Compressed Sensing :
Exact Signal Reconstruction from Highly Incomplete Frequency Information
Review and Experiments

Geoffrey Chinot, Benjamin Muller

April 4^{st}, 2017

The purpose of this project is to demonstrate our understanding of a major publication *Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information* published in 2006 by *Candes, Romberg and Tao* [1]. To do so, we'll first focus our attention on the main result of the papers. We'll give detailed proof of the theorems when needed and describe more briefly the main outlines the rest of the time. We'll finish by describing our own experiments of the main algorithm described in the paper.

1 Introduction

1.1 Compressed Sensing Framework and Purposes

Compressed sensing is a signal processing technique for efficiently acquiring and reconstructing a signal, by finding solutions to underdetermined linear systems. The core notion of *Compressed Sensing* is sparsity which stands that a given vector/signal is sparse with respect to a certain basis if it admits lots of null coordinates when expressed in this basis (or low rank for matrix problems).

This article is a specific case of this problem. It's about understanding the conditions under which *perfect reconstruction* of a finite *discrete-time signal* is possible from partial information of its Fourier coefficients. To sum up we are going to show that exact recovery of a discrete-time signal $f \in C^N$ can be obtained by solving a simple convex optimization problem provided we observe "enough" Fourier coefficients. However this procedure is not optimal in the sense that the method is working with probability at least $1-O(N^{-M})$, where M is linked to the number of Fourier coefficients.

1.2 Definitions and Properties

We are now going to highlight a few notions that we will need to go further in the description of the paper. We consider a discrete-time signal $f \in C^N$ and we observed partial information if its Fourier coefficients randomly chosen on a set Ω. Moreover we assume that f is sparse, that is to say is supported on a subset T. This assumption turns out to be correct in different fields such as audio signals. Let us remind the classical discrete-Fourier transform $\mathcal{F}f = \hat{f} : C^N \to C^N$.

$$\hat{f}(w) = \sum_{t=0}^{N-1} f(t) e^{\frac{-2\pi i \omega t}{N}}$$

It is a well known fact that one can recover f exactly from the observation of the Fourier coefficients with the inversion formula:

$$f(t) = \sum_{\omega=0}^{N-1} f(\omega) e^{\frac{2\pi i \omega t}{N}}$$

Along this paper we will assume that we observe the Fourier coefficients $\hat{f}_{|\Omega}$ sampled on some partial subset Ω. We will show sufficient conditions on $\hat{f}_{|\Omega}$ and Ω to reconstruct exactly f.

Let us introduce the restricted Fourier transform $\mathcal{F}_{T\to\Omega} : l_2(T) \to l_2(\Omega)$ defined as:

$$\mathcal{F}_{T\to\Omega} f = \hat{f}_{|\Omega} \quad \forall f \in l_2(T)$$

$l_2(T)$ is the space of signals that are zero outside T. If $|T| = |\Omega|$ then $\mathcal{F}_{T\to\Omega}$ is bijective, if $|T| \leq |\Omega|$ then $\mathcal{F}_{T\to\Omega}$ is injective and if $|T| \geq |\Omega|$ $\mathcal{F}_{T\to\Omega}$ is surjective

1.3 Outline

- l_0 procedure to recover f
- l_1 precedure and the main theorem to recover f
- Sketch of proof
- Experiments with Douglas-Rachford algorithm

2 Procedures to recover f

2.1 l_0 procedure

The first idea to recover f from partial Fourier coefficients relies on the following combinatorial optimization problem:

$$\min_{g \in C^N} ||g||_{l_0}$$
$$\hat{g}_{|\Omega} = \hat{f}_{|\Omega}$$

Theorem 1 : *Suppose that the signal length N is prime integer. Let Ω be a subset of $\{0, ..., N-1\}$, and let f be a vector supported on T such that $|T| \leq \frac{|\Omega|}{2}$. Then f can be reconstructed uniquely from Ω and $\hat{f}_{|\Omega}$* by resolving the previous combinatorial optimization problem

However solving the previous optimization problem is NP-Hard and hence computationally infeasible.

