

Rethinking Compressive Sensing

Giuseppe Campobello

Department of Engineering - University of Messina (Italy), gcampobello@unime.it, +390903977378

Abstract—In this paper we show that Compressive Sensing (CS) can be casted as an impulse response estimation problem. Using this interpretation we re-obtain some theoretical results of CS in a simple manner. Moreover, we prove that in the case of a randomly generated sensing matrix, reconstruction probability depends on the kurtosis of the distribution used for its generation.

I. INTRODUCTION

Compressive Sensing (CS) was first introduced about ten years ago by Candés, Tao [1] and Donoho [2] and today it is one of the hottest topics in signal processing with several applications ranging from image processing [3] to data gathering in sensor networks [4]. Contrary to other compression paradigms [5][6], CS theory states that compression and sampling can be merged and carried out at the same time and, moreover, that a signal can be recovered from a few random measurements by using a sampling rate lower than that imposed by the Shannon-Nyquist theorem [7].

More formally, CS theory states that a sparse signal can be reconstructed from a small number of linear measurements by solving a l_1 -based convex optimization problem [8].

This is the usual manner to introduce CS theory; however, whilst words such as “compression” and “linear” are in the dictionary of undergraduate students, this is not necessarily true for “ l_1 -norm” and “convex optimization”. Therefore, when CS is introduced in first level courses on signal processing, it appears as a kind of “magic tool” (it is not a case that l_1 -magic [9] is the name of one of the most used tool).

This paper has its own origin exactly from this problem and one of its aim is to introduce CS by means of mathematical concepts that are familiar to undergraduate students of first level courses in signal processing.

More precisely, after a formal statement of the CS problem in Sec. II, we will show in Sec. III that a CS problem can be casted as an impulse response estimation problem. This new interpretation lead us to a new framework described in Secs.IV and V that is able to solve the CS problem. Moreover, in Sec. VI we show that the proposed framework allows to re-obtain some theoretical results of CS in a simple manner. In particular, we prove that, in the case of a randomly generated sensing matrix, reconstruction probability depends on the kurtosis of the distribution used for its generation.

With the aim of validating the proposed framework, a few simulation results are provided in Sec. VII. Finally, some concluding remarks and future works are drawn in Sec. VIII.

II. THE CS PROBLEM

According to [7] a real-valued, one-dimensional, discrete-time signal is represented as a $N \times 1$ column vector \mathbf{x} with

elements $x_i \in \mathbb{R}$ and $i \in [1, N]$; moreover, a signal is defined K -sparse if it is a linear combination of K basis vectors, i.e. if a vector \mathbf{h} with only $K \ll N$ nonzero components and a $N \times N$ matrix Ψ , named basis matrix, exist such that $\mathbf{x} = \Psi\mathbf{h}$.

Let us assume that a second $M \times N$ matrix Φ , named measurement matrix, is used to obtain the measurement vector

$$\mathbf{y}_M = \Phi\mathbf{x} = \Theta\mathbf{h} \quad (1)$$

where $\Theta = \Phi\Psi$. When $M < N$ the vector \mathbf{y}_M can be seen as a compressed representation of \mathbf{x} and it is referred to as compressed vector.

With the above notation, a CS problem consists of *designing a measurement matrix Φ and a reconstruction algorithm such that the original/uncompressed vector \mathbf{x} can be recovered from the measurement/compressed vector \mathbf{y}_M .*

Solving the above problem is similar to solving the linear system $\mathbf{y} = A\mathbf{x}$. However, CS theory addresses the case where the number of measurements is less than the number of unknown parameters showing that a unique solution exists in the case of K -sparse signals and properly designed measurement matrices [10]. In particular, good measurement matrices Φ can be obtained when elements ϕ_{ij} are normally distributed [11]. In this case the following lower bound for M exists [4],[12]:

$$M > 2K \log\left(\frac{N}{K}\right) \quad (2)$$

In this paper eq.(2) will be re-obtained in a simpler way and the above bound will be generalized to any other distribution for which the kurtosis is known.

Throughout the paper we will indicate with Ω the index set of the nonzero elements of \mathbf{h} , i.e. what is named the “support” in the algebra theory; moreover, we will indicate with \mathbf{z}_Ω the elements of a vector \mathbf{z} corresponding to the support Ω and with $\mathbf{z}_{\bar{\Omega}}$ the elements in the complementary set (note that $\mathbf{h}_{\bar{\Omega}} = \mathbf{0}$). Finally, we will indicate with Θ_Ω the $M \times K$ matrix obtained considering the columns of Θ belonging to Ω .

