Compressed Sensing: a new paradigm for Rapid MR Imaging

Eleonora Mastrorilli
Registration Number: 614076
Supervisor: Professor Michele Pavon
University of Padova
Bachelor Degree in Information Engineering

September 24^{th}, 2012

Randomness is too important to
be left to Chance.

Robert R. Coveyou,
Oak Ridge National Laboratory

Introduction

The world we know would not be the same without the ease with wich we store and transmit images and signals in nowadays applications. We make extensive use of compression algorithms, which have been developed to reduce irrelevance and redundancy of image data in order to store or transmit them in an efficient form. However, image compression is still based on a complete data acquisition process, that obeys the traditional Nyquist-Shannon sampling theorem. Only later, when the most relevant information are known, the desired image can be converted into a small digital data set. In other words, we are stuck with acquiring a lot of redundant information, just to throw part of them away during compression.

A recent mathematical theory, known as *Compressed Sensing* (CS), tries to do much more: reduces the acquisition times and costs, thus lowering the number of data acquired, while maintaining high reconstruction fidelity. Taking advantage of some common properties of signals, like their *sparsity* and compressibility, this theory uses some "hard maths" to acquire just as many data as the most important ones, still ensuring exact reconstruction of the desired signal. The most surprising fact is that its efficiency relies on the possibility to choose *randomly* which information to gather from the acquisition process, without any previous information about the signal we are sensing. That is, we are bringing the compression right to the sensing process: and that's not only theory.

Many different applications have been proposed in the last few years, but the most intresting results have risen from the diagnostic imaging field. Magnetic Resonance Imaging (MRI), which is an essential medical imaging tool, seems to be optimal for successful application of Compressed Sensing. Often current medical imagery suffers from slow data acquisition processes and long timing, and MRI is an important example of this. Even if this field of science is at an open stage, applying CS to MRI offers potentially significant scan time reductions, with clear benefits for patients and operating costs reduction.

In this thesis we will give a brief introduction to this scenario. We firstly revise some useful concepts about signal acquisition and elaboration. In Chapter 2 we review the requirements for successful CS and its main results, while in the next chapter we describe its natural fit to MRI. Lastly,

we give some interesting examples of practical applications of CS in MRI. We try to emphasize an intuitive (yet precise) understanding of CS and of its potential, and a general understanding of the the driving factors and limitations in its application in MRI.

Contents

Chapter 1

Useful Concepts on Signals and MRI

In this chapter, we revise some basic concepts useful to understand the Compressive Sampling Theory and its innovative power in the field of Diagnostic Imaging. Although biological signals are often continuous-time by their nature, computerized analysis makes extensive use of their discrete form obtained through sampling. That's why we introduce discrete time signals, the Nyquist-Shannon sampling theorem and Fourier's Series and Transform, which are fundamental tools for signals' frequency analysis [13]. Lastly, we will revise also some basic Physics concepts about electronic and nuclear spin resonance in order to understand how Magnetic Resonance Imaging (MRI) actually works [10],[12].

1.1 Discrete Time Signals and Support of a Signal

Definition 1.1 (Discrete-Time Signal)**.** A discrete-time signal $x[n]$ is a time series consisting of a sequence of values. It is a function defined only for integer values of the indipendent variable n, that is

$$x\colon [n_1, n_2] \mapsto \mathbb{R} \quad \text{or} \quad (\mathbb{C}) \quad \text{with} \quad -\infty < n_1 < n_2 < +\infty. \tag{1.1}$$

Such a signal $x[n]$ is said to be *periodic* with period N, where N is a positive integer, if it is unchanged by a time shift of N, i.e. if

$$x[n] = x[n+N], \quad \forall n \in \mathbb{N}. \tag{1.2}$$

The *fundamental period* N_0 is the smallest positive value of N for which eq.(1.2) holds.

Fundamental discrete-time signals are the *Complex Exponential Signal*, defined by

$$x[n] = e^{j\theta_0 n} \tag{1.3}$$

and *Sinusoidal Signals*, defined by

$$x[n] = A\cos(\theta_0 n + \phi). \tag{1.4}$$

These signals are closely related to each other through the Euler's Formula,

$$e^{j\theta_0 n} = \cos(\theta_0 n) + j\sin(\theta_0 n). \tag{1.5}$$

We note that the fundamental complex exponential $e^{j(2\pi/N)n}$ is periodic with period N. Furthermore, the set of all discrete-time complex exponential signals that are periodic with period N is given by

$$\phi_k[n] = e^{jk\theta_0 n} = e^{jk(2\pi/N)n}, \quad k = 0, \pm 1, \pm 2, \ldots \tag{1.6}$$

All of these signals have fundamental frequencies that are multiples of $2\pi/N$ and thus are *harmonically related*. There are only N distinct signals in the set given by eq.(1.6). This is a consequence of the fact that discrete-time complex exponentials which differ in frequency by a multiple of 2π are identical. Hence, it suffices to take $k = 0, 1, \ldots, N-1$.

Definition 1.2 (Support of a Signal)**.** The *support* of a signal $x[n]$ is the smallest set of values $[m_x, M_x]$ for which

$$x[n] = 0 \quad \text{if} \quad n < m_x \quad \text{or} \quad n > M_x. \tag{1.7}$$

1.2 Fourier Series Representation of Discrete Time Signals

Joseph Fourier introduced in the study of trigonometric series the fundamental idea that a periodic signal can be decomposed into the sum of a (possibly infinite) set of oscillating functions, namely sines and cosines or complex exponentials. In particular, the Fourier series representation of a discrete-time periodic signal is a *finite* series. Using the set of harmonically related complex exponentials defined in (1.6), we can consider the representation of a periodic sequence in terms of linear combinations of the sequences $\phi_k[n]$. It has the form

$$x[n] = \sum_{k=0}^{N-1} a_k e^{jk\theta_0 n}, \quad \text{with} \quad \theta_0 = \frac{2\pi}{N}. \tag{1.8}$$

This equation is referred to as the *discrete-time Fourier series* and the coefficients a_k as the *Fourier series coefficients*. Since the exponentials $\phi_k[n]$ are linearly indipendent, we can solve (1.7) backwards, obtaining the coefficients a_k, as

$$a_k = \frac{1}{N}\sum_{n=0}^{N-1} x[n]e^{-jk(2\pi/N)n} \tag{1.9}$$

We note that, if we consider more than N sequential values of k, the values a_k repeat periodically with period N as a consequence of eq.(1.6). All the information about a periodic signal $x[n]$, with period N, is thus contained in its N *Fourier series coefficients*. In fact, as long as we know these N complex numbers, we can recover the original signal perfectly thanks to eq.(1.8). In other words, once we fix the set of $\phi_k[n]$, $x[n]$ is equally described by its discrete-time representation (in the time domain) and its Fourier coefficients (in the frequency domain). This means that we can identify a relationship between a periodic signal and its Fourier series coefficient, i.e.

$$x[n] \overset{\mathcal{F}_s}{\longleftrightarrow} a_k \tag{1.10}$$

described by the *Discrete-Time Fourier Series, (DTFS)*. It is an isometric map

$$\mathcal{F}_s \colon \mathbb{C}^N \to \mathbb{C}^N \tag{1.11}$$

with several properties:

- Linearity:
$$Ax[n] + By[n] \overset{\mathcal{F}_s}{\longleftrightarrow} Aa_k + Bb_k; \tag{1.12}$$
- Time Shifting:
$$x[n - n_0] \overset{\mathcal{F}_s}{\longleftrightarrow} a_k e^{-jk(2\pi/N)n_0}; \tag{1.13}$$
- Time Reversal:
$$x[-n] \overset{\mathcal{F}_s}{\longleftrightarrow} a_{-k}; \tag{1.14}$$
- Multiplication:
$$x[n]y[n] \overset{\mathcal{F}_s}{\longleftrightarrow} \sum_{l=0}^{N-1} a_l b_{k-l} = a_k * b_k; \tag{1.15}$$
- Periodic Convolution:
$$\sum_{r=0}^{N-1} x[r]y[n-r] = x[n] * y[n] \overset{\mathcal{F}_s}{\longleftrightarrow} Na_k b_k; \tag{1.16}$$
- Conjugation:
$$\overline{x[n]} \overset{\mathcal{F}_s}{\longleftrightarrow} \overline{a}_{-k}; \tag{1.17}$$

where $x[n]$ and $y[n]$ are periodic signals with period N and a_k and b_k are their Fourier Series coefficients. Since the DTFS is a linear and isometric map form $\mathbb{C}^N$ to $\mathbb{C}^N$ it can be described by a matrix F, named *Fourier*

Matrix. F is a square, complex-valued, symmetric matrix whose elements, $F_{jk} = e^{\frac{2\pi jk}{N}}$, are primitives Nth roots of unity. Hence, if we define two arrays

$$x = \begin{pmatrix} x(0) \\ x(1) \\ \vdots \\ x(N-1) \end{pmatrix} \quad \text{and} \quad A = \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_{N-1} \end{pmatrix} \tag{1.18}$$

we can describe the DTFS as $x = FA$. Moreover, if we multilpy F by a normalization factor $\frac{1}{\sqrt{N}}$, the resulting matrix is *unitary*, that is the conjugate and the inverse matrix coincide. Formally, if we define $U = \frac{1}{\sqrt{N}}F$ then $\overline{U}^T = U^{-1}$, so we can invert U obtaining

$$U^{-1} = \sqrt{N}F^{-1} = \frac{1}{\sqrt{N}}\overline{U}^T \longrightarrow F^{-1} = \frac{1}{N}\overline{F}^T. \tag{1.19}$$

Thus, inverting the DTFS equation $x = FA$, we obtain $A = F^{-1}x = \frac{1}{N}\overline{F}^T x$. We note two important results:

- the DTFS can be described as a linear transformation associated to the Fourier Matrix;
- DTFS and its inverse are almost identical, as a consequence of U being unitary.

1.3 Fourier Series Representation of Continuous-Time Signals

Fourier series analysis evaluates also which periodic continuous-time signals can be represented as a linear combination of complex exponentials. As we did for discrete-time signals, we define the fundamental complex exponential signal as

$$x(t) = e^{j\omega_0 t}, \quad \text{with} \quad t \in \mathbb{R}. \tag{1.20}$$

It is always periodic with fundamental period $T = \frac{2\pi}{\omega_0}$ and it can be expressed as a combination of sines and cosines through Euler's Formula (1.5). A set of complex exponential

$$\phi_k(t) = e^{jk\omega_0 t} \quad \text{with} \quad k = 0, \pm 1, \pm 2, \ldots \tag{1.21}$$

is said to be *harmonically related* if each one of these signals has a fundamental frequency that is multiple of ω_0, and therefore, each one is periodic with period T (although their fundamental period is a fraction of T for $|k| > 1$).

Thus, a linear combination of harmonically related complex exponentials of the form

$$x(t) = \sum_{k=-\infty}^{+\infty} a_k e^{jk\omega_0 t} = \sum_{k=-\infty}^{+\infty} a_k e^{jk(\frac{2\pi}{T})t} \tag{1.22}$$

is also periodic with period T. As the Riesz-Fischer teorem guarantees [13], eq.(1.22) describes a signal that converges to the original one, $x(t)$, in a mean square limit sense (yet not punctually). The term in (1.22) for $k = 0$ is a constant defined by the average value of x(t) over one period, while the terms with $k = \pm 1$ both have fundamental frequency equal to ω_0, and are referred to as *first harmonic components*. The representation of a periodic signal in the form of eq.(1.22) is referred to as its *Fourier Series* representation. The coefficients a_k, referred to as the *Fourier series coefficients*, can be obtained by

$$a_k = \frac{1}{T}\int_T x(t)e^{-jk\omega_0 t}\,dt = \frac{1}{T}\int_T x(t)e^{-jk(\frac{2\pi}{T})t}\,dt. \tag{1.23}$$

Note that, in eq.(1.23), we will obtain the same result if we integrate over *any* interval of length T. We want to stress that the relations obtained by eq.(1.22) and eq.(1.23) are well defined (that is, the integral or the series do not diverge) for a large class of signals, including those with finite energy over a single period. The Fourier Series representation for continuous-time signals has several properties, analogous to that we have already defined in the discrete-time case, useful to reduce the complexity of the Fourier series of many signals.

1.4 Continuous-time Fourier Transform

The results seen so far about *Fourier Series* apply only to periodic signals. In this paragraph, we see how these concepts can be applied to signals that are not periodic. The idea is to represent an aperiodic signal $x(t)$ by first constructing a periodic signal $\tilde{x}(t)$ that is equal to $x(t)$ over one period. Then, as this period approaches infinity, $\tilde{x}(t)$ is equal to $x(t)$ over larger and larger intervals of time, and the Fourier series representation for $\tilde{x}(t)$ converges to the so called *Fourier Transform* representation of $x(t)$. A rather large class of signals, including all signals with finite energy, can be represented through a linear combination of complex exponentials close in frequency. The resulting spectrum of coefficients in this representation is called the *Fourier Transform.*

Definition 1.3 (Fourier Transform)**.** Let $x(t)$ be a continuous-time signal with finite energy, i.e.

$$\int_{-\infty}^{+\infty} |x(t)|^2\,dt < +\infty. \tag{1.24}$$

Then its *Fourier Transform* $X(j\omega)$ is defined as a mean-square limit by

$$X(j\omega) = \int_{-\infty}^{+\infty} x(t)e^{-j\omega t}\,dt. \tag{1.25}$$

Moreover, we can derive the original signal $x(t)$ from its transform thanks to the following theorem:

Theorem 1.4.1 (Inversion Theorem)**.** *Let* x(t) *be a continuous-time signal for which (1.24) holds, and let* $X(j\omega)$ *be its Fourier Transform. If both are absolutely integrable, we can define a function* g(t)*, that is*

$$g(t) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} X(j\omega)e^{j\omega t}\,d\omega. \tag{1.26}$$

This function is well-defined, continuous and satisfies

$$\int_{-\infty}^{+\infty} |x(t) - g(t)|\,dt = 0. \tag{1.27}$$

Moreover, if x(t) *satisfies the Dirichelet Conditions, namely*

- $x(t)$ *is absolutely integrable;*
- $x(t)$ *has a finite number of maxima and minima within any finite interval*
- $x(t)$ *has a finite number of discontinuities within any finite interval,*

then

$$g(t) = x(t) \tag{1.28}$$

except where x(t) *has a discontinuity.*

The *Fourier Transform* has several properties. We underline that it is *linear* and *isometric*, i.e. it preserves scalar product. It is worth to emphasize that, although *Fourier Transform* has been introduced referring to aperiodic signals, we can develop it for periodic signals too. In particular, *Fourier Transform* of a periodic signal with Fourier coefficients a_k consists of a train of impulses occurring at the related frequencies and with area that is proportional to the Fourier series coefficients, as we will see in the next example. We want to stress that, since periodic signals do not satisfy (1.24), their *Fourier Transform* is a generalized function obtained by analysis of the *inversion formula* (1.26).

