Adaptive Matrix Design for Boosting Compressed Sensing

Adaptive Matrix Design for Boosting Compressed Sensing

Mauro Mangia, *Member, IEEE,* Fabio Pareschi, *Member, IEEE,* Riccardo Rovatti, *Fellow, IEEE,* Gianluca Setti, *Fellow, IEEE,*

***Abstract*—Compressed Sensing (CS) has been proposed to reduce operating cost (e.g., energy requirements) of acquisition devices by leveraging its capability of sampling and compressing an input signal at the same time. This paper aims at increasing CS performance (i.e., either achieving a better compression or allowing a higher signal reconstruction quality) and proposes two novel methods. Our first approach (*Nearly-Orthogonal CS*) is based on a geometric constraint enforcing diversity between compressed measurements, while the second one (*Maximum-Energy CS*) on a heuristic screening of candidate measurements that acts as a run-time self-adapted optimization technique. Intensive simulation results show that the proposed approaches have different applications, and ensure an appreciable performance boost with respect to the state of the art.**

I. Introduction

COMPRESSED Sensing (CS) is a signal processing technique that allows, at the same time, acquisition and compression of sparse input signals. This approach has been proposed [1], [2] to replace Analog-to-Digital Converters (ADCs) with appealing and resource-efficient Analog-to-Information Converters (AICs). Under the hypothesis of sparse input signals, it has been proven that a number of linear *measurements* potentially smaller than the corresponding number of Nyquist-rate samples, is enough to correctly reconstruct the input signal.

A number of AIC prototypes have recently been proposed in the literature. Applications range from biomedical signals [3], [4], [5] to radio-frequency signals [6], [7]. The peculiarity of all designs is that resources needed for the acquisition are tied to the information rate of the input signals and not to their bandwidth. Implementations are either analog or based on a pure digital approach [8]. In particular, once can show [9], [10] that CS may lead to clear practical advantage when used in early-digital compression stages of full digital architectures.

The cost of using such a simple and convenient sub-Nyquist sampling stage is paid at the recovery stage, where signal reconstruction from compressed measurements requires the solution of a convex optimization problem [11]. Nevertheless, many particular scenarios exist where this peculiarity makes CS extremely appealing: a typical example is given by Body Area Sensor Networks (BASNs) [12], [13] where a large number of low-complexity, miniaturized, battery-powered sensor nodes used for the acquisition of biological signals exploit the capabilities offered by the CS approach at the sensing stage, while the more power-hungry decoding stage is executed on a central gateway where energy is not a critical issue.

In this paper, we focus on the *optimization* of CS performance, either as an increase of the compression capability (i.e., a decrease in the amount of information produced by the sampling process) required for achieving a target quality, or an increase of the reconstruction quality given an information amount budget. A few works have recently been proposed on this topic. Some of them focus on the decoder side, proposing a reconstruction algorithm adapted to the peculiarities of the considered signal [14], [15] or a run-time adaptation of the sparsity basis to the input signal (*dictionary learning*) [16]. Others focus on the encoder side [17], [18], [19], [20], [21]. Note that, in the most general case, optimization at the encoder and at the decoder side is independent, and both techniques can be applied at the same time.

The aim of this paper is to make a step further in the field of encoder-side optimization. The state of the art is currently focused on two different approaches. From the one side there are some signal-agnostic techniques known as *mutual coherence minimization* [19], [20], [21], where the CS acquisition is designed to reinforce theoretical requirements [2] without any additional assumption on the characterization of the acquired class of signals. A second line of research shows that better results can be achieved when some priors on the input signal are used. The *rakeness* approach [17] aligns second-order statistical properties of the CS sensing stage with that of the input signal, while at the same time preserving requirements for a correct signal reconstruction needed by the standard CS theory.

This paper is a follow-up of [18] and proposes, following the same strategy of aligning sensing stage to the input signals features, two simple performance boosting strategies. The first one (*Maximum-Energy CS*) has been introduced in [18] and it is based on the computation of a number of measurements much higher than what strictly necessary, and on the selection of those with largest energy. The second (*Nearly-Orthogonal CS*), similarly to the rakeness-based approach [17], design the sensing process on the statistics of the acquired class of signals. With respect to rakeness-based CS, Nearly orthogonal CS ensures a minimum angle between each couple of directions along which the signal is projected during the acquisition. Roughly speaking, this reduces the overlap

M. Mangia and R. Rovatti are with the Department of Electrical, Electronic, and Information Engineering, University of Bologna, 40136 Bologna, Italy, and also with the Advanced Research Center on Electronic Systems, University of Bologna, 40125 Bologna, Italy (e-mail: mauro.mangia2@unibo.it; riccardo.rovatti@unibo.it).

F. Pareschi and G. Setti are with the Department of Engineering, University of Ferrara, 44122 Ferrara, Italy, and also with the Advanced Research Center on Electronic Systems, University of Bologna, 40125 Bologna, Italy (e-mail: fabio.pareschi@unife.it; gianluca.setti@unife.it).

in terms of information carried by each measurement thus minimizing redundancy.

The paper is organized as follows. Aim of Section II is to summarize the main concepts of CS and to introduce the notation. Section III describes some optimization techniques, including the Nearly-Orthogonal CS and the Maximum-Energy CS approaches. In Section IV, a comparison between results achieved by the considered techniques is provided by means of extensive simulations. Finally, we draw the conclusion.

II. Overview of Compressed Sensing

The CS formulation adopted here is the discrete-time one, where the input signal is an instance of a given stochastic process, and it is represented by the set of its n samples $x = (x_0, \dots, x_{n-1})^\top \in \mathbb{R}^n$ collected at Nyquist rate.

The key assumption of the entire CS framework is *sparsity*. Namely, a process is κ-sparse if a proper n-dimensional *sparsity basis* $S \in \mathbb{R}^{n \times n}$ exists, in which any possible signal instance $x = S\xi$ is represented by a vector $\xi \in \mathbb{R}^n$ with no more than $\kappa \ll n$ non-zero components.

It is intuitive that, for a sparse signal, the number of degrees of freedom in x is considerably smaller than n. This property has been leveraged by theory [11], which shows that any input signal instance can be captured by a set of $m < n$ properly designed linear *measurements*. By indicating with y the m-dimensional vector $y = (y_0, \dots, y_{m-1})^\top \in \mathbb{R}^m$ collecting such measurements, a *sensing matrix* $A \in \mathbb{R}^{m \times n}$ can be defined as the linear operator generating y from x, i.e., $y = Ax = AS\xi$.

Formal results [11], [22] guarantee that ξ (and thus x) can be recovered from y provided that $m = \mathcal{O}(\kappa \log n)$, despite the fact that A (and thus AS) yields a dimensionality reduction. Roughly speaking, there is a twofold rationale behind these guarantees. First, sensing rows of A should not be aligned with any vector in the sparsity basis S, thus avoiding the possibility of an almost-zero energy measurement. This is known as *incoherence*. Then, generic κ-sparse vectors need to be mapped *almost isometrically* into the measurements, i.e., the energy of ξ has to be almost the same of y [1].

Under these hypotheses, the recovery of the original x from y is possible by enforcing the *a priori* knowledge that its representation ξ is sparse. Mathematically, many algorithms have been explored to do so and improved in recent years [11], [23], [24]. Indicating with $\hat{x} = S\hat{\xi}$ the recovered signal, most of them are based on the solution of the convex optimization problem

$$\begin{aligned} \hat{\xi} = \arg\min_{\xi \in \mathbb{R}^n} \ & \|\xi\|_1 \\ \text{s.t.} \ & \|AS\xi - y\|_2^2 \leq \varepsilon^2 \end{aligned} \tag{1}$$

where the 1-norm $\|\xi\|_1$ is used to promote sparsity, and $\|AS\xi - y\|_2^2$ is the usual Euclidean norm indicating the accuracy with which the measurements y are matched by the solution. The value of the constant $\varepsilon \geq 0$ should be chosen proportionally to the amount of noise expected on y. Note that, according to the constraint in (1), the knowledge of S is required at the decoding stage, but not at the encoding one. Conversely, A needs to be a shared knowledge between encoder and decoder.

Most of the practical interest in CS is due to the fact that the aforementioned mathematical conditions required for the correct signal reconstruction are satisfied (in *probabilistic* terms) by simply drawing sensing matrix A at random. Although theoretical guarantees depend on the choice of the specific distribution [25], a wide class of random matrices allows effective signal recovery [22]. In practical cases, the most common choice is to draw elements of A by means of an antipodal Bernoulli distribution, i.e., $A \in \{-1, +1\}^{m \times n}$ where -1 and $+1$ occur with the same probability. This hardware-friendly choice is typically adopted in actual implementations of AIC [4], [5], [6], [7] and allow to replace expensive multiply-and-accumulate operations required in the computation of $y = Ax$ with much simpler sum and sign inversion operations. For this reason, from now on we will focus on antipodal sensing matrix only.