Hereafter we assume that the following hypotheses hold:

- H_1) ϕ_{ij} are i.i.d. random variables of whatever distribution with zero mean ($\mu_\phi = 0$) and unit variance ($\sigma_\phi^2 = 1$);
- H_2) \mathbf{x} is a K -sparse vector;
- H_3) The matrix Ψ is normalized so that $\sum_{i=1}^N \psi_{ij}^2 = 1 \forall j$;
- H_4) The matrix Θ_Ω has full column rank.

It is worth noting that H_4 is a necessary condition for several CS algorithms [10].

Note that *if the support Ω of \mathbf{h} is known and Θ_Ω has full column rank than the solution of the CS problem is straightforward.* In fact, considering that the elements of $\mathbf{h}_{\bar{\Omega}}$ are all zeros, the linear system $\mathbf{y}_M = \Theta\mathbf{h}$ can be rewritten as $\mathbf{y}_M = \Theta_\Omega\mathbf{h}_\Omega$. Moreover, under the hypothesis that Θ_Ω has

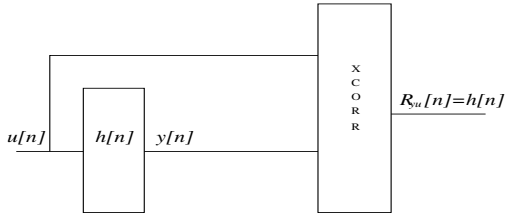


Fig. 1. Impulse response estimation by means of cross-correlation.

full column rank, the system $\mathbf{y}_M = \Theta_\Omega \mathbf{h}_\Omega$ admits a unique solution, i.e. $\mathbf{h}_\Omega = \Theta_\Omega^+ \mathbf{y}_M$ where $\Theta_\Omega^+ = (\Theta_\Omega^T \Theta_\Omega)^{-1} \Theta_\Omega^T$ is the pseudo-inverse of Θ_Ω [13]. Thus, considering that $\mathbf{h}_{\bar{\Omega}} = \mathbf{0}$, it is possible to reconstruct \mathbf{h} from \mathbf{h}_Ω . Finally, the original signal \mathbf{x} can be recovered as $\mathbf{x} = \Psi \mathbf{h}$.

As a consequence we can state that *a CS problem is solved if there exists a reconstruction algorithm able to recover the support Ω of \mathbf{h} from the measurement vector \mathbf{y}_M .*

III. BASIC IDEA

With the aim of introducing our basic idea we consider a discrete time-invariant linear system whose finite impulse response is $h[n]$ with $n \in [0, \dots, N-1]$. In this case the input signal, $u[n]$, and the output signal, $y[n]$, are related by the convolution product

$$y[n] = u[n] * h[n] = \sum_{m=0}^{N-1} h[m]u[n-m]. \quad (3)$$

According to the CS notation, the signal $h[n]$ will be represented by the vector \mathbf{h} whose elements coincide with the set $\{h_n = h[n-1] : n \in [1, \dots, N]\}$.

By introducing the cross-correlation function between two signals $x[n]$ and $y[n]$ of the same length L as

$$R_{xy}[n] = \frac{1}{L} \sum_{i=0}^{L-1} x[i]y[i-n] \quad (4)$$

it is straightforward to prove that the following equation holds:

$$R_{yu}[n] = \frac{1}{L} \sum_{i=0}^{L-1} \sum_{m=0}^{N-1} h[m]u[i-m]u[i-n] \quad (5)$$

In particular, let us consider a linear system excited by an ergodic random process with an ideal impulse-like autocorrelation $R_{uu}[n] = \delta[n]$. In this case, for sufficiently large values of L , the cross-correlation function becomes $R_{yu}[n] \approx h[n] * R_{uu}[n] = h[n] * \delta[n] = h[n]$.

The previous result can be exploited as shown in Fig.1 to obtain the impulse response of a linear system [14].

In practice, ideal impulse-like autocorrelation functions cannot be achieved for signals of finite length L . Nevertheless, with a proper choice of L , the difference between $R_{yu}[n]$ and $h[n]$ can be small enough to distinguish between zero and non-zero values of $h[n]$ (and thus to obtain the support Ω). This observation leads us to a new framework able to solve the CS problem that can be explained with the aid of Fig.2.

