ACTAS DEL XII CONGRESO
DR. ANTONIO A. R. MONTEIRO (2013)
2014, Páginas 23–34

A SHORT INTRODUCTION TO COMPRESSIVE SENSING

CARLOS CABRELLI

ABOUT THESE NOTES

These notes correspond to a four hour course given at the XII Congreso Dr. Antonio Monteiro, that took place in the city of Bahía Blanca, Argentina in May 2013. I am grateful to the organizers that invited me to lecture at that meeting. In particular I am enormously grateful for the strong (friendly) pressure exerted by Sheldy Ombrosi in order that this work was completed. I do not think that otherwise it would have been successful. I also want to thank Diego Castaño from the Universidad Nacional del Sur who helped me, very efficiently, with the LaTeX of the notes.

The course was aimed at advanced undergraduates and graduate students with no previous background on the subject. It actually required not more than some knowledge of linear algebra and some maturity in mathematics. While writing these notes, I refrained the natural tendency to expand them adding more material. Independent of the limitation of time, I think it would have been a mistake. In the actual form it is a very basic and short introduction that I believe does not scare prospective students and on the other hand it provides a general idea of what this subject is about. This is at least my intention.

People that are interested to go deeper in the subject can browse the extensive bibliography on the subject. I recommend the recent book [1], the tutorial papers listed at Rice University DSP website [3], and the 2008 lecture by Terence Tao available in video [2]. The present notes have benefitted from these references and from a seminar organized by Akram Aldroubi at Vanderbilt University on compressive sensing.

INTRODUCTION

Assume that x is a finite signal or vector in $\mathbb{C}^N$ that is unknown. However we are able to extract some information from x. This information should be enough to be able to reconstruct the signal. So usually we will need N samples or measurements. Assume now that we add some extra information about the vector x. We know that it is sparse in some orthonormal basis (i.e. most of the coefficients of x in the basis are zero), and assume also that our measurements take the form of linear functionals on x. Then we can reduce dramatically the number of samples needed. This is basically the main idea behind compressive sensing. We will describe in these notes some basic facts of this theory that, I think, will give some insight of why it actually works.

NOTATION

Throughout this article we will fix N to be a positive integer. In applications N is usually very large (of the order of 10^6). It will be convenient for the reader to have this in mind throughout the paper.

If B is a set, $|B|$ will denote its cardinality.

We will write $\|x\|_p$ for the p-norm of the vector x, $1 \leq p < +\infty$, i.e. $\|x\|_p = (\sum_{i=1}^N |x_i|^p)^{1/p}$.

As usual $\mathbb{Z}_N = \{0,\dots,N-1\}$ will denote the cyclic group of integers with addition modulus N, or equivalently the multiplicative group of roots of unity of order N.

Sometimes in the text we will consider the vectors in $\mathbb{C}^N$ as belonging to the space $l^2(\mathbb{Z}_N)$, i.e. functions defined in $\mathbb{Z}_N$ with the norm $\|x\|_{\ell^2(\mathbb{Z}_N)} = \left(\sum_{i=0}^{N-1}|x(i)|^2\right)^{1/2}$.

For $x \in \mathbb{C}^N$, define the **support** of x as $\operatorname{supp}(x) = \{j : x(j) \neq 0\}$ and

$$\|x\|_0 = |\{i : x(i) \neq 0\}|$$

the 0-norm of x. Note that $\|.\|_0$ is not a norm in the usual sense. The notation is justified by the fact that $\lim_{p\to 0^+} \|x\|_p = \|x\|_0$.

If $s \leq N$, $\Sigma_s := \{x \in \mathbb{C}^N : \|x\|_0 \leq s\}$ is the set of s-**sparse** vectors in $\mathbb{C}^N$.

If $T \subseteq \{1,\dots,N\}$ and v is a vector in $\mathbb{C}^N$, v_T will denote the vector in $\mathbb{C}^N$ whose components corresponding to indexes not in T are zero and the components with indexes in T remain the same, i.e. $v_T(i) = v(i)$ for $i \in T$, $v_T(i) = 0$ for $i \notin T$.

If $A \in \mathbb{C}^{m\times N}$ is a matrix, A_T will denote the matrix in $C^{m\times|T|}$ where we only keep from A the columns with indexes in T.

SOME TOOLS

The discrete Fourier Transform in $\mathbb{C}^N$. Define $\mathbf{F}_N$ the Fourier matrix of order N with coefficients:

$$F_{jk} = \frac{1}{\sqrt{N}} e^{-2\pi ijk/N}, \qquad j,k = 0,\dots,N-1.$$

That is, $\mathbf{F}_N = \mathbf{F} = [F_{jk}]$.

Using properties of the roots of unity of order N we see that:

- $\mathbf{F}^*\mathbf{F} = \mathbf{F}\mathbf{F}^* = I_N$ (here $\mathbf{F}^*$ denotes the transpose conjugate of $\mathbf{F}$).
- $\|\mathbf{F}x\|_2 = \|x\|_2$ for all $x \in \mathbb{C}^N$.
- $\mathbf{F}$ is a Vandermonde matrix.

