Compressed Sensing: "When sparsity meets sampling"*

Laurent Jacques [†] and Pierre Vandergheynst [‡]

January 6, 2010

The recent theory of Compressed Sensing (Candès, Tao & Romberg, 2006, and Donoho, 2006) states that a signal, e.g. a sound record or an astronomical image, can be sampled at a rate much smaller than what is commonly prescribed by Shannon-Nyquist. The sampling of a signal can indeed be performed as a function of its "intrinsic dimension" rather than according to its cutoff frequency.

This chapter sketches the main theoretical concepts surrounding this revolution in sampling theory. We emphasize also its deep affiliation with the concept of "sparsity", now ubiquitous in modern signal processing. The end of this chapter explains what interesting effects this theory may have on some Compressive Imaging applications.

Contents

1 Introduction **2**

2 In Praise of Sparsity **4**

3 Sensing and Compressing in a Single Stage **6**
3.1 Limits of the Shannon-Nyquist Sampling 6
3.2 New Sensing Model 7

4 Reconstructing from Compressed Information: a Bet on Sparsity **9**

*This is the preprint of a chapter written for the book "Optical and Digital Image Processing - Fundamentals and Applications", Edited by G. Cristbal; P. Schelkens; H. Thienpont. 2010. ©Wiley-Blackwell.

[†] Communications and Remote Sensing Laboratory (TELE), Université catholique de Louvain (UCL), B-1348 Louvain-la-Neuve, Belgium. LJ is funded by the Belgian National Science Foundation (F.R.S.-FNRS).

[‡] Signal Processing Laboratory (LTS2), Swiss Federal Institute of Technology (EPFL), CH-1015 Lausanne, Switzerland.

5 Sensing Strategies Market **12**

6 Reconstruction Relatives **16**

7 Some Compressive Imaging Applications **19**
7.1 Compressive Imagers 19
7.2 Compressive Radio-Interferometry 21

8 Conclusion and the "Science 2.0" Effect **24**

1 Introduction

The 20$^{\text{th}}$ century has seen the development of a huge variety of sensors acquiring measurement in a faithful representation of the physical world (e.g. optical sensors, radio receivers, seismic detector, ...). Since the purpose of these systems was to directly acquire a meaningful "signal", a very fine sampling of this latter had to be performed. This was the context surrounding the famous Shannon-Nyquist condition stating that every continuous (*a priori*) band-limited signal can be recovered from its discretization if its sampling rate is at least two times bigger than its cutoff frequency.

But a recent theory named Compressed Sensing (or Compressive Sampling) [26, 9] states that this lower bound on the sampling rate can be highly reduced, as soon as, first, the sampling is generalized to any linear measurement of the signal, and second, specific a priori hypotheses on the signal are realized. More precisely, the sensing pace is reduced to a rate equals to a few multiple of the intrinsic signal dimension rather than to the dimension of the embedding space.

Technically, this simple statement is a real revolution both in the theory of reliable signal sampling and in the physical design of sensors. It means that a given signal does not have to be acquired in its initial space as previously, but it can really been observed through a "distorting glass" (providing it is linear) with fewer measurements. The couple *encoder* (sensing) and *decoder* (reconstruction) are also completely asymmetric: the encoder is computationally light and linear, and so completely independent of the acquired signal (non-adaptive), while the decoder is non-linear and requires high CPU power.

The (short) history of Compressed Sensing has started in 2006 by the seminal works of D. Donoho, E. Candès, T. Tao and J. Romberg [26, 9, 11], even if some of its founding concepts, e.g. sparse recovery by convex optimization, were known from several decades. CS has actually emerged and grown from the rich multidisciplinary hotbed of Information and Sampling Theory, Statistics and Measure Concentration, Inverse Problems solving, High-Dimensional (Polytope) Geometry and Graph theory.

In this Chapter, we will emphasize how the Compressed Sensing theory may be interpreted as an evolution of the Shannon-Nyquist sampling theorem. We will explain that what characterizes this new theory is the generalization of the *a priori* made on the signal. In other words, we will affiliate the CS theory to the important concept of "*sparsity*" expressing the signal as the linear combination of few elements taken in a particular basis (orthogonal or redundant).

Since the growing CS community is very active, this chapter cannot hope to be a comprehensive presentation of the field. However, we aim at providing a global overview of the topic from its mathematical foundations to its practical implementations reviewing the most successful algorithms. For educational purposes, we have however selected the "axiomatic" theory of Compressed Sensing, i.e. the one that relies on the so-called Restricted Isometry Property of the sensing matrix. Other efficient approaches exist like those describing the signal reconstruction as a stochastic process [7], or those using geometry thanks to Graph theory [3] or polytope projections [25], or finally those exploiting the properties of the null space of the sensing matrix [74]. Since this chapter is dedicated to Compressed Sensing developments, we will not speak either of the parallel (continuous) theory aiming at sampling signals of "Finite Rate of Innovation", as explained in the work of M. Vetterli, T. Blu, P.L. Dragotti and collaborators [71, 30].

Our presentation of the CS theory will come with a "dense" collection of useful tutorials, bibliographic references and internet links, since, as explained later, the Compressed Sensing theory is truly a "Science 2.0" by-product.

Finally, we will show how CS breaks the conventional way to tackle the problem of sensing and compression in some imaging applications.

Conventions: In this Chapter, we will use extensively the following mathematical notations. A discrete *signal* $\boldsymbol{x}$ denotes a N-dimensional vector $\boldsymbol{x} \in \mathbb{R}^N$ of components x_j for $1 \leq j \leq N$. The support of $\boldsymbol{x}$ is $\operatorname{supp} \boldsymbol{x} = \{1 \leq j \leq N : x_j \neq 0\}$, i.e. a subset of the index set $\{1, \cdots, N\}$.

When $\mathbb{R}^N$ is seen as a Hilbert space, the scalar product between two vectors in $\mathbb{R}^N$ is denoted as $\langle \boldsymbol{u}, \boldsymbol{v} \rangle = \boldsymbol{u}^* \boldsymbol{v} = \sum_{j=1}^N u_j v_j$, where $\langle \boldsymbol{u}, \boldsymbol{u} \rangle = \|\boldsymbol{u}\|_2^2$ is the square of the Euclidean ℓ_2-norm of $\boldsymbol{u}$. The ℓ_p-norm of $\boldsymbol{x}$ for $p > 0$ is $\|\boldsymbol{x}\|_p = (\sum_{j=1}^N |x_j|^p)^{1/p}$, and by extension, the ℓ_0 "quasi" norm of $\boldsymbol{x}$ is $\|\boldsymbol{x}\|_0 = \# \operatorname{supp} \boldsymbol{x}$, i.e. the number of non-zero elements in $\boldsymbol{x}$.

If $1 \leq K \leq N$, $\boldsymbol{x}_K \in \mathbb{R}^N$ is the best K-term approximation of $\boldsymbol{x}$ in a given basis (see Sections 2 and 3). If T is a subset of $\{1, \cdots, N\}$ of size $\#T$, according to the context, $\boldsymbol{x}_T$ is either the restriction of $\boldsymbol{x}$ to T, or a thresholded copy of $\boldsymbol{x}$ to the indices in T, i.e. $(\boldsymbol{x}_T)_j = x_j$ for $j \in T$ and $(\boldsymbol{x}_T)_j = 0$ elsewhere.

The identity matrix, or equivalently the canonical basis in $\mathbb{R}^N$, is written $\boldsymbol{I} \in \mathbb{R}^{N \times N}$. Given a matrix $\boldsymbol{A} \in \mathbb{R}^{M \times N}$, $\boldsymbol{A}^*$ is its transposition (or adjoint), and if $M < N$, $\boldsymbol{A}^\dagger$ is the pseudoinverse of $\boldsymbol{A}^*$, i.e., $\boldsymbol{A}^\dagger = (\boldsymbol{A}\boldsymbol{A}^*)^{-1}\boldsymbol{A}$ with $\boldsymbol{A}^\dagger \boldsymbol{A}^* = \boldsymbol{I}$. The Fourier basis is denoted by $\boldsymbol{F} \in \mathbb{R}^{N \times N}$.

In this chapter, we define also many convex optimization techniques. In that context, for a *convex* function $f : \mathbb{R}^N \to \mathbb{R}$, $\arg\min_{\boldsymbol{x}} f(\boldsymbol{x})$ returns the $\boldsymbol{x}$ that minimizes f. For constrained minimization, "*s.t.*" is a shorthand for "*subject to*", e.g. $\arg\min_{\boldsymbol{x}} f(\boldsymbol{x}) \,\text{s.t.}\, \|\boldsymbol{x}\|_2 \leq 1$. The typical big-O notation $A = O(B)$ means that there exists a constant $c > 0$ such that $A \geq cB$.

2 In Praise of Sparsity

The concept of sparse representations is one of the central methodologies of modern signal processing and it has had tremendous impact on numerous application fields. Despite its power, that idea is genuinely simply and intuitive. Given a N-dimensional signal $\boldsymbol{x}$, it is often easy to express it by means a linear superposition of $K \ll N$ elementary signals, called atoms:

$$\boldsymbol{x} = \sum_{k=1}^{K} \alpha_k \boldsymbol{\psi}_k \,. \tag{1}$$

The equality in (1) may not need to be reached, in which case a K-term approximant $\tilde{\boldsymbol{x}}_K$ is found:

$$\tilde{\boldsymbol{x}}_K = \sum_{k=1}^{K} \alpha_k \boldsymbol{\psi}_k, \quad \text{with} \quad \|\boldsymbol{x} - \tilde{\boldsymbol{x}}_K\|_2 \leq \epsilon(K)\,, \tag{2}$$

for some approximation error ϵ. Such an approximant is sometimes called (ϵ, K)-sparse. These K atoms $\boldsymbol{\psi}_k$ are chosen from a large collection called a *dictionary*, which can be conveniently represented by a large $N \times D$ matrix $\boldsymbol{\Psi}$, with $D \geq N$, where each column is an atom. Strictly speaking, there is no restriction on the dictionary but usually the atoms are chosen normalized $\|\boldsymbol{\psi}_k\|_2 = 1$. With these conventions, Eq. (1) can be written $\boldsymbol{x} = \boldsymbol{\Psi}\boldsymbol{\alpha}$, where $\boldsymbol{\alpha} \in \mathbb{R}^D$. Note that an exact sparse representation of the form (1) may be a very strong requirement. In many cases, this assumption is replaced by a weaker notion of sparsity called *compressibility*. A vector $\boldsymbol{\alpha}$ is termed compressible if its entries sorted in decreasing order of magnitude decay like a power law: $|\boldsymbol{\alpha}_k| \leq c\,k^{-c'}$ for some constants c, $c' > 0$. Alternatively, this compressibility may be characterized by the decay of the ℓ_2-approximation error $e(K)$ obtained by the best K-term approximation $\boldsymbol{x}_K = \boldsymbol{\Psi}\boldsymbol{\alpha}_K$ with $\|\boldsymbol{\alpha}_K\|_0 = K$. This error is such that

$$e(K) = \|\boldsymbol{x} - \boldsymbol{x}_K\|_2 \leq \|\boldsymbol{x} - \boldsymbol{x}'\|_2,$$

for any other $\boldsymbol{x}' = \boldsymbol{\Psi}\boldsymbol{\alpha}'$ such that $\|\boldsymbol{\alpha}'\|_0 \leq K$.