Note that, with the exception of a decimator block in front of the cross-correlation, the upper part of the scheme in Fig.2

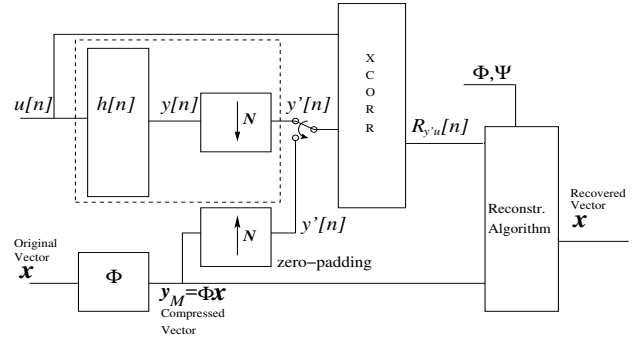


Fig. 2. Proposed framework.

coincides with Fig.1. As a consequence, it can be argued that it is possible to obtain \mathbf{h} from the cross-correlation $R_{y'u}[n]$.

However, differently from the previous scenario, in this case the estimation exploits a sub-sampled signal, i.e. the output of the decimator block, which can be expressed as

$$y'[n] = \begin{cases} y[n] & \text{if } (n+1) \equiv 0 \pmod{N} \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

where *mod* is the modulo operation (i.e. the remainder of the Euclidean division).

Let us assume that \mathbf{h} is related to a K -sparse vector \mathbf{x} by $\mathbf{x} = \Psi \mathbf{h}$ and that the input signal $u[n]$ is a sequence of $L = M \times N$ values obtained from the elements θ_{ij} of the matrix $\Theta = \Phi \Psi$ so that $u[iN-j] = \theta_{ij}$.

In this case, the sub-sampled signal $y'[n]$ defined by eq.(6) has only M non-zero values, i.e.

$$y_i = y[i \cdot N - 1] \text{ with } i \in [1, \dots, M], \quad (7)$$

that can be represented by the $M \times 1$ vector \mathbf{y}_M . Moreover, it is possible to prove that

$$\mathbf{y}_M = \Theta \mathbf{h} = \Phi \mathbf{x} \quad (8)$$

Since eq.(8) coincides with eq.(1), the problem of recovery of \mathbf{x} from \mathbf{y}_M in our framework is exactly the CS problem stated in Section II. Considering that \mathbf{x} can be straightforwardly obtained from \mathbf{h} as $\mathbf{x} = \Psi \mathbf{h}$, we can state that a CS problem is solved if \mathbf{h} can be obtained from \mathbf{y}_M .

In the next Sections we will show how it is possible to recover \mathbf{h} (and its support Ω) using the cross-correlation

$$R_{y'u}[n] = \frac{1}{L} \sum_{i=0}^{L-1} y'[i]u[i-n] \quad (9)$$

It is important to note that the signal $y'[n]$ can be obtained directly by zero-padding the vector \mathbf{y}_M , as shown in the bottom part of Fig.2. As a consequence, it is not necessary to know \mathbf{h} to evaluate $R_{y'u}[n]$, i.e. the blocks surrounded by the dashed box in Fig.2 are not necessary at all.

Summarizing, we propose to recover the original vector \mathbf{x} in two phases: in the first phase we evaluate $R_{y'u}[n]$ directly from \mathbf{y}_M and Θ ; thus, in the second phase, $R_{y'u}[n]$ is exploited to obtain Ω , \mathbf{h} and finally \mathbf{x} .

IV. PRELIMINARY RESULTS

Now it will be shown that the support Ω can be easily obtained from the cross-correlation $R_{y'u}[n]$. To this end, let us introduce the vector \mathbf{X} whose entries are $X[n] = L \cdot R_{y'u}[n]$ with $n \in [1, \dots, N]$.