III. Sensing Optimization

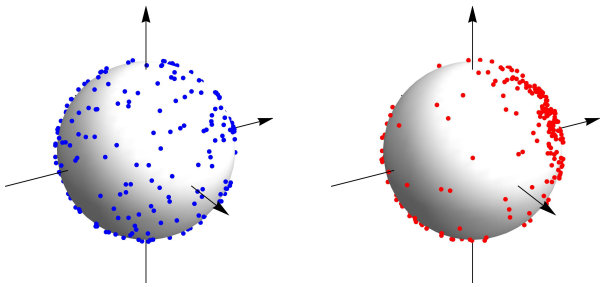
Designing sensing matrices A as instances of a random antipodal process ensures the highest possible generality as this applies equally well to any possible class of sparse input signals, i.e., any stochastic process modeling the generation of sparse instances x. However, performance may not be optimal if one has to deal with a CS system specialized to some particular class of signals. In this case, even if x is actually unknown a-priori, CS performance can be boosted by properly adopted matrices A.

A first step along this way has been proposed in the literature, introducing design guidelines that aim to increase as much as possible the mutual coherence of AS [26], [14]. Note however that the relationship between system performance and properties like the mutual coherence is related only to a bound on the reconstruction performance, and not to the actual average system behavior. In other words, there is almost no relation between reconstructed quality observed in practical CS implementations and any numerical evaluation of the aforementioned properties.

Another interesting approach is presented in [27] that proposes deterministic sensing matrix where A is composed by rows of an Hadamard transformation matrix, aiming at maximizing the entropy of the measurement vector. Although this is a very promising approach, the adoption of deterministic sensing sequences does not guarantee correct reconstruction of signals sparse in any possible domain.

In this section we propose a set of three sensing optimization procedures that exploit statistical properties of the process generating x. More precisely, we investigate *localization*, i.e., the property of the class of sparse input signals to present an energy distribution that is not uniform over the whole signal space. This property is quite common among real signals [17].

To provide an intuitive idea of localization, consider the simple toy case shown in Figure 1. The focus is on a class of signals whose instances are points on the surface of a sphere. The case in Figure 1(a) refers to a signal x that uniformly spans the whole surface, while Figure 1(b) is the case of a



(a) (b)

Fig. 1. Points on the sphere surface as instances of a purely random signal (a) and as instances of a localized signal (b).

localized signal, where the probability of a point on the surface to be a signal instance is not uniform.

Mathematically, a class of signal is localized when its correlation matrix $\mathcal{X} = \mathbf{E}[xx^\top]$ is not a multiple of the $n \times n$ identity matrix $\boldsymbol{I}_n$, or in case one deals with slices of a random process, when the power spectrum is not flat. We evaluate the localization property by computing the deviation of the eigenvalues of $\mathcal{X}$ from the isotropic case as in the following definition

$$\mathfrak{L}_x = \sum_{j=0}^{n-1} \left(\frac{\mu_j}{\mathrm{tr}(\mathcal{X})} - \frac{1}{n} \right)^2 = \frac{\mathrm{tr}\left(\mathcal{X}^2\right)}{\mathrm{tr}^2(\mathcal{X})} - \frac{1}{n} \tag{2}$$

where μ_j is the j-th eigenvalue of $\mathcal{X}$ and $\mathrm{tr}(\cdot)$ stands for matrix trace.

Note that non-localized signals come as instances of a white random stochastic process. In this case, all eigenvalues are equal to each others, with $\mathfrak{L}_x = 0$. The opposite case is when signals are non-white random, and concentrated along one direction only, i.e., a single eigenvalue is not null. In this case the localization value is maximum and equal to $\mathfrak{L}_x = 1 - \frac{1}{n}$.

Note also that sparsity and localization are different and independent priors. An example is represented by the class of signals of Figure 1(b): it is localized but not sparse. Sparsity implies that x lies in a union of κ-dimensional subspaces, while localization refers to a non-flat energy profile.

A. Rakeness-based CS

In [17], [28] a technique able to adapt the sensing matrix A to the localization profile of the class of signals to be acquired is introduced. The approach is known as rakeness-based CS (Rak-CS), and its basic underlying idea is to adapt the correlation profile of the rows of A to the one of the input signals[1]. When dealing with a localized signal, some directions exist along which the signal is most likely lying, easily identifiable from $\mathcal{X}$. Rak-CS uses this prior to design sensing matrix rows that focus on those directions. Note That, also other less energetic directions must still be explored, in order to correctly reconstruct uncommon signal instances.

[1]Infinitively this concept is similar to what we used in [29][30] to optimize (chaos-based) DS-CDMA communication, in which hip waveforms, spreading sequence statistics and rake receivers taps were jointly chosen to rake more energy at the receiver.

In other words, one should not limit to observe privileged directions, but all the signal space has still to be considered.

More precisely, the design procedure is as follows. Let us indicate with $a = (a_0, \ldots, a_{n-1})^\top$ the vector corresponding to a generic row of A. The correlation matrix $\mathcal{A} = \mathbf{E}_a[aa^\top]$ identifying the process to be used for generating a is given by the solution of the following optimization problem.

$$\max_{\mathcal{A}} \ \mathrm{tr}\left(\mathcal{A}\mathcal{X}\right) \tag{3a}$$
$$\text{s.t. } \mathrm{tr}(\mathcal{A}) = e \tag{3b}$$
$$\text{s.t. } \mathcal{A}_{j,k} = \mathcal{A}_{k,j} \qquad j,k \in \{0,\ldots,n-1\} \tag{3c}$$
$$\text{s.t. } \mathcal{A} \succeq 0 \tag{3d}$$
$$\text{s.t. } \mathfrak{L}_a \leq \ell \mathfrak{L}_x \tag{3e}$$

Since $\mathbf{E}_{a,x}\left[\left(a^\top x\right)^2\right] = \mathrm{tr}\left(\mathcal{A}\mathcal{X}\right)$, (3a) aims to maximize the expected energy of each of the entry of the measurement vector evaluated as $a^\top x$. This is how Rak-CS is able to focus the sensing process along directions where signal concentrates its energy. Conversely, constraint (3e) ensures the spanning of the whole signal space by imposing an upper bound for the localization of the sensing sequences, that has to be smaller (more precisely, scaled by a parameter ℓ) with respect to the input signal localization. A common choice is $\ell = 0.25$ [31]. Aim of (3b) is to impose that each sensing sequences has the same finite energy e, while Equations (3c) and (3d) impose matrix $\mathcal{A}$ to be symmetric and positive semi-definite..

According to [28], the correlation matrix $\mathcal{A}$ solving the above constrained optimization problem is given by

$$\mathcal{A} = e\left(\frac{\mathcal{X}}{\mathrm{tr}(\mathcal{X})}\sqrt{\ell} + \frac{1}{n}\boldsymbol{I}_n\left(1-\sqrt{\ell}\right)\right)$$

Interestingly, assuming to be capable of generating antipodal sequences with a prescribed correlation, Rak-CS is perfectly compatible with the hardware-friendly choice of having antipodal sensing matrix. In this case, $a \in \{-1,+1\}^n$, and it is easy to see that $e = n$ and that $\mathcal{A}$ can be written as

$$\mathcal{A} = \frac{n\mathcal{X}}{\mathrm{tr}(\mathcal{X})}\sqrt{\ell} + \boldsymbol{I}_n\left(1-\sqrt{\ell}\right) \tag{4}$$

The generation of antipodal sequences with an assigned profile has been already discussed in [17], [28]. Here it is enough to highlight that different approaches can be used according to the peculiarity of the actual CS system implementation.

Assuming a CS encoder whose hardware is designed so that sensing sequences (i.e., the rows of A) are pre-computed and stored in a local memory, the clipping of Gaussian random sequences with a correlation profile that is a pre-distorted version of the desired one [32] is the easiest approach. Note that in this setting, the adoption of Rak-CS is completely transparent to the encoding stage. No additional costs are required in terms of energy requirement or hardware complexity, since it is enough to load sequences generated according to (4) into memory.

If instead an on-the-fly generaton of A is necessary, the use of a linear probability feedback generator [33] along with a lookup table approach [34] represents a possible replacement of an independent and identically distributed antipodal random

number generator required by standard CS. A simple hardware implementing this solution is the linear feedback shift register (see, f.i., [4], [8]) even if sometimes a more complex hardware is adopted [3].