Sparse representations have been traditionally used in signal processing as a way to compress data by trying to minimize the number of atoms K in the representation. However, sparsity has recently appeared as a defining property of signals and sparse signal models are by now very common as we shall see. The success of these sparse models started out with wavelet non-linear approximations. Indeed, many interesting signal models are sparse models involving wavelet series. For example, piecewise smooth signals or images yield wavelet decomposition coefficients that are compressible in the sense defined above: most of the information is concentrated in few big coefficients that characterize the discontinuities in the signal, or edges in the image. The main intuitive idea behind wavelet de-noising for example is to realize that while the signal is represented by sparse wavelet coefficients, noise will induce a lot of small coefficients.

They can be removed by enforcing sparsity via thresholding. There are many other signal models involving sparsity: for example locally oscillatory signals are sparse on the MDCT[1] basis and are widely used in audio for tonal components or to model textures in images. This example also shows that complex signals cannot be well modeled by a single basis: an image contains edges, but it often contains textures as well and the latter are not represented in a sparse way by wavelets. Generating sparsity often requires the use of a collection of bases, or a dictionary.

The ultimate goal of sparse representation techniques would be to find the best, that is the sparsest, possible representation of a signal, in other words to solve the following problem:

$$\arg\min_{\boldsymbol{u}} \|\boldsymbol{u}\|_0 \quad \text{s.t.} \quad \boldsymbol{x} = \boldsymbol{\Psi}\boldsymbol{u}. \tag{\textbf{Exact Sparse}}$$

If the dictionary is well adapted to the signal, there are high hopes that a very sparse representation or approximation may exist. When $\boldsymbol{\Psi}$ is an orthonormal basis, there is a unique solution to that problem: the coefficients are computed by projections on the basis $\boldsymbol{\alpha} = \boldsymbol{\Psi}^*\boldsymbol{x}$. Unfortunately, the problem of finding a sparse expansion of a signal in a generic dictionary leads to a daunting NP-complete combinatorial optimization problem [55]. In [17], Chen, Donoho and Saunders proposed to solve the following slightly different problem coined Basis Pursuit (BP):

$$\arg\min_{\boldsymbol{u}} \|\boldsymbol{u}\|_1 \quad \text{s.t.} \quad \boldsymbol{x} = \boldsymbol{\Psi}\boldsymbol{u}. \tag{\textbf{BP}}$$

Minimizing the ℓ_1-norm helps finding a sparse approximation, because it prevents diffusing the energy of the signal over a lot of coefficients. While keeping the essential property of the original problem, this subtle modification leads to a tremendous change in the very nature of the optimization challenge. Indeed, this ℓ_1 problem, called *Basis Pursuit* or BP, is a much simpler convex problem, that can be efficiently solved by various classical optimization techniques. Note that the same ideas can be applied when an approximation of the signal is more suited, i.e solving the following convex quadratic problem known as Basis Pursuit Denoising (BPDN):

$$\arg\min_{\boldsymbol{u}} \|\boldsymbol{u}\|_1 \quad \text{s.t.} \quad \|\boldsymbol{x} - \boldsymbol{\Psi}u\|_2 \leq \epsilon. \tag{\textbf{BPDN}}$$

To understand how solving this problem promotes sparsity, let us consider its augmented lagrangian form:

$$\boldsymbol{\alpha}^* \;=\; \arg\min_{\boldsymbol{u}} \; \|\boldsymbol{x} - \boldsymbol{\Psi}\boldsymbol{u}\|_2^2 + \lambda\|\boldsymbol{u}\|_1,$$

where the first term is squared to be differentiable.

If the dictionary is an orthonormal basis, we can use the Parseval equality to write everything in terms of coefficients:

$$\boldsymbol{\alpha}^* \;=\; \arg\min_{\boldsymbol{u}} \; \|\boldsymbol{\beta} - \boldsymbol{u}\|_2^2 + \lambda\|\boldsymbol{u}\|_1.$$

[1] Modified Discrete Cosine Transform.

where $\boldsymbol{\beta} = \Psi^* \boldsymbol{x}$. It is easy to see that this problem decouples into independent problems for each coefficient:

$$\alpha_i^* \;=\; \arg\min_{u_i \in \mathbb{R}} \,(\beta_i - u_i)^2 + \lambda |u_i|,$$

and the solution to this problem is given by soft-thresholding the projection coefficients $\boldsymbol{\beta}$ of the original signal, which shows that minimizing the ℓ_1-norm enforces sparsity. When the dictionary is not an orthonormal basis, a general solution has been provided by Daubechies, Defrise and Demol in [22].

Note that the link between the Exact Sparse and the BP problems is quite strong and has been well studied. First, let us introduce the following useful characterization of the dictionary. The coherence of $\boldsymbol{\Psi}$ is defined as:

$$\mu(\boldsymbol{\Psi}) \;=\; \sup_{i \neq j} |\langle \boldsymbol{\psi}_i, \boldsymbol{\psi}_j \rangle| \,. \tag{3}$$

Intuitively, Eq. (3) shows that $\boldsymbol{\Psi}$ is not too far from being an orthonormal basis when its coherence is sufficiently small (although it may be highly overcomplete). Building on early results of Donoho and Huo [28], Elad and Bruckstein [33] and later Gribonval and Nielsen [37] have shown that if a signal has a sufficiently sparse representation, i.e

$$\|\boldsymbol{\alpha}\|_0 \;<\; \tfrac{1}{2}\big(1 + \mu(\boldsymbol{\Psi})^{-1}\big)$$

then this representation is the unique solution solution of both the Exact Sparse and Basis Pursuit problems.

In the next sections we will show how sparse signal models are central to the idea of Compressive Sensing and how this induced new ways to "compressively" record images.

3 Sensing and Compressing in a Single Stage

3.1 Limits of the Shannon-Nyquist Sampling

Whatever the field of application, most of the acquisition systems built during the last 50 years have been designed under the guiding rules of the Nyquist-Shannon sampling theorem [64, 69]. These devices implicitly relied on collecting discrete samples from the continuous reality of the signal domain, e.g. in the time or spatial domain for sounds or images respectively, either from the knowledge of this function on specific locations or, by averaging it on very localized domains (for instance, CCD[2] cameras integrate light over each pixel area).

By the Shannon-Nyquist sampling theorem, assuming that a signal is bandlimited, i.e. that it does not contain frequencies higher than a certain limit ν, it is indeed possible to faithfully sample the signal at a period $\Delta T = 1/2\nu$ so that there exists a perfect interpolation procedure rebuilding the continuous

[2] Charged-Coupled Device.

signal. In short, no information has been lost during the sampling process since the initial continuous signal can be recovered.

As explained in Section 2, the concept of *sparsity* in the representation of signals in certain bases or dictionaries has provided a way to compress the acquired information. However, there is one problem: **the process aiming at sampling and then representing the signal with few coefficients in a given basis from the recorded signal samples is wasteful.**

This may be observed on the following idealized example. Let $x(t)$ be a 1-D signal with a cutoff frequency $\nu_s > 0$. Within a certain interval of time $[0, T)$, let us say that we collect $N > 2T\nu_s$ samples $x_n = x(n\Delta T)$ every $\Delta T = T/N$ seconds.

If $x(t)$ is piecewise continuous, an orthonormal wavelet basis $\mathbf{\Psi} = \{\boldsymbol{\psi}_j \in \mathbb{R}^N, 1 \leq j \leq N\}$ represents the vector $(x_1, \cdots, x_N)^T \in \mathbb{R}^N$ with few non-zero elements $\alpha_j = \langle \boldsymbol{\psi}_j, \boldsymbol{x} \rangle = \sum_n \psi_{jn} x_n$ [51]. There exists therefore a $K \ll N$ such that the K strongest coefficients in $\boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_N)^T$ of indices $\{j_k : 1 \leq k \leq K\}$ suffice to provide a good approximation $\boldsymbol{x}_K = \mathbf{\Psi}\boldsymbol{\alpha}_K = \sum_{k=1}^{K} \boldsymbol{\psi}_{j_k} \alpha_{j_k}$ of $\boldsymbol{x}$, i.e. $\|\boldsymbol{x} - \boldsymbol{x}_K\|_2$ is small compared to $\|\boldsymbol{x}\|_2$. We can establish that the global process that led from $\boldsymbol{x}$ to the K significant values α_{j_k} demands, first, to discretize the signal over N samples, and second, to compute the coefficients α_j and to sort them by decreasing absolute amplitude. The computation of the α_j's is realized in $O(N^2)$ operations, or, at best, in $O(N)$ operations if the sparsity basis $\mathbf{\Psi}$ is provided with a fast decomposition/reconstruction algorithm. However, at the end of the day, only $K \ll N$ coefficients α_j are recorded and the others thrown away!

Is it possible to simplify this procedure? Can we avoid wasting time and computations to record N samples and process them, for finally keeping only $K \ll N$? This is what Compressed Sensing, also called Compressive Sampling, is all about.

As we will show in this section, CS answers positively to the merging of sampling and compression thanks to three main changes:

- the switch in the a priori *sparsity* knowledge: the signal may be assumed sparse in any kind of sparsity basis $\mathbf{\Psi}$ (not only in the Fourier domain as for bandlimited signal).
- the generalization of the sampling procedure to any linear (and thus non-adaptive) measurement of the signal, i.e. represented by a correlation of the signal with a *sensing basis*;
- the use of non-linear reconstruction techniques, e.g. relying on convex optimization or on greedy methods, to recover the initial signal by using the signal sparsity a priori.

3.2 New Sensing Model

Our signal of interest from now on is finite, i.e. we work with a vector $\boldsymbol{x}$ in the N-dimensional space $\mathbb{R}^N$. This vectorial representation can be adapted to any space, e.g. by concatenating all the columns of a $\sqrt{N} \times \sqrt{N}$ image into a vector of N components[3].

As for the example given in the end of Section 3.1, we assume that $\boldsymbol{x}$ has a certain structure, or a geometrical content. In other words, there exists a *sparsity basis* $\boldsymbol{\Psi} \in \mathbb{R}^{N\times D}$, made of $D \geq N$ elements $\boldsymbol{\Psi}_j \in \mathbb{R}^N$ with $1 \leq j \leq D$, such that $\boldsymbol{x}$ can be represented as

$$\boldsymbol{x} \;=\; \boldsymbol{\Psi}\boldsymbol{\alpha} \;=\; \sum_{j=1}^{D} \boldsymbol{\Psi}_j\, \alpha_j, \tag{4}$$

with few non-zero or important coefficients α_j in $\boldsymbol{\alpha} \in \mathbb{R}^D$. This kind of signal transformations are implicitly used everyday when you listen MP3 compressed songs or JPEG2000 coded images by using sparsity basis like the Discrete Cosine Transform (DCT) or the 2-D Discrete Wavelet Transform [51, 23].

In this framework, a signal is said K-*sparse* if $\|\boldsymbol{\alpha}\|_0 = K$, and *compressible* if the coefficients of $\boldsymbol{\alpha}$ decay rapidly when sorted by decreasing order of magnitude.

For simplicity, we will always assume that the sparsity basis is orthonormal, i.e. $D = N$ and $\boldsymbol{\Psi}^* = \boldsymbol{\Psi}^{-1}$, leading to the relation $\alpha_j = \langle \boldsymbol{\Psi}_j, \boldsymbol{x}\rangle$. The theory is however validated for dictionaries and frames as explained here [59].