2.2 l_1 procedure

Instead of solving the l_0 optimization problem, we are going to use its convex relaxation which is the l_1 one. The problem becomes:

$$\min_{g \in C^N} ||g||_{l_1} = \sum_{t=0}^{N-1} |g(t)| \qquad (P_1)$$
$$\hat{g}_{|\Omega} = \hat{f}_{|\Omega}$$

The counterpart of this relaxation is that we need more Fourier coefficients to recover exactly f. It is necessary to have $|T| \leq \alpha \frac{|\Omega|}{\log N}$. We need $\log N$ more observations which seems to be reasonable.
To establish this upper bound the are going to assume that the observed Fourier coefficients are randomly sampled. The subset Ω of size N_ω is chosen uniformly at random. It is important to highlight that the set Ω is random that's why we will use probabilist arguments to show the following main theorem of the paper:

Theorem 2 : *Let $f \in C^N$ be a discrete signal supported on an unknown set T, and choose Ω of size $|\Omega| = N_\omega$ uniformly at random. For a given accuracy parameter M, if:*

$$|T| \leq C_M (\log N)^{-1} |\Omega|,$$

then with probability at least $1 - O(N^{-M})$, the minimizer of (P_1) is unique and equal to f.

The proof gives an explicit value for C_M. $C_M = \frac{1-\tau}{22.6(M+1)}(1 + o(1))$ with $\tau = \frac{N_\omega}{N}$ We can give some remarks on the previous theorem:

- It is working for almost any set Ω. A counter example shows that the procedure does not recover correctly f: the discrete Dirac comb.
- The bound on $|T|$ is optimal in the sense that no recovery can be successful for all signals using significantly fewer observations.
- This procedure is closed from finding the sparsest decomposition of a signal f using a matrix of measures A which was exactly the scope of the course.

3 Proof of the main theorem

Let us begin by remark that (P_1) is a convex optimization problem, hence there exist a unique minimizer to (P_1). In this section we are going to present the sequence of arguments to show that this minimizer equals f.

- We will use to duality of Fourier transform to present a necessary and sufficient condition for f to be the solution to (P_1) which is the existence of a certain trigonometric polynomial.
- We will restrict our-self to the Bernoulli model to choose Ω and show that it is sufficient to study the uniformly random case.
- We will construct a certain trigonometric polynomial which is random (because it depends on the random set Ω)
- We will show that with probability at least $1 - O(N^{-M})$ this polynomial is well defined
- We will show that with probability at least $1 - O(N^{-M})$ this polynomial obeys to the constraints defined in the necessary and sufficient condition.

3.1 Duality of Fourier transform

The following lemma shows a necessary and sufficient condition for the solution f to be the solution of (P_1).

Lemma 1: *Let $\Omega \subseteq \{0, ..., N-1\}$. For a vector $f \in C^N$ with $T = supp(f)$, define the sign vector $sgn(f)(t) = \frac{f(t)}{|f(t)|}$ when $t \in T$ and $sgn(f)(t) = 0$ otherwise. Suppose there exists a vectors P whose Fourier transform $\hat{P}$ is supported on Ω such that:*

$$P(t) = sgn(f)(t) \forall t \in T \tag{1}$$

$$|P(t)| < 1 \quad \forall t \in T^c \tag{2}$$

Then if $\mathcal{F}_{T\to\Omega}$ is injective, the minimizer $f*$ to the problem (P_1) is unique and is equal to f. Conversely, is f is the unique minimizer of (P_1), then there exists a vector P with the above properties.

The proof is quite simple and based on the Fourier duality (Parseval Theorem).

3.2 Construction of the trigonometric polynomial

When $|T| \leq |\Omega|$, $\mathcal{F}_{T\to\Omega}$ is injective. Then there exist a lot of polynomials supported on Ω satisfying (1). We choose the following one:

$$P = \mathcal{F}_\Omega^* \mathcal{F}_{T\to\Omega} (\mathcal{F}_{T\to\Omega}^* \mathcal{F}_{T\to\Omega})^{-1} \iota^* sgn(f) \tag{3}$$

Where $\mathcal{F}_\Omega$ is the Fourier transform followed by a restriction to the set Ω, ι is the embedding operator which extends a vector on $l_2(T)$ to $l_2(Z_N)$ by placing zeros outside T and ι^* is the dual restriction map $\iota^* f = f_{|T}$.