If $x \in \mathbb{C}^N$ we define $\widehat{x} \in \mathbb{C}^N$ by

$$\widehat{x}(j) = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} x(k) e^{-2\pi ijk/N}.$$

$\widehat{x}$ is the Fourier transform of x. In matrix form we have

$$\widehat{x} = \mathbf{F}x.$$

Since $\mathbf{F}$ is unitary, we have an inversion formula:

$$x = \mathbf{F}^*\widehat{x},$$

that is

$$x(k) = \frac{1}{\sqrt{N}} \sum_{j=0}^{N-1} \widehat{x}(j) e^{2\pi ijk/N}.$$

For a vector in $y \in \mathbb{C}^N$ we denote by $y^\vee$ its inverse Fourier transform, that is $y^\vee = \mathbf{F}^*y$.

Example 1. *For $N = 4$ we have*

$$\mathbf{F}_4 = \frac{1}{2}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & i & -1 & -i \\ 1 & -1 & 1 & -1 \\ 1 & -i & -1 & i \end{bmatrix}.$$

Note that $\mathbf{F}$ *is a Vandermonde matrix corresponding to* $[1, i, -1, -i]$.

Another important property of this matrix is that

$$\sum_{j=0}^{N-1} F_{kj} = \frac{1}{\sqrt{N}} \sum_{j=0}^{N-1} (e^{-2\pi ik/N})^j = \frac{1}{\sqrt{N}} \frac{1-(e^{-2\pi ik/N})^N}{1-e^{-2\pi ik/N}} = 0, \quad \text{if } k \neq 0.$$

That is, the sum of the last $N-1$ rows is zero as well as the sum of the last $N-1$ columns of $\mathbf{F}$.

The sampling theorem in $\mathbb{C}^N$. There exists a version of the sampling theorem for the case of functions defined in $\ell^2(\mathbb{Z}_N)$.

Theorem 1 (**The Sampling Theorem**). *Let $N = kd$ and $x \in \mathbb{C}^N$. Assume that* $\operatorname{supp}(\widehat{x}) \subseteq \Omega_k = \{0, \ldots, k-1\}$. *Then x can be reconstructed from the values $\{x(0), x(d), x(2d), \ldots, x((k-1)d)\}$ and the following reconstruction formula holds:*

$$x(t) = \sqrt{d/k} \sum_{j=0}^{k-1} x(jd)\, \chi^{\vee}_{\Omega_k}(t + jd).$$

The theorem says that a vector in $\mathbb{C}^N$ whose Fourier transform is supported in a small set around the origin (band limited) can be recovered from its samples in a subgroup with rate N/k, that is, the inverse of the bandwidth.

Proof. Let $y \in \mathbb{C}^k$ defined by $y(j) = \widehat{x}(j)$, $0 \leq j \leq k-1$.
Using the Fourier inversion formula on $\mathbb{C}^k$ we can write:

$$y(s) = (1/\sqrt{k}) \sum_{j=0}^{k-1} \alpha(j)\, e^{2\pi i \frac{sj}{k}}, \quad 0 \leq s \leq k-1. \tag{1}$$

That is,

$$\widehat{x}(s) = (1/\sqrt{k}) \sum_{j=0}^{k-1} \alpha(j)\, e^{2\pi i \frac{sj}{k}}\, \chi_{\Omega_k}(s), \quad 0 \leq s \leq N-1.$$

Applying again the Fourier inversion on $\mathbb{C}^N$, in both sides of the equation and using the linearity, we get,

$$x(t) = (1/\sqrt{k}) \sum_{j=0}^{k-1} \alpha(j)\, \chi^{\vee}_{\Omega_k}(t + jd), \quad 0 \leq t \leq N-1. \tag{2}$$

Now we compute the coefficients $\alpha(j)$, $0 \leq j \leq k-1$. From (1) we obtain that $0 \leq j \leq k-1$,

$$\begin{aligned}\alpha(j) =& (1/\sqrt{k}) \sum_{s=0}^{k-1} y(s)\, e^{-2\pi i \frac{js}{k}} = (1/\sqrt{k}) \sum_{s=0}^{N-1} \widehat{x}(s)\, e^{-2\pi i \frac{js}{k}} \\ =& (N/k)^{1/2} \left((1/\sqrt{N}) \sum_{s=0}^{N-1} \widehat{x}(s)\, e^{-2\pi i \frac{djs}{N}} \right) = d^{1/2} x(jd).\end{aligned}$$

Substituting in (2) we get,

$$x(t) = (d/k)^{1/2} \sum_{j=0}^{k-1} x(jd)\, \chi^{\vee}_{\Omega_k}(t + jd), \quad 0 \leq t \leq N-1,$$

which completes the proof. □

The uncertainty principle in $\mathbb{C}^N$. This principle asserts in general that a function cannot be well localized in time and frequency. There is a version in $\mathbb{C}^N$ of this principle:

Theorem 2. *Let $x \in \mathbb{C}^N$, and denote by $T = \operatorname{supp}(x)$ and $\Omega = \operatorname{supp}(\widehat{x})$, then*

$$|T||\Omega| \geq N.$$

For the proof of (1) we need the following Lemma.

