_T d_T = f = A_T c_T$. So by Lemma 3 we can conclude that $d = c$. ☐

The following are the two lemmas needed to show (i) and (ii)

Lemma 5. *Let $S, S' \geq 1$ be such that $\delta_S < 1$, and c be a real vector supported on $T \subset N$ such that $|T| \leq S$. Then there exists a vector ω such that $\langle \omega, a_j \rangle = c_j$ for all $j \in T$. Furthermore, there is an "exceptional set" $E \subset N$ which is disjoint from T, of size at most $|E| \leq S'$, and with the properties*

$$|\langle \omega, a_j \rangle| \leq \frac{\theta_{S,S'}}{(1 - \delta_S)\sqrt{S'}} \|c\| \textit{ for all } j \notin T \cup E$$

and

$$\left(\sum_{j \in E} |\langle \omega, a_j \rangle|^2 \right)^{1/2} \leq \frac{\theta_S}{1 - \delta_S} \|c\|.$$

In addition, $\|\omega\| \leq K \|c\|$ for some constant $K > 0$ dependent only upon δ_S.

Proof. Let A_T^* be the adjoint transformation of A_T. Then it is easy to see that $A_T^* \omega = (\langle \omega, a_j \rangle)_{j \in T}$. From linear algebra we know that

$$\begin{aligned}
&\sup_c \|A_T c\|_2^2 = \sigma_{max}^2(A_T) = \lambda_{max}(A_T^* A_T), \text{ and} \\
&\inf_c \|A_T c\|_2^2 = \sigma_{min}^2(A_T) = \lambda_{min}(A_T * A_T),
\end{aligned}$$

where $\sigma_{max}(A_T)$ and $\sigma_{min}(A_T)$ denote the largest and smallest singular value of A_T and $\lambda_{max}(A_T^* A_T)$ and $\lambda_{min}(A_T^* A_T)$ denote the largest and smallest eigenvalues of $A_T^* A_T$. Hence, combining (3) with these facts gives

$$1 - \delta_S \leq \lambda_{min}(A_T^* A_T) \leq \lambda_{max}(A_T^* A_T) \leq 1 + \delta_S.$$

Since $\delta_{|T|} < 1$ we have $\lambda_{min}(A_T^* A_T) > 0$, hence $A_T^* A_T$ is invertible with

$$\|(A_T^* A_T)^{-1}\|_2 = \sup_c \|(A_T^* A_T)^{-1} c\|_2 = \frac{1}{\sigma_{min}^2(A_T} \leq \frac{1}{1-\delta_S}. \tag{8}$$

The second equality in the line above follows from the fact that the singular values of A_T^* coincide with the singular values of A_T, and that the singular values of the inverse of any matrix M are just the inverse of the singular values of M. Also note that $\|A_T(A_T^* A_T)^{-1}\| \leq \|A_T\| \|(A_T^* A_T)^{-1}\| \leq \frac{\sqrt{1+\delta_S}}{1-\delta_S}$ and set $\omega := A_T(A_T^* A_T)^{-1} c_T$. It is then clear that $A_T^* \omega = c_T$, i.e. $\langle \omega, a_j \rangle = c_j$ for all $j \in T$. Moreover, $\|\omega\| \leq K \|c_T\|$ with $K = \frac{\sqrt{1+\delta_S}}{1-\delta_S}$. Thus we have established the first and last parts of the theorem. Now, if T' is any set in N disjoint from T with $|T'| \leq S'$ and $d_{T'} = (d_j)_{j \in T}$ is any sequence of real numbers, then basic properties of the adjoint,(4), and (8) give

$$\begin{aligned}
|\langle A_{T'}^* \omega, d_{T'} \rangle| &= |\langle \omega, A_{T'} d_{T'} \rangle| \\
&= |\langle A_T(A_T^* A_T)^{-1} c_T, A_{T'} d_{T'} \rangle| \\
&\leq \theta_{S,S'} \|(A_T^* A_T)^{-1} c_T\| \|d_{T'}\| \\
&\leq \frac{\theta_{S,S'}}{1-\delta_S} \|c_T\| \|d_{T'}\|.
\end{aligned}$$