It is obvious that $\iota^* \mathcal{F}_\Omega^* = \mathcal{F}_{T\to\Omega}^*$. Hence $\iota^* P = \iota^* sgn(f)$ and (1) is satisfied.

First we are going to show that $\mathcal{F}_{T\to\Omega}^* \mathcal{F}_{T\to\Omega}$ is invertible with probability at least $1 - O(N^{-M})$ and hence is well defined. Then we will show that (2) is satisfied with probability at least $1 - O(N^{-M})$.

3.3 Restriction to the Bernoulli model

The Bernoulli model is much more simple to analyse than if Ω is chosen uniformly at random. A set Ω' of Fourier coefficients is sampled using Bernoulli model with parameter $\tau = \frac{N_\omega}{N}$. For all $\omega \in \{1, ..., N-1\}$, $\omega \in \Omega'$ with probability τ (Bernoulli distribution). Then the size of Ω' is random following a binomial with $\mathbb{E}[|\Omega'|] = \tau N$.

The following theorem ensures that the previous trigonometric polynomial is well defined.

Theorem 3: *Let T be a fixed subset, and choose Ω using the Bernoulli model with parameter τ and suppose that:*

$$|T| \leq C_M (\log N)^{-1} \tau N$$

where C_M is the same as in theorem 2. Then $\mathcal{F}^*_{T\to\Omega}\mathcal{F}_{T\to\Omega}$ is invertible with probability at least $1 - O(N^{-M})$. Moreover P in (3) obeys $|P(t)| \leq 1, \forall t \in T^c$ with probability at least $1 - O(N^{-M})$.

This theorem claims that with probability at least $1 - O(N^{-M})$ the trigonometric polynomial presented in (3) is well defined and obeys (1) and (2) and hence with probability at least $1 - O(N^{-M})$ f is the unique solution of the l_1 optimization problem.

So far this theorem is working with the Bernoulli model. However we want to prove the theorem (2) with Ω chosen uniformly at random. In fact, it can be shown that controlling the probability of failure in the Bernoulli model (probability that there is no dual polynomial P, supported on Ω obeying (1) and (2)) allows to control the failure probability for the uniform model.

To summarize, we have shown that with the Bernoulli model, the trigonometric polynomial defined in (3) exists and obeys (1) and (2) with high probability. Then with Ω chosen uniformly at random f is the unique solution to (P_1) with high probability.

3.4 Proof invertibility

In this subsection we are going to present the proof of the invertibility of $\mathcal{F}^*_{T\to\Omega}\mathcal{F}_{T\to\Omega}$.

To begin we can rewrite the polynomial P by introducing:

$$\begin{aligned}
&Hf(t) = -\sum_{\omega\in\Omega}\sum_{t'\in T: t'\neq t} e^{2\pi i \frac{w(t'-t)}{N}} \\
&H_0 = \iota^* H \\
&P = (\iota - \frac{1}{|\Omega|}H)(I_T - \frac{1}{|\Omega|}H_0)^{-1}\iota^* sgn(f)
\end{aligned}$$

So, to show that $\mathcal{F}^*_{T\to\Omega}\mathcal{F}_{T\to\Omega}$ is invertible it is sufficient to show that the largest eigen value of H_0 (operator norm of H_0) is less than $|\Omega|$. To do this the authors propose to bound the operator norm by the Frobenius one and use an estimate of the 2nth moment of H_0. The key argument, which will not be proven here is:

$$\begin{aligned}
&\text{Suppose that } \tau \leq \frac{1}{1+e} \text{ and } n \leq \frac{\tau N}{4|T|(1-\tau)} \\
&\text{Then } \mathbb{E}[Tr(H_0^{2n})] \leq 2\left(\frac{4}{e(1-\tau)}\right)^n n^{n+1}|\tau N|^n |T|^{n+1}
\end{aligned}$$