Lemma 1. *Let $x \neq 0$ be a vector in $\ell^2(\mathbb{Z}_N)$, and $s \in \{1,\ldots,N\}$. If $\|x\|_0 \leq s$, then $\widehat{x}$ cannot have s consecutive zeros.*

Proof. Let $T \subseteq \{0,\ldots,N-1\}$, $T = \{t_1,\ldots,t_s\}$ and $x \in \ell^2(\mathbb{Z}_N)$ with $x_j = 0$ if $j \notin T$. Assume that there exists ℓ, $0 \leq \ell \leq N-1$ such that $\widehat{x}(\ell) = \widehat{x}(\ell+1) = \ldots = \widehat{x}(\ell+s-1) = 0$.

Let $\Omega = \{\ell,\ldots,\ell+s-1\}$ and $\mathbf{F}_{\Omega,T} = \{\frac{1}{\sqrt{N}} e^{-2\pi i n m/N}\}_{\substack{n\in\Omega \\ m\in T}}$ the submatrix of $\mathbf{F}$ in $\mathbb{C}^{|\Omega|\times|T|}$ with rows in Ω and columns in T. We have,

1) $\mathbf{F}_{\Omega,T} \in \mathbb{C}^{s\times s}$.
2) $\mathbf{F}_{\Omega,T}\widetilde{x}_T = \widetilde{\widehat{x}}_\Omega$, with $\widetilde{x}_T = (x_{t_1},\ldots,x_{t_s}) \neq 0$ and $\widetilde{\widehat{x}}_\Omega = (\widehat{x}_\ell,\ldots,\widehat{x}_{\ell+s-1}) = 0$.
3) $\mathbf{F}_{\Omega,T}$ is invertible.

To verify 3), we observe that the column j of $\mathbf{F}_{\Omega,T}$ is

$$\begin{aligned}\frac{1}{\sqrt{N}}(e^{-2\pi i\ell t_j/N},\ldots,e^{-2\pi i(\ell+s-1)t_j})^t &= \frac{1}{\sqrt{N}}e^{-2\pi i\ell t_j/N}(1,e^{-2\pi i t_j/N},\ldots,(e^{-2\pi i t_j/N})^{s-1})^t \\ &= \frac{1}{\sqrt{N}}e^{-2\pi i\ell t_j/N}(1,w_{t_j},\ldots,w_{t_j}^{s-1})^t,\end{aligned}$$

with $w_{t_j} = e^{-2\pi i t_j/N}$.

So $\mathbf{F}_{\Omega,T}$ is a Vandermonde matrix $[w_{t_1},\ldots,w_{t_s}]$ where each column has been multiplied by a non-zero complex number.

Since the $\{w_{t_j}\}$ are different roots of unity of order N, the corresponding Vandermonde matrix is invertible and then $\mathbf{F}_{\Omega,T}$ is invertible. □

We will now prove Theorem 2.

Proof of Theorem 2. Let $T = \operatorname{supp}(x)$ and $|T| = s$. Assume first that $N = sd$ for some integer d. So

$$\{0,\ldots,N-1\} = \{0,\ldots,s-1\}\cup\{s,\ldots,2s-1\}\cup\ldots\cup\{(d-1)s,\ldots,(d-1)s+s-1\}. \quad (3)$$

Since $\widehat{x}$ cannot have s-consecutive zeros, there should be at least one j, for each of the sets in (3) such that $\widehat{x}(j) \neq 0$. This implies that $|\Omega| = |\operatorname{supp}(\widehat{x})| \geq d$.

Since by hypothesis $|T| = |\operatorname{supp}(x)| = s$, we have $|\Omega|.|T| \geq d\, s = N$.

This completes the proof in case N is a multiple of s. We leave as an exercise the general case. □

THE HEART OF THE MATTER

Assume that $x \in \mathbb{C}^N$. We want to extract some information from x. That is, we will take certain measurements of x. Each measurement will give us some partial information about the vector x. We would like to perform enough measurements in order that the information that we obtain be sufficient to reconstruct x.

On the other hand we want the number of measurements to be very small. And finally we want the measurements not to depend on x, that is, we want to use the same measurements for all the vectors that we are interested in.

Assume that each measurement is given by taking the product $\langle a,x\rangle = \sum_{i=1}^{N} a_i x_i$, where a is some vector in $\mathbb{C}^N$. So, in order to perform m measurements we will need m vectors, $a_1,\dots,a_m \in \mathbb{C}^N$, $m < N$. Then we can write:

$$y_1 = \langle a_1, x\rangle,$$
$$\vdots$$
$$y_m = \langle a_m, x\rangle.$$

We want to recover x from the samples $y_1, y_2, \dots, y_m$.

If we collect the measurement vectors $\{a_j\}_{j=1}^m$ as rows of a matrix $A \in \mathbb{C}^{m\times N}$ and $y = (y_1,\dots,y_m) \in \mathbb{C}^m$, we can write in matrix notation

$$Ax = y. \tag{4}$$

Here A and y are known and we want to find x from $y = Ax$. We assume that A is a full rank matrix, since linear dependent measurements do not provide extra information.

If $m \geq N$, the system is overdetermined and there exists a unique solution given by $x = (A^*A)^{-1}A^*y$.