In the Appendix we prove the following theorem:

Theorem 1. *Under the assumptions H_1, H_2, H_3 in Sec.II and for sufficiently large values of N , the values $X[n]$ are normally distributed with*

$$X[n] \sim \begin{cases} N(0, ME_h) & \text{if } n \notin \Omega \\ N(Mh_n, ME_{hn}) & \text{if } n \in \Omega \end{cases} \quad (10)$$

where $E_h = \sum_{m=1}^N h_m^2$, $E_{hn} = E_h + (\kappa - 2)h_n^2$ and $\kappa = \frac{E(\theta_{ij}^4)}{\sigma_\theta^4}$ is the kurtosis of the distribution of the random variable θ_{ij} .

From Theorem 1 it follows that

Theorem 2. *For sufficiently large values of M , the support Ω of \mathbf{h} coincides, with high probability, with the index set of the first K maximum absolute values of \mathbf{X} .*

Proof: Note that the theorem provides a method for obtaining the support Ω and is proven if all absolute values of \mathbf{X}_Ω are greater than the absolute values of $\mathbf{X}_{\bar{\Omega}}$. The last statement can be roughly proven as follows: from eq.(10) it follows that $E(X[n])$ is zero when $n \notin \Omega$ and it is proportional to M when $n \in \Omega$; therefore, by increasing M , we can arbitrarily increase the “distance” between the values of \mathbf{X} that are in Ω and those that are in $\bar{\Omega}$; in particular, we can choose M so that all absolute values of \mathbf{X}_Ω will be greater than the absolute values of $\mathbf{X}_{\bar{\Omega}}$.

More formally, considering $w \in \mathbf{X}_\Omega$ and $z \in \mathbf{X}_{\bar{\Omega}}$, from Theorem 1 it is possible to prove that

$$P(|w| > |z|) \geq \text{erf} \left(\frac{\sqrt{M} \cdot |h_n|}{\sqrt{2(E_h + E_{hn})}} \right) \quad (11)$$

where $\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$ and $n \in \Omega$ (i.e. $|h_n| > 0$). As a consequence, we can state that, for sufficiently large values of M , any element of \mathbf{X}_Ω will be greater (in modulus) than any element of $\mathbf{X}_{\bar{\Omega}}$ with high probability, i.e. at least equal to $\text{erf} \left(\frac{\sqrt{M} \cdot |h_n|}{\sqrt{2(E_h + E_{hn})}} \right)$. For the sake of space, the proof of eq.(11) is here omitted. ■

V. RECONSTRUCTION ALGORITHM

A reconstruction algorithm based on the previous results is shown in Fig.3. Note that it can be divided in two phases.

In the first phase (steps 1-4) the matrix Θ and the vector \mathbf{y}_M are used in order to calculate $R_{y'u}$. In practice, this phase implements the zero-padding operator and the cross-correlation in Fig.2. In the first step of the second phase (step 5), the support Ω is obtained accordingly to Theorem 2, i.e. considering the first K maximum absolute values of \mathbf{X} ; finally, the other steps (6-10) are used to recover the original vector \mathbf{x} following the procedure already discussed at the end of Sec. II.

*Reconstruction Algorithm:***Inputs:**

- measurement/compressed vector \mathbf{y}_M
- Matrix Φ, Ψ

Output: original/uncompressed vector \mathbf{x} .

Phase I:

- 1) Calculate the matrix $\Theta = \Phi\Psi$
- 2) Generate the signal $u[n]$ so that $u[iN - j] = \theta_{ij}$;
- 3) Generate the signal $y'[n]$ from \mathbf{y}_M through zero-padding (see eqs.(6) and (7));
- 4) Calculate $R_{y'u}[n]$ where $n \in [1, N]$ on the basis of eq.(9);

Phase II:

- 5) Find the index set Ω corresponding to the first K maximum values of $|X[n]| = L|R_{y'u}[n]|$;
- 6) Extract the sub-matrix Θ_Ω from Θ ;
- 7) Calculate the pseudo-inverse $\Theta_\Omega^+ = (\Theta_\Omega^T \Theta_\Omega)^{-1} \Theta_\Omega^T$;
- 8) Calculate the vector $\mathbf{h}_\Omega = \Theta_\Omega^+ \mathbf{y}_M$;
- 9) Reconstruct the vector \mathbf{h} from \mathbf{h}_Ω and Ω ;
- 10) Reconstruct the original signal \mathbf{x} as $\mathbf{x} = \Psi\mathbf{h}$.