In Compressed Sensing theory, following a process natural in quantum physics, we get knowledge of $\boldsymbol{x}$ by "asking" a certain number of independent questions, or *linear measurements*. For M measurements, the signal is thus "sampled" in M values $y_j = \langle \boldsymbol{\varphi}_j, \boldsymbol{x}\rangle$ $(1 \leq j \leq M)$, where the vectors $\boldsymbol{\varphi}_j \in \mathbb{R}^N$ form the *sensing matrix* $\boldsymbol{\Phi} = \big(\boldsymbol{\varphi}_1, \cdots, \boldsymbol{\varphi}_M\big)^* \in \mathbb{R}^{M\times N}$. Using matrix algebra, the sensing model is thus

$$\boldsymbol{y} \;=\; \boldsymbol{\Phi}\boldsymbol{x} \;=\; \boldsymbol{\Phi}\boldsymbol{\Psi}\,\boldsymbol{\alpha}, \tag{5}$$

where the last equality follows from (4). Notice that the sparsity and the sensing basis can be merged into the global sensing basis $\boldsymbol{\Theta} = \boldsymbol{\Phi}\boldsymbol{\Psi}$.

Equation (5) really models the sensing of $\boldsymbol{x}$ as if it was obtained by a physical sensor outputting $\boldsymbol{y}$. In that model, we have no-access to the components of $\boldsymbol{x}$ or those of $\boldsymbol{\alpha}$. What we have is only $\boldsymbol{y}$ and the knowledge of the rectangular sensing matrix $\boldsymbol{\Phi}$.

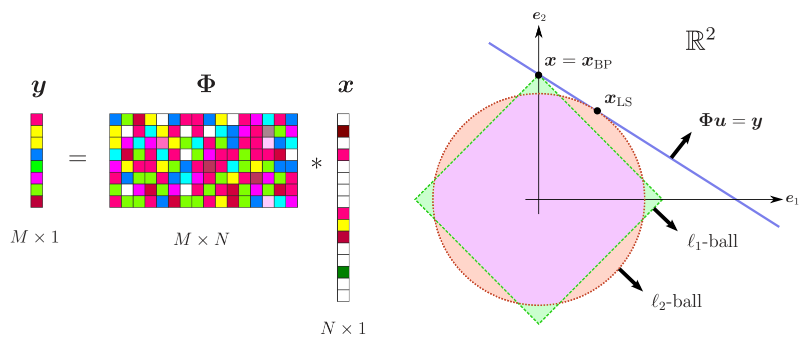
Following this framework, it is often more appropriate to consider a more realistic sensing scenario where $\boldsymbol{y}$ is plagued by different noises:

$$\boldsymbol{y} \;=\; \boldsymbol{\Theta}\,\boldsymbol{\alpha} + \boldsymbol{n}. \tag{6}$$

This equation integrates an additional noise $\boldsymbol{n} \in \mathbb{R}^M$ representing for instance the digitization (quantification) of $\boldsymbol{\Phi}\boldsymbol{x}$, for further storage and/or transmission, and the unavoidable instrumental noises (Poisson, thermal, ...). Most of the time this noise is assumed identically and independently distributed over its components with a Gaussian distribution.

[3] Implicitly, this "vectorization" process must be realized correspondingly on the basis elements that serve to sparsify the signal.

4 Reconstructing from Compressed Information: a Bet on Sparsity



(a) (b)

Figure 1: (a) Visual interpretation of the "compressed" sensing of a signal $\boldsymbol{x}$ sparse in the canonical basis (i.e. $\boldsymbol{\Psi} = \boldsymbol{I}$). (b) Explanation of the recovery of 1-sparse signals in $\mathbb{R}^2$ with BP compared to a least square (LS) solution.

Let us now study how to reconstruct the signal from its measurements. The fundamental theorem of algebra – "as many equations as unknowns" – teaches us that the recovery of $\boldsymbol{\alpha}$ or $\boldsymbol{x} = \boldsymbol{\Psi\alpha}$ from $\boldsymbol{y}$ (and from the knowledge of the sensing model) is possible if $M \geq N$. This is true in all generality, whatever the properties of $\boldsymbol{x}$. Can we reduce M if $\boldsymbol{x}$ has a priori a certain structure? After all, if $\boldsymbol{x} = \boldsymbol{\Psi\alpha}$ was exactly K-sparse, with $K \ll N$, and if the support $S \subset \{1, \cdots, N\}$ of $\boldsymbol{\alpha}$ was known, we would have

$$\boldsymbol{y} = \boldsymbol{\Phi\Psi\alpha} = \boldsymbol{\Theta}_S \boldsymbol{\alpha}_S,$$

where $\boldsymbol{\Theta}_S \in \mathbb{R}^{M \times K}$ is the restrictions of the columns of $\boldsymbol{\Theta}$ to those of index in S, and $\boldsymbol{\alpha}_S$ is the vector $\boldsymbol{\alpha}$ restricted to its support. Therefore, if $K \leq M$, the recovery problem stops being ill-posed.

An evident gap exists therefore between the number of measurements required for solving the known-support problem ($M \geq K$) and the one needed for the general case ($M = N$). This missing link is found by regularizing the problem, i.e. by adding a prior information on the sensed $\boldsymbol{x}$, somewhere between the full knowledge of $\operatorname{supp} \boldsymbol{x}$ and the general "no prior information" case. In short, we must now assume the *sparsity* of $\boldsymbol{x}$ to recover it.

The signal reconstruction problem has been recasted as the recovery of the sparse vector $\boldsymbol{\alpha}$ from the observed (compressed) signal $\boldsymbol{y} = \boldsymbol{\Phi x} = \boldsymbol{\Theta\alpha}$ given the sparsity basis $\boldsymbol{\Theta} = \boldsymbol{\Phi\Psi}$. According to Section 2, we may therefore use the ideal non-linear recovery technique (or *decoder*)

$$\Delta_0(\boldsymbol{y}) \;\triangleq\; \arg\min_{\boldsymbol{u}} \|\boldsymbol{u}\|_0 \;\text{ s.t. }\; \boldsymbol{y} = \boldsymbol{\Theta u}. \tag{7}$$

In spite of its combinatorial complexity, we can study theoretically the properties of this decoder Δ_0 and observe later how we can simplify it. For that purpose, we need to introduce a certain regularity on the matrix $\mathbf{\Theta}$. After all, if the support of $\boldsymbol{\alpha}$ was known and equals to S, we should impose that $\mathbf{\Theta}_S$ is rank K, i.e. the K columns of $\mathbf{\Theta}_S$ are linearly independent, so that $\boldsymbol{\alpha} = \mathbf{\Theta}_S^\dagger \boldsymbol{y} = (\mathbf{\Theta}_S^* \mathbf{\Theta}_S)^{-1} \mathbf{\Theta}_S^* \boldsymbol{y}$, with $\mathbf{\Theta}_S^\dagger$ the Moore-Penrose pseudo-inverse of $\mathbf{\Theta}_S$. The property we need now is a generalization of this concept for all the possible support S of a given size.

Definition 1 ([13]) *A matrix $\mathbf{\Theta} \in \mathbb{R}^{M\times N}$ satisfies the Restricted Isometry Property (or RIP(K,δ)) of order $K < M$ and isometry constant $0 \leq \delta < 1$ if, for all K-sparse signal $\boldsymbol{u} \in \mathcal{S}_K = \{\boldsymbol{v} \in \mathbb{R}^N : \|\boldsymbol{v}\|_0 = K\}$,*

$$(1-\delta)\|\boldsymbol{u}\|_2^2 \;\leq\; \|\mathbf{\Theta}\boldsymbol{u}\|_2^2 \;\leq\; (1+\delta)\|\boldsymbol{u}\|_2^2. \tag{8}$$

This definition clearly amounts to impose $\mathbf{\Theta}$ of rank K over all the possible support $S \subset \{1, \cdots, N\}$ of size K. We will see later what kind of matrices $\mathbf{\Theta}$ respect the RIP.

The RIP implies the following key result: if $\boldsymbol{y} = \mathbf{\Theta}\boldsymbol{\alpha}$ and $\|\boldsymbol{\alpha}\|_0 \leq K$, then

$$\boxed{\mathbf{\Theta} \text{ is } \mathrm{RIP}(2K,\delta) \qquad \Rightarrow \qquad \boldsymbol{\alpha} = \Delta_0(\boldsymbol{y}).}$$

The proof of this result is simple and enlightening. Denoting $\boldsymbol{\alpha}^* = \Delta_0(\boldsymbol{y})$, we must show that $\boldsymbol{x} = \mathbf{\Psi}\boldsymbol{\alpha} = \mathbf{\Psi}\boldsymbol{\alpha}^\star$, i.e. $\boldsymbol{\alpha} = \boldsymbol{\alpha}^\star$. Since $\boldsymbol{\alpha}^\star$ is the solution of the minimization problem (7), $\|\boldsymbol{\alpha}^\star\|_0 \leq \|\boldsymbol{\alpha}\|_0$ and $\|\boldsymbol{\alpha} - \boldsymbol{\alpha}^\star\| \leq 2K$. Since $\mathbf{\Theta}$ is RIP$(2K,\delta)$, $(1-\delta)\,\|\boldsymbol{\alpha} - \boldsymbol{\alpha}^\star\|_2^2 \leq \|\mathbf{\Theta}\boldsymbol{\alpha} - \mathbf{\Theta}\boldsymbol{\alpha}^\star\|_2^2 = \|\boldsymbol{y} - \boldsymbol{y}\|_2^2 = 0$, proving that $\boldsymbol{\alpha} = \boldsymbol{\alpha}^\star$ since $1 - \delta > 0$.

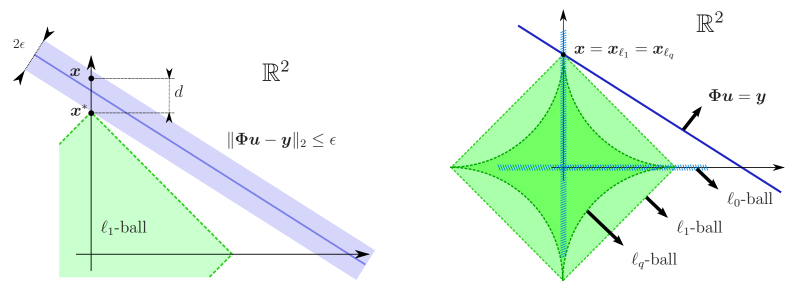
We may notice that the ideal decoder (7) is not guaranteed to provide a meaningful reconstruction in the cases where the signal $\boldsymbol{x}$ deviates from the exact sparsity, i.e. if it is just compressible, or when measurements are corrupted by a noise. These two problems are solved in the relaxed decoders proposed in the next section.