B. Nearly-Orthogonal CS

Rak-CS administrates the trade-off between the ability to focus sensing on a proper region of the signal space and the need to span the whole space. Nearly-Orthogonal CS (NeO-CS) relies on a different trade-off to allow the whole space to be spanned. More precisely, let first adopt the correlation matrix of the input signal as that of the sensing matrix rows, excepts for a scaling factor due to the antipodality constraint. Mathematically we have

$$\mathcal{A} = n\frac{\mathcal{X}}{\mathrm{tr}(\mathcal{X})} \tag{5}$$

Note that this automatically implies $\mathfrak{L}_a = \mathfrak{L}_x$.

The constraint for allowing this approach to span the entire signal space is based on the observation that, for large $\mathfrak{L}_a = \mathfrak{L}_x$, the generation of A could result in a matrix with many very similar rows, whose associated measurements would bring almost the same information. To avoid this, we impose a bound on the minimum angles between each couples of rows composing A.

Assume that the rows of A are made of samples from an antipodal zero-mean and mixing stochastic process with power spectrum $S_a(f)$ and autocorrelation profile

$$C_a(\tau) = \int_{-1/2}^{1/2} S_a(f) e^{2\pi i f \tau} \mathrm{d}\, f$$

so that the entries of the correlation matrix $\mathcal{A} = \mathbf{E}[aa^\top]$ are $\mathcal{A}_{j,k} = C_a(k-j)$ where a is the generic A row.

If two independent rows a' and a'' are taken, the cosine of their angle $\alpha = \widehat{a'a''}$ is

$$\cos(\alpha) = \frac{(a')^\top a''}{\|a'\|_2 \|a''\|_2} = \frac{1}{n}\sum_{j=0}^{n-1} a'_j a''_j \tag{6}$$

where we have exploited the fact that for n-dimensional antipodal vectors $\|a'\|_2 = \|a''\|_2 = \sqrt{n}$.

Now, let us consider $b_j = a'_j a''_j$. The corresponding process is mixing, with a correlation function $C_b(\tau) = \mathbf{E}[a'_0 a''_0 a'_\tau a''_\tau] = \mathbf{E}[a'_0 a'_\tau]\mathbf{E}[a''_0 a''_\tau] = C_a^2(\tau)$. Hence, by applying [35, Theorem 27.4] we know that

$$\frac{1}{\sqrt{n}}\sum_{j=0}^{n-1} b_j \underset{n\to\infty}{\sim} \mathcal{N}\left(0, \sigma^2\right) \tag{7}$$

i.e., it is asymptotically Gaussian with zero mean and variance

$$\sigma^2 = \sum_{\tau=-\infty}^{\infty} C_a^2(\tau) \tag{8}$$

Thanks to the Parseval's equality we also know that $\sigma^2 = \sum_{\tau=-\infty}^{\infty} C_a^2(\tau) = \int_{-1/2}^{1/2} S_a^2(f)\mathrm{d}\, f$. Moreover, if one considers the eigenvalues λ_j of the correlation matrix $\mathcal{A}$, the main theorem in [36, Chapter 5] states that

$$\sigma^2 = \int_{-1/2}^{1/2} S_a^2(f)\mathrm{d}\, f = \lim_{n\to\infty}\frac{1}{n}\sum_{j=0}^{n-1}\lambda_j^2$$

from which, given that for antipodal processes $\mathrm{tr}(\mathcal{A}) = n$, we infer that for large values of n

$$\sigma^2 \simeq n\frac{\mathrm{tr}\mathcal{A}^2}{\mathrm{tr}^2(\mathcal{A})} = n\mathfrak{L}_a + 1$$

where the last equality directly descends from the definition of localization (2). Since $\cos(\alpha)$ is $1/\sqrt{n}$ times the normalized sum in (7), we finally have

$$\cos(\alpha) \underset{n\to\infty}{\sim} \mathcal{N}\left(0, \mathfrak{L}_a + \frac{1}{n}\right) \tag{9}$$

This ascertained, the probability p that two rows of A can be considered orthogonal up to a certain tolerance θ is

$$p = \Pr\left\{|\cos(\alpha)| \le \theta\right\} = \mathrm{erf}\left(\frac{\theta}{\sqrt{2\left(\mathfrak{L}_x + 1/n\right)}}\right) \tag{10}$$

where $\mathfrak{L}_x = \mathfrak{L}_a$ is imposed according to the proposed method.

Equation (10) can be also inverted to determinate the threshold θ on the maximum value of $|\cos(\alpha)|$ with a certain probability p such that

$$\theta = \sqrt{2\left(\mathfrak{L}_x + \frac{1}{n}\right)}\,\mathrm{erf}^{-1}(p) \tag{11}$$

Clearly, (10) holds only for two rows of A, while the proposed method imposes a maximum value of $|\cos(\alpha)|$ to any possible couple of rows. In this sense (11) could be used only as a guideline in the evaluation of θ to be imposed to the entire matrix, and the probability that the imposed minimum angle condition holds between any couple of rows rapidly decreases with m.

Overall, the design of a sensing matrix based on NeO-CS can be summarized as follow:

1) a first antipodal row is generated with a correlation profile as in (5);
2) to obtain the second row, antipodal sequences with the same correlation matrix are generated until $|\cos(\alpha)| \le \theta$ where α is the angle between the current sequences and the first row and θ is evaluated by (11) given the value of p;
3) the rows generation goes ahead as in the previous step, where the constraint on the minimum angle is imposed with respect to all previously generated rows.

Note that the above algorithm is quite complex, and not designed for the on-the-fly generation of the rows of A. NeO-CS finds therefore natural applications in encoders where sensing sequences are pre-computed and stored in a local memory. In this case, as in the Rak-CS approach, optimization comes with no additional costs in terms of encoder energy requirement or complexity, since it is enough lo load sequences generated with the described procedure into memory.

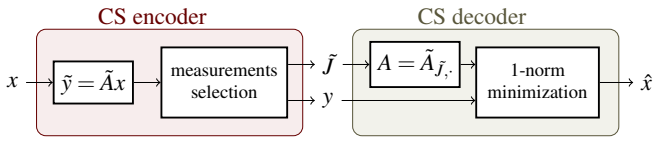


Fig. 2. System implementing the Max-CS.

C. Maximum-Energy CS

Both Rak-CS and NeO-CS rely on off-line procedures to tune $\mathcal{A}$ to $\mathcal{X}$. This section describes a run-time approach: here the alignment of the rows of A with directions of x is achieved by looking at the m rows that, among M candidates (with $M > m$), have the largest energy. We refer to this technique as Maximum-Energy CS (Max-CS).

Limiting to the hardware-friendly antipodal case, and indicating with $\tilde{A} \in \{-1, +1\}^{M \times n}$ a matrix with white random antipodal entries, we can formalize the approach as follows.

1) Given the actual input signal instance x, the whole set of the M candidate rows memorized in $\tilde{A}$ is used to compute the set of M measurements by means of the vector
$$\tilde{y} = \tilde{A}x$$
2) The m elements of $\tilde{y}$ with the highest absolute values are identified.
Let $\tilde{J}$ be the set of indexes of these elements, and let $A \in \{-1, +1\}^{m \times n}$ be the matrix composed by the rows of $\tilde{A}$ addressed by $\tilde{J}$. In this way $y = Ax$ is the m-size measurement vector with the highest energy.
3) For each signal instance, the decoder stage retrieves from the encoder, both $\tilde{J}$ and y. Assuming that $\tilde{A}$ is known both by the encoder and by the decoder, retrieving $\tilde{J}$ allows the decoder to compose A and to reconstruct $\hat{x}$ from y by (1) without any additional requirement.

A visual representation of the system implementing the described approach is shown in Figure 2.

The main advantage of the Maximum-Energy approach is the adaptability of the sensing procedure to each signal instance without any requirements on the knowledge of the statistic of input signal class. In other words, Max-CS is a run-time self-adapting CS encoder.

Note that, with this approach $\tilde{J}$ must be transmitted along with y. The overhead due to the transmission of $\tilde{J}$ must be compensated by a proper reduction of the number of measurement needed to correctly reconstruct x. In order to determine if the trade-off is profitable, we have to compare Max-CS behavior with that of others CS approaches by taking into account the total amount of transmitted information from the encoder to the decoder. To this aim, let us compute the total amount of bits used for each signal instance by a Max-CS based encoder.

The number of bits required to encode $\tilde{J}$ depends both on M and m. Even if plenty of different coding techniques can be considered, the smartest one is to identify which specific subset of m elements among M possible candidates is the correct one among the $\binom{M}{m}$ possibilities. As a consequence, the corresponding minimum amount of bits $b_{\tilde{J}}(M, m)$ required to represent $\tilde{J}$ is

$$b_{\tilde{J}}(M, m) = \left\lceil \log_2 \binom{M}{m} \right\rceil \tag{12}$$

where $\lceil \cdot \rceil$ is the smallest integer not smaller than its argument.