So, the interesting case is when $m < N$.

Actually we are interested in the case where $m \ll N$.

When $m < N$, there are infinitely many solutions. Denote by S the set of solutions

$$S = \{z \in \mathbb{C}^N : Az = Ax = y\},$$

that is if $z_0 \in S$ then $S = z_0 + \ker(A)$.

We see that since there are infinitely many solutions, we cannot determine x, unless we assume some extra hypothesis on our vector x.

Sparsity. Let $x \in \mathbb{C}^N$, $s \in \mathbb{N}$, $s < N$. We will say that x is *s-sparse* if $\|x\|_0 \leq s$. We usually will assume that s is very small compared with N. In other words, a vector is sparse if it has a lot of zero components.

We will see in what follows that with the hypothesis of sparsity we can solve our system (4) with certain conditions on the matrix A. That is, if we know that our x is s-sparse for some $s < N$, we want to identify x out of all the solutions in the set S.

Basic conditions. Let $A \in \mathbb{C}^{m\times N}$ be a measurement matrix, $x \in \mathbb{C}^N$ an s-sparse vector, $y = Ax$, $y = (y_1,\dots,y_m)$. Here $2s \leq m < N$.

Clearly a necessary condition to recover x from y is that the map

$$\phi_A : \mathbb{C}^N \to \mathbb{C}^m, \quad \phi_A(x) = Ax$$

is one to one on the set Σ_s of s-sparse vectors in $\mathbb{C}^N$.

(Note that Σ_s is not a subspace of $\mathbb{C}^N$, it is a finite union of subspaces.)

If ϕ_A is one to one on Σ_s then $z = x$ is the unique s-sparse solution of the system $Az = Ax$.

Proposition 1. *The following statements are equivalent:*

i) *ϕ_A is one to one on Σ_s.*

ii) $\Sigma_{2s} \cap \ker(A) = \{0\}$.

iii) *Every choice of $2s$ columns of A are linearly independent.*

Proof. $i \Rightarrow ii$) If $z \in \Sigma_{2s} \cap \ker(A)$, then $z = z_1 + z_2$, $z_1, z_2 \in \Sigma_s$ and $0 = Az = Az_1 + Az_2$. So $Az_1 = A(-z_2)$ implies $z_1 = -z_2$, so $z = z_1 + z_2 = 0$.

$ii \Rightarrow i$) Let z_1, z_2 be s-sparse and assume that $Az_1 = Az_2$. Then we have $A(z_1 - z_2) = 0$ and $z_1 - z_2 \in \Sigma_{2s}$. Then $z_1 - z_2$ must be zero, that is, $z_1 = z_2$.

We leave as an exercise for the reader the proof of $ii \iff iii$. □

Assume now that the matrix A satisfies

$$\ker(A) \cap \Sigma_{2s} = \{0\}. \tag{5}$$

Then if x is s-sparse, it can theoretically be recovered from y.

So several questions arise.

1) Does there exist matrices A with the property that any choice of $2s$ columns are linearly independent? This is true for example if A is a Vandermonde matrix. However it is a very difficult problem to construct such matrices for very large values of N.
2) If A has property (5) and $Ax = y$, how can we find x from y? More exactly, which algorithm can we use to recover x?

Let $x \in \mathbb{C}^N$, $x = (x_1, \dots, x_N)$ and $\operatorname{supp}(x) = T$ with $|T| = s$.

Denote by A_T the submatrix of A with m rows and $|T|$ columns (we only keep the columns with indices in T), and $\widetilde{x}_T \in \mathbb{C}^{|T|}$ where we only keep the coordinates with indexes on T. So we can write

$$Ax = A_T \widetilde{x}_T = y.$$

Then $x_T = (A_T^* A_T)^{-1} A_T^* y$.

The problem is that we don't know T. We only know that $|T| = s$.

Since ϕ_A is one to one on Σ_s, the system $A_{T'} z = y$ does not have solutions for $z \in \mathbb{C}^s$ if $T \neq T'$ and $|T'| = s$.

We have $\binom{N}{s}$ possible choices for T'.

The algorithm consists in solving $A_{T'} z = y$ for every choice of T' with $|T'| = s$ until we reach the right T'.

Clearly this algorithm reconstructs x, but it is computationally intractable for large values of N!! (This is actually an NP-complete problem.)

The Null Space Property (NSP). We will impose now strong sufficient conditions on the measurement matrix A in order to obtain computationally efficient algorithms to recover x from $y = Ax$.

From now on we assume that the number of measurements m is at least the double of the sparsity s, that is, we assume that $m \geq 2s$.

Definition 1.

1) *Let $A \in \mathbb{C}^{m \times N}$. We say that A satisfies the* Null Space Property *(NSP) of order s if for all $v \in \ker(A)$, $v \neq 0$, we have that for all $T \subseteq \{1, \dots, N\}$, $|T| \leq s$, $\|v_T\|_1 < \|v_{T^c}\|_1$.*

   Note that $\|v_T\|_1 < \|v_{T^c}\|_1$ if and only if $\|v_T\|_1 < \frac{1}{2}\|v\|_1$.