From the first papers about Compressed Sensing [12, 26], inspired by similar problems in the quest for sparse representation of signals in orthonormal or redundant basis, researcher have used a “relaxation” of (7). As explained in Section 2, the convex ℓ_1-norm can replace favorably its non-convex ℓ_0 counterpart, and in the CS formalism, the two relaxed optimizations BP and BPDN can thus be rephrased as:

$$\Delta(\boldsymbol{y}) \;\triangleq\; \arg\min_{\boldsymbol{u}} \|\boldsymbol{u}\|_1 \;\text{ s.t. }\; \boldsymbol{y} = \mathbf{\Theta}\boldsymbol{u}, \tag{\textbf{BP}}$$

$$\Delta(\boldsymbol{y},\epsilon) \;\triangleq\; \arg\min_{\boldsymbol{u}} \|\boldsymbol{u}\|_1 \;\text{ s.t. }\; \|\boldsymbol{y} - \mathbf{\Theta}\boldsymbol{u}\|_2 \leq \epsilon. \tag{\textbf{BPDN}}$$

The Basis Pursuit decoder is suited for sparse or compressible signal reconstruction in case of pure sensing model (5), while the Basis Pursuit DeNoise (BPDN) program adds the capability to handle noisy measurements (6) with a noise power assumed bounded, i.e. $\|\boldsymbol{n}\|_2 \leq \epsilon$.



(a) (b)

Figure 2: (a) Explanation of the robustness of BPDN for 1-sparse signals in $\mathbb{R}^2$. (b) Geometrical illustration of the non-convex reconstruction minimizing the ℓ_q-norm ($0 < q \leq 1$).

In Figure 1(b), we provide a common illustration of why in the pure sensing case the Basis Pursuit is an efficient way to recover sparse signals from their measurements. On this Figure, the signal $\boldsymbol{x}$ is assumed 1-sparse in the canonical basis of $\mathbb{R}^2$, i.e. $\boldsymbol{x}$ lives on one of the two axis $\boldsymbol{e}_1$ or $\boldsymbol{e}_2$ of this space. The constraint of BP is the line $D_y = \{\boldsymbol{u} \in \mathbb{R}^2 : \boldsymbol{\Phi u} = \boldsymbol{y}\}$ intersecting one of the two axis, here $\boldsymbol{e}_2$, in $\boldsymbol{x}$. For a different reconstruction based on a regularization with a ℓ_2-norm, i.e. the Least Square method (LS), the solution $\boldsymbol{x}_{\rm LS} = \arg\min_{\boldsymbol{u}} \|\boldsymbol{u}\|_2$ s.t. $\boldsymbol{\Phi u} = \boldsymbol{y}$ corresponds to the intersection of D_y with the smallest ℓ_2-ball $B_2(r) = \{\boldsymbol{u} \in \mathbb{R}^2 : \|\boldsymbol{u}\|_2 \leq r\}$ intersecting D_y. Clearly, this point does not match the initial signal $\boldsymbol{x}$. However, in non-degenerated situations, i.e. when D_y is not oriented at 45° with $\boldsymbol{e}_1$ in our illustration, the solution $\boldsymbol{x}_{\rm BP}$ of BP, which is provided by the common point between D_y and the smallest ℓ_1-ball $B_1(r) = \{\boldsymbol{u} \in \mathbb{R}^2 : \|u\|_1 \leq r\}$ intersecting $D_y = D_y(0)$, is precisely the original $\boldsymbol{x}$.

In Figure 2(a), the previous illustration is adapted to noisy measurement. The constraint of BPDN is now a tube $D_{\boldsymbol{y}}(\epsilon) = \{\boldsymbol{u} \in \mathbb{R}^2 : \|\boldsymbol{\Phi u} - \boldsymbol{y}\|_2 \leq \epsilon\}$ of thickness 2ϵ around the line $D_{\boldsymbol{y}}$. The solution of BPDN, i.e. $\boldsymbol{x}^\star$, is now the common point between $D_{\boldsymbol{y}}(\epsilon)$ and the smallest ℓ_1-ball touching this tube. Geometrically, it is clear that, for most of the configuration, the distance $d = \|\boldsymbol{x} - \boldsymbol{x}^\star\|_2$ between the original signal $\boldsymbol{x}$ and the reconstruction $\boldsymbol{x}^\star$ is proportional to 2ϵ, since $\boldsymbol{x}^\star$ and $\boldsymbol{x}$ are in $D_{\boldsymbol{y}}(\epsilon)$.

These two intuitive explanations, i.e. perfect recovery of sparse signals and the approximate signal reconstruction from noisy measurements, have been theoretically proved in [11, 8] again from the essential RIP. Mathematically, if $\boldsymbol{y} = \boldsymbol{\Theta\alpha} + \boldsymbol{n}$ with $\|\boldsymbol{n}\|_2 \leq \epsilon$ (noisy sensing), then:

$$\boxed{\begin{array}{c}\mathbf{\Theta} \text{ is RIP}(2K,\delta) \quad \text{and} \quad \delta < \sqrt{2}-1 \\ \Rightarrow \\ \|\boldsymbol{\alpha} - \Delta(\boldsymbol{y},\epsilon)\|_2 \;\le\; C\epsilon \;+\; D\tfrac{1}{\sqrt{K}}\|\boldsymbol{\alpha} - \boldsymbol{\alpha}_K\|_1.\end{array}} \tag{9}$$

for C and D function of δ only. For instance, for $\delta = 0.2$, $C < 4.2$ and $D < 8.5$.

In summary, (9) proves the robustness of the Compressed Sensing setup. Perfect recovery in the pure sensing model ($\epsilon = 0$) of a K-sparse signal is still achievable but at a higher price, i.e. we must have $\delta < \sqrt{2} - 1 \simeq 0.41$ in the RIP. However, the result is fairly general since now it allows one to reconstruct the signal from noisy measurements (with a linear dependence in ϵ), and this even if $\boldsymbol{x}$ is not exactly sparse in $\mathbf{\Psi}$. The deviation to the exact sparsity is indeed measured by the compressibility error $e_0(K) = \|\boldsymbol{\alpha} - \boldsymbol{\alpha}_K\|_1/\sqrt{K}$. For compressible signals, K is then a parameter that must be tuned to the desired accuracy of the decoder. Since $\mathbf{\Theta}$ must be RIP, we will see later that increasing K comes with an increase of the number of measurements M.

5 Sensing Strategies Market

In Section 4 we detailed the required property of the sensing matrix $\mathbf{\Theta} = \mathbf{\Phi}\mathbf{\Psi}$ to guarantee a faithful reconstruction of the signal $\boldsymbol{x}$. The obvious question is now: Do such RIP matrices exist? The answer is "yes, with very high probability". Indeed, deterministic construction of RIP matrices exists [24] but they suffer from very high lower bound on M with respect to the RIP(K,δ). We will not detail them here. However, as a striking result of the Concentration of Measure theory [48, 49], stochastic constructions of RIP matrices exist with a precise estimation of their probability of success. Once generated, these sensing matrices have of course to be stored, or at least they must be exactly reproducible, for sensing and reconstruction.

This section aims at guiding the reader through the market of available random constructions, with a clear idea of the different dimensionality conditions ensuring the RIP.

Random sub-Gaussian Matrices: A first case of random construction of sensing matrices is the one provided by the class sub-Gaussian distributions [53]. For instance, the matrix $\mathbf{\Phi} \in \mathbb{R}^{M\times N}$ can be generated as the realization of a normalized Gaussian random variable of variance $1/M$, identically and independently (*iid*) for each entry of the matrix, i.e.

$$\Phi_{ij} \underset{\text{iid}}{\sim} \mathcal{N}(0, 1/m).$$

Another possibility is to select one of the discrete Bernoulli distributions, e.g. $\Phi_{ij} \sim_{\text{iid}} \pm 1/\sqrt{M}$ with probability 1/2, or $\Phi_{ij} \sim_{\text{iid}} \pm\sqrt{3/M}$ or 0 with probability 1/6 or 2/3 respectively. Notice that the randomness of the procedure above helps only at the creation of the matrices. These three cases, and matrices generated by other sub-Gaussian distributions [53], have the two following interesting properties.

First, if $\mathbf{\Phi}$ is sub-Gaussian, then it can be shown that, for any orthonormal sparsity basis $\mathbf{\Psi}$, $\mathbf{\Theta} = \mathbf{\Phi\Psi}$ is again sub-Gaussian. In particular, if the RIP is proved for any sub-Gaussian matrix $\mathbf{\Phi}$, it will hold then for any $\mathbf{\Theta} = \mathbf{\Phi\Psi}$. The previous stability results are thus guaranteed in this sensing strategy.

Second, in [2], it is proved that a sub-Gaussian matrix $\mathbf{\Phi}$ satisfies the RIP(K, δ) with a controlled probability as soon as

$$M \;=\; O(\delta^{-2}\, K \ln N/K).$$

More precise asymptotic estimations of this last relation have been realized in [25].

Even if they correspond to very general sensing strategy independent of the sparsity basis, there are two main problems with the random sub-Gaussian matrices above.

First, their randomness makes them difficult to generate. The stochastic construction above is therefore often replaced by a pseudo-random procedure. In short, this provides a sequence of numbers with the desired distribution of occurrences that seems random but is in fact deterministic, i.e. it is determined by a set of few initial values also called the *seeds*. This has the merit to reduce to storage of the sensing matrix to the knowledge of the pseudo-random generator and its initial seeds.

Second, the numerical reconstruction of $\boldsymbol{\alpha}$ from $\boldsymbol{y} = \mathbf{\Theta}\boldsymbol{\alpha}$ in greedy methods like Orthogonal Matching Pursuit or in ℓ_1-minimization programs as BPDN often involves iterative methods relying on the application of $\mathbf{\Theta}$ and $\mathbf{\Theta}^*$ onto vectors at each iteration. Even with the use of the pseudo-randomness trick above, the absence of structure in $\mathbf{\Phi}$ makes the complexity of these computations of order $O(MN)$. For Compressed Sensing of images or of videos, this makes the reconstruction of reasonable size objects often very slow.

Fortunately, as explained below, other sensing strategies exist, even if they are often less general with respect to the class of RIP-compatible signal sparsity basis $\mathbf{\Psi}$.

Random Fourier Ensemble: The first possibility to obtain a "fast" sensing matrix, i.e. offering both fast encoding and faster decoding techniques, is to use the Fourier transform. The sensing matrix $\mathbf{\Phi}$ is here given by

$$\mathbf{\Phi} \;=\; \boldsymbol{S}\boldsymbol{F},$$

where $\boldsymbol{F} \in \mathbb{R}^{N\times N}$ is the (real) Discrete Fourier Transform on $\mathbb{R}^N$ (or on the 2-D plane $\mathbb{R}^{N_1\times N_2}$ with $N_1 N_2 = N$ for vectorized $N_1 \times N_2$ images) and the rectangular matrix $\boldsymbol{S} \in \mathbb{R}^{M\times N}$ picks randomly M elements of any N-dimensional vector.

The point is that, when $\mathbf{\Phi}$ or $\mathbf{\Phi}^*$ are applied to a vector, the Fast Fourier Transform (FFT) Cooley-Tukey algorithm can be applied instead of the corresponding matrix multiplication. This decreases the complexity from $O(NM)$ to $O(N \ln N)$.

For a canonical sparsity basis $\mathbf{\Psi} = \boldsymbol{I}$, $\mathbf{\Theta} = \mathbf{\Phi}$ is RIP with overwhelming probability as soon as [12, 62]

$$M \;=\; O(K \ln^4 N)$$

About the exponent of the ln factor, some converging experimental results, for instance for CS applied to Magnetic Resonance Imaging (MRI) or in CS for Radio-Interferometric data sensing (see Section 7.2), seem to validate a nicer $M \;=\; O(K \ln N)$ requirement.