Example 1.4.1. Let $x(t) = \cos(\omega_0 t)$ be the signal we want to transform. It is clearly periodic, so we cannot use (1.25) to obtain its Fourier Transform.

If we impose the complex exponential $z(t) = e^{j\omega_0 t}$ to have as its generalized Fourier Transform $Z(j\omega) = 2\pi\delta(\omega - \omega_0)$, we obtain

$$\begin{aligned} e^{j\omega_0 t} &= \frac{1}{2\pi}\int_{-\infty}^{+\infty} Z(j\omega)e^{j\omega t}\,dt \\ &= \frac{1}{2\pi}\int_{-\infty}^{+\infty} 2\pi\delta(\omega - \omega_0)e^{j\omega t}\,dt \end{aligned}$$

that satisfies (1.26). So we can now write

$$x(t) = \cos(\omega_0 t) = \frac{1}{2}\left[e^{j\omega_0 t} + e^{-j\omega_0 t}\right]$$

and, thanks to linearity, we can finally find $X(j\omega)$ as

$$\begin{aligned} X(j\omega) &= \frac{1}{2}[2\pi\delta(\omega - \omega_0) + 2\pi\delta(\omega + \omega_0)] \\ &= \pi[\delta(\omega - \omega_0) + \delta(\omega + \omega_0)] \end{aligned}$$

As we mentioned earlier, the Fourier Transform of the periodic signal $x(t)$ consists of two impulses.

1.5 Discrete-time Fourier Transform

As in the previous paragraph, we introduce *Fourier Transform* for aperiodic discrete-time signals too.

Definition 1.4 (Discrete-time Fourier Transform)**.** Let $x[n]$ be a discrete-time signal with finite energy, i.e.

$$\sum_{n=-\infty}^{+\infty} |x[n]|^2 < +\infty. \tag{1.29}$$

Then its *discrete-time Fourier Transform* $X(e^{j\theta})$ is defined as a mean-square limit by

$$X(e^{j\theta}) = \sum_{n=-\infty}^{+\infty} x[n]e^{-j\theta n}. \tag{1.30}$$

The inversion forumla is

$$x[n] = \frac{1}{2\pi}\int_{2\pi} X(e^{j\theta})e^{j\theta n}\,d\theta. \tag{1.31}$$

Hence, even an aperiodic discrete-time signal can be thought as a linear combination of a continuoum complex exponentials. In particular, eq.(1.31) is a representation of $x[n]$ as a linear combination of complex exponentials infinitesimally close in frequency and with amplitude proportional to

$X(e^{j\theta})$. For this reason, $X(e^{j\theta})$ is often referred to as the *spectrum* of $x[n]$, because it provides the information on how $x[n]$ is composed at different frequencies. The discrete-time Fourier transform shares many similarities with the continuous-time case. The major difference lies in the periodicity of discrete-time transform $X(e^{j\theta})$, since it is always 2π periodic. As in the continuous-time case, discrete-time periodic signals can de described in the transform domain by interpreting their transform as an impulse train in the frequency domain.We note that their Fourier Transform is a generalized function obtained by inspection of the *inversion formula* (1.31).

1.6 The Sampling theorem

We can always see a discrete-time signal as the sampled version of a particular continuous-time signal. In fact, we can *sample* a continuous-time signal, that is, we can evaluate it only on a sequence of equally spaced values of the indipendent variable t. The simplest way to do this is through the use of a periodic impulse train multiplied by the continuous-time signal $x(t)$ that we wish to sample. This mechanism is known as *impulse-train sampling*, where the periodic impulse train $p(t)$ is the *sampling function*, the period T is the *sampling period* and the fundamental frequency of $p(t)$, $\omega_s = 2\pi/T$ is the *sampling frequency*. However, in the absence of any additional conditions, we would not expect the original signal to be uniquely specified by its samples. This means that we cannot generally reconstruct it perfectly because of the loss of information that sampling introduces. *The Nyquist-Shannon Sampling Theorem* introduces some conditions that guarantee the possibility to recover perfectly certain signals from their samples. This result is extremely important in practical applications of signal and system analysis. Infact it defines a lower bound to the amount of "information" of a signal we need to acquire.

Theorem 1.6.1 (The Nyquist-Shannon Sampling Theorem)**.** *Let $x(t)$ be a band-limited signal, with $X(j\omega) = 0$ for $|\omega| > \omega_M$. Then $x(t)$ is uniquely determined by its samples $x(nT)$, $n = 0, \pm 1, \pm 2, \dots$ if*

$$\omega_s > 2\omega_M \tag{1.32}$$

where

$$\omega_s = \frac{2\pi}{T}. \tag{1.33}$$

Given these samples, we can reconstruct $x(t)$ by generating a periodic impulse train in wich successive impulses have amplitudes that are successive sample values. This impulse train is then processed through an ideal low-pass filter with gain T and cutoff frequency greater than ω_M and less than $\omega_s - \omega_M$. The resulting output signal will exactly equal $x(t)$.

The frequency $2\omega_M$, which, under the sampling theorem, must be exceeded by the sampling frequency, is commonly referred to as the *Nyquist rate*. We underline that ideal filters are generally not used (or not avaliable) in practice for a variety of reasons. In fact, they are typically non-causal and not input-output stable. In any practical application, the ideal lowpass filter used in the theorem would be replaced by a nonideal filter $H(j\omega)$ that approximates the ideal one in the frequency band of interest. Interpolation, that is, the fitting of a continuous signal to a set of sample values, is a commonly used procedure to reconstruct a function, either approximately or exactly, from its samples. When the Sampling Theorem conditions aren't met, it is impossible to reconstruct correctly the original signal by an ideal low-pass filter. In this case there is a superposition, in the frequency domain, of sampled signal's Fourier Tranform that prevents the antitrasformated filtered signal to be correct, as we will see in the next example. This effect is referred to as *aliasing*, and it introduces unwanted artifacts in signals reconstruction.