2) *We will say that a matrix $A \in \mathbb{C}^{m \times N}$ has the property "P_s" if for all $x \in \Sigma_s$, $\|x\|_1 < \|z\|_1$ if $Az = Ax$ and $z \neq x$.*

   Note that property P_s says that the s-sparse vectors minimize the ℓ_1-norm between all vectors z such that $Az = Ax$. In other words A has property P_s if and only if for all $x \in \Sigma_s$, $x = \arg\min\{\|z\|_1 : z \in x + \ker(A)\}$.

Minimization of the ℓ_1-norm is computationally efficient since there exist convex optimization algorithms that can be used.

The next question is now which matrices have property P_s? Do they exist at all? Can we actually construct these matrices?

The next theorem comes to our help.

Theorem 3. *Let $m \geq 2s$ and $A \in \mathbb{C}^{m\times N}$. The following statements are equivalent.*

i) A satisfies NSP of order s.
ii) A satisfies P_s.

Proof. $i \Rightarrow ii$) Let $x \in \Sigma_s$ and $T = \operatorname{supp}(x)$. For $z \neq x$ and $Az = Ax$ we have that $x - z \in \ker(A)$ and $x - z \neq 0$, so

$$\begin{aligned}\|x\|_1 &\leq \|x - z_T\|_1 + \|z_T\|_1\\ &= \|(x-z)_T\|_1 + \|z_T\|_1\\ &< \|(x-z)_{T^c}\|_1 + \|z_T\|_1\\ &= \|z_{T^c}\|_1 + \|z_T\|_1\\ &= \|z\|_1.\end{aligned}$$

Note that we have proved that x is unique, that is, there is only one minimizer.

$ii \Rightarrow i$) Let $v \in \ker(A) \setminus \{0\}$ and $T \subseteq \{1,\dots,N\}$ with $|T| \leq s$. Write $v = v_T + v_{T^c}$, then $0 = Av = Av_T + Av_{T^c}$, so $Av_T = -Av_{T^c}$. Since $v_T \in \Sigma_s$, using property P_s we have that $\|v_T\| < \|z\|$ for all z such that $Av_T = Az$ and $z \neq v_T$.

In particular $\|v_T\|_1 < \|v_{T^c}\|_1$. (Note that $v_T \neq v_{T^c}$ unless $v = 0$.) □

Stability of the reconstruction. Now that we have an efficient method for the reconstruction, we want to know how it performs when our signal x is not sparse or it is contaminated with noise. So we want to analyze the stability of the method.

Exercise: Let $x \in \mathbb{C}^N$, $y = Ax$ and $x^\sharp \in \arg\min\{\|z\|_1 : Az = Ax\}$. Prove that there always exists a minimum.

Theorem 4. *Let $A \in \mathbb{C}^{m\times N}$, $s \in \mathbb{N}$ such that $2s \leq m < N$. The following statements are equivalent:*

i) A satisfies NSP of order s.
ii) For all $x \in \mathbb{C}^N$, $\|x - x^\sharp\|_1 \leq c\sigma_s(x)$, where $x^\sharp \in \arg\min\{\|z\|_1 : Az = Ax\}$ and $\sigma_s(x) = \inf\{\|x - z\|_1 : z \in \Sigma_s\} = d_1(x, \Sigma_s)$.

Proof. $i \Rightarrow ii$) For every $T \subseteq \{1,\dots,N\}$, $|T| \leq s$, we have that $\|v_T\|_1 < \frac{1}{2}$ for every $v \in \ker(A) \cap S_1^N$ with $S_1^N = \{x \in \mathbb{C}^N : \|x\|_1 = 1\}$, the sphere in $\mathbb{C}^N$ of radius 1.

Since $\ker(A) \cap S_1^N$ is compact, $\sup_{v\in\ker(A)\cap S_1^N} \|v_T\|_1 < \frac{1}{2}$, and since there are finitely many sets T with cardinality s in $\{1,\dots,N\}$,

$$c := 2 \sup_{|T|\leq s} \left(\sup_{v\in\ker(A)\cap S_1^N} \|v_T\|_1 \right) < 1.$$

Note that $c = c(s, \ker(A))$.

Then, for every $|T| \leq s$ and $v \in \ker(A) \setminus \{0\}$ we have

$$\frac{\|v_T\|_1}{\|v\|_1} \leq \frac{c}{2} \quad \text{or} \quad \|v_T\|_1 \leq \frac{c}{2}\|v\|_1,$$

which we can write as

$$\|v_T\|_1 \leq \frac{1}{2}\|v\|_1 - \frac{1-c}{2}\|v\|_1$$

and subtracting $\frac{1}{2}\|v_T\|_1$ we get

$$\frac{1}{2}\|v_T\|_1 \leq \frac{1}{2}\|v_{T^c}\| - \frac{1-c}{2}\|v\|_1.$$

Finally

$$\|v_T\|_1 \le \|v_{T^c}\|_1 - (1-c)\|v\|_1 \quad \text{for every } v \in \ker(A)\setminus\{0\},\ |T| \le s. \tag{6}$$

Let now $x \in \mathbb{C}^N$, $v \in \ker(A)$, $v := x - x^\sharp$ with

$$x^\sharp \in \arg\min\{\|z\|_1 : Az = Ax\}, \tag{7}$$

and choose T to be the indices corresponding to the coefficients of x of maximum absolute value (i.e. if $j \in T$, $j' \in T^c$, then $|x_j| \ge |x_{j'}|$) and $|T| \le s$.