Random Fourier ensemble matrices are however less general than the sub-Gaussian random constructions. Indeed, as explained in the next subsection, the proportionality constant implicitly involved in the last relation grows when a non-trivial sparsity basis is used.

This problem can be bypassed by altering the sparsity term of the BPDN reconstruction and replacing it by the TV-norm. This is detailed in the Section 6.

Random Basis Ensemble: It is possible to generalize the previous sensing basis construction to any orthonormal basis. In other words, given an orthonormal basis $\boldsymbol{U} \in \mathbb{R}^{N\times N}$, we construct $\mathbf{\Phi}$ as

$$\mathbf{\Phi} \;=\; \boldsymbol{S}\boldsymbol{U},$$

with the previous random selection matrix $\boldsymbol{S} \in \mathbb{R}^{M\times N}$.

In that case, the sparsity basis $\mathbf{\Psi}$ and the basis $\boldsymbol{U}$ from which $\mathbf{\Phi}$ is extracted must be sufficiently incoherent, i.e. it must be difficult to express one element of $\mathbf{\Psi}$ as a sparse representation in $\boldsymbol{U}$, and conversely.

More precisely, this fact is measured by the concept of *coherence* between two basis: similarly to the definition (3), given an orthonormal sparsity basis $\boldsymbol{V} = \{\boldsymbol{V}_1, \cdots, \boldsymbol{V}_N\} \in \mathbb{R}^{N\times N}$ and another basis $\boldsymbol{U} = \{\boldsymbol{U}_1, \cdots, \boldsymbol{U}_N\} \in \mathbb{R}^{N\times N}$, the *mutual coherence* between $\boldsymbol{U}$ and $\boldsymbol{V}$ is the value

$$\mu(\boldsymbol{U}, \boldsymbol{V}) \;=\; \max_{i,j} |\langle \boldsymbol{U}_i, \boldsymbol{V}_j \rangle|.$$

The coherence corresponds also to the highest amplitude entry of the correlation matrix $\boldsymbol{U}^*\boldsymbol{V}$. By construction, μ is always greater than $1/\sqrt{N}$ for orthonormal basis[4]

The general result is then that $\mathbf{\Theta} = \boldsymbol{S}\boldsymbol{U}\,\mathbf{\Psi}$ is RIP(K, δ) if [12, 62]

$$M \;\geq\; C\; \mu(\boldsymbol{U}, \mathbf{\Psi})^2 N\, K\, (\ln N)^4, \tag{10}$$

for a certain constant $C > 0$.

[4]Indeed, $1 = \|\boldsymbol{U}_i\|_2^2 = \sum_j |\langle \boldsymbol{V}_j, \boldsymbol{U}_i \rangle|^2$ involves $\max_j |\langle \boldsymbol{V}_j, \boldsymbol{U}_i \rangle| \geq 1/\sqrt{N}$.

Interestingly, the combination of the Fourier basis with the canonical sparsity basis is a perfect dual pair, i.e. it saturates the mutual coherence lower bound with $\mu(\boldsymbol{F}, \boldsymbol{I}) = 1/\sqrt{N}$. The requirement above meets then the one of Section 5.

A useful orthonormal basis choice is provided by the Noiselet transform [18]. This basis is maximally incoherent with the wavelet basis [51]. The coherence is for instance given by $\sqrt{2}$ for the Haar system, and it corresponds to 2.2 and 2.9 for the Daubechies wavelets D4 and D8 respectively [14].

Random Convolution: Recently, J. Romberg introduced another fast sensing strategy called *random convolution* [61]. Working again in the (complex) Fourier domain, the idea is simply to disturb the phase of the Fourier coefficients of a signal by multiplying them with a random complex sequence of unit amplitude; M samples of the result are subsequently randomly selected in the spatial domain. Mathematically, this amounts to consider

$$\boldsymbol{\Phi} \;=\; \boldsymbol{S}\,\boldsymbol{F}^*\boldsymbol{\Sigma}\,\boldsymbol{F},$$

with $\boldsymbol{\Sigma} \in \mathbb{R}^{N\times N}$ a complex diagonal matrix made of unit amplitude diagonal element and random phase (their generation respects a certain symmetry to guarantee a real measurement vector) [61].

Interestingly, given $0 < \gamma < 1$, it is proved then that the coherence between the orthonormal basis $\boldsymbol{H} = \boldsymbol{F}^*\boldsymbol{\Sigma}\boldsymbol{F}$ and any sparsity basis $\boldsymbol{\Psi}$ is

$$\mu(\boldsymbol{H}, \boldsymbol{\Psi}) \;\le\; 2\sqrt{\ln(2N^2/\gamma)/N},$$

with probability exceeding $1-\gamma$.

Equation (10) involves then that $\boldsymbol{\Theta} = \boldsymbol{\Phi}\boldsymbol{\Psi}$ is $\mathrm{RIP}(K,\delta)$ with high probability if

$$M \;\ge\; C\;K\,(\ln N)^5,$$

for a certain constant $C > 0$.

Despite a stronger requirement on the number of measurements, and contrarily to the Random Fourier Ensemble, Random Convolution sensing works with any kind of sparsity basis with possibly very fast implementation. In addition, its structure seems very adapted to analog implementations in the optical domain [60], or even for CMOS Imager implementing random convolution on the focal plane [42], as detailed in Section 7.1.

Other Sensing Strategies: Nowadays, more and more different sensing strategies are developed in the field [15, 34].

Under various but similar requirements on the number of measurements, we can mention the sensing matrix of the Analog-to-Information Converter (AIC) designed in [47, 44] to compressively acquire a 1-D signal of infinite length by temporal blocks. It is also possible to design Toeplitz-structured sensing matrices [1], or to play with scrambled block Hadamard ensemble [35, 36], or with the use of a spread-spectrum sequence preceding an Hadamard partial transform adapted to pulse train sensing [54], ... Possible choices are now numerous, and the final selection of a sensing matrix depends on criteria like: the existence of an analog model corresponding to this sensing, the availability of a fast implementation of the numerical sensing at the decoding stage, the storage of the sensing matrix in the sensors or/and in the decoder, the coherence between this sensing and the (sparsity) class of signal to be observed, ...

6 Reconstruction Relatives

Previous sections have explained the main concepts of the Compressed Sensing theory where the reconstruction (or decoding) task is solved through an optimization program relying on the ℓ_1-norm sparsity measure. In this section, we present briefly some variations around the reconstruction task. These either alter the sparsity measure of the BP or the BPDN reconstruction, or complement their constraints by adding signal priors, or replace the often heavy optimization process by less optimal but fast iterative (greedy) algorithms.

Be Sparse in Gradient: As explained in Section 5, sensing based on Random Fourier Ensemble is perhaps fast but it lacks of universality, i.e. it is adapted mainly to signal sparse in the canonical basis $\boldsymbol{\Psi} = \boldsymbol{I}$. In certain applications however, the sensing strategy is not a choice, it can be imposed by the physics of the acquisition. For instance, in Magnetic Resonance Imaging [50] or in Radio-Interferometry [72] (Section 7.2), the signal (an image) is acquired in the k-space, i.e. in the Fourier frequency domain, on a subset of the whole frequency plane.

To face this problem, researcher have introduced a variation around the previous reconstructions algorithms. Rather than expressing the signal sparsity in the spatial domain using the ℓ_1-norm measure, it is indeed possible to impose the sparsity of the gradient of the image, leading to the "Total Variation" (TV) quasi-norm [63].

In its discrete formulation, the TV-norm of a signal $\boldsymbol{x}$ representing an image is given by

$$\|\boldsymbol{x}\|_{\mathrm{TV}} \;=\; \sum_i \big[\, (x_{p+1,q} - x_{p,q})_i^2 + (x_{p,q+1} - x_{p,q})_i^2 \,\big]^{1/2},$$

where (p,q) are the coordinates of the i^{th} pixel of $\boldsymbol{x}$. This norm is small for "cartoon" images composed of smooth areas separated by curved edges (i.e. C^2 smooth), i.e. a good approximation of many natural images showing of distinct objects not too textured. [63]

For a noisy sensing model, the TV norm leads then to the new program

$$\arg\min_{\boldsymbol{u}} \|\boldsymbol{u}\|_{\mathrm{TV}} \;\text{ s.t. }\; \|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{u}\|_2 \le \epsilon. \qquad \textbf{(BPDN-TV)}$$

In [10], it is proved that BPDN-TV recovers with overwhelming probability the initial image for random Fourier ensemble matrices and in the absence of noise. To our knowledge, nothing has been proved for general sensing matrix and for noisy measurements. Experimentally however, BPDN-TV provides always an increasing in reconstruction quality compared to the results obtained with the ℓ_1-norm, even for other sensing strategies, e.g. for random convolutions [61, 42].

Add or Change Priors: In addition to the sparsity prior on the signal, other priors can enforce the stability of a reconstruction program when they are available.

For instance, if the signal is known to be positive in the spatial domain, we may alter BPDN into

$$\arg\min_{\boldsymbol{u}} \|\boldsymbol{u}\|_1 \ \text{ s.t. } \begin{cases} \|\boldsymbol{y} - \boldsymbol{\Theta u}\|_2 \leq \epsilon, \\ \boldsymbol{\Psi u} \geq 0. \end{cases} \tag{$\mathbf{BPDN^+}$}$$

Some stability guarantees have been established in this formalism thanks to particular requirement on the null space of the sensing matrix [74, 43].

Sometimes, it is difficult to estimate the noise power on the measurements but easy to have a bound on the signal sparsity, e.g. there exist a τ such that $\|\boldsymbol{\alpha}\|_1 \leq \tau$. In that case the Lasso formulation [66] can be useful:

$$\arg\min_{\boldsymbol{u}} \|\boldsymbol{y} - \boldsymbol{\Theta u}\|_2 \ \text{ s.t. } \ \|\boldsymbol{u}\|_1 \leq \tau, \tag{$\mathbf{Lasso}$}$$

In [70], it is explained that there is a continuous and differentiable mapping between the parameter τ of Lasso and the noise power ϵ of BPDN, i.e. the Pareto frontier. It is therefore possible to solve the latter from the first by probing this Pareto curve. In addition, for the right $\epsilon(\tau)$, the stability results of BPDN holds therefore for Lasso.

In certain situation, the noise level on the measurements depends on the component of the measurement vector. For Random Fourier Ensemble, this amounts to say that the noise on the measurement is the result of a "colored" noise on the initial signal, i.e. a non-flat spectrum noise. From the knowledge of the noise spectrum, a *whitening* of the measurement can be obtained by weighting the ℓ_2-norm of the BPDN constraint so that each weighted-components of the measurement vector has a normalized standard deviation, i.e.

$$\arg\min_{\boldsymbol{u}} \|\boldsymbol{u}\|_1 \ \text{ s.t. } \ \|\boldsymbol{W}\,(\boldsymbol{y} - \boldsymbol{\Theta u})\|_2 \leq \epsilon,$$

where $\boldsymbol{W} \in \mathbb{R}^{M\times M}$ is non-negative diagonal matrix. The direct effect of this treatment is to give more confidence to (perhaps small) measurements with low noise level and less to those with higher standard deviation [72].