Example 1.6.1. Let $x(t)$ be the signal we want to sample and then reconstruct.

$$x(t) = \cos\omega_0 t \overset{\mathcal{F}_s}{\longleftrightarrow} X(j\omega_0) = \pi[\delta(\omega-\omega_0)+\delta(\omega+\omega_0)]$$

so $X(j\omega)$ is band-limited and we can apply the sampling theorem. Now we choose the sampling rate as

$$\omega_s = 6\omega_0 = 6\omega_M > 2\omega_M$$

that satisfies the Sampling Theorem condition, and we use a lowpass ideal filter with gain T and cutoff frequency of $\omega_c = \frac{\omega_s}{2}$. In this way, we can perfectly reconstruct the signal from its sampled version, as the teorem guarantees (see also Fig.1.1), that is $x_r(t) = x(t)$.

Now we choose a new sampling rate as

$$\omega_s = \frac{3}{2}\omega_0 = \frac{3}{2}\omega_M < 2\omega_M$$

so that the Sampling Theorem condition is violated. In this case, choosing the cutoff frequency of the lowpass filter as in the previous case, $\omega_c = \frac{\omega_s}{2}$, the reconstructed signal is

$$\begin{aligned} x_r(t) &= \mathcal{F}_s^{-1}[X_r(j\omega)] = \mathcal{F}_s^{-1}[\pi\delta(\omega-\omega_0+\omega_s)+\pi\delta(\omega+\omega_0-\omega_s)] \\ &= \cos(\omega_s-\omega_0 t) = \cos\left(\frac{1}{2}\omega_0 t\right) \neq x(t). \end{aligned}$$

We can thus see the destructive effect of aliasing (see also Fig.1.2).

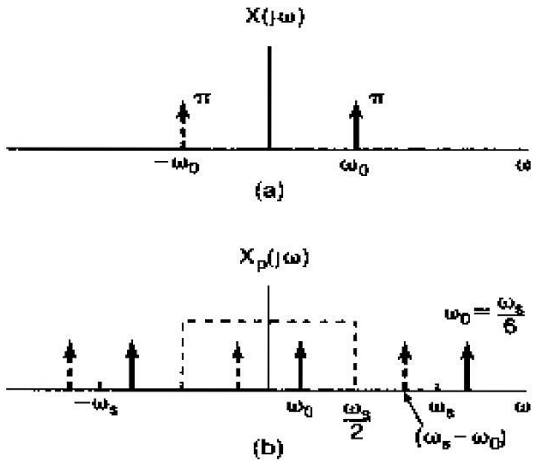

Figure 1.1: (a) spectrum of the sinusoidal signal and (b) spectrum of the sampled signal with $\omega_s > 2\omega_0$

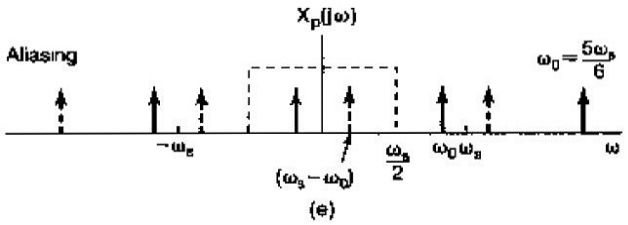

Figure 1.2: spectrum of the sampled signal with $\omega_s < 2\omega_0$: we notice that the impulses falling within the passband of the lowpass filter differs from the spectrum of the original signal, fig1.1(a), thus the reconstructed signal is affected by aliasing

One of the best-known example of aliasing is the principle on which the stroboscopic effect is based. In this case a disc, with a single radial line marked on, rotates at a constant rate. A flashing strobe acts as a sampling system, since it illuminates the disc for brief time intervals at a periodic rate. When the strobe frequency is much higher than the rotational speed of the disc, the speed of rotation of the disc is perceived correctly. When the strobe frequency becomes less than twice the rotation frequency of the disc, the rotation appears to be at a lower frequency than its actual one. When the strobe frequency becomes less than that of rotation, the disc appears to be rotating in the opposite direction! The stroboscopic effect is thus an example of useful application of aliasing due to undersampling.

1.7 Electronic Spin and Nuclear Magnetic Resonance

For the sake of simplicity, in this paragraph we will refer to the Hydrogen atom, although the concepts we will revise still remain valid for more complex atoms or molecules. The hydrogen atom is the simplest element in nature. It has just one proton (+) and one electron (-). Because of its physical structure, the hydrogen atom's proton spins on its axis. This generates a magnetic field that interacts with external magnetic fields. Because of this spin rotation, a magnetic dipole is created along the axis of rotation. The entire atom also spins around a second axis, like a top, moving within a conelike trajectory (precession). If the spin and the precession axis of an hydrogen atom rotate in the same direction, we define it a *low-energy nucleum*. Otherwise, if the spin and the precession axis rotate in opposite direction, we call it a *high-energy nucleum*. Under normal conditions, a bunch of hydrogen atoms will have their precession axes randomly oriented in different directions. When a strong magnetic field is applied, all the atoms line up their precession axes in the same directions. In this case, if the atoms are stimulated through the application of radio waves of a particular frequency (called the *resonance frequency*), low-energy protons will absorb it to become high-energy protons. When the trasmission of radio waves stops, the low-energy protons return to their previous state. While they relax, they release the energy they have absorbed in the form of a wave of precise frequency that can be captured and analyzed. These waves describe precisely the magnetic and chemical properties of the atoms they were released by.

1.8 Magnetic Resonance Imaging

Thanks to a technology that combines magnetic fields and radio waves, it is possible to render high-quality images of soft tissues in the human body. To do this, the Magnetic Resonance Scanner scans for the hydrogen atoms

in these tissues. To detect the atoms, the area is initially subjected to a powerful magnetic field and later stimulated using radio-frequency waves. This process causes the atoms to release energy that is then detected by the scanner and converted into images. The scanner is thus composed by a superconducting magnet and its cooling system, and by a radio-frequency transmitter (coil) that stimulates the atoms. The revolutionary features of this tecnique is that it has no inconvenience to the patient, other than the requirement to remain still for a while. Moreover, it does not require the use of contrast agents or the use of X-rays, as in the case of radiography or computerized tomography.

Chapter 2

A new way of sampling: Compressed Sensing

As we saw in the previous chapter, the Sampling Theorem defines a strict boundary to signal acquisition. That is, if we want to acquire a signal, we have to satisfy Nyquist condition or we won't be able to reconstruct it correctly. However, we often treat signals (or images) whose information content can be described by far less data than what Nyquist Theorem states. That's why, with the increasing interest in data storage and elaboration, compression algorithms are now fundamental. They can reduce data sets by order of magnitude, making systems that acquire extremely high resolution images (or signals) feasible. There is an extensive body of literature on image compression, but the central concept is to transform the image into an appropriate basis and then code only the important expansions coefficients. We underline that, according to this method, we have to acquire a certain number of samples in order to satisfy the Sampling Theorem only to throw part of them away with a compression algorithm. But is there a way to avoid the large data acquisition? In other words, is it possible to build the data compression directly into the acquisition? The answer is yes, and it is what *Compressed Sensing (CS)* is all about. Since 2004 Emmanuel Candés, Terence Tao and David Donoho developed the CS theory starting from some simple ideas:

- if we know the image structure, we match the sensing method to it, in order to minimize the number of measurements m needed to reconstruct it faithfully; if we don't know anything about the image we want to acquire, the best way is to sense the signal in a complete random, inchoerent and unstructured way;

- therefore we can reconstruct the signal only in a *probabilistic* sense, meaning that with a certain probability we can recover the original image from the samples we have randomly acquired;

- we have to use a non-linear algorithm in order to recover with high probability the correct signal, since we are trying to solve an undetermined problem.