So, the overall encoding of m measurements, each one with b_y bits, and of $\tilde{J}$ requires $mb_y + b_{\tilde{J}}(M, m)$ bits. Comparing this with the straightforward option of a Nyquist approach, where n samples are encoded with b_x bits, gives rise to a bit-wise compression ratio

$$CR_{\text{bit}} = \frac{nb_x}{mb_y + b_{\tilde{J}}(M, m)} \tag{13}$$

In the next section, by means of a few practical examples, the CR_{bit} in (13) will be compared with that achieved by standard CS, Rak-CS and NeO-CS, that is simply given by (13) when assuming $b_{\tilde{J}}(M, m) = 0$. The role of M as degree of freedom in the design of a Max-CS system will also be investigated. In fact, clearly, increasing M increases the possibility to have rows of $\tilde{A}$ with a good alignment with the principal signal instance components (i.e., to have measurements with a high absolute value), and this could increase performance. However, M should not be too large to avoid a drastic reduction of CR_{bit}.

As a final comment, it may be interesting to evaluate what is the overhead in terms of complexity of the Max-CS approach with respect to the standard CS. Two aspects have to be considered. First, M measurements have to be computed instead of m. Accordingly, the complexity increase of the projection stage is M/m. Then, measurements has to be sorted according to their absolute value. The most common sorting algorithms (*quicksort*, *heapsort*, ...) present an average complexity equal to $M \log_2 M$. However, Max-CS does not require a fully sorted measurement list, but only the identification of the m elements with highest value. Without entering into details, it is possible to see that, for $m \ll M$, the asymptotic complexity for this task is $\mathcal{O}(m \log M)$.

IV. Results

A. Framework for Performance Evaluation

To compare the performance of the considered approaches, a set of results based on extensive Montecarlo simulations is provided. Two figures of merit are used. First, let us define the Reconstruction Signal-to-Noise-Ratio (RSNR) as the ratio (expressed in dB) between the 2-norm of the input signal instance x and the 2-norm of the reconstruction error $x - \hat{x}$. The first figure of merit considered is the Average RSNR (ARSNR) expected over any possible input signal instance and any possible sensing matrix

$$\text{ARSNR} = \mathop{\mathbf{E}}_{A,x} \left[20 \log_{10} \left(\frac{\|x\|_2}{\|x - \hat{x}\|_2} \right) \right]$$

that we evaluate as the average RSNR observed in the trials.

Alternatively, reconstructions performance can be evaluated by assuming that the reconstruction is correct when the corresponding RSNR exceeds a certain minimum value

$\mathrm{RSNR}_{\min}$. As such, the second figure of merit considered is the Probability of Correct Reconstruction (PCR) defined as

$$\mathrm{PCR} = \Pr\{\mathrm{RSNR} \geq \mathrm{RSNR}_{\min}\}$$

that we estimated using the success rate over the trials.

To cope with possible different class of signals, a simulation environment has been developed to generate n-dimensional instances x that are both localized and κ-sparse with respect to a certain orthonormal basis S. The signal generation starts from an instance of a zero-mean Gaussian random vector x' with covariance/correlation matrix $\mathcal{X}'$ and follows these steps:

1) take $x' \sim \mathcal{N}(0, \mathcal{X}')$ as initial vector;
2) compute ξ' as the representation of x' over S such that
$$\xi' = S^{-1}x' = S^{\top}x';$$
3) select the κ element of ξ' with the higher absolute values to obtain the vector ξ;
4) finally, compute $x = S\xi$.

The approach aims to formalize the intuitive idea of taking a possibly non-white vector (x'), project it onto the basis along which we want the signal to be sparse, sparsify it, and map it back into its original basis.

A proper localization can be imposed by using a non-diagonal $\mathcal{X}'$ whose features will approximately be translated into those of $\mathcal{X}$, i.e., the correlation matrix of the generated signals. Despite the fact that the relationship between $\mathcal{X}'$ and $\mathcal{X}$ is difficult to model analytically, we may expect to have only a relative low difference in the statistics of x with respect to that imposed to x'. In fact, since the κ largest components of ξ' are carried over to ξ, ξ is the best possible κ-sparse approximation of ξ', and the same relationship holds between x and x'. Hence, the larger the κ, the more similar the behavior of x to that of x'.

In the following, we will consider $\mathcal{X}'$ such that $\mathcal{X}'_{j,k} = r^{|j-k|}$ for some $-1 < r < 1$. This means that x' is a chunk of a stationary stochastic process that assumes a high pass profile for $-1 < r < 0$ and a low-pass profile for $0 < r < 1$. Furthermore, we set $n = 128$, $\kappa = 6$ and S to be either the orthonormal Discrete Cosine Transform (DCT) basis or the Daubechies-8 Wavelet transformation basis (DAUB). With this setting, and by generating a large amount of sample vectors x' and thus of x, it is possible to estimate the localization $\mathfrak{L}_x$ for any given value of r.

The proposed simulation setting is reported in Table I, and includes signals with either a low-pass (LP) or a high-pass (HP) profile, and signals that are highly localized (HL) or that are lowly localized (LL). Although this is not representative for any possible class of signals, it includes cases of sparse signals with opposite power spectrum profiles, with different localization values and different sparsity matrices. In particular, the choice of S aims at including a case (the DCT) presenting columns that have only a limited number of non-consecutive zero elements, and a case (DAUB) where the columns present a compact support. Note that Table I also includes the estimation of the localization values.

Finally, in order to model possible system non-idealities, an additional white Gaussian noise is added to each signal instance x such that the corresponding SNR is $60\,\mathrm{dB}$. This value has to be considered as a reference asymptotic level when discussing performance of the proposed optimization approaches. For this reason, we set to $\mathrm{RSNR}_{\min} = 55\,\mathrm{dB}$ in the computation of the PCR.

TABLE I
CLASSES OF SIGNALS IN THE PROPOSED SIMULATION ENVIRONMENT.

Sparsity basis	r	Power spectrum profile	$\mathfrak{L}_x$	Localization profile
DCT	0.58	LP	0.02	LL
DCT	0.96	LP	0.2	HL
DCT	-0.58	HP	0.02	LL
DCT	-0.96	HP	0.2	HL
DAUB	-0.78	HP	0.013	LL
DAUB	0.97	LP	0.25	HL

Signals reconstructions, i.e., the solution of (1), are obtained by means of SPGL1 toolbox[2].

B. CS approaches comparison

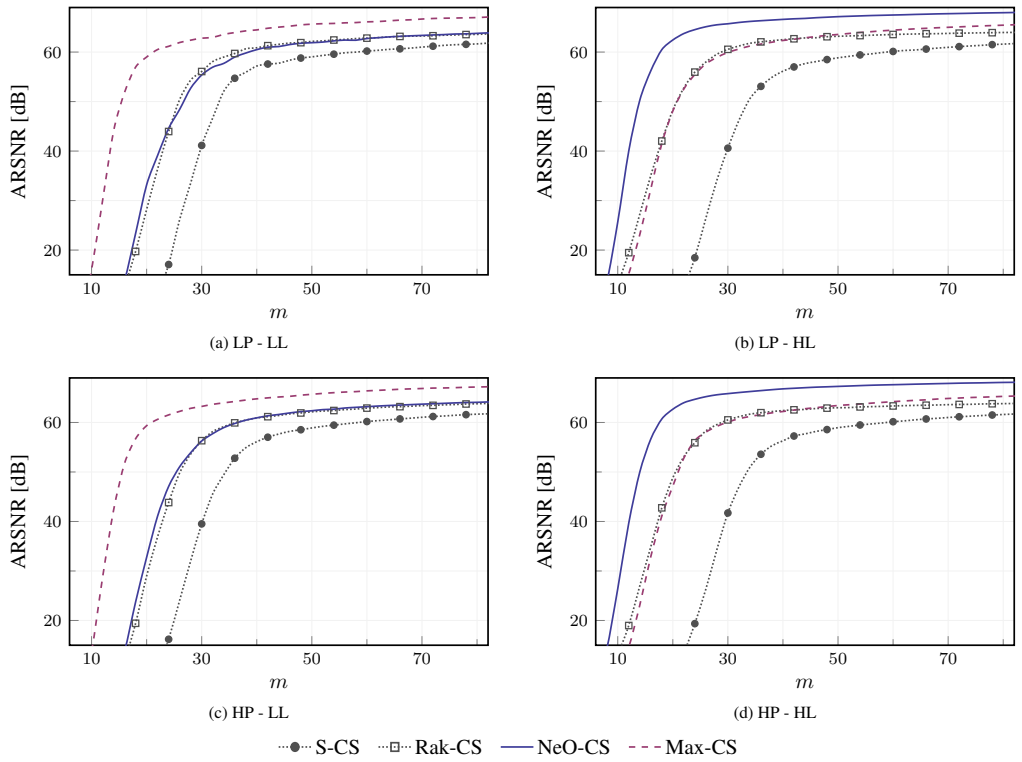
The simulation setting is designed to allow the comparison of results when different class of signals in terms of localization values, sparsity basis and power spectrum profiles are taken into account. Performance results in terms of both ARSNR and PCR for different values of m are considered, comparing what is achieved by the standard non-optimized CS, with independent and identically distributed antipodal random matrices (S-CS) with what is achieved by using the proposed optimization approaches. Figures 3 and 4 refer to the DCT case, showing ARSNR and PCR, respectively. Figure 5 refers to the DAUB case. Every point in the figures has been obtained by 1000 Montecarlo trials, each of them considering a different combination of input signal instance x and sensing matrix A.