Note that since $\|x^\sharp\|_1 \le \|x\|_1$, using (7)

$$\begin{aligned} \|x_T\|_1 + \|x_{T^c}\|_1 = \|x\|_1 &\ge \|x^\sharp\|_1 = \|x - v\|_1 \\ &= \|(x-v)_T\|_1 + \|(x-v)_{T^c}\|_1 \\ &= \|x_T - v_T\|_1 + \|x_{T^c} - v_{T^c}\|_1 \\ &\ge \|x_T\|_1 - \|v_T\|_1 + \|v_{T^c}\|_1 - \|x_{T^c}\|_1. \end{aligned}$$

That is

$$\|v_{T^c}\|_1 \le \|v_T\|_1 + 2\|x_{T^c}\|_1 = \|v_T\|_1 + 2\sigma_s(x). \tag{8}$$

Finally from (6) and (8)

$$\begin{aligned} \|v\|_1 &= \|v_T\|_1 + \|v_{T^c}\|_1 \\ &\le (\|v_{T^c}\|_1 - (1-c)\|v\|_1) + (\|v_T\|_1 + 2\sigma_s(x)) \\ &= c\|v\|_1 + 2\sigma_s(x). \end{aligned}$$

That is,

$$\|v\|_1 \le \frac{2}{1-c}\sigma_s(x)$$

for every $x \in \mathbb{C}^N$.

ii $\Rightarrow$ *i*) Assume that for every $x \in \mathbb{C}^N$

$$\|x - x^\sharp\|_1 \le c\sigma_s(x) \tag{9}$$

where $x^\sharp \in \arg\min\{\|z\|_1 : Az = Ax\}$.

If $x \in \Sigma_s$ then $\sigma_s(x) = d(x, \Sigma_s) = 0$.

So $x = x^\sharp$ because of (9). Then A has the P_s property and then A satisfies the NSP of order s. ☐

Robustness. Assume now that our measurements of x are contaminated with noise, that is, we don't know Ax; instead we have only an approximation y of Ax, that is,

$$\|y - Ax\| \le \varepsilon.$$

for some norm in $\mathbb{C}^N$ that doesn't need to be specified. We want to see that the reconstruction in that case is possible with an error that depends on ε and the sparsity.

That is, we want to have an inequality

$$\|x - x^\sharp\|_1 < c\varepsilon$$

for some constant $c > 0$ that does not depend on x, and $x^\sharp \in \arg\min\{\|z\|_1 : \|Az - y\| \le \varepsilon\}$.

Using a stronger version of the NSP it is posible to get the desired result as we see in the following theorem.

Theorem 5. *Let $A \in \mathbb{C}^{m\times N}$ and $1 \le s \le N$ such that for every $v \in \mathbb{C}^N$ and $|T| \le s$,*

$$\|v_T\|_1 \le c\|v_{T^c}\|_1 + d\|Av\| \tag{10}$$

for some norm $\| \ \|$ in $\mathbb{C}^N$, for some $0 < c < 1$ and $d > 0$. Then for every $x \in \mathbb{C}^N$ such that $\|x\|_0 \le s$ and for every $y \in \mathbb{C}^m$ such that $\|y - Ax\| \le \varepsilon$, we have

$$\|x - x^\sharp\|_1 \le c_1\varepsilon,$$

where c_1 depends only on c and d, and $x^\sharp \in$ arg min$\{\|z\|_1 : \|Az - y\| \le \varepsilon\}$.

Proof. If $v = x - x^\sharp$, then

$$\|Av\| = \|Ax - Ax^\sharp\| \le \|Ax - y\| + \|y - Ax^\sharp\| \le 2\varepsilon. \tag{11}$$

Let $T := \text{supp}(x)$. Using now the minimality of $x^\sharp$

$$\|x\|_1 \ge \|x^\sharp\|_1 = \|(x-v)_T\|_1 + \|(x-v)_{T^c}\|_1 \ge \|x\|_1 - \|v_T\|_1 + \|v_{T^c}\|_1.$$

So $\|v_{T^c}\|_1 \le \|v_T\|_1$ and using condition (10)

$$\|v_T\|_1 \le c\|v_{T^c}\|_1 + d\|Av\| \le c\|v_T\|_1 + d\|Av\|.$$

Using now (11) we have

$$\|v_T\|_1 \le \frac{2d}{1-c}\varepsilon,$$

that is,

$$\|v\|_1 = \|v_T\|_1 + \|v_{T^c}\|_1 \le 2\|v_T\|_1 \le \frac{4d}{1-c}\varepsilon.$$

Setting $c_1 = \dfrac{4d}{1-c}$ and using that $v = x - x^\sharp$ we have

$$\|x - x^\sharp\|_1 \le c_1\varepsilon. \quad \square$$

Application to vectors that are sparse in an orthonormal basis. So far we learned that the exact condition in order that an s-sparse vector x can be reconstructed using ℓ^1-minimization is that the measurement matrix A satisfy the NSP of order s.