The mathematical theory underlying CS is deep and beautiful, and draws from different fields, but the moral is very general: a good signal representation can aid the acquisition process and outrun the concept of information acquisition we commonly have.

2.1 A new notion of "sampling"

Compressed Sensing and compression algorithms both take advantage of the inner property of most signals to contain some redundancy. In fact, the analysis in the transform domain of many signals and most of images shows that a few information (i.e., coefficients in the transform domain) could represent the entire signal, while the rest of the samples, with high probability, will be zero-valued. For example, if we consider the simple cosinusoidal signal of Example1.4.1 we can see how it is completely described by just two impulses in the Fourier Transform domain, while its band, that depends linearly on ω_0, could increase at discretion (and thus could increase the number of samples necessary to satisfy the Nyquist theorem). In other words, often a signal contains less information then what we are constrained to acquire by the Nyquist Sampling theorem. This is particularly true for signals that, in some representation, are *sparse*, that is they can be usefully described by a few coefficients.

Definition 2.1 (Sparsity of a signal)**.** A signal $x[n]$ is said to be *S-sparse* if its coefficient series (when it is represented in a certain basis) has at most S non-zero elements.

In the previous example, the cosinusoidal signal is 2-sparse in the Fourier Transform domain. Clearly, if we could build an acquisition system that samples only these two Fourier coefficients we would be able to rebuild the entire signal without sampling it at high rate as we did in Example 1.6.1. What Compressed Sensing tries to do is then to reduce the number of samples acquired, going far behind the Nyquist theorem and approaching the S-sparsity of the analyzed signal.

Until now we have considered just one sampling method, the *impulse-train sampling*, but we can simply generalize our notion of sampling. In our acquisition system, we can obtain each mesurement y_k as an inner product against a different *test function* ϕ_k:

$$y_k = \langle x, \phi_k \rangle \quad \text{for} \quad k = 1, \dots, m, \tag{2.1}$$

where x is the discrete-time signal expressing the image. In this way, the choice of the ϕ_k allows us to choose in which domain we gather the information about the image. For example, if the ϕ_k are sinusoids at different

frequencies, we are directly collecting Fourier coefficients, while if they are indicator functions on squares, we are just collecting pixels. We stress that, so far, we have assumed that we have complete control over which ϕ_k to use, while it is not always true in real image acquisition. Moreover, since the measurements $y_1, \dots, y_m$ are in some sense a *coded version* of the image x, we will refer to these generalized kinds of samples as *coded imaging*. Shortly, if we combine all the m ϕ_k in a matrix Φ, then we can describe the vector Y of the samples acquired through a linear transformation. Formally, if

$$Y = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} \tag{2.2}$$

then we can rewrite eq.(2.1) as

$$Y = \Phi X, \quad \text{where} \quad \Phi = \begin{pmatrix} \phi_1 \\ \vdots \\ \phi_m \end{pmatrix} \tag{2.3}$$

is an $m \times n$ matrix whose rows are the test functions ϕ_i, while X is the vector (with dimension $n \geq m$) of the complete signal x. Our goal is then to choose the test functions ϕ_k in order to minimize the number of measurement m we need to reconstruct X faithfully. The simplest idea is to match these test functions to the image structure. In other words, we try to make measurements in the same domain in which we would compress the image, and in which it is possibly sparse. With the ϕ_k matched to the signal structure, we reconstruct our image using a simple regression algorithm, the *least squares*[1], finding the closest image that matches the observed projection onto the span of $[\phi_1, \dots, \phi_m]$:

$$\hat{X} = \Phi^*(\Phi\Phi^*)^{-1}Y, \tag{2.4}$$

where Φ is the linear operator we introduced, that maps the image to a set of m measurements, Φ^* is its adjoint, and Y is the m-vector of the observed values. The effectiveness of this strategy is determined by how well images of interest can be approximated in the *fixed* linear subspace spanned by the ϕ_k. We now demonstrate why this method fails in most of cases, even if it is an interesting starting point to understand the CS theory. Unfortunately this strategy isn't *adaptive* because of its simplistic approach: even if the image changes, and so does its structure, we are stuck recording the same m transform coefficients for every image, while the important coefficients to

[1]The method of least squares is a standard approach to the approximate solution of overdetermined systems, i.e., sets of equations in which there are more equations than unknowns. "Least squares" means that the overall solution minimizes the sum of the squares of the errors made in the results of every single equation.

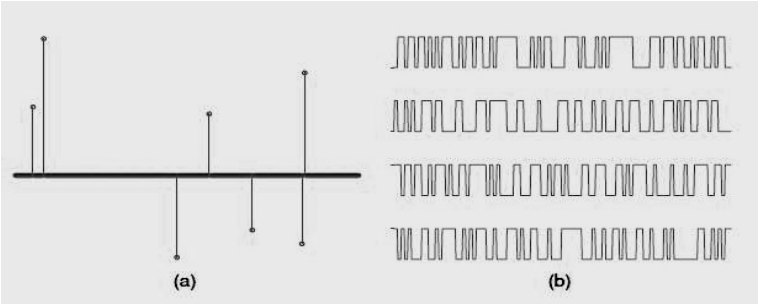

Figure 2.1: (a) A sparse vector. If we try to sample it with no knowledge of its structure, we will see zero-valued samples most of the time. (b) Examples of pseudorandom incoherent test vectors ϕ_k. With each measurement of (a) with a test vector from (b) we gather a little bit of information about which components are active.

reconstruct the image could vary. Since compression algorithms have the entire image and its transform to examine, they can abridge the complexity of this problem with ease [6]. The same advantage is not afforded to an acquisition system, that has no way to judge which transform coefficients are important until after the image is sampled! To outrun this problem, we modify the test function ϕ_k and the matrix Φ while we persist using a linear acquisition system.

The best way to achieve an adaptive approximation performance with a predetermined set of linear mesurement is to choose the ϕ_k to be completely unstructured and *noise-like*, although we have then to modify our reconstruction algorithm. The central result of CS, as we will see deeply in the next paragraphs, is that from m of these noise-like, incoherent measurement, we can renconstruct the image as well as if we had observed its most important coefficients. In order to do so, while we know that the image we are trying to acquire is sparse in some domain, it is critical to choose our measurement functions (and their linear combinations) not to be. In fact, if we don't know where the important values are, sampling the transform coefficients directly (that is, using sinusoids for the ϕ_k we introduced, in order to collect Fourier coefficients) will be chiefly fruitless effort, since we will see most of the sampled values that are very close to zero. Instead, if we take combinations of the transform coefficients, using incoherent ϕ_k, we gather a little bit of information about the whole signal with each measurement (see also Fig.2.1). We now have to formalize how to modify the test vectors ϕ_k in order to obtain noise-like, incoherent measurements. First of all, we have

to rewrite Φ as an *orthogonal matrix*[2], with

$$\Phi^*\Phi = nI. \tag{2.5}$$

In this way, the observed transform coefficients are still described by eq.(2.3). Of course, if all of the n coefficients of X are observed, we simply have to apply $\frac{1}{n}U^*$ to the vector of observations Y to recover X. Instead, we are analyzing the case in which only a small part of the components of Y are actually observed. Given a subset $\Omega \subset \{1, \ldots, n\}$ of size $|\Omega| = m$, the challenge is to infer the n-dimensional vector X from the shorter m-dimensional vector of observations $Y = \Phi_\Omega X$, that is we wish to solve an undetermined system of equations. Published results ([8]) take ϕ_k to be a realization of Gaussian white noise, or a sequence of Bernoulli random variables taking values ± 1 with equal probability (similarly to Fig.2.1 (b)).