Finally, in case where the measurement noise is non-Gaussian, the fidelity term of BPDN can be altered. For instance, if the noise comes from a uniform quantization of the measurements, i.e. the kind of digitization process that is implicitly used by any Compressed Sensing sensor [31, 47, 42], or if it follows a Generalized Gaussian Distribution (GGD) of shape parameter p, we can use the Basis Pursuit DeQuantization program

$$\arg\min_{\boldsymbol{u}} \|\boldsymbol{u}\|_1 \ \text{ s.t. } \ \|\boldsymbol{y} - \boldsymbol{\Theta u}\|_p \le \epsilon, \qquad (\text{with } p \ge 2). \tag{$\mathbf{BPDQ}_p$}$$

For uniformly quantized measurements, as suggested initially in [12] for $p = \infty$, it is shown in [41, 40] that if M is higher than a bound growing with p, i.e. in oversampled situation compared to $p = 2$ (i.e. BPDN), BPDQ_p improves the quality of the reconstructed signal. Other variations of BPDN in the same context exist [57, 20].

In the same quantization context, we may notice that a realistic quantization model must tackle the problem of *saturation*. A quantizer has indeed a limited number of bits to digitize its continuous input value. Therefore, every values outside of a certain range $[-B, B]$ for $B \ge 0$, will be assigned to the same digital number (up to a sign bit). Considering the matrix $\tilde{\boldsymbol{\Theta}}$ obtained by retaining the rows of $\boldsymbol{\Theta}$ leading to the measurements that did not saturate, and the matrix $\boldsymbol{\Theta}_\pm$ of those that saturated over and under B and $-B$ respectively, the two following optimization program can be considered [45, 46]

$$\arg\min_{\boldsymbol{u}} \|\boldsymbol{u}\|_1 \ \text{ s.t. } \ \|\boldsymbol{y} - \tilde{\boldsymbol{\Theta}}\boldsymbol{u}\|_2 \le \epsilon,$$

$$\arg\min_{\boldsymbol{u}} \|\boldsymbol{u}\|_1 \ \text{ s.t. } \ \begin{cases} \|\boldsymbol{y} - \tilde{\boldsymbol{\Theta}}\boldsymbol{u}\|_2 \le \epsilon, \\ \boldsymbol{\Theta}_+\boldsymbol{u} \ge B \\ \boldsymbol{\Theta}_-\boldsymbol{u} \le -B, \end{cases}$$

the first method consisting simply in the rejection of saturated measurements, and the second to the consistency of the rejected ones with respect to the range limit B. The *democratic property* of RIP matrices, i.e. the fact that they remain RIP (with higher constant) if some of their rows are removed, proves the theoretical stability of these two reconstruction methods [45, 46].

Outside Convexity: As explained in Section 4, the ideal reconstruction method using the ℓ_0 sparsity measure is usually *relaxed* by preferring to it the convex ℓ_1-norm. This allows one to efficiently solve the optimization programs BP and BPDN and to reach their global minimum.

Recently, some researches have focused on non-convex relaxation, where the ℓ_0-norm is rather replaced by the ℓ_q-norm, i.e. the sparsity term on $\boldsymbol{u} \in \mathbb{R}^N$ is $\|\boldsymbol{u}\|_q = (\sum_i |u_i|^q)^{1/q}$ with $0 < q \le 1$:

$$\arg\min_{\boldsymbol{u}} \|\boldsymbol{u}\|_q \ \text{ s.t. } \ \|\boldsymbol{y} - \boldsymbol{\Theta u}\|_2 \le \epsilon, \tag{ℓ_q–$\mathbf{BPDN}$}$$

Even if the non-convexity of the new optimization program prevents us to reach each a global minimum of ℓ_q-BPDN, several authors in [16, 6] introduced sub-optimal reconstruction based on reweightings of the ℓ_1 or the ℓ_2-norm.

Roughly speaking, the interest of this non-convex setting relies on the fact that an ℓ_q-ball is close to the "ℓ_0-ball" characterizing exactly sparse signals in $\mathbb{R}^N$ (Figure 2(b)).

Be greedy: From the advent of the sparsity concept in signal representations, and therefore also from the beginning of the Compressed Sensing theory, researchers have tried to find fast algorithms to solve the different convex optimization programs presented so far. However, even if optimal in the way they provide a global minimum, the complexity of these methods is generally high and function of the dimension of the signal space. Consequently, some suboptimal iterative methods, coined *greedy*, have quickly been proposed. Matching Pursuit (MP) [52], Orthogonal Matching Pursuit (OMP) [58], Compressive Sampling Matching Pursuit (CoSaMP) [56], Iterative Hard Thresholding (IHT) [4], Subspace Pursuit (SP) [19], ... are such examples of iterative procedures. They all follow the same principle.

Considering the noiseless case to simplify our explanation, given a signal $\boldsymbol{x}$ assumed to be sparse or compressible in the basis $\boldsymbol{\Psi}$, for a measurement vector $\boldsymbol{y} = \boldsymbol{\Phi x}$ sensed by $\boldsymbol{\Phi}$, a greedy algorithm applied to CS tries to find a sparse vector $\boldsymbol{\alpha}$ that explains the measurements through the lens of the sparsity basis, i.e. a vector $\boldsymbol{\alpha}$ such that $\boldsymbol{y} \approx \boldsymbol{\Phi\Psi\alpha}$. The signal $\boldsymbol{x}$ is subsequently recovered or approximated by computing $\boldsymbol{x} = \boldsymbol{\Psi\alpha}$.

Practically, we can say that a greedy procedure starts by setting a residual $\boldsymbol{r}$ to $\boldsymbol{y}$ and a current coefficient vector $\boldsymbol{\alpha}$ to 0. It checks then amongst all the columns of $\boldsymbol{\Theta} = \boldsymbol{\Phi\Psi}$, or amongst all the subset of columns of fixed size in this matrix, the one that is best correlated with $\boldsymbol{r}$. This selection is included in the current coefficient vector $\boldsymbol{\alpha}$ following various procedures, and the residual is updated by removing from it the influence of the selection. The process is then repeated on the new residual and the algorithm stops either naturally or after the minimal number of iterations needed to significantly reduce a certain score function, e.g. the energy of the residual.

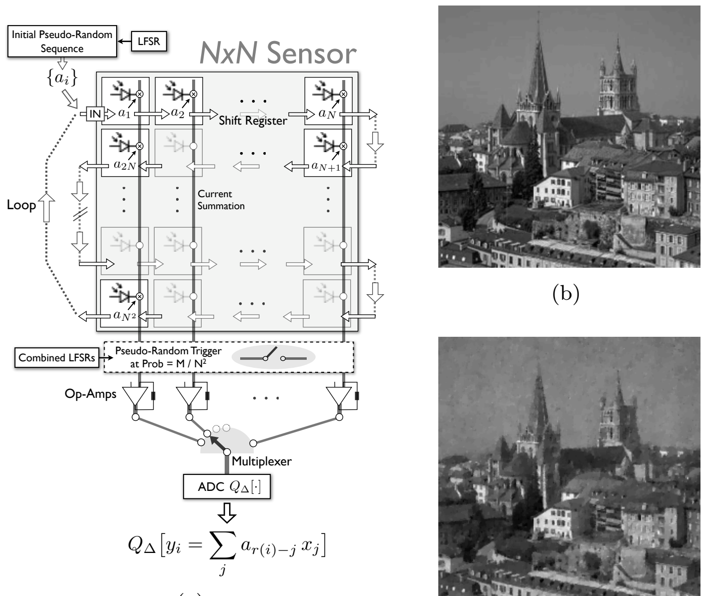
The first greedy methods used in CS were MP and OMP. The approximation error obtained with these was unfortunately not very well controlled. Recently however, CoSaMP, IHT and SP filled this gap and provided stability guarantees similar to the one given in (9).

7 Some Compressive Imaging Applications

Compressed Sensing is a very general concept reported to be adaptable to a tremendous amount of applications covering astronomy, radars, seismology, satellite telemetry, ... In this section we restrict our attention to only two of them, namely Compressive Imagers and Radio-Interferometry. We refer the readers to Chapters 14 and 23 of this book, or to parse the web references [15, 34] to get a complementary overview of the huge activity developped nowadays around applicative CS.

7.1 Compressive Imagers

The single pixel camera [31] was the first device to implement an optical Compressed Sensing directly acquiring compressive measurements. It relies on the



(a)

(b)

(c)

Figure 3: CMOS Compressive Imager. (a) The imager scheme. (b) and (c), simulated example of image sensing and reconstruction. In (b) original image (Lausanne Cathedral, Switzerland). In (c), reconstruction from a sensing of $M = \lfloor N/3 \rfloor$ 11-bits quantized measurements, corrupted by an additional Gaussian white noise (PSNR 27.3 dB).

use of a Digital Micromirror Device (DMD), a key element of a lot of Digital Light Processing (DLP) projectors. This DMD is composed of a grid of mirrors that may have only two different orientations, and one of them is used to focus scene light onto a single photodetector (PD). By setting electronically the grid arrangement according to a random configuration, the photodector actually collects a linear combination of the spatial light intensity with binary weights. Repeating the acquisition M times with different random patterns provides M measurements. In other words, M photodetector analog signals are generated, digitized and sent to a receiver which can reconstruct the image following one of the non-linear reconstructions described before. Pictures of this camera may be found in the Chapter 23 of this book.

Other Compressive Imagers have been realized since the single pixel camera. We may cite the Georgia Tech Analog Imager of R. Robucci et al. [60], and some multispectral and coded aperture extensions as described in much more details by R. Willet et al. in Chapter 23. The CMOS Compressive Imager presented in Figure 3 developed in the Swiss Federal Institute of Technology (EPFL) is one of them [42].

As for the Georgia Tech imager, this imager proceeds by embedding in the analog domain the compressed sensing of images. The selected sensing strategy relies on the Random Convolutions (RC) (Sec. 5) performed in the focal plane of the imager. This sensing has indeed a very simple electronic translation. Indeed, in the spatial domain, a RC of an image $\boldsymbol{x} \in \mathbb{R}^N$ is equivalent to

$$y_i = (\boldsymbol{\Phi x})_i = \sum_i a_{r(i)-j}\, x_j \;=\; (\boldsymbol{x} * \boldsymbol{a})_{r(i)}, \tag{11}$$

where a is the random filter and the indices $r(i)$ are selected uniformly at random in $\{1, \cdots, N\}$. For this compressive imager, the filter is not defined in the frequency domain but it corresponds to a suboptimal pseudorandom Rademacher sequence, i.e. $a_i = \pm 1$ with equal probability. Practically, a one-bit flip-flop memory is integrated in each pixel of the camera. Its binary state alters the direction of the electric current outgoing from the photodiode electronics before being gathered on column wires thanks to the Kirchoff law. Therefore, if the N flip-flop memories are set into a random pattern driven by the filter $\boldsymbol{a}$, the sum of all the column currents provides one CS measurement of the image seen on the focal plane. The next measurement are obtained easily since the one-bit memory are connected in a big chain represented in Fig. 3(a), creating a N shift register closed by connecting the last pixel memory to the first. In few clock signals, the whole random pattern can therefore be shifted on its chain, hence performing the desired convolution operation for the next measurements. We refer the reader to [42] for more details about the electronic design of this imager.