Let us see how this can be applied to vectors that are not necessarily sparse but are sparse in some orthonormal basis. That is, the method is still valid for a class of vectors that are sparse in some fixed othonormal basis. Assume that f is s-sparse in some orthonormal basis $\{\psi_1, \dots, \psi_N\}$ of $\mathbb{C}^N$. So $f = \sum_{j=1}^N x_j\psi_j$ with $x = (x_1, \dots, x_N) \in \Sigma_s$.

Using matrix notation we can write $f = \Psi x$, with $\Psi = [\psi_1|\psi_2|\dots|\psi_N] \in \mathbb{C}^{N\times N}$; that is, Ψ is the $\mathbb{C}^{N\times N}$ matrix whose columns are the vectors in the basis. The matrix Ψ is unitary, then $x = \Psi^* f$.

Let $A \in \mathbb{C}^{m\times N}$ satisfying the NSP of order s. Define the matrix $\Phi = [\varphi_1|\dots|\varphi_m]$, $\varphi_j \in \mathbb{C}^N$, $j = 1, \dots, N$, by $\Phi = A\Psi^* \in \mathbb{C}^{m\times N}$.

Then $y := \Phi f = A\Psi^* f = Ax$. So we can solve $y = Ax$ with s-sparse x and A satisfying the NSP of order s. The solution will give us the coefficientes of f in the basis $\{\psi_1, \dots, \psi_N\}$.

Coherence of a matrix

It seems at this point that the problem of recovering a vector x from a few measurements is completely solved theoretically and complutationally with the NSP condition. However it is not an easy task, when N is large, to find matrices with this property.

So we try to find stronger conditions on A that imply the NSP but actually allow the construction of A. One of these conditions is given by the notion of “coherence” of a matrix.

Definition 2. *Let $A \in \mathbb{C}^{m\times N}$ be a matrix with normalized columns with respect to the ℓ^2-norm:*

$$A = [a_1|\dots|a_N], \quad a_j = (a_{j1},\dots,a_{jm})^t \quad \text{and} \quad \|a_j\|_2 = 1, j = 1,\dots,N.$$

We define $\mu(A)$, the coherence *of A, as*

$$\mu(A) = \max_{i\neq j} |\langle a_i, a_j\rangle|.$$

It can be proved that

$$\sqrt{\frac{N-m}{m(N-1)}} \leq \mu(A) \leq 1$$

for every $A \in \mathbb{C}^{m\times N}$ with normalized columns.

For our purposes we want matrices with low coherence.

Theorem 6. *Let $A \in \mathbb{C}^{m\times N}$ be a matrix with ℓ^2-normalized columns. If $\mu(A) < \dfrac{1}{2s-1}$ then A satisfies the NSP of order s.*

Proof. Let $A = [a_1|\dots|a_N]$ with $a_j \in \mathbb{C}^m$ and $\|a_j\|_2 = 1$, $j = 1,\dots,N$. Let $v \in \ker(A)$, $v = \{v_1,\dots,v_N\}$, $T \subseteq \{1,\dots,N\}$, $|T| \leq s$. So $\sum_{j=1}^N v_j a_j = 0$, and for every k

$$0 = \langle \sum_{j=1}^N v_j a_j, a_k\rangle = v_k + \sum_{j\neq k} v_j \langle a_j, a_k\rangle.$$

So

$$|v_k| \leq \sum_{j\neq k} |v_j| \mu(A) = (\|v\|_1 - v_k)\mu(A).$$

Now, summing over $k \in T$

$$\|v_T\|_1 \leq (s\|v\|_1 - \|v_T\|_1)\mu(A),$$

which gives

$$\|v_T\|_1 \leq \|v\|_1 \frac{\mu(A)}{1+\mu(A)} s.$$

Then if $\dfrac{\mu(A)}{1+\mu(A)} s < \frac{1}{2}$, A satisfies the NSP of order s. This inequality is equivalent to $\mu(A) < \dfrac{1}{2s-1}$. □

THE RESTRICTED ISOMETRY PROPERTY (RIP)

Another sufficient condition for the NSP is the Restricted Isometry Property. Let us begin with a definition:

Definition 3. *Let $A \in \mathbb{C}^{m\times N}$, $1 \leq s \leq N$. Let $\delta_s = \delta_s(A)$ be the smallest $\delta \geq 0$ such that*

$$(1-\delta)\|x\|_2^2 \leq \|Ax\|_2^2 \leq (1+\delta)\|x\|_2^2$$

for all $x \in \Sigma_s$.

An equivalent definition for $\delta_s(A)$ is,

$$\delta_s(A) = \max_{\substack{T\subseteq\{1,\dots,N\}\\ |T|\leq s}} \|A_T^* A_T - I\|_{2\to 2}$$

We will say loosely that A satisfies the RIP of order s if δ_s is "small" when s is "reasonably big".

Note that $\delta_1 \leq \delta_2 \leq \ldots \leq \delta_N$.