In all considered cases, Rak-CS is capable of outperforming the standard CS approach. For this reason, we consider from now on results of the Rak-CS optimization as a benchmark to evaluate both NeO-CS and Max-CS.

Simulations with DCT-sparse signals of Figures 3 and 4 include all four corner cases (high- and low-localization, high- and low-pass profile). In this case, it is easy to understand the behavior of the proposed optimization techniques: independently of the high-pass or low-pass profile, when considering lowly localized signals, the NeO-CS approach has performance not different from the benchmark Rak-CS, while the Max-CS provides a considerable boost. Conversely, when considering highly localized signals, the NeO-CS is the approach providing a considerable performance boost, while performance of Max-CS are aligned with the benchmark. A very similar behavior can be observed when the DAUB sparsity basis is considered.

In conclusion it can be inferred that, when dealing with highly localized signal, the NeO-CS is capable of exploiting all possibilities offered by the localization thus outperforming all other approaches, while low localized signal take advantages by the run-time optimization offered by the Max-CS.

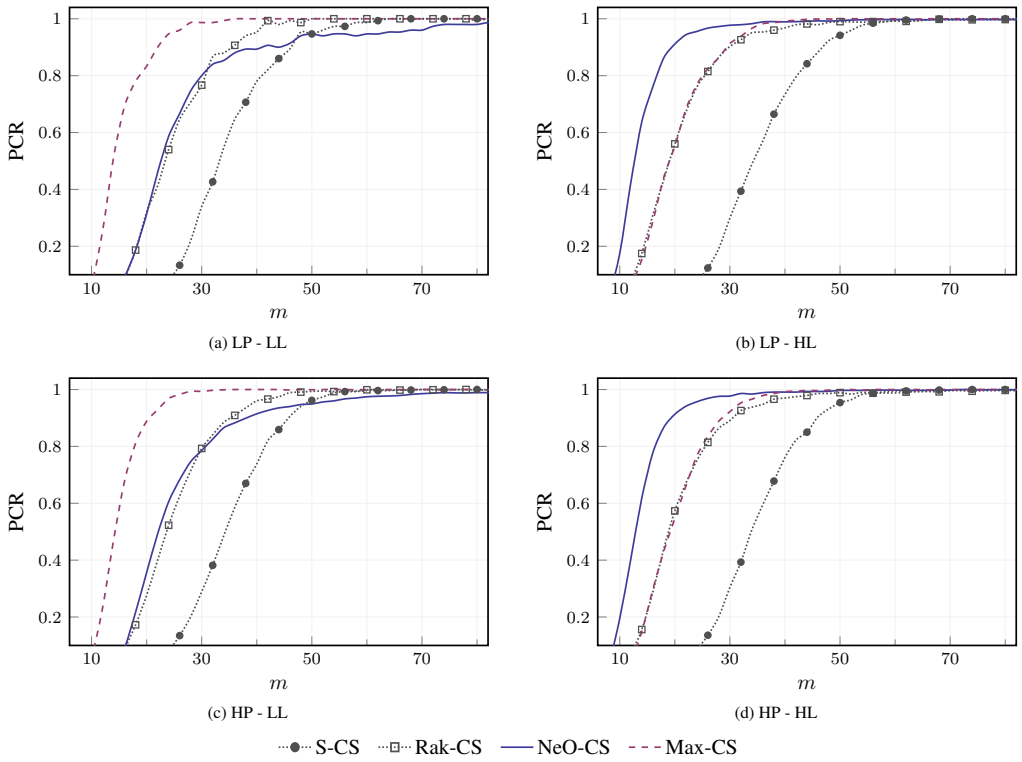
[2]online available at http://www.cs.ubc.ca/~mpf/spgl1/



(a) LP - LL (b) LP - HL

(c) HP - LL (d) HP - HL

Fig. 3. Performance comparison in terms of ARSNR between the standard CS and of the considered optimization technique for different input signals sparse on the DCT basis where $\ell = 0.25$ for Rak-CS, $M = 512$ for Max-CS and $p = 0.7$ for NeO-CS. (a): low-pass profile, lowly localized (LP-LL); (b): low-pass profile, highly localized (LP-HL); (c): high-pass profile, lowly localized (HP-LL); (d): high-pass profile, highly localized (HP-HL).



(a) LP - LL (b) LP - HL

(c) HP - LL (d) HP - HL

Fig. 4. Performance comparison in terms of PCR between the standard CS and of the considered optimization technique for different input signals sparse on the DCT basis where $\ell = 0.25$ for Rak-CS, $M = 512$ for Max-CS and $p = 0.7$ for NeO-CS. (a): low-pass profile, lowly localized (LP-LL); (b): low-pass profile, highly localized (LP-HL); (c): high-pass profile, lowly localized (HP-LL); (d): high-pass profile, highly localized (HP-HL).

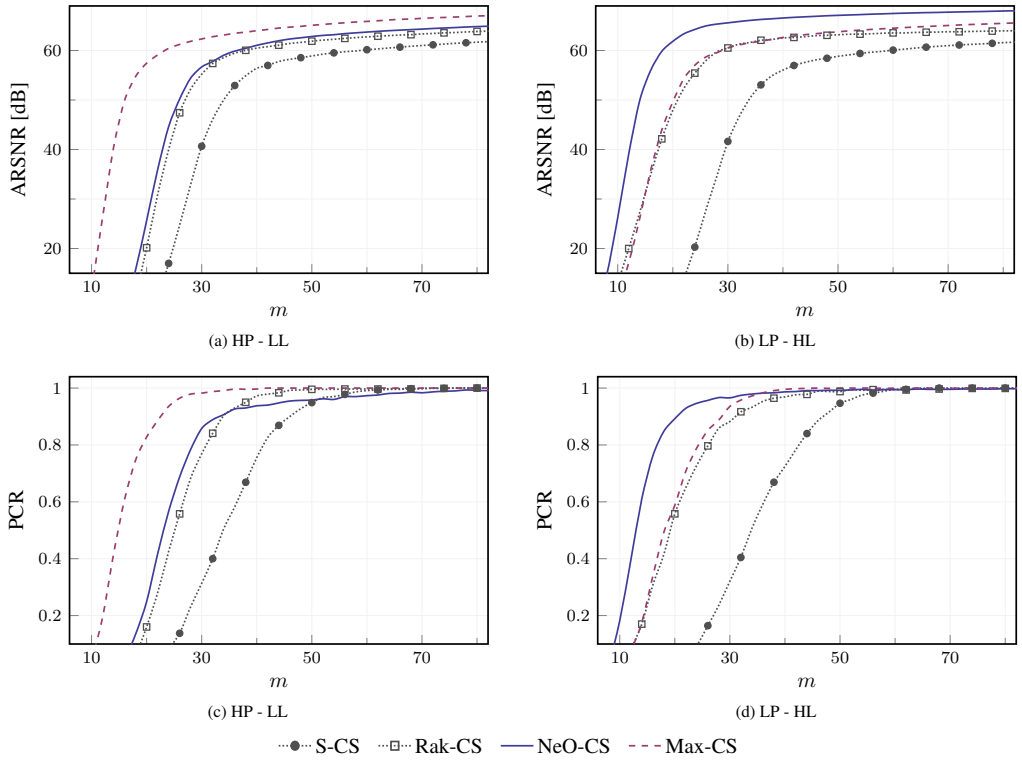


Fig. 5. Performance comparison in terms of both ARSNR and PCR between the standard CS and of the considered optimization technique for different input signals sparse on the DAUB basis where $\ell = 0.25$ for Rak-CS, $M = 512$ for Max-CS and $p = 0.7$ for NeO-CS. (a): ARSNR of high-pass profile, lowly localized signal (HP-LL); (b): ARSNR of low-pass profile, highly localized signal (LP-HL); (c): PCR of high-pass profile, lowly localized signal (HP-LL); (d): PCR of low-pass profile, highly localized signal (LP-HL).