Sometimes a signal we want to acquire could have a sparse representation in a basis different from the one defined by the measurement system Φ [4]. In these cases we introduce a pair of orthonormal basis (Φ, Ψ), where the former is the measurement basis introduced earlier, while the latter is the one we use to describe the signal $x[n]$. For instance, if the signal we wish to recover from m measurement is not sparse in the time domain, its expansion in the basis Ψ could be sparse, that is

$$f(t) = \sum_{j=1}^{n} x_j \psi_j(t), \quad f = \Psi x, \quad \text{and} \quad Y = \Phi f, \tag{2.6}$$

(where the waveforms ψ_j are the columns of Ψ). Our goal is then to search for the coefficient sequence in the Ψ-domain that explains the samples obtained in the domain Φ, enjoying the properties

$$\Psi^*\Psi = I, \quad \Phi^*\Phi = nI, \tag{2.7}$$

(as always, I represents the identity matrix). Hence, we can define a new matrix U as

$$U = \Phi\Psi, \tag{2.8}$$

whereupon we can define a new parameter $\mu(U)$, also referred to as *mutual coherence*.

[2]In linear algebra, an orthogonal matrix is a square matrix with real entries whose columns and rows are orthogonal unit vectors (i.e., orthonormal vectors). Equivalently, a matrix Q is orthogonal if its transpose is equal to its inverse: $Q^T = Q^{-1}$, which entails $Q^TQ = QQ^T = I$, where I is the identity matrix. An orthogonal matrix Q is necessarily invertible (with inverse $Q^{-1} = Q^T$), unitary and normal ($Q^*Q = QQ^*$). As a linear transformation, an orthogonal matrix preserves the dot product of vectors, and therefore acts as an isometry of Euclidean space, such as a rotation or reflection. In other words, it is a unitary transformation.

Definition 2.2 (mutual coherence)**.** Given two bases Ψ and Φ we define their mutual coherence as

$$\mu(\Phi, \Psi) = \max_{1 \leq k,j \leq n} | \langle \phi_k, \psi_j \rangle |, \tag{2.9}$$

$$\mu(U) = \mu(\Phi, \Psi) = \max_{k,j} |U_{k,j}|. \tag{2.10}$$

This parameter is a rough characterization of the degree of similarity between the sparsity and measurement system, thus it can be interpreted as a measure of how concentrated the rows of U are. Since each row of U has an l_2-norm equal to $\sqrt{n}$, μ will take values between 1 and $\sqrt{n}$. When the rows of U are perfectly flat, that is $|U_{k,j}| = 1$ for each k, j (as in the case when U is the discrete Fourier Transform), we will have $\mu(U) = 1$, and we will achieve the maximum incoherence possible. If a row of U is maximally concentrated (all the row entries but one vanish), then $\mu(U) = \sqrt{n}$ and we have obtained an (undesiderable) coherent measurement system. For μ to be close to its minimum value of 1 (that is, to attain *incoherent* measurements) each of the measurement vectors ϕ_k must be "spread out" in the Ψ domain. This parameter is particularly relevant for the CS theory, since it expresses the relationship between the sensing modality (Φ) and the signal model (Ψ), that affects the number of measurements required to reconstruct a sparse signal.

2.2 The recovery method

We just saw how, measuring a series of random combinations of the entries of x observing its inner product with the random vectors ϕ_k, we measure the signal globally, learning something new about the sparse vector with every measurement. This random sensing strategy works because each sparse signal will have a unique set of measurements.

To recover an S-sparse vector X from $Y = \Phi X$, inverting the measurement process, we have to solve an optimization problem. Since we have less samples than the required ones to reconstruct the signal perfectly (according to the Nyquist Sampling theorem) we are trying to solve an undetermined problem (that is, a system which has multiple or infinite solutions, in opposition to a system with a unique solution). At first glance, solving the undetermined system of equations appears hopeless; but suppose now that the signal is *compressible*, meaning that it depends on a number of degrees of freedom which is smaller than N. For instance, if our signal is sparse, then the problem changes radically, making the search for the solutions feasible. The fundamental idea that allows us to overcome this problem is to look for the sparsest signal that satisfies the sampled values. In other words, we assume that we are dealing with a signal x that we know sparse in some

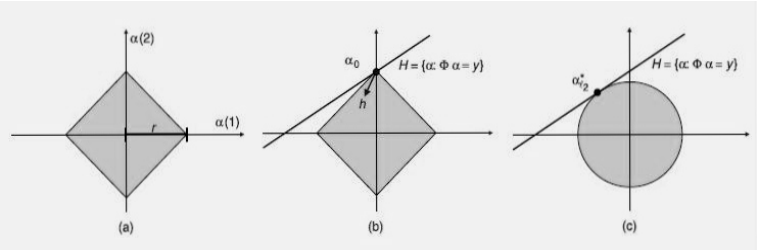

Figure 2.2: (a) l_1 ball with radius r. (b) l_1 minimization recovers the sparsest vector of all the set of vectors that share the same measurement values Y; (c) l_2 minimization isn't effctive because it can't recover spase vectors.

basis and that we sampled it with a set of m incoherent, noise-like test function ϕ_k, obtaining a measurement vector Y. Thus we search for a signal that matches exactly the sampled values at their frequencies, and that is zero-valued elsewhere, so that it is the sparsest of the infinite set of signals which maps to Y. At first, the best recovery algorithm seems to be

$$\min_{\tilde{X}} \#[i : \tilde{X}(i) \neq 0] \quad \text{subject to} \quad \Phi\tilde{X} = Y. \tag{2.11}$$

The functional $\#[i : \tilde{X}(i) \neq 0]$ is simply the number of nonzero terms in the candidate vector $\tilde{X}$, and it is sometimes referred to as the l_0 norm. The problem with eq.(2.10) is that solving it directly is infeasible, since it is combinatorial and NP-hard[3] [6]. If we relax the boundaries defined by eq.(2.10) we obtain a convex program that works almost as well and that is far easier to solve:

$$\min_{\tilde{X}} ||\tilde{X}||_{l_1} \quad \text{subject to} \quad \Phi\tilde{X} = Y, \tag{2.12}$$

where

$$||\tilde{X}||_{l_1} = \sum_{i=1}^{n} |\tilde{X}_i|. \tag{2.13}$$

The main diffrence between these two strategies is the substitution of sum of magnitudes in place of size of the support; even though they are fundamentally different, they produce the same results in many interesting situations.