Theorem 7. *Let $A \in \mathbb{C}^{m\times N}$, $1 \leq s \leq N$. If $\delta_{2s} < \frac{1}{3}$, then A satisfies the NSP of order s.*

Proof. We need to prove that:

$$\text{for all } v \in \ker(A),\ v \neq 0, \text{ and for all } T \subseteq \{1,\ldots,N\},\ |T| \leq s,\ \|v_T\|_1 < \frac{1}{2}\|v\|_1.$$

We will actually prove a stronger statement:

$$\text{for every } v \in \ker(A)\setminus\{0\} \text{ and } T \subseteq \{1,\ldots,N\} \text{ with } |T| \leq s,\ \|v_T\|_2 \leq \frac{\beta}{2\sqrt{s}}\|v\|_1, \qquad (12)$$

where $\beta := \dfrac{2\delta_{2s}}{1-\delta_{2s}}$ satisfies $\beta < 1$ iff $\delta_{2s} < \frac{1}{3}$.

Fix $v \in \ker(A)\setminus\{0\}$. We choose $T_0 \subseteq \{1,\ldots,N\}$ such that $|v_i| \geq |v_j|$ for every $i \in T_0$, $j \in T_0^c$. Since $\|v_T\|_2 \leq \|v_{T_0}\|_2$ for every $T \subseteq \{1,\ldots,N\}$ we only need to prove (12) for $T = T_0$.

We now partition T_0^c as $T_0^c = T_1 \cup T_2 \cup \ldots$, where T_j is the index set corresponding to the s largest absolute value entries of $(T_0 \cup T_1 \cup \ldots \cup T_{j-1})^c$. So $v = v_{T_0} + v_{T_1} + v_{T_2} + \ldots$ and using that $v \in \ker(A)$

$$Av_{T_0} = A(-v_{T_1}) + A(-v_{T_2}) + \ldots$$

so that

$$\|v_{T_0}\|_2^2 \leq \frac{1}{1-\delta_{2s}}\|Av_{T_0}\|_2^2 = \frac{1}{1-\delta_{2s}}\langle Av_{T_0}, Av_{T_0}\rangle = \frac{1}{1-\delta_{2s}}\sum_{k\geq 1}\langle Av_{T_0}, -Av_{T_k}\rangle. \qquad (13)$$

Now we use that if u,w are vectors in $\mathbb{C}^N$ of sparsity at most s and $\operatorname{supp}(u)\cap\operatorname{supp}(w) = \emptyset$ then $|\langle Au, Aw\rangle| \leq \delta_{2s}\|u\|_2\|w\|_2$. For, if $T = \operatorname{supp}(u)\cup\operatorname{supp}(w)$, we have

$$\begin{aligned}
|\langle Au, Av\rangle| &= |\langle Au_T, Aw_T\rangle - \langle u_T, w_T\rangle| \\
&= |\langle (A_T^*A_T - I)u_T, w_T\rangle| \\
&\leq \|(A_T^*A_T - I)u_T\|_2\|w_T\|_2 \\
&\leq \|A_T^*A_T - I\|_{2\to 2}\|u\|_2\|w\|_2 \\
&\leq \delta_{2s}\|u\|_2\|w\|_2.
\end{aligned}$$

Using this property in each term of (13) and dividing by $\|v_{T_0}\|_2 > 0$ we have

$$\|v_{T_0}\|_2 \leq \frac{\delta_{2s}}{1-\delta_{2s}}\sum_{k\geq 1}\|v_{T_k}\|_2 = \frac{\beta}{2}\sum_{k\geq 1}\|v_{T_k}\|_2. \qquad (14)$$

Since

$$\begin{aligned}
\frac{\|v_{T_k}\|_2}{\sqrt{s}} &\leq \left(\frac{1}{s}\sum_{j\in T_k}|v_j|^2\right)^{1/2} \leq \max_{j\in T_k}|v_j| \\
&\leq \min_{j\in T_{k-1}}|v_j| \leq \frac{1}{s}\sum_{j\in T_{k-1}}|v_j|,
\end{aligned}$$

that is,

$$\|v_{T_k}\|_2 \leq \frac{1}{\sqrt{s}}\|v_{T_{k-1}}\|_1.$$

Substituting in (14) we have

$$\|v_{T_0}\|_2 \leq \frac{\beta}{2\sqrt{s}}\sum_{k\geq 1}\|v_{T_{k-1}}\|_1 \leq \frac{\beta}{2\sqrt{s}}\|v\|_1. \quad \square$$

REFERENCES

[1] S. Foucart and H. Rauhut. *A Mathematical Introduction to Compressive Sensing.* Applied and Numerical Harmonic Analysis. Birkhäuser/Springer, New York, 2013. xviii+625 pp. MR 3100033.
[2] *Terence Tao Guest Lecture, Norway, 9 December 2008.* `http://nuit-blanche.blogspot.com/2008/12/cs-terence-tao-video-presentation-of.html`
[3] *Compressive Sensing Resources, Digital Signal Processing group at Rice University* `http://dsp.rice.edu/cs`

DEPARTAMENTO DE MATEMÁTICA, FCEYN, UNIVERSIDAD DE BUENOS AIRES, AND IMAS-CONICET, ARGENTINA