To begin let us remark that H_0 is self-adjoint, then:

$$||H_0||^{2n} = ||H_0^n||^2 \leq ||H_0^n||_F^2 = Tr(H_0^{2n})$$

The most natural way to establish to bound the probability of events is to used deviation inequality. We are going to use the Markov one, let α be a positive number such as $0 < \alpha < 1$:

$$\begin{aligned}
\mathbb{P}(||H_0^n||_F \geq \alpha|\tau N|^n) &\leq \frac{\mathbb{E}[||H_0^n||_F^2]}{\alpha^{2n}|\tau N|^{2n}} \\
&\leq 2ne^{-n}\left(\frac{4n}{\alpha^2(1-\tau)}\right)^n \left(\frac{|T|}{|\tau N|}\right)^n |\tau|
\end{aligned}$$

Moreover if we assume that:

$$|T| \leq \frac{\alpha_M^2(1-\tau)}{4}\frac{|\tau N|}{n}, \text{ for some } \alpha_M \leq \alpha \leq 1 \tag{4}$$

We can show that:

$$\mathbb{P}(||H_0^n||_F \geq \alpha|\tau N|^n) \leq \frac{1}{2}\alpha^2 e^{-n}|\tau N| \tag{5}$$

We remark that the last inequality holds for any n provided (4) is satisfied. Let us select $n = (M+1)\log N$. The condition (4) becomes:

$$|T| \leq \frac{\alpha_M^2(1-\tau)|\tau N|}{4(M+1)\log N} = C_M(\log N)^{-1}|\tau N| \text{ which is the condition of the theorem 3.}$$

Now let us use a typical large deviation theorem (Markov + Hoeffding):

$$\mathbb{P}(|\Omega| < \mathbb{E}|\Omega| - t) \leq \exp(-\frac{t^2}{2\mathbb{E}|\Omega|})$$

$$\mathbb{P}(|\Omega| < |\tau N|(1-\epsilon_M)) := \mathbb{P}(B_M) \leq N^{-M} \text{ by taking } \epsilon_M = \sqrt{\frac{2M\log N}{|\tau N|}}$$

Moreover, let us take $n = (M+1)\log N$ and $\alpha = \frac{1}{\sqrt{2}}$ and let us introduce $A_M := \{||H_0|| \geq \frac{|\tau N|}{\sqrt{2}}\}$ then, using (5) we obtain:

$$\mathbb{P}(A_M) \leq \frac{1}{4}|\tau N|N^{-(M+1)} \leq \frac{1}{4}N^{-M}$$

On the complement of $B_M \cap A_M$ we have:

$$||H_0|| \leq \frac{\tau N}{\sqrt{2}} \leq \frac{|\Omega|}{\sqrt{2}(1-\epsilon_M)} \leq |\Omega|$$

The last inequality enforces to correctly choose M. In that case $I_T - \frac{1}{|\Omega|}$ is invertible with probability at least $1-O(N^{-M})$

3.5 Proof magnitude of P on T^c

In this subsection we are going to show that with probability at least $1-O(N^{-M})$ we have $\max_{t\in T^c}|P(t)| < 1$. We first use the following algebraic identity:

$$(1-M)^{-1} = (1-M^n)^{-1}(1+M+...+M^{n-1}) \quad \forall n \in \mathbb{N}$$

Then we can write:

$$\left(I_T - \frac{1}{|\Omega|^n}H_0^n\right)^{-1} = I_T + \sum_{p=1}^{\infty}\frac{1}{|\Omega|^{pn}}H_0^{pn} := I_T + R$$

$$\left(I_T - \frac{1}{|\Omega|}H_0\right)^{-1} = (I_T + R)\sum_{m=0}^{n-1}\frac{1}{|\Omega|^m}H_0^m$$