We will now show why the geometry of l_1 norm minimization works well as a substitute for sparsity. Referring to Fig(2.2) we see how the l_1

[3]NP-hard (non-deterministic polynomial-time hard), in computational complexity theory, is a class of problems that are, informally, "at least as hard as the hardest problems in NP". NP-hard problems have a deep theoric and practical relevance. In fact, if a problem P is provable to be equivalent to a well-known NP-hard problem, then it is demonstrated that is nearly impossible to find an efficient way to solve it.

"ball", that is the set of points of the space with equal l_1-norm from a given point (the centre), is clearly anisotropic. If we compare it with the standard Euclidean l_2 ball, which is spherical and thus completely isotropic, we note that it is pointy along the axis, and this property is the fundamental key that favors sparse vectors. In fact, the l_1 ball of radious r contains all the points of $\mathbb{R}^2$ such that $|\alpha(1)| + |\alpha(2)| \leq r$. If we now search for a sparse α_0 that satisfies $Y = \Phi\alpha_0$, the only intersection point between the l_1 ball (that represents the sparsity hypotesis) and the line H (that represents the congruence with the samples acquired) recovers the sparsest vector that satisfies eq.(2.12) (see also Fig.(2.2), (b)). The point labeled α_0 is a sparse vector, since only one of its two components are nonzero, of which we make one measurement, while the line labeled H is the set of all α that share the same measurement value. To pick the point with minimum l_1 norm we can imagine the l_1 ball to start with tiny radius and then expanding it until it touches H. This first point of intersection is by definition the solution to eq.(2.12). The anisotropy of the l_1 ball, combined with the flatness of the space H results in a unique solution that coincides with the original sparse signal α_0. On the contrary, minimizing the l_2 norm does not recover α_0, because the l_2 ball is isotropic and thus recovers a solution ${\alpha^*}_{l_2}$ that will in general not be sparse at all.

Intuitively, the l_2 norm penalizes large coefficients heavily, therefore solutions tendo to have many smaller coefficients, hence not to be sparse. In the l_1 norm, many small coefficients tend to carry a larger penalty than a few large coefficients, therefore small coefficients are suppressed and solutions are often sparse [1].

2.3 The new sampling theorem

In this paragraph we will exactly formalize the intuitions about CS we have introduced so far.

Let x be a signal defined with respect to a certain orthonormal basis Ψ, so that $x = \Psi X$ (where X is thus the vector of coeffcients that describes the signal on Ψ), and let be $y_k = \langle x, \phi_k \rangle = \langle \Psi X, \phi_k \rangle$. Then, as we saw in the previous paragraph, we can recover the original signal x starting from the samples y_k thanks to the l_1 minimization. We will now discuss the number of samples necessary to recover the signal with a negligible error.

Theorem 2.3.1 (The New Sampling Theorem)**.** *[3] Assume that a signal $x \in \mathbb{R}^N$ has its coefficient vector X that is S-sparse and that we are given m of its samples in the generic domain Φ with frequencies selected uniformly at random. Suppose that the number of observations obeys*

$$m \geq C \cdot \mu^2(\Psi, \Phi) \cdot S \cdot \log n, \tag{2.14}$$

for some small constant C.Then minimizing l_1 reconstructs x exactly with overwhelming probability. In details, if the constant C is of the form $22(\delta+1)$ in (2.14), then the probability of success exceeds $1 - O(N^{-\delta})$.

The first conclusion is that one suffers no information loss by measuring just about any set of m frequency coefficients. The second is that the signal x can be exactly recovered by minimizing a convex functional which does not assume any knowledge about the number of nonzero coordinates of X, their locations, and their amplitudes which we assume are all completely unknown a priori. In fact, the theorem does not require any knowledge about the position of the nonzero elements of X but only their number S. Moreover, the importance of the coerence parameter is now clear: if we choose a couple of basis (Ψ, Φ) such that $\mu(\Psi, \Phi) = 1$ (that is, we have maximum incoherence between the sensing and the sparsity basis), the number of samples we need to reconstruct X correctly is the order of magnitude of $S \cdot \log n$.

To illustrate the deep innovation this theorem introduces over the classic Nyquist-Shannon Theorem we will develop a brief comparison [7]. Suppose that a signal x has support Ω in the frequency domain and has $B = |\Omega|$. If Ω is a connected set, we can think of B as the bandwidth of x. If in addition the set Ω is known, then the classical Nyquist-Shannon sampling theorem states that x can be reconstructed perfectly from B equally spaced samples in the time domain, while the reconstruction is simply obtained by linear interpolation. Now suppose that the set Ω, still of size B, is unknown and not necessarily connected. In this situation, the Nyquist-Shannon theory is unhelpful. However, Theorem(2.3.1) asserts that far fewer samples are necessary. Solving eq(2.12) will recover X perfectly from about $B \cdot \log N$ time samples. What is more, these samples do not have to be carefully chosen; almost any sample set of this size will work. Thus we have a nonlinear analog to the classical sampling theorem: we can reconstruct a signal with arbitrary and unknown frequency support of size B from about $B \cdot \log N$ samples arbitrarily chosen.

While this seems to be a great achievement, one could still ask whether this is optimal, or if we could do it with even fewer samples. The answer is that in general, we cannot reconstruct S-sparse signals with fewer samples. This is a consequence of a *uniform uncertainty principle* (UUP) introduced by Candés and Tao ([3]). The UUP essentially states that the $m \times n$ sensing matrix Φ obeys a "restricted isometry hypothesis. Let Φ_T, with $T \subset 1, \ldots, n$ be the submatrix (with dimensions $m \times |T|$) obtained extracting the columns of Φ corresponding to the indices chosen in T. Then we define the S-restricted isometry constant δ_S of Φ as the smallest quantity that satisfies

$$(1 - \delta_S)||X||^2{}_{l_2} \leq ||\Phi_T X||^2{}_{l_2} \leq (1 + \delta_S)||X||^2{}_{l_2} \tag{2.15}$$

for all subsets T with $|T| \leq S$. This property essentially requires that every

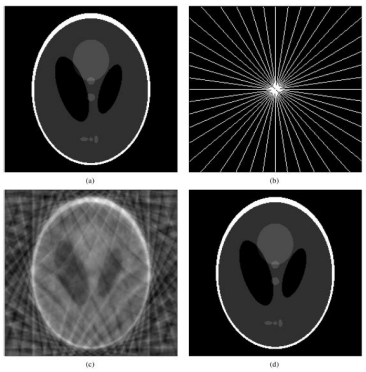

Figure 2.3: (a) The logan-Shepp phantom test image. (b) Sampling grid in the frequency plane. (c) Minimum energy reconstruction. (d) Reconstruction obtained by l_1 minimization. The recostructed image is an exact replica of (a).