Fig. 3. Reconstruction Algorithm for the proposed framework

The reader could observe that the proposed algorithm has some similarities with the OMP algorithm [15], and its derivatives, e.g. the CoSaMP algorithm [16]. In fact, all these algorithms exploit the same input signal $\mathbf{X} = \Theta^T \mathbf{y}_M$. For lack of space, we cannot further discuss this aspect.

Finally note that the proposed algorithm allows a perfect reconstruction when the following condition holds

$$\min_{n \in \Omega} \{|X[n]|\} > \max_{n \in \bar{\Omega}} \{|X[n]|\} \quad (12)$$

VI. ANALYTICAL RESULTS

Eq.(12) is verified with a probability P_c , henceforward named reconstruction probability, that in general depends on M, N, K, κ and \mathbf{h} . However, it is possible to argue that eq.(12) is verified with a not too small probability only if $E(\min_{n \in \Omega} \{|X[n]|\}) > E(\max_{n \in \bar{\Omega}} \{|X[n]|\})$, where $E()$ is the expectation operator. This can be formally stated as follows:

Theorem 3. *Let*

$$D = E(\min_{n \in \Omega} \{|X[n]|\}) - E(\max_{n \in \bar{\Omega}} \{|X[n]|\}) \quad (13)$$

there exists a monotone non-decreasing function $F : \mathbb{R} \rightarrow [0, 1]$ such that

$$P_c = F(D) \quad (14)$$

Proof: Let us introduce a random variable q such that

$$\min_{n \in \Omega} \{|X[n]|\} - \max_{n \in \bar{\Omega}} \{|X[n]|\} = D - q. \quad (15)$$

Note that the probability that eq.(12) is satisfied can be expressed as $P_c = P(q < D) = F_q(D)$, where F_q is the cumulative distribution function of the random variable q . We recall that a cumulative distribution function is always a monotone non-decreasing function with values in the interval $[0, 1]$, thus the theorem is proven for $F = F_q$. ■

Theorem 4. *Let D and F be defined as in Theorem 3 and*

$$M^* = \frac{(\sqrt{2 \log(N - K)} E_h + \sqrt{2 \log(K)} E_h')^2}{h_{min}^2} \quad (16)$$

where $h_{min} = \min_{n \in \Omega} \{|h_n|\}$ and $E'_h = \min_{n \in \Omega} \{E_{hn}\}$ then when $M \geq M^*$ we have $D > 0$ and

$$P_c \geq F \left(h_{min} \cdot \left(M - \sqrt{M \cdot M^*} \right) \right). \quad (17)$$

Proof: We recall that the expected maximum of m samples of a standard normal variable $x \sim N(0, 1)$ is bounded by $\sqrt{2 \log(m)}$ [17]. As a consequence the expected minimum and maximum of m samples of a normal variable $y = \mu + \sigma x \sim N(\mu, \sigma^2)$ lie in the interval $(\mu - \sigma \sqrt{2 \log(m)}, \mu + \sigma \sqrt{2 \log(m)})$.

We can use the above result to bound $E(\max_{n \in \Omega} |X[n]|)$ and $E(\min_{n \in \Omega} |X[n]|)$. In particular, from Theorem 1 we have $X_{\Omega}[n] \sim N(0, \sqrt{M E_h})$, and therefore we can state that

$$E(\max_{n \in \Omega} |X[n]|) < \sqrt{2 \log(N - K) M E_h}. \quad (18)$$

Now let us consider $E(\min_{n \in \Omega} |X[n]|)$. From eq.(16) it follows that $M^* > \frac{2 \log(K) E'_h}{h_n^2} \geq \frac{2 \log(K) E_{hn}}{h_n^2}$; thus when $M \geq M^*$ we have $h_n^2 M^2 \geq h_n^2 M \cdot M^* > 2M \log(K) E_{hn}$, i.e.

$$|h_n| M - \sqrt{2 \log(K) M E_{hn}} > 0 \quad \forall n \in \Omega \quad (19)$$

In this case we have $E(\min_{n \in \Omega} |X[n]|) > \min_{n \in \Omega} \{|h_n| M - \sqrt{2 \log(K) M E_{hn}}\}$ and considering that the right term is an increasing function of $|h_n|$, we can write

$$E(\min_{n \in \Omega} |X[n]|) > h_{min} M - \sqrt{2 \log(K) M E'_h} \quad (20)$$