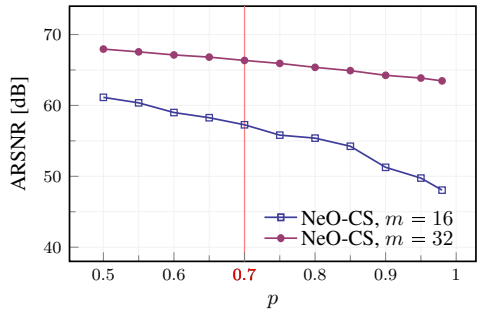


Fig. 6. Neo-CS performance in terms of ARSNR as a function of p for LP-HL signals sparse on DCT. The highlighted vertical line refer to performances shown in Figure 3(b).

C. NeO-CS performances

In the examples of Figures 3, 4 and 5 the threshold θ is such that $p = 0.7$. It may be interesting to determine how performance changes according to p.

This can be seen in Figure 6, showing the ARSNR as a function of p in the case of LP-HL signals sparse on DCT, with $m = 16$ and $m = 32$. The value $p = 0.7$ has been highlighted to allow a better identification of the working point used in Figure 3(b) (both figures refer to the same class of signals). As expected, the stronger the condition on the minimum angle (i.e., the lower p), the better the performance. The price to pay for increased performance is that generating A is more difficult due to the low value of p.

As noted in the previous section, the difficulty of imposing a minimum angle between each couple of rows of A is also strongly increasing with m. Although sensing sequences are off-line computed and locally stored in the device performing compression, this increase in complexity could not be balanced by a sufficient increase in performance. For $m = 32$, the difference in performance is small, while when $m = 16$ the difference is large and decreasing p is certainly worth. A motivation behind this phenomena can be found in how the sensing matrix is able to span the signal space. Although we use localized sequences, increasing m correspond to an increase in the ability of the encoder to capture the information content of x such that the increasing in the minimum angle required between the rows of A is less effective.

To prove the effectiveness of NeO-CS approach we consider also performance related to the acquisition and reconstruction of a real-life signal as an electrocardiogram (ECG). In particular an ECG track is generated according to [37] with same setting used in [17] where sampling rate is $256\,\mathrm{Hz}$ and the heart rate ranging from 40 to $80\,\mathrm{bpm}$. Signal instances are also perturbed by a white additive Gaussian noise with a $50\,\mathrm{dB}$ SNR to emulate non-idealities as quantization.

Since the ECG is known to be a HL signal [31], according to the synthetic example of Section IV-B, we expect that NeO-CS would outperform both standard CS and Rak-CS. With 1 second windows ($n = 256$) and the Symlet-6 wavelet transformation as sparsity basis, results in terms of ARSNR

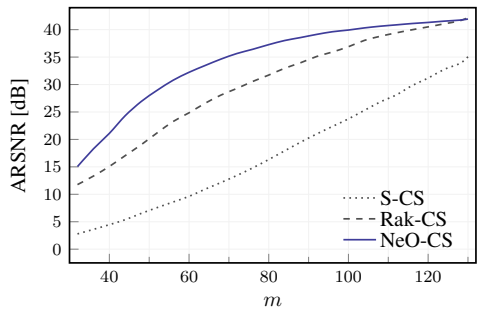


Fig. 7. Performance comparison in terms of ARSNR as a function of m between S-CS, Rak-CS (with $\ell = 0.5$) and Neo-CS (with $p = 0.7$) in the ECG signal testing.

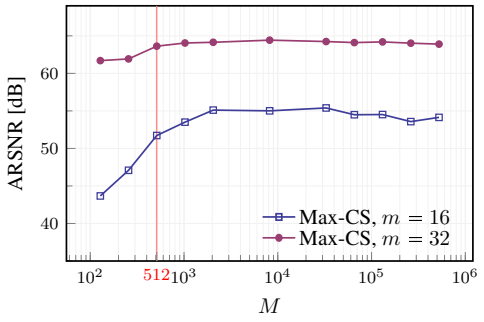


Fig. 8. Max-CS performance in terms of ARSNR as a function of M for HP-LL signals sparse on DCT. The highlighted vertical line refer to performances shown in Figure 3(c).

as a function of m are shown in Fig. 7 for S-CS, Rak-CS (with $\ell = 0.5$) and NeO-CS (with $p = 0.7$). Simulation results confirm that NeO-CS strongly outperforms S-CS while the introduced advantage with respect to Rak-CS decrease whem m increase.

D. Max-CS performances

When focusing on Max-CS approach, the role of M must be discussed. The idea to select m sequences over M candidates implies that an increase in M corresponds to an increase in the probability to select the rows of A properly aligned with the actual signal instance, thus increasing performance. However,

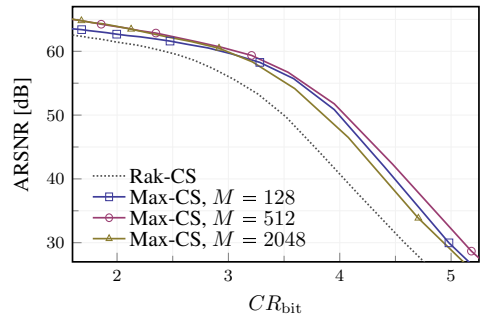


Fig. 9. Max-CS performance in terms of ARSNR as a function of CR_{bit} for HP-LL signals sparse on DCT including Rak-CS performance as reference.

increasing M implies that $b_{\tilde{J}}$ in (13) also increases implying a tighter a trade-off.

Figure 8 shows the impact of M in the ARSNR for two different m values in the reconstruction of HP-LL signals sparse on DCT. In the figure the value of M used in Figure 3(c) has been highlighted. Remarkably, performance saturates for large values of M and, considering the impact of M on the total amount of information transmitted from the encoder to the decoder, an optimum value can be found.

In the adopted simulation setting, each x instance has been perturbed by a white Gaussian noise with a corresponding signal-to-noise ratio equal to 60 dB to emulate non-idealities of the sensing stage. This can be approximated as 10 bit quantization noise, so b_x in (13) can be set to 10. To encode the measurement vector we adopt the conservative choice $b_y = \lceil 1/2 \log_2(n) \rceil = 14$ as suggested in [38] that does not implies the re-quantizzation of y. In this setting, Figures 9 shows the performance in terms of ARSNR as a function of CR_{bit} for three different values of M and for the Rak-CS as reference. Results confirm the discussed trade-off by showing how, in this case, the intermediate value $M = 512$ is able to outperform higher or lower considered values.

Max-CS has also been tested on image reconstructions task. Without any limitation to a particular class of images, i.e., without focusing on specific cases such as, f.i., Magnetic Resonance Imaging or Barcode images, it is not possible to estimate the statistic of the acquired class of signals. In this scenario both Rak-CS and NeO-CS cannot be applied, while Max-CS is expected to provide a performance improvement with respect to S-CS.

As test image X (8-bit 256×256 pixels) we used the gray scale version of the *pears* image from the Matlab demo images repository of the image processing toolbox. X has been partitioned into 16×16 blocks x ($n = 256$), and reconstructions were preformed by the 16×16 Daubechies-8 Wavelet as 2D sparsity basis. After that, the $N \times N$ reconstructed images $\tilde{X}$ (with $N = 256$) were compared with the original picture for different values of m and M. To evaluate the reconstruction quality, as common in this field, we use the average observed value of the mean square error (MSE), defined as

$$\text{MSE} = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{\left(X_{i,j} - \tilde{X}_{i,j}\right)^2}{L^2}$$

where L is the maximum value that each $X_{i,j}$ can assume. This allows us to get a MSE value independent of the gray-scale encoding type. Note that, differently from the SNR case, lower values correspond to a higher quality.

Results are reported in Fig. 10 as a function of the achieved CR_{bit}, i.e., taken into account the overhead given by the coding of $\tilde{J}$ in the Max-CS approach as in (13). The advantage of the Max-CS is clear also in this case. However, note that performance is almost aligned with that of the standard CS for high quality reconstructions, revealing how in this case the advantage of the Max-CS is significant only when low-quality reconstruction are tolerable in the specific application.

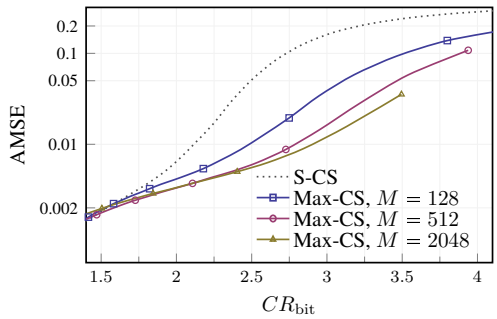


Fig. 10. Max-CS performance in terms of AMSE, as a function of CR_{bit} in the decoding of the considered test image.