Now let us remind that we are working on T^c. On the complement of T we have ι which vanishes outside of T:

$$P = \frac{1}{|\Omega|}H(1 - \frac{1}{|\Omega|}H_0)^{-1}\iota^* sgn(f)$$

$$P(t) = P_0(t) + P_1(t) \quad \forall t \in T^c$$

where,

$$P_0 = S_n sgn(f), \quad P_1 = \frac{1}{|\Omega|}HR\iota^*(I + S_{n-1})sgn(f)$$

$$With\ S_n = \sum_{m=1}^{n}|\Omega|{-}m(H\iota^*)^m$$

The quantity to bound is:

$$\mathbb{P}\Big(\sup_{t\in T^c}|P(t)|>1\Big)=\mathbb{P}\Big(\sup_{t\in T^c}|P_0(t)+P_1(t)|>1\Big)=\mathbb{P}\Big(\sup_{t\in T^c}|P(t)|>a_0+a_1\Big) \text{ with } a_0,a_1>0 \text{ such as } a_1+a_1=1$$
$$\leq \mathbb{P}\Big(||P_0||_\infty>a_0\Big)+\mathbb{P}\Big(||P_1||_\infty>a_1\Big) \text{ by definition of the } l_\infty \text{ norm}$$

The idea is to bound separately each term. Put $Q_0=S_{n-1}sgn(f)$, with this notation we have:

$$\begin{aligned}||P_1||_\infty &\leq \frac{1}{|\Omega|}||HR||_\infty(1+||\iota^*Q_0||\infty) \qquad \text{(sgn(f) bounded by 1)}\\ &\leq \frac{1}{|\Omega|}||HR||_\infty(1+||Q_0||_\infty)\end{aligned}$$

Let $t\in T^c$,

$$P_0(t)=\sum_{m=1}^{n}|\Omega|^{-m}X_m(t), \qquad X_m=(H\iota^*)^m sgn(f)$$

The idea is the same as for the existence of the polynomial, we use the estimate of the moment to control the size of each term $X_m(t)$.
Let us admit the following lemma:

Lemma 2: *Set $n=km$. Then $\mathbb{E}[|X_m(t_0)|^{2k}]$ obeys:*

$$\mathbb{E}[|X_m(t_0)|^{2k}]\leq\frac{1}{|T|}B_n=2\Big(\frac{4}{e(1-\tau)}\Big)^n n^{n+1}|\tau N|^n|T|^n$$
$$\textit{Provided } n\leq\frac{\tau N}{4|T|(1-\tau)}$$

Using this lemma and the same reasoning as previously we can show $||P_0||_\infty$ is bounded with probability at least $1-O(N^{-M})$. *A fortiori* the result is the same for $||Q_0||_\infty$. The term $||HR||_\infty$ can be easily bounded. This is mainly numeric approximation that why we have decided to not present the computation. The final result is:

$$\forall t\in T^c|P(t)|<1 \text{ with probability at least } (1-O(N^{-M}) \text{ if T obeys:}$$

$$|T|\leq C_M\frac{|\tau N|}{logN}$$
$$C_M=\frac{1-\tau}{22.6(1+M)}(1+o(1))$$

One can remark that on one hand high value of M ensures that the polynomial P exists and obeys (2) and (3) with high probability. On the other hand high value for M implies to have more observations because we need $|T|\leq C_M(\log N)^{-1}|\Omega|$. There is a trade-off to find.

4 Experiments

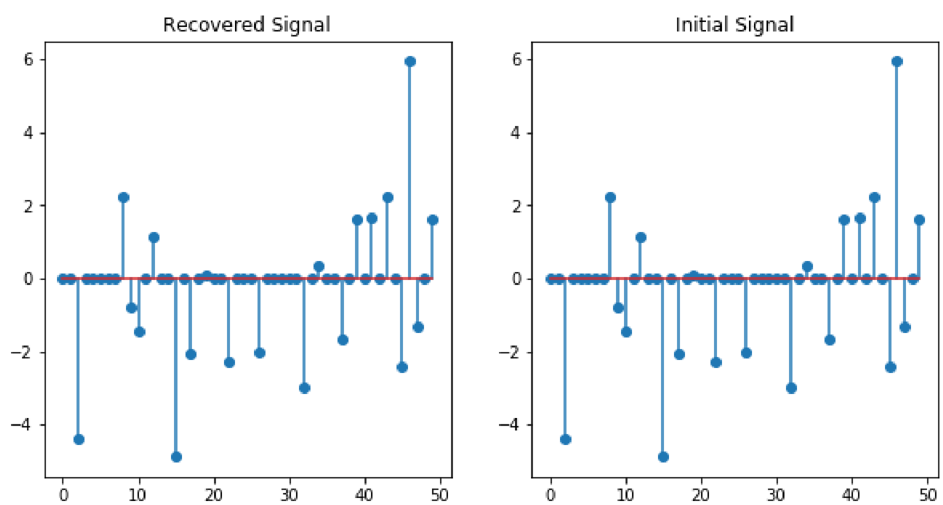
In this section we present our own implementation to recover a signal from partial Fourier coefficient. Let us remind that the problem is equivalent to solve:

$$\left| \begin{array}{l} \min_{g \in C^N} ||g||_{l_0} \\ \\ \hat{g}_{|\Omega} = \hat{f}_{|\Omega} \end{array} \right.$$

All details are presented in the Jupyter notebook. To summarize we have rewritten the optimization problem in order to use the famous Douglas-Rachford [2] algorithm which allows to minimize functionals of the form:

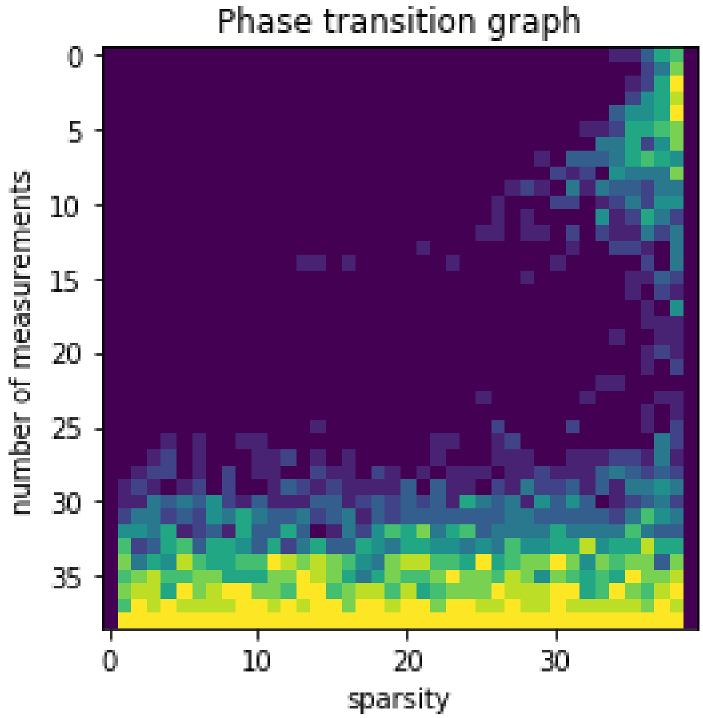
$$\left| \begin{array}{l} \min_x f(x) + g(x) \\ \\ \text{f and g are convex and we can compute the proximal operators} \end{array} \right.$$

With some convex arguments it can be shown that the algorithm converges toward a minimizer of $f + g$. One can notice that this is a splitting algorithm in the sense of we are able to work with N problem into $\mathbb{R}$ instead of one into $\mathbb{R}^N$. It is even more interesting because we can parallelize the computations.



With a signal of length 50, a support of size 5 and only 35 Fourier coefficient we have succeeded in recover the signal f perfectly.

We have done a transition phase diagram. The results are quite surprising. One can observe an area where there is no recovering.



5 Conclusion

The summary of this article is finally quite simple, this is the theorem 3: exact recovery of a discrete-time signal $f \in C^N$ can be obtained by solving a simple convex optimization problem provided we observe "enough" Fourier coefficients. We coded the procedure using a Douglas-Rachford algorithm. Other algorithms more efficient could be considered.
With more time, it would have been interesting to code that algorithm to more concrete example such as audio signal.

References

[1] Emmanuel J Candès, Justin Romberg, and Terence Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on information theory*, 52(2):489–509, 2006.

[2] Patrick L Combettes and Jean-Christophe Pesquet. Proximal splitting methods in signal processing. In *Fixed-point algorithms for inverse problems in science and engineering*, pages 185–212. Springer New York, 2011.