By combining eqs.(18) and (20) and using the definition of M^* provided in eq.(16) we have

$$E(\min_{n \in \Omega} |X[n]|) - E(\max_{n \in \Omega} |X[n]|) > h_{min} \left(M - \sqrt{M \cdot M^*} \right) \quad (21)$$

i.e. $D > h_{min} \cdot \left(M - \sqrt{M \cdot M^*} \right)$. Consequently, when $M \geq M^*$ we have $D > 0$. Finally, on the basis of Theorem 3, F is a non-decreasing function and thus we have $P_c = F(D) \geq F \left(h_{min} \cdot \left(M - \sqrt{M \cdot M^*} \right) \right)$, i.e. eq.(17). ■

On the basis of Theorem 4, we can state that P_c is an increasing function of M and a decreasing function of M^* .

A lower bound for M^* is given by the following theorem.

Theorem 5. Let M^* defined by eq.(16), it holds

$$M^* > 2K [\log(N - K) + \log(K)] + 2(\kappa - 2) \log(K) \quad (22)$$

Proof: From eq.(16), considering that $(\sqrt{\alpha} + \sqrt{\beta})^2 > \alpha + \beta$, we have

$$M^* > \frac{(2 \log(N - K) E_h + 2 \log(K) E'_h)}{h_{min}^2}$$

Finally, by recalling that $E'_h = E_h + (\kappa - 2) h_{min}^2$ and considering that $E_h \geq K h_{min}^2$, we can write

$$M^* > 2K [\log(N - K) + \log(K)] + 2(\kappa - 2) \log(K). \quad \blacksquare$$

The previous bound generalizes other lower bounds known in CS theory. For instance, considering that the last term in eq.(22) can be neglected for small values of κ and introducing the quantity $a = \frac{\log(N - K) + \log(K)}{\log(\frac{N}{K})}$, eq.(22) can be rewritten as $M^* > 2aK \log(\frac{N}{K})$. Finally, considering that $a \approx 1$ when

$N \gg K$, we have $M^* > 2K \log(\frac{N}{K})$, that is the same lower bound in eq.(2).

From the previous results it follows that

Theorem 6. For a fixed value of M , the reconstruction probability P_c is greater for distributions with lower kurtosis.

Proof: From eq.(17) we can state that the reconstruction probability P_c increases when M^* decreases. Moreover, from Theorem 5 we can state that M^* will be lower for lower values of κ . Accordingly, P_c is a decreasing function of κ . ■

The previous theorem justifies why commonly used distributions in CS-related problems are Gaussian distribution ($\kappa = 3$) and Bernoulli distribution ($\kappa = 1$).

Simulation results provided in the next Section confirm the above statements.

VII. SIMULATION RESULTS

In this Section a few simulation results are provided with the aim of validating the proposed framework and the related analytical results. For the sake of simplicity we assume that Ψ is an unitary matrix of dimension N and $x_i \in \{-1, 0, +1\}$. We start by considering the case $\phi_{ij} \sim N(0, 1)$. In particular, in Table I we report a few experimental values of the reconstruction probability P_c achieved for different values of $K \in \{5, 10, 15\}$ and $N \in \{500, 1000\}$ and averaged over 10000 trials.

$c = \frac{M}{M^*}$	N	K	M	P_c
0.5	500	5	76	0.2308
	500	10	167	0.1320
	1000	5	81	0.1943
	1000	10	178	0.1055
	1000	15	281	0.0736
1	500	5	151	0.9364
	500	10	333	0.9313
	1000	5	162	0.9360
	1000	10	356	0.9334
	1000	15	562	0.9265
1.5	500	5	227	0.9981
	500	10	500	0.9983
	1000	5	243	0.9985
	1000	10	534	0.9988
	1000	15	843	0.9987

TABLE I
EXPERIMENTAL RECONSTRUCTION PROBABILITY P_c FOR DIFFERENT VALUES OF K , N AND M WHEN $\phi_{ij} \sim N(0, 1)$ AND $x_i \in \{-1, 0, 1\}$.

As it is possible to observe, according to Theorem 4, reconstruction probability P_c starts to be high when $M = M^*$ (i.e. $c = \frac{M}{M^*} = 1$) and increases for greater values of M (i.e. $c = 1.5$). Our simulation results show that this result is true also if different distributions are considered.

Now we wish to investigate the impact of the kurtosis on the reconstruction probability. With this aim, reconstruction probabilities obtained with different distributions but with the same number of measurements M have been considered and are shown in Tab. II.