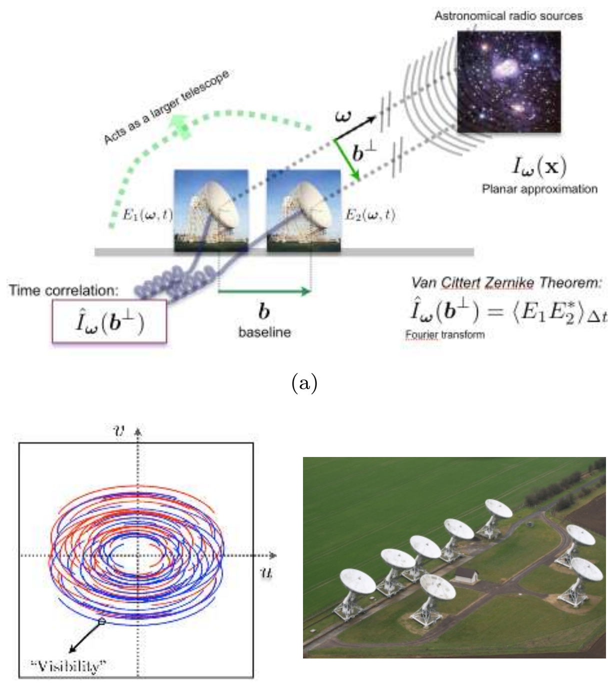
Compared to common CMOS imagers, this compressive acquisition has some advantages: *(i)* data are compressed at a reasonable compression ratio without to integrate a Digital Signal Processing (DSP) block in the whole scheme, *(ii)* no complex strategies must be developed to address numerically the pixels/columns of the grid before data transmission, and *(iii)* the whole system has a low power consumption since it benefit of the parallel analog processing (e.g. for filter shifting in the Shift Register). The proposed sensor is of course not designed for end-user systems (e.g. mobile phone). It meets however the requirements of technological niches with strong constraints (e.g. low power consumption) since the adopted CS coding involves a low computational complexity compared to systems embedding transform-based compression (e.g. JPEG 2000).

7.2 Compressive Radio-Interferometry

Recently, the Compressed Sensing paradigm has been successfully applied to the field of Radio-Interferometry detailed in Chapter 14.

In radio-interferometry, radio-telescope arrays[5], synthesize the aperture of a unique telescope of size related to the maximum distance between two telescopes. This allows observations with otherwise inaccessible angular resolutions and sensitivities in radio astronomy. The small portion of the celestial sphere

[5] As the one of Arcminute Microkelvin Imager (AMI) (Fig. 4(c))



(a)

(b) (c)

Figure 4: Radio-Interferometry by aperture synthesis. (a) General principles. (b) Typical visibility maps in the Fourier plane. (c) AMI radio-telescopes configuration

accessible to the instrument around the pointing direction $\boldsymbol{\omega}$ tracked during observation defines the original image $I_{\boldsymbol{\omega}}$ (or *intensity field*) to be recovered. As a matter of fact, thanks to the Van Cittert Zernike Theorem [72], the time correlation of the two electric fields E_1 and E_2 recorded by two radio-telescopes separated by a baseline vector $\boldsymbol{b}$ corresponds to $\hat{I}_{\boldsymbol{\omega}}(\boldsymbol{b}^{\perp})$, i.e. the evaluation of the Fourier transform of $I_{\boldsymbol{\omega}}$ on the frequency vector $\boldsymbol{b}^{\perp} = \boldsymbol{b} - (\boldsymbol{\omega} \cdot \boldsymbol{b})\boldsymbol{\omega}$ obtained by projecting $\boldsymbol{b}$ on the plane of observation perpendicular to $\boldsymbol{\omega}$. Since there are $\binom{N}{2}$ different pairs in a group of N telescope, and since the Earth is rotating, radio-interferometry amounts finally to sample the image of interest $I_{\boldsymbol{\omega}}$ in the Fourier domain on a list of frequencies, or *visibilities*, similar to one represented in Fig. 4(b).

As explained in Chapter 14, many techniques have been designed by radio-astronomers to reconstruct $I_{\boldsymbol{\omega}}$ from the set of visibilities. One commonly used method is the CLEAN algorithm [38] and the many variants generated since its first definition in 1974. Interestingly, CLEAN is actually nothing else but a Matching Pursuit greedy procedure (Sec. 6) that iteratively subtract from the visibilities the elements of the sky that are the most correlated with them.

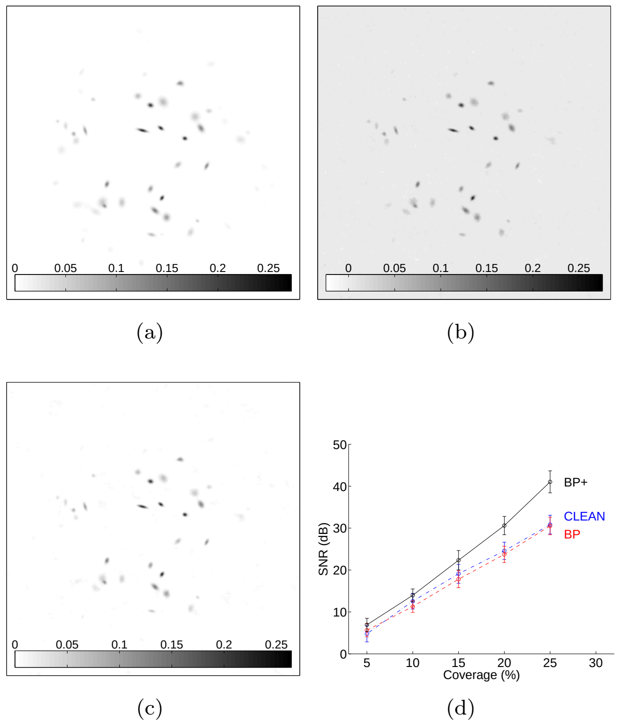


Figure 5: (a) Compact object intensity field I in some arbitrary intensity units. In (b) and (c), CLEAN and BP$^+$ (i.e. BPDN$^+$ with $\epsilon = 0$) reconstructions for $M = N/10$ and random visibilities. (d) The graph of the mean SNR with 1σ error bars over 30 simulations for the CLEAN, BP, and BP+ reconstructions of I as a function of M (in percentage of N).

CLEAN suffers however from a lack of flexibility and it is for instance not obvious to impose the positivity of the intensity field to be reconstructed. In [72], it is shown that, for random arrangements of visibilities in the frequency plane, Compressed Sensing reconstruction procedures like BPDN$^+$ or BPDN-TV (Sec. 6) provide significative gains in the quality of the reconstructions, as reported by Fig. 5 for one synthetic example.

In spirit, the sampling in radio-interferometry is very similar to this obtained in Magnetic Resonance Imaging (MRI) [50] where Compressed Sensing provides also significative improvements over the previous reconstruction schemes.

Recently, it has been shown in [73] that the modulating wave arising from the non-planar configuration of large radio-telescopes arrays on the Earth surface can advantageously be used as a "spread-spectrum" technique improving farther the image reconstruction quality.

8 Conclusion and the "Science 2.0" Effect

Let us conclude this chapter by one important, and somehow "sociologic", observation. The Compressed Sensing paradigm is born in what is often called the "Science 2.0." evolution. This term represents a real boom in the exchange of information in the scientific community, a byproduct of the steady Internet development. This recent trend take several forms. It consists for instance in the online publication of preprint before their acceptance in scientific journals (through ArXiv [68] or directly on personal homepages), or, more remarkably, in the writing of scientific blogs regularly fed by researchers and often the source of fruitful online discussions with other scientists. Another aspect of this trend is the free exchange of numerical codes guaranteeing the reproducibility of the results.

Information Sources: For Compressed Sensing, the two most visited sources of information are the Compressed Sensing Resource webpage maintained by Richard Baraniuk and his team at the Rice University [34] and the "Nuit Blanche" blog of Igor Carron [15]. The first link lists and classifies new preprints, papers, tutorial and softwares in the field, often from the personal announcements of the authors.

Thanks to his untiring writer Igor Carron, Nuit Blanche is realizing the same task than the first link but in addition his author comments the content of added research topics, trying to link certain of these to others, observing some new trends in the theory, disseminating the different softwares and toolboxes guaranteeing the reproducibility of the experiments.

Many other blogs exist in related topics to CS and it could be difficult to list them all here. We can mention for instance the applied mathematics blog "What's new" by Terence Tao [65], one of the founders of CS; "La vertu d'un LA" of L. Duval [32] about "*lets" bases, mathematics and anagrams; the "Geomblog" of Piotr Indyk and Suresh Venkatasubramanian about computational geometry and algorithms [39]; or the blog of D. Brady [5] on optical imaging and spectroscopy. Actually, many other interesting links can be found by surfing on these blogs since most of them maintain a "blogroll" section listing websites related to their own contents.

Reproducible Research: The advent of compressed sensing theory is also related to a free dissemination of the numerical codes helping the community to reproduce experiments. A lot of toolboxes have therefore been produced in different languages: C, C++, Matlab©, Python, ... One of the first tool designed to solve optimizations program as BP, BPDN or BPDN-TV was the ℓ_1-magic toolbox[6] in Matlab. Now, a lot of different implementation allow one to perform the same operation with more efficient methods, to give few names, CVX, SPGL1, Sparco, SPOT, TwIST, YALL, . . . are such examples of recent toolboxes. Once again, a comprehensive list is maintained on [34, 15].

[6] `http://www.acm.caltech.edu/l1magic`

References

[1] W.U. Bajwa, J.D. Haupt, G.M. Raz, S.J. Wright, and R.D. Nowak. Toeplitz-Structured Compressed Sensing Matrices. *Stat. Sig. Proc., 2007. SSP'07. IEEE/SP 14th Work.*, pp. 294–298, 2007.

[2] R. G. Baraniuk, M. A. Davenport, R. A. DeVore, and M. B. Wakin. A simple proof of the restricted isometry property for random matrices. *Constructive Approximation*, 28(3):253–263, Dec. 2008.

[3] R. Berinde, A.C. Gilbert, P. Indyk, H. Karloff and M.J. Strauss. Combining geometry and combinatorics: A unified approach to sparse signal recovery. In *Allerton Conference*, 2008.

[4] T. Blumensath and M.E. Davies. Iterative Hard Thresholding for Compressed Sensing. *App. Comp. Harm. Anal.*, 27(3):265–274, 2009.

[5] D. Brady. Optical Imaging and Spectroscopy Blog. `http://opticalimaging.org/OISblog`.

[6] E.J. Candès, M. Wakin, and S. Boyd. Enhancing sparsity by reweighted ℓ_1 minimization. *J. Fourier Anal. Appl.*, 14(5):877–905, 2008.

[7] E.J. Candès. Compressive sampling. In *Proc. Int. Cong. Math.*, vol. 1, 2006.

[8] E.J. Candès. The restricted isometry property and its implications for compressed sensing. *Compte Rendus Acad. Sc., Paris, Serie I*, 346:589–592, 2008.

[9] E.J. Candès and J. Romberg. Quantitative Robust Uncertainty Principles and Optimally Sparse Decompositions. *Found. Comp. Math.*, 6(2):227–254, 2006.

[10] E.J. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction fromhighly incomplete frequency information. *IEEE Tran. Inf. Th.*, 52(2):489–509, 2006.

[11] E.J. Candès, J. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Comm. Pure Appl. Math*, 59(8):1207–1223, 2006.

[12] E.J. Candès and T. Tao. Near-optimal signal recovery from random projections: universal encoding strategies. *IEEE Trans. Inf. Th.*, 52:5406–5425, 2004.