Choosing $d_{T'} = A_{T'}^* \omega$, the above gives

$$\|A_{T'}^* \omega\| = \left(\sum_{j \in T'} |\langle \omega, a_j \rangle|^2 \right)^{1/2} \leq \frac{\theta_{S,S'}}{1-\delta_S} \|c_T\|, \tag{9}$$

whenever $T' \subset N \backslash T$ and $|T'| \leq S'$. If we set

$$E := \{ j \in N \backslash T : |\langle \omega, a_j \rangle| > \frac{\theta_{S,S'}}{(1-\delta_S)\sqrt{S'}} \|c_T\| \},$$

then the claim is that E must be such that $|E| \leq S'$. For a contradiction, suppose to the contrary that $|E| > S'$. Then take any subset of T' of E with cardinality S'. Then we have

$$\begin{aligned}
\left(\sum_{j \in T'} |\langle \omega, a_j \rangle|^2 \right)^{1/2} &> \left(\sum_{j \in T'} \frac{\theta_{S,S'}^2}{(1-\delta_S)^2 S'} \|c_T\|^2 \right)^{1/2} \\
&= \frac{\theta_{S,S'}}{(1-\delta_S)} \|c_T\|,
\end{aligned}$$

which contradicts (9). This completes the proof. □

Lemma 6. *Let $S \geq 1$ be such that $\delta_S + \theta_{S,2S} < 1$ and let c be a real vector supported on $T \subset J$ obeying $|T| \leq S$. Then there exists a vector ω such that $\langle \omega, a_j \rangle = c_j$ for all $j \in T$. Furthermore, ω obeys*

$$|\langle \omega, a_j \rangle| \leq \frac{\theta_S}{(1-\delta_S-\theta_{S,2S})\sqrt{S}} \|c\| \text{ for all } j \notin T.$$

Proof. We may scale c so that $\sum_{j \in T} |c_j|^2 = \sqrt{S}$. We will use Lemma 5 inductively to create a sequence of sets T_n and a sequence of vectors ω_n.

Let $T_0 := T$. Taking $S' = S$, then applying Lemma 5, we can find a vector ω_1 and an "exceptional set" $T_1 \subset N$ such that

$$\begin{aligned}
T_0 \cap T_1 &= \varnothing \\
|T_1| &\leq S \\
\langle \omega_1, a_j \rangle &= c_j \text{ for all } j \in T_0 \\
|\langle \omega_1, a_j \rangle| &\leq \frac{\theta_S}{1-\delta_S} \text{ for all } j \notin T_0 \cup T_1 \\
(\sum_{j \in T_1} |\langle \omega_1, a_j \rangle|^2)^{1/2} &\leq \frac{\theta_S}{1-\delta_S}\sqrt{S} \\
\|\omega_1\| &\leq K.
\end{aligned}$$

Now we will apply Lemma 5 inductively. For all $n \geq 1$, the set T described in Lemma 5 will be $T_n \cup T_0$. The vector c from the lemma will now contain the entries $c_j = \langle \omega_n, a_j \rangle$ for $j \in T_n$, and $c_j = 0$ for $j \in T_0$. Finally the set S' from the lemma will be S. Thus by Lemma 5 there exists an "exceptional set" T_{n+1} and a vector ω_{n+1} with the following properties

$$\begin{aligned}
(T_n \cup T_0) \cap T_{n+1} &= \varnothing \\
|T_{n+1}| &\leq S \\
\langle \omega_{n+1}, a_j \rangle &= \langle \omega_n, a_j \rangle \text{ for all } j \in T_n \\
\langle \omega_{n+1}, a+j \rangle &= 0 \text{ for all } j \in T_0 \\
|\langle \omega_{n+1}, a_j \rangle| \leq \|c\| \frac{\theta_{S,2S}}{1-\delta_S} &\leq \frac{\theta_S}{1-\delta_S}\left(\frac{\theta_{S,2S}}{1-\delta_S}\right)^n \text{ for all } j \notin T_n \cup (T_0 \cup T_{n+1}) \\
(\sum_{j \in T_{n+1}} |\langle \omega_{n+1}, a_j \rangle|^2)^{1/2} &\leq \frac{\theta_S}{1-\delta_S}\left(\frac{\theta_{S,2S}}{1-\delta_S}\right)^n \sqrt{S} \\
\|\omega_{n+1}\| &\leq \left(\frac{\theta_S}{1-\delta_S}\right)^{n-1} K
\end{aligned}$$