As it is possible to observe from Tab. II, according to Theorem 6, for a fixed value of M the distribution with lower kurtosis (i.e. Bernoulli) allows to obtain a greater reconstruction probability.

N	K	M	$P_{c,Normal}$ ($\kappa = 3$)	$P_{c,Bernoulli}$ ($\kappa = 1$)	$P_{c,Laplace}$ ($\kappa = 6$)
500	5	151	0.9364	0.9765	0.8400
500	10	333	0.9313	0.9574	0.8727
1000	5	162	0.9360	0.9789	0.8397
1000	10	356	0.9334	0.9632	0.8793

TABLE II
RECONSTRUCTION PROBABILITIES FOR DIFFERENT DISTRIBUTIONS.

VIII. CONCLUSIONS

We presented a new framework capable of justifying some results of CS theory with mathematical concepts familiar to undergraduate students. Moreover, we proved that reconstruction probability depends on the kurtosis of the distribution used for generating the sensing matrix. As future work, an analytical expression for the reconstruction probability will be derived and we will extend the framework to the case $\mathbf{h}_{\bar{\Omega}} \neq \mathbf{0}$.

APPENDIX A: PROOF OF THEOREM 1

Due to the fact that $y'[n]$ is a sub-sampled signal, $R_{y'u}$ can be obtained from eq.(5) by replacing the index i with kN :

$$R_{y'u}[n] = \frac{1}{L} \sum_{m=0}^{N-1} \sum_{k=1}^M h[m]u[kN-m]u[kN-n] \quad (23)$$

$$\text{Thus, } X[n] = LR_{y'u}[n] = \sum_{m=1}^M h_m w_{mn} \quad (24)$$

$$\text{where } w_{mn} = \sum_{k=1}^M u[kN-m]u[kN-n] = \sum_{k=1}^M \theta_{km}\theta_{kn}. \quad (25)$$

Now we want to calculate mean and variance of w_{mn} and thus of $X[n]$ under the hypotheses H_1 - H_3 introduced in Sec. II. Firstly, we prove that $\theta_{km} = \sum_{i=1}^N \phi_{ki}\psi_{im}$ is a random variable with zero mean and unit variance. In fact, by considering that the expectation operator $E()$ is a linear operator and that $\text{var}(\sum_i x_i) = \sum_i \text{var}(x_i)$ when x_i are i.i.d. random variables, it follows that

$$E(\theta_{km}) = \sum_{i=1}^N E(\phi_{ki})\psi_{im} = 0 \quad (\text{see } H_1) \quad (26)$$

$$\text{var}(\theta_{km}) = \sum_{i=1}^N \text{var}(\phi_{ki})\psi_{im}^2 = 1 \quad (\text{see } H_1 \text{ and } H_3) \quad (27)$$

The previous results can be used to obtain mean and variance of the random variable w_{mn} defined in eq.(25); however, the cases $m \neq n$ and $m = n$ must be distinguished:

- when $m \neq n$, considering that θ_{km} and θ_{kn} are i.i.d., we have $E(w_{mn}) = 0$ and $\text{var}(w_{mn}) = M$;
- when $m = n$, by observing that $w_{nn} = \sum_{m=1}^M \theta_{mn}^2$ and using the fact that $E(\theta_{mn}^2) = \text{var}(\theta_{mn}) + [E(\theta_{mn})]^2 = 1$, we have $E(w_{nn}) = M$.

Furthermore, using the property $\text{var}(\theta) = E(\theta^2) - [E(\theta)]^2$, we have $\text{var}(\theta_{mn}^2) = E(\theta_{mn}^4) - [E(\theta_{mn}^2)]^2 = \kappa - 1$, where $\kappa = E(\theta_{mn}^4)/[E(\theta_{mn}^2)]^2$ is the kurtosis of the distribution of the random variable θ_{mn} , and thus

$$\text{var}(w_{nn}) = \text{var}\left(\sum_m \theta_{mn}^2\right) = M \text{var}(\theta_{mn}^2) = M(\kappa - 1) \quad (28)$$