E. Algorithms complexity

According to results in Sec. IV, Max-CS and NeO-CS are able to reduce the number of digital words needed to correctly reconstruct each input signal window with respect to the standard CS approach (and, consequently, also to a classical Nyquist-based system) by a quantity that depends on the actual class of input signals. So, energy requirements for measurements dispatch are relaxed. Yet, as anticipated in Sec. III, this may come at additional cost in terms of memory footprint and/or computational complexity for either generating sensing matrices or computing measurements.

The aim of this section is to give additional details on a possible comparison between the computational costs of the proposed techniques using standard CS as reference case. We evaluate both space and time complexities using test signals generated according to the setting described in Section IV-A. In particular, and referring to Table I, we use the DCT LP-LL and the DCT LP-HL cases, considering $\ell = 0.25$ for Rak-CS, $M = 512$ for Max-CS and $p = 0.7$ for NeO-CS. The value of m is the smallest one ensuring $\text{ARSNR} \geq 60\,\text{dB}$.

Two different contributions are computed: i- the cost of generating/memorizing the sensing matrix A; ii- the cost for computing the measurements vector y. The reason is that the cost for the generation of A needs to be considered only if a different A is generated for each time window, while this cost is negligible if A is fixed, and therefore generated on-line or pre-computed only once. In the latter case, the memory required to store A is the most important figure of merit. Costs for sensing matrix generation/storage are reported in Table II. Space complexity refers to the amount of memory required to store A, while time complexity to the amount of CPU time required to generate it and has been estimated as the execution time of Matlab code on a standard PC. Both space and time complexities have been normalized with respect to the value required by the standard CS case. As expected, in the NeO-CS approach the time complexity of generating A is the highest one, but the space complexity (in particular in the HL case) is the lowest one. Conversely, the space complexity for the Max-CS is always the highest one due to the need of memorizing the $\tilde{A}$ matrix.

Costs for measurements computation in terms of normalized time complexity are reported in Table III. In this case, the lower the m, the faster the measurements computation, except for the case of the Max-CS that presents a higher complexity mainly due to both the pre-computation of M candidates for y and the additional sorting step required to select the m measurements.

TABLE II
SENSING MATRICES GENERATION/STORAGE COMPARISONS IN TERMS OF NORMALIZED SPACE AND TIME COMPLEXITY OF RAK-CS ($\ell = 0.25$), NEO-CS ($p = 0.7$) AND MAX-CS ($M = 512$) WITH RESPECT TO S-CS. INPUT SIGNALS ARE LP AND SPARSE ON DCT BASIS, WITH m ENSURING $\text{ARSNR} \geq 60\,\text{dB}$.

		S-CS	Rak-CS	NeO-CS	Max-CS
LL	$\min_m(\text{ARSNR} \geq 60\,\text{dB})$	59	37	39	22
	normal. space complexity	1	0.63	0.66	8.7
	normal. time complexity	1	3.3	720	15
HL	$\min_m(\text{ARSNR} \geq 60\,\text{dB})$	59	29	18	31
	normal. space complexity	1	0.49	0.31	8.7
	normal. time complexity	1	2.6	37	15

TABLE III
MEASUREMENTS EVALUATION COMPARISONS IN TERMS OF NORMALIZED TIME COMPLEXITY OF RAK-CS ($\ell = 0.25$), NEO-CS ($p = 0.7$) AND MAX-CS ($M = 512$) WITH RESPECT TO S-CS. INPUT SIGNALS ARE LP SPARSE ON DCT BASIS, WITH m ENSURING $\text{ARSNR} \geq 60\,\text{dB}$.

		S-CS	Rak-CS	NeO-CS	Max-CS
LL	$\min_m(\text{ARSNR} \geq 60\,\text{dB})$	59	37	39	22
	normal. time complexity	1	0.88	0.88	1.42
HL	$\min_m(\text{ARSNR} \geq 60\,\text{dB})$	59	29	18	31
	normal. time complexity	1	0.84	0.78	1.42

V. CONCLUSION

Two new sensing matrix optimization techniques are proposed with the aim of increasing compressed sensing performance. The first one (named Nearly-Orthogonal CS) exploits a geometrical constraint on the rows of the sensing matrix, while the second one (Maximum-Energy CS) is based on a run-time screening of the compressed measurements.

By means of intensive numerical simulation, performance of the proposed approaches has been compared with a benchmark given by the rakeness optimization technique. Furthermore, different classes of signal have been tested, in order to investigate the applications for which the proposed methods achieve better results.

Results shows that, when the input signal is highly localized, the Nearly-Orthogonal CS is capable to achieve much better results with respect to the benchmark. Conversely, if the input signal is lowly localized, it is possible to get much better results with respect to the state-of-the-art by using the Maximum-Energy CS approach.

REFERENCES

[1] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *Signal Processing Magazine, IEEE*, vol. 25, no. 2, pp. 21–30, Feb. 2008.

[2] D. L. Donoho, "Compressed Sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[3] D. Gangopadhyay, E. G. Allstot, A. M. R. Dixon, K. Natarajan, S. Gupta, and D. J. Allstot, "Compressed sensing analog front-end for bio-sensor applications," *IEEE Journal of Solid-State Circuits*, vol. 49, no. 2, pp. 426–438, Feb 2014.

[4] M. Shoaran, M. H. Kamal, C. Pollo, P. Vandergheynst, and A. Schmid, "Compact low-power cortical recording architecture for compressive multichannel data acquisition," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 8, no. 6, pp. 857–870, Dec 2014.

[5] F. Pareschi, P. Albertini, G. Frattini, M. Mangia, R. Rovatti, and G. Setti, "Hardware-algorithms co-design and implementation of an analog-to-information converter for biosignals based on compressed sensing," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 10, no. 1, pp. 149–162, Feb. 2016.

[6] J. Yoo, S. Becker, M. Loh, M. Monge, E. Candès, and A. Emami-Neyestanak, "A 100mhz-2ghz 12.5x sub-nyquist rate receiver in 90nm cmos," in *2012 IEEE Radio Frequency Integrated Circuits Symposium*, Jun. 2012, pp. 31–34.

[7] X. Chen, E. A. Sobhy, Z. Yu, S. Hoyos, J. Silva-Martinez, S. Palermo, and B. M. Sadler, "A sub-nyquist rate compressive sensing data acquisition front-end," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 2, no. 3, pp. 542–551, Sep. 2012.

[8] J. Zhang, Y. Suo, S. Mitra, S. P. Chin, S. Hsiao, R. F. Yazicioglu, T. D. Tran, and R. Etienne-Cummings, "An efficient and compact compressed sensing microsystem for implantable neural recordings," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 8, no. 4, pp. 485–496, Aug 2014.

[9] F. Chen, A. P. Chandrakasan, and V. M. Stojanovic, "Design and analysis of a hardware-efficient compressed sensing architecture for data compression in wireless sensors," *IEEE Journal of Solid-State Circuits*, vol. 47, no. 3, pp. 744–756, Mar. 2012.

[10] D. Bortolotti, M. Mangia, A. Bartolini, R. Rovatti, G. Setti, and L. Benini, "An ultra-low power dual-mode ecg monitor for healthcare and wellness," in *2015 Design, Automation Test in Europe Conference Exhibition (DATE)*, Mar. 2015, pp. 1611–1616.

[11] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE Transactions on Information Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.

[12] A. M. R. Dixon, E. G. Allstot, D. Gangopadhyay, and D. J. Allstot, "Compressed sensing system considerations for ecg and emg wireless biosensors," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 6, no. 2, pp. 156–166, April 2012.

[13] M. Mangia, D. Bortolotti, F. Pareschi, A. Bartolini, L. Benini, R. Rovatti, and G. Setti, "Zeroing for hw-efficient compressed sensing architectures targeting data compression in wireless sensor networks," *Microprocessors and Microsystems*, vol. 48, pp. 69 – 79, 2017.

[14] J. Zhang, Z. Gu, Z. L. Yu, and Y. Li, "Energy-efficient ecg compression on wireless biosensors via minimal coherence sensing and weighted ℓ_1 minimization reconstruction," *IEEE Journal of Biomedical and Health Informatics*, vol. 19, no. 2, pp. 520–528, March 2015.

[15] Z. Zhang, T. P. Jung, S. Makeig, and B. D. Rao, "Compressed sensing for energy-efficient wireless telemonitoring of noninvasive fetal ecg via block sparse bayesian learning," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 2, pp. 300–309, Feb 2013.