By hypothesis, we have $\frac{\theta_{S,2S}}{1-\delta_S} < 1$. Thus if we define

$$\omega := \sum_{n=1}^{\infty} (-1)^{n-1} \omega_n$$

then the series is absolutely convergent, and, therefore, ω is a well-defined vector. We now study the coefficients

$$\langle \omega, a_j \rangle = \sum_{n=1}^{\infty} (-1)^{n-1} \langle \omega_n, a_j \rangle \tag{10}$$

for $j \in N$. First consider $j \in T_0$. From the construction $\langle \omega_1, a_j \rangle = c_j$ (here we are refering to the c_j from the original vector c, not the unductively defined c_j) and $\langle \omega_n, a_j \rangle = 0$ for all $n \geq 2$, so then it is clear that

$$\langle \omega, a_j \rangle = c_j \text{ for all } j \in T_0.$$

Second, fix j with $j \notin T_0$ and we will consider the set of coefficients given by $I_j := \{n \geq 1 : j \in T_n\}$. By construction T_n and T_{n+1} disjoint, so the integers in each I_j cannot be consecutive. Now if$n \in I_j$, then by definition $j \in T_n$, and by construction we have $\langle \omega_{n+1}, a_j \rangle = \langle \omega_n, a_j \rangle$. Due to the alternation of sign, it follows that the n and $n+1$ terms of (10) cancel each other, and we are left with

$$\langle \omega, a_j \rangle = \sum_{n \geq 1; n, n-1 \notin I_j} (-1)^{n-1} \langle \omega_n, a_j \rangle. \tag{11}$$

Lastly, if $n, n-1 \notin I_j$ and $n \neq 0$, then again by construction we have $j \notin T_n \cup T_{n-1}$ and $|\langle \omega_n, a_j \rangle| \leq \frac{\theta_S}{1-\delta_S}\left(\frac{\theta_{S,2S}}{1-\delta_S}\right)^{n-1}$. Thus from (11) we obtain

$$\begin{aligned} \Big|\sum_{n\geq 1; n, n-1 \notin I_j} (-1)^{n-1}\langle \omega_n, a_j \rangle\Big| &\leq \sum_{n\geq 1; n, n-1 \notin I_j} |(-1)^{n-1}\langle \omega_n, a_j \rangle| \\ &\leq \frac{\theta_S}{1-\delta_S}\sum_{n=1}^{\infty}\left(\frac{\theta_{S,2S}}{1-\delta_S}\right)^{n-1} \\ &= \frac{\theta_S}{1-\delta_S}\left(\frac{1}{1-\frac{\theta_{S,2S}}{1-\delta_S}}\right) = \frac{\theta_S}{1-\delta_S-\theta_{S,2S}}. \end{aligned}$$

Recalling the $\|c\| = \sqrt{S}$, the proof is complete. $\square$

To see how Lemma 6 gives us (i) and (ii), take $\tilde{c} = sign(c)$ and apply the lemma to $\tilde{c}$. Then (i) follows immediately. To realize (ii) we simply note that $\|\tilde{c}\| = \sqrt{|T|}$ and so

$$\begin{aligned} |\langle \omega, a_j \rangle| &\leq \frac{\theta_S}{(1-\delta_S-\theta_{S,2S})\sqrt{S}}\|\tilde{c}\| \\ &= \frac{\theta_S}{1-\delta_S-\theta_{S,2S}}\sqrt{\frac{|T|}{S}} \\ &\leq \frac{\theta_S}{1-\delta_S-\theta_{S,2S}}. \end{aligned}$$

Given that $\delta_S + \theta_S + \theta_{S,2S} < 1$, the last line above is less than 1, therefore (ii) holds.

Now this leads us to a major issue. If our sensing matrix has good restricted isometry constants, then we may use ℓ_1 minimization for our problem. But how do we construct these sensing matrices so that they have good isometry constants? It turns out that just using randomness to create these matrices usually gives good results. In particular, it can be seen [4] that with overwhelming probablility, Gaussian random matrices have good isometry constants.

References

[1] E. Candés, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inform. Theory,* vol. 52, no. 2, pp. 489-509, Feb. 2006.
[2] D. Donoho, "Compressed sensing," *IEEE Trans. Inform. Theory* vol. 52, no. 4, pp. 1289-1306, Apr. 2006.
[3] S. Boyd, and L. Vandenberghe, "Convex Optimization," Cambridge University Press, 2004.
[4] E. Candés, T. Tao, "Decoding by Linear Programming," *IEEE Trans. Infrom. Theory,* Dec. 2004.
[5] R. Baraniuk, "Compressive Sensing [Lecture Notes]," *Signal Processing Magazine, IEEE*, pp. 118-121, 2007.
[6] E. Candés, M. Wakin, "An Introduction to Compressive Sampling," *Signal Processing Magazine, IEEE*, pp. 21-30, March 2008.