[13] E.J. Candès and T. Tao. Decoding by linear programming. *IEEE Trans. Inf. Th.*, 51(12):4203–4215, 2005.

[14] E.J. Candès and M.B. Wakin. An introduction to compressive sampling. *IEEE Sig. Proc. Mag.*, 25(2):21–30, 2008.

[15] I. Carron. "Nuit Blanche" blog. `http://nuit-blanche.blogspot.com`.

[16] R. Chartrand. Exact reconstruction of sparse signals via nonconvex minimization. *Sig. Proc. Lett.*, 14(10):707–710, 2007.

[17] S. Shaobing Chen, D.L. Donoho, and M.A. Saunders. Atomic decomposition by basis pursuit. *SIAM J. Sc. Comp.*, 20(1):33–61, 1998.

[18] R. Coifman, F. Geshwind, and Y. Meyer. Noiselets. *App. Comp. Harm. Anal.*, 10(1):27–44, 2001.

[19] W. Dai and O. Milenkovic. Subspace pursuit for compressive sensing signal reconstruction. *IEEE Tran. Inf. Th.*, Jan 2009.

[20] Wei Dai, Hoa Vinh Pham, and Olgica Milenkovic. Distortion-rate functions for quantized compressive sensing. *Arxiv preprint: arXiv:0901.0749,* 2009.

[21] I. Daubechies. *Ten Lectures on Wavelets.* Society for Industrial and Applied Mathematics, 1992.

[22] I. Daubechies, M. Defrise, and C. De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Communications on Pure and Applied Mathematics*, 57(11):1413–1457, 2004.

[23] A. Descampe, C. De Vleeschouwer, L. Jacques, and F. Marques. How does digital cinema compress images? *In "Applied Signal Processing - A MATLAB-based approach" by T. Dutoit and F. Marquez*, Chap. 10, Springer Verlag, June 2009.

[24] R.A. DeVore. Deterministic constructions of compressed sensing matrices. *J. Complexity.*, 23(4-6):918–925, 2007.

[25] D. Donoho and J. Tanner. Counting faces of randomly-projected polytopes when the projection radically lowers dimension. *J. AMS*, 22(1):1–15, January 2009.

[26] D.L. Donoho. Compressed Sensing. *Information Theory, IEEE Transactions on*, 52(4):1289–1306, 2006.

[27] D.L. Donoho and M. Elad. Optimally sparse representation in general (nonorthogonal) dictionaries via l 1 minimization. *Proc. Nat. Aca. Sci*, 100(5):2197–2202, 2003.

[28] D.L. Donoho and X. Huo. Uncertainty principles and ideal atom decomposition. *IEEE T. Inform. Theory.*, 47(7):2845–2862, November 2001.

[29] D.L. Donoho, M. Vetterli, R.A. DeVore, and I. Daubechies. Data compression and harmonic analysis. *IEEE Transactions on Information Theory*, 44(6):2435–2476, 1998.

[30] P.L. Dragotti, M. Vetterli, and T. Blu. Sampling moments and reconstructing signals of finite rate of innovation: Shannon meets Strang-Fix. *IEEE Tran. Sig. Proc.*, 55(5):1741–1757, May 2007. Part 1.

[31] M.F. Duarte, M.A. Davenport, D. Takbar, J.N. Laska, T. Sun, K.F. Kelly, and R.G. Baraniuk. Single-Pixel Imaging via Compressive Sampling [Building simpler, smaller, and less-expensive digital cameras]. *IEEE Sig. Proc. Mag.*, 25(2):83–91, 2008.

[32] L. Duval. "La vertu d'un LA" blog. `http://laurent-duval.blogspot.com`.

[33] M. Elad and A.M. Bruckstein. A generalized uncertainty principle and sparse representation in pairs of bases. *IEEE Transactions on Information Theory*, 48(9):2558–2567, 2002.

[34] R. Baraniuk et al. Compressive Sensing Resources. `http://www-dsp.rice.edu/cs`.

[35] L. Gan. Block compressed sensing of natural images. In *15th Int. Conf. Dig. Sig. Proc.*, pp. 403–406, 2007.

[36] L. Gan, T. Do, and T.D. Trans. Fast compressive imaging using scrambled block hadamard ensemble. In *Proc. EUSIPCO, Lausanne, Switzerland*, Sep. 2008.

[37] R. Gribonval and M. Nielsen. Sparse representations in unions of bases. *IEEE Transactions on Information Theory*, 49(12):3320–3325, 2003.

[38] JA Högbom. Aperture Synthesis with a Non-Regular Distribution of Interferometer Baselines, 1974. *A&AS*, 15:417–426.

[39] P. Indyk and S. Venkatasubramanian. The "Geomblog". `http://geomblog.blogspot.com/`.

[40] L. Jacques, D. K. Hammond, and M.J. Fadili. Dequantizing compressed sensing with non-gaussian constraints. In *Proc. of IEEE International Conference on Image Processing (ICIP)*, Cairo, November 2009. accepted.

[41] L. Jacques, D.K. Hammond, and M.J. Fadili. Dequantizing compressed sensing: When oversampling and non-gaussian constraints combine. *IEEE Trans. Trans. Inf. Th.*, 2009. (submitted).

[42] L. Jacques, P. Vandergheynst, A. Bibet, V. Majidzadeh, A. Schmid, and Y. Leblebici. CMOS Compressed Imaging by Random Convolution. In *IEEE Int. Conf. Ac. Speech Sig. Proc., 2009.* , pp. 1113–1116, 2009.

[43] M.A. Khajehnejad, A.G. Dimakis, W. Xu, and B. Hassibi. Sparse Recovery of Positive Signals with Minimal Expansion. *Arxiv preprint: arXiv:0902.4045*, 2009.

[44] J. Laska, S. Kirolos, Y. Massoud, R. Baraniuk, A. Gilbert, M. Iwen, and M. Strauss. Random sampling for analog-to-information conversion of wideband signals. *Proc. IEEE Dallas Circuits and Systems Workshop (DCAS)*, 2006.

[45] J.N. Laska, P. Boufounos, and R.G. Baraniuk. Finite-range scalar quantization for compressive sensing. In *Conf. on Sampling Theory and Applications (SampTA)*, 2009.

[46] J.N. Laska, P.T. Boufounos, M.A. Davenport, and R.G. Baraniuk. Democracy in Action: Quantization, Saturation, and Compressive Sensing. 2009. submitted.

[47] J.N. Laska, S. Kirolos, M.F. Duarte, T.S. Ragheb, R.G. Baraniuk, and Y. Massoud. Theory and implementation of an analog-to-information converter using random demodulation. In *IEEE International Symposium on Circuits and Systems, 2007. ISCAS 2007*, pp. 1959–1962, 2007.

[48] M. Ledoux. The concentration of measure phenomenon, volume 89 of Mathematical Surveys and Monographs. *American Mathematical Society, Providence, RI*, 208, 2001.

[49] M. Ledoux and M. Talagrand. *Probability in Banach Spaces: Isoperimetry and Processes*. Springer, 1991.

[50] M. Lustig, D. Donoho, and J.M. Pauly. Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magnetic Resonance in Medecine*, 58(6):1182, 2007.

[51] S. Mallat. *A Wavelet Tour of Signal Processing*. Academic Press., 2nd ed, 1999.

[52] S. Mallat and Z. Zhang. Matching pursuit with time-frequency dictionaries. *IEEE Tran. Sig. Proc.*, 41(12):3397–3415, Dec. 1993.

[53] S. Mendelson, A. Pajor, and N. Tomczak-Jaegermann. Uniform uncertainty principle for Bernoulli and subgaussian ensembles. *Constructive Approximation*, 28(3):277–289, 2008.

[54] F. M. Naini, R. Gribonval, L. Jacques, and P. Vandergheynst. Compressive sampling of pulse trains: Spread the spectrum! In *IEEE Int. Conf. Ac. Speech Sig. Proc., 2009*, Taipei, Taiwan, April 2009. in press.

[55] B.K. Natarajan. Sparse Approximate Solutions to Linear Systems. *SIAM J. Comp.*, 24:227, 1995.

[56] D. Needell and J. A. Tropp. CoSaMP: Iterative signal recovery from incomplete and inaccurate samples. *App. Comp. Harm. Anal.*, 26(3):301–321, 2009.

[57] Boufounos P. and Baraniuk R. G. 1-bit compressive sensing. In *42nd Conf. Inf. Sc. Sys. (CISS)*, pp. 19–21, Princeton, NJ, Mar. 2008.

[58] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad. Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition. In *In Proc. 27th Annual Asilomar Conf. Sig. Sys. Comp.*, Nov. 1993.

[59] H. Rauhut, K. Schnass, and P. Vandergheynst. Compressed Sensing and Redundant Dictionaries. *IEEE Tran. Inf. Th.*, 54(5):2210–2219, 2008.

[60] R. Robucci, L.K. Chiu, J. Gray, J. Romberg, P. Hasler, and D. Anderson. Compressive sensing on a CMOS separable transform image sensor. In *IEEE Int. Conf. Ac. Speech Sig. Proc., 2008*, pp. 5125–5128, 2008.

[61] Justin Romberg. Sensing by Random Convolution. *2nd IEEE Int. Work. Comp. Adv. Multi-Sensor Adap. Proc., 2007*, pp. 137–140, 2007.

[62] M. Rudelson and R. Vershynin. Sparse reconstruction by convex relaxation: Fourier and Gaussian measurements. In *40th An. Conf. Inf. Sc. Sys. 2006* , pp. 207–212, 2006.

[63] L. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal. *Physica D.*, 60:259–268, 1992.

[64] C.E. Shannon. Communication in the presence of noise. *Proc. IRE*, 37(1):10–21, 1949.

[65] T. Tao. "What's new" blog. `http://terrytao.wordpress.com`.

[66] R. Tibshirani. Regression shrinkage and selection via the lasso. *J. Royal Stat. Soc. Ser. B (Meth.)*, pp. 267–288, 1996.

[67] J. A. Tropp. Just relax: Convex programming methods for identifying sparse signals. *IEEE Tran. Inf. Th.*, 51(3):1030–1051, 2006.

[68] Cornell University. arxiv.org e-print archive. `http://www.arxiv.org`.

[69] M. Unser. Sampling-50 years after Shannon. *Proc. IEEE*, 88(4):569–587, 2000.

[70] E. van den Berg and M. P. Friedlander. Probing the pareto frontier for basis pursuit solutions. *SIAM J. Sc. Comp.*, 31(2):890–912, 2008.

[71] M. Vetterli, P. Marziliano, and T. Blu. Sampling signals with finite rate of innovation. *IEEE Tran. Sig. Proc.*, 50(6):1417–1428, 2002.

[72] Y. Wiaux, L. Jacques, G. Puy, AMM Scaife, and P. Vandergheynst. Compressed sensing imaging techniques for radio interferometry. *Mon. Not. R. Astron. Soc.*, 395(3):1733–1742, 2009.

[73] Y. Wiaux, G. Puy, Y. Boursier, and P. Vandergheynst. Spread spectrum for imaging techniques in radio interferometry. *Mon. Not. R. Astron. Soc.*, 400:1029, 2009.

[74] Y. Zhang. On theory of compressive sensing via ℓ_1-minimization: Simple derivations and extensions. Tech. Rep. TR08-11, Dep. Comp. Appl. Math., Rice Univ., Houston, Texas, 2008.