[16] Y. Suo, J. Zhang, T. Xiong, P. S. Chin, R. Etienne-Cummings, and T. D. Tran, "Energy-efficient multi-mode compressed sensing system for implantable neural recordings," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 8, no. 5, pp. 648–659, Oct 2014.

[17] M. Mangia, R. Rovatti, and G. Setti, "Rakeness in the design of analog-to-information conversion of sparse and localized signals," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 59, no. 5, pp. 1001–1014, May 2012.

[18] M. Mangia, F. Pareschi, R. Rovatti, and G. Setti, "Countering the false myth of democracy: Boosting compressed sensing performance with maximum-energy approach," in *2017 IEEE International Symposium on Circuits and Systems (ISCAS)*, May 2017.

[19] M. Elad, "Optimized projections for compressed sensing," *Signal Processing, IEEE Transactions on*, vol. 55, no. 12, pp. 5695–5702, 2007.

[20] J. M. Duarte-Carvajalino and G. Sapiro, "Learning to sense sparse signals: Simultaneous sensing matrix and sparsifying dictionary optimization," *Image Processing, IEEE Transactions on*, vol. 18, no. 7, pp. 1395–1408, 2009.

[21] J. Xu, Y. Pi, and Z. Cao, "Optimized projection matrix for compressive sensing," *EURASIP Journal on Advances in Signal Processing*, vol. 2010, no. 1, p. 560349, 2010.

[22] E. J. Candes and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?" *IEEE Transactions on Information Theory*, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.

[23] D. Needell and J. A. Tropp, "Cosamp: Iterative signal recovery from incomplete and inaccurate samples," *Applied and Computational Harmonic Analysis*, vol. 26, no. 3, pp. 301–321, 2009.

[24] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," *Proceedings of the National Academy of Sciences*, vol. 106, no. 45, pp. 18 914–18 919, Nov. 2009.

[25] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constructive Approximation*, vol. 28, no. 3, pp. 253–263, 2008.

[26] N. Cleju, "Optimized projections for compressed sensing via rank-constrained nearest correlation matrix," *Applied and Computational Harmonic Analysis*, vol. 36, no. 3, pp. 495–507, May 2014.

[27] V. Cambareri, R. Rovatti, and G. Setti, "Maximum entropy hadamard sensing of sparse and localized signals," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2014, pp. 2357–2361.

[28] M. Mangia, F. Pareschi, V. Cambareri, R. Rovatti, and G. Setti, "Rakeness-based design of low-complexity compressed sensing," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 64, no. 5, pp. 1201–1213, May 2017.

[29] R. Rovatti, G. Mazzini, and G. Setti, "Enhanced rake receivers for chaos-based ds-cdma," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 44, pp. 818–829, 2001.

[30] G. Setti, R. Rovatti, and G. Mazzini, "Performance of chaos-based asynchronous ds-cdma with different pulse shapes," *IEEE Communications Letters*, vol. 8, pp. 416–418, 2004.

[31] V. Cambareri, M. Mangia, F. Pareschi, R. Rovatti, and G. Setti, "A rakeness-based design flow for analog-to-information conversion by compressive sensing," in *2013 IEEE International Symposium on Circuits and Systems (ISCAS2013)*. IEEE, May 2013, pp. 1360–1363.

[32] G. Jacovitti, A. Neri, and G. Scarano, "Texture synthesis-by-analysis with hard-limited gaussian processes," *Image Processing, IEEE Transactions on*, vol. 7, no. 11, pp. 1615–1621, 1998.

[33] R. Rovatti, G. Mazzini, and G. Setti, "Memory-m antipodal processes: spectral analysis and synthesis," *Circuits and Systems I: Regular Papers, IEEE Transactions on*, vol. 56, no. 1, 2009.

[34] A. Caprara, F. Furini, A. Lodi, M. Mangia, R. Rovatti, and G. Setti, "Generation of antipodal random vectors with prescribed non-stationary 2-nd order statistics," *Signal Processing, IEEE Transactions on*, vol. 62, no. 6, pp. 1603–1612, Mar. 2014.

[35] P. Billingsley, *Probability and Measure*. Wiley, 2008.

[36] G. S. U. Grenander, *Toeplitz Forms and Their Applications*. Chelsea Publishing Company, 1984.

[37] P. E. McSharry, G. D. Clifford, L. Tarassenko, and L. A. Smith, "A dynamical model for generating synthetic electrocardiogram signals," *IEEE Transactions on Biomedical Engineering*, vol. 50, no. 3, pp. 289–294, March 2003.

[38] V. Cambareri, M. Mangia, F. Pareschi, R. Rovatti, and G. Setti, "A case study in low-complexity ecg signal encoding: How compressing is compressed sensing?" *IEEE Signal Processing Letters*, vol. 22, no. 10, pp. 1743–1747, Oct 2015.

Mauro Mangia (S'09-M'13) received the B.Sc. and M.Sc. in Electronic Engineering and the Ph.D. degree in Information Technology from the University of Bologna (Bologna, Italy), respectively in 2005, 2009 and 2013. He is currently a Postdoctoral Researcher in the statistical signal processing group of ARCES - University of Bologna. In 2009 and 2012, he was a visiting Ph.D. student at the Ecole Polytechnique Federale de Lausanne (EPFL). His research interests are in nonlinear systems, compressed sensing, ultra-wideband systems, and systems biology. He was the recipient of the 2013 IEEE CAS Society Guillemin-Cauer Award and best student paper award at ISCAS2011. He is also the Web and Social Media Chair for ISCAS2018

Fabio Pareschi (S'05-M'08) received the Dr. Eng. degree (with honours) in Electronic Engineering from University of Ferrara, Italy, in 2001, and the Ph.D. in Information Technology under the European Doctorate Project (EDITH) from University of Bologna, Italy, in 2007. He is currently an Assistant Professor in the Department of Engineering, University of Ferrara. He is also a faculty member of ARCES - University of Bologna, Italy. He served as Associate Editor for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS - PART II (2010-2013). His research activity focuses on analog and mixed-mode electronic circuit design, statistical signal processing, random number generation and testing, and electromagnetic compatibility. He was recipient of the best paper award at ECCTD 2005 and the best student paper award at EMC Zurich 2005.

Riccardo Rovatti (M'99-SM'02-F'12) received the M.S. degree in Electronic Engineering and the Ph.D. degree in Electronics, Computer Science, and Telecommunications both from the University of Bologna, Italy in 1992 and 1996, respectively. He is now a Full Professor of Electronics at the University of Bologna. He is the author of approximately 300 technical contributions to international conferences and journals, and of two volumes. His research focuses on mathematical and applicative aspects of statistical signal processing and on the application of statistics to nonlinear dynamical systems. He received the 2004 IEEE CAS Society Darlington Award, the 2013 IEEE CAS Society Guillemin-Cauer Award, as well as the best paper award at ECCTD 2005, and the best student paper award at EMC Zurich 2005 and ISCAS 2011. He was elected IEEE Fellow in 2012 for contributions to nonlinear and statistical signal processing applied to electronic systems.

Gianluca Setti (S'89-M'91-SM'02-F'06) received a Ph.D. degree in Electronic Engineering and Computer Science from the University of Bologna in 1997. Since 1997 he has been with the School of Engineering at the University of Ferrara, Italy, where he is currently a Professor of Circuit Theory and Analog Electronics and is also a permanent faculty member of ARCES, University of Bologna. His research interests include nonlinear circuits, implementation and application of chaotic circuits and systems, electromagnetic compatibility, statistical signal processing and biomedical circuits and systems. Dr. Setti received the 2013 IEEE CAS Society Meritorious Service Award and co-recipient of the 2004 IEEE CAS Society Darlington Award, of the 2013 IEEE CAS Society Guillemin-Cauer Award, as well as of the best paper award at ECCTD2005, and the best student paper award at EMCZurich2005 and at ISCAS2011. He held several editorial positions and served, in particular, as the Editor-in-Chief for the IEEE Transactions on Circuits and Systems - Part II (2006-2007) and of the IEEE Transactions on Circuits and Systems - Part I (2008-2009). Dr. Setti was the Technical Program Co-Chair ISCAS2007, ISCAS2008, ICECS2012, BioCAS2013 as well as the General Co-Chair of NOLTA2006 and ISCAS2018. He was Distinguished Lecturer of the IEEE CAS Society (2004-2005 and 2014-2015), a member of its Board of Governors (2005-2008), and he served as the 2010 President of CASS. He held several other volunteer positions for the IEEE and in 2013-2014 he was the first non North-American Vice President of the IEEE for Publication Services and Products.