Finally, from eq.(24) it follows that

$$E(X[n]) = h_n E(w_{nn}) + \sum_{m \neq n} h_m E(w_{mn}) = M h_n \quad (29)$$

$$\begin{aligned} \text{var}(X[n]) &= h_n^2 \text{var}(w_{nn}) + \sum_{m \neq n} h_m^2 \text{var}(w_{mn}) = \\ &= h_n^2 M(\kappa - 1) + M \sum_{m \neq n} h_m^2 = M E_{h_n} \quad (30) \end{aligned}$$

Considering that $h_m w_{mn}$ are independent random variables and that N is usually a large value, we can use the Central Limit Theorem [18] to state that $X[n] = \sum_{m=1}^N h_m w_{mn}$ is a normal variable. Thus, using the previous results, we have

$$X[n] \sim N(M h_n, M E_{h_n}) \quad (31)$$

Finally, note that the previous equation is valid for all n , thus in particular for $n \in \Omega$, but can be simplified as $X[n] \sim N(0, M E_h)$ when $h_n = 0$ (i.e. for all $n \notin \Omega$).

REFERENCES

- [1] E. Candes and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?" *Information Theory, IEEE Transactions on*, vol. 52, no. 12, pp. 5406–5425, Dec 2006.
- [2] D. Donoho, "Compressed sensing," *Information Theory, IEEE Transactions on*, vol. 52, no. 4, pp. 1289–1306, April 2006.
- [3] N. Eslahi, A. Aghagolzadeh, "Compressive sensing image restoration using adaptive curvelet thresholding and nonlocal sparse regularization," *IEEE Tr. on Image Processing*, vol. 25, no. 7, pp. 3126–3140, July 2016.
- [4] G. Campobello, A. Segreto, and S. Serrano, "Data gathering techniques for wireless sensor networks," *Int. J. Distributed Sensor Networks*, vol. 2016, pp. 31–47, Mar. 2016.
- [5] G. Campobello, O. Giordano, A. Segreto, and S. Serrano, "Comparison of local lossless compression algorithms for wireless sensor networks," *J. Network and Computer Applications*, vol. 47, pp. 23–31, 2015.
- [6] T. Srisooksai et al., "Practical data compression in wireless sensor networks: A survey," *Journal of Network and Computer Applications*, vol. 35, pp. 37–59, 2012.
- [7] R. G. Baraniuk, "Compressive sensing [lecture notes]," *IEEE Signal Processing Magazine*, vol. 24, no. 4, pp. 118–121, July 2007.
- [8] C. Karakus, A. Gurbuz, and B. Tavli, "Analysis of energy efficiency of compressive sensing in wireless sensor networks," *Sensors Journal, IEEE*, vol. 13, no. 5, pp. 1999–2008, May 2013.
- [9] E. Candes. (2005) 11-magic: Recovery of sparse signals via convex programming. [Online]. Available: <https://statweb.stanford.edu/~candes/11magic/>
- [10] H. Zhang, W. Yin, and L. Cheng, "Necessary and sufficient conditions of solution uniqueness in 1-norm minimization," *J. Optim. Theory Appl.*, vol. 164, no. 1, pp. 109–122, Jan. 2015.
- [11] K. Hayashi, M. Nagahara, and T. Tanaka, "A users guide to compressed sensing for communications systems," *IEICE Trans. Comm.*, vol. 96, no. 3, pp. 685–712, 2013.
- [12] V. Chandrasekaran, B. Recht, P. A. Parrilo, and A. Willsky, "The convex geometry of linear inverse problems," *Foundations of Computational Mathematics*, vol. 12, no. 6, pp. 805–849, 2012.
- [13] M. James, "The generalized inverse," *Mathematical Gazette*, vol. 62, no. 420, pp. 109–114, June 1978.
- [14] L. Ljung, *System Identification: Theory for the User*. PrenticeHall, 1999.
- [15] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Tr. on Information Theory*, vol. 53, pp. 4655–4666, 2007.
- [16] J. A. Tropp and D. Needell, "Cosamp: Iterative signal recovery from incomplete and inaccurate samples," *Communications of the ACM*, vol. 53, no. 12, pp. 93–100, 2010.
- [17] G. Kamath. Bounds on the expectation of the maximum of samples from a gaussian. [Online]. Available: http://www.gautamkamath.com/writings/gaussian_max.pdf
- [18] W. Feller, *An introduction to probability theory and its applications. Volume 2, 3rd ed.*, ser. Wiley series in probability and mathematical statistics. John Wiley & sons, 1971.