set of columns with cardinality less than S approximately behaves like an orthonormal matrix. Then our matrix Φ obeys the UUP since, for any S-sparse vector X, the energy of the measurements ΦX will be comparable to the energy of X itself:

$$\frac{1}{2} \cdot \frac{m}{n} \cdot ||X||^2{}_2 \leq ||\Phi X||^2{}_2 \leq \frac{3}{2} \cdot \frac{m}{n} \cdot ||X||^2{}_2, \tag{2.16}$$

where, as usual, m is the number of sampled values observed, while n is the dimension of the complete vector X. We call this an uncertainty principle because the proportion of the energy of X that appears as energy in the measurement is roughly the same as the undersampling ratio m/n. To understand how the UUP affects sparse recovery, we demonstrate that it is fundamental for the unicity of the recovery algorithm's solution [6]. In fact, if we suppose that eq.(2.16) holds for sets of size $2S$, while we keep measuring our S-sparse vector as above $Y = \Phi X$, we could ask if is there any other S-sparse (or sparser) vector X' that shares the same measurements. If there were such a vector, then the difference $h = X - X'$ would be $2S$-sparse and have $\Phi h = 0$: these two properties are though incompatible with the UUP. In short, if the UUP holds at about the level S, the minimum l_1-norm reconstruction is provably exact.

The theoretical power of the results shown so far has several practical experiments to support it. The true power of Compressed Sensing isn't its complete mathematical unified theory, but its stability, robustness (i.e. versus measurement errors) and its generality, since it can be applied widely to any basis used to describe the signal. Probably the most famous example to show how well it applies to sense and recovery sparse signal is the so

called *Longan-Shepp phantom test image* (Fig 2.3). This example considers the problem of reconstructing a two-dimensional image from samples of its discrete Fourier Transform on a star-shaped domain. To recover the original signal, (c) assumes that the Fourier coefficients at all the unobserved frequencies are zero (thus reconstructing the image with minimum energy). This l_2 strategy doesn't work well, since it suffers from severe artifacts. On the contrary, a strategy based on convex optimization (d), gathers *exact* reconstruction of the image.

Although the acquisition process is simple, solving the recovery program is computatively burdersome. Fortunately there have been drastic advances in the field of convex optimization that make it tractable on the image quality and dimension we are usually intrested in. As a general rule, solving an l_1 minimization program is about 30 or 50 times more expensive than solving an l_2 problem. Improving the algorithmic efficiency is thus one of the CS research goals in the last few years.

Chapter 3

The Natural Fit Between CS and MRI

Compressed Sensing clearly offers some useful properties for signal acquisition and elaboration, since it can drastically reduce sampling time and dataset. Consequently it has found serveral applications in many different fields, from astronomy to communication networks. We will focus our attention on its potential in the diagnostic imaging field, in particular on its application to Magnetic Resonance Imaging. In fact, MR images are inherently sparse and incoherent, two properties we now know fundamental in order to apply efficiently the CS theory. Moreover, as we saw in the first chapter, MRI presents no inconvenience for the patient, but has an inherently slow data acquisition process which causes long scanning time. CS can overcome this weakness by undersampling the desired images, thus scanning faster, or alternatively, improving MR imagery resolution. Consequently, applying CS to MRI offers potentially significant scan time reductions, with benefits for patients and health care economics [2]. Another follow-up is that dynamic-3D imaging (such as acquisition of heart motion) could improve image definition, avoiding blur and out-focusing due to unintended motions.

In this chapter we will revise at first the properties of MRI that are particularly significant for CS and describe their natural fit; then we will see how to implement the CS theory in a realistic and effective way.

3.1 MRI properties and constraints

The MRI signal is generated by protons in the body, mostly those in water molecules. The application of a strong static magnetic field and a radio frequency excitation field produces a net magnetic moment that precesses at a frequency proportional to the static field strenght. The transverse magnetization $m(\vec{r})$ and its corresponding emitted RF signal (detected by a receiver coil) can be made proportional to many different physical prop-

erties of tissues [2]. MR images reconstruction attempts to visualize the spatial distribution of the transverse magnetization. This practice, called *Spatial Encoding*, directly samples the spatial frequency domain of the image, thus extracting a Fourier relation between the received MR signal and the magnetization distribution. Since encoded information are obtained by superimposing additional gradient magnetic fields on top of the strong static field, the total magnetic field will vary with position as $B(x) = B_0 + G_x x$ (where G_x is the gradient field that varies linearly in space). If we take into account all the three Cartesian axes, we will have G_x, G_y, G_z to be our gradient field's components, that could vary independently from each other. These gradient are limited in amplitude and slew rate[1], which are both system specific, by physical constraints. In fact, high gradient amplitudes and rapid switching can produce peripheral nerve stimulation and must be avoided, then providing a physiological limit to system speed and performance.

Gradient-induced variation in precession frequency causes the development of a linear phase dispersion. Therefore the receiver coil detects a signal encoded by the linear phase, in the form of a Fourier Integral [9],

$$s(t) = \int_R m(\vec{r}) e^{-j2\pi \vec{k}(t)\cdot\vec{r}}\, dr, \tag{3.1}$$

where $m(\vec{r})$ is the transverse magnetization at position $\vec{r}$, while $k(t) \propto \int_0^t G(s)\, ds$ describes formally the sampling trajectory. In other words, the received signal at time t is the Fourier Transform of $m(\vec{r})$ sampled at the spatial frequency $\vec{k}(t)$ [2].

Since the spatial frequency space is crossed by our sampling trajectory $\vec{k}(t)$, that depends on the gradient waveforms $\vec{G}(t)$, it can be alternatively called k-space. What we are free to develop in this contest is $\vec{G}(t) = [G_x(t), G_y(t), G_z(t)]^T$ and thus the k-space sampling pattern. Traditionally it is designed to meet the Nyquist criterion, depending on the desired resolution and field of view of the image. The most popular trajectory used is a set of lines from a Cartesian grid, which guarantees robustness to many sources of system imperfections. Of course, violation of the Nyquist criterion in this acquisition scenario causes image artifacts in linear reconstruction.

If we abandon the traditional method in favor of the CS approach, seeking a drastic performance growth, we have to satisfy some basic criteria. In fact, the CS approach requires that: (a) the desired image has a sparse representation in a known transform domain (i.e. it is compressible), (b) the aliasing artifacts due to k-space undersampling are incoherent (noise like) in that transform domain, (c) a non-linear reconstruction method is used both to enforce sparsity of the image and consistency with the acquired data [1].

[1]The slew rate of a system is defined as the maximum rate of change (expressed in time units) of the output signal, when the whole system has an impulse as its input.

As we will see in the next paragraphs, each of these requirements is fully satisfied by MR imaging.

3.2 Sparsity and Sparsifying Transform

Most MR images are sparse in an appropriate transform domain. In fact, when an image does not already look sparse in time, frequency or pixel domain, we can always apply a sparsifing transform to it. As we saw in the previous chapter, it is an operator that maps a vector of image data to a sparse vector in some bases. We currently possess a set of transform that can sparsify many different type of images. Piecewise constant images can be well sparsified by spatial finite-differencies, that is computing the difference between close pixels (it applyies well to images where boundaries is what we are most interested in). Real-life images are known to be often sparse in the discrete cosine transform (DCT) or in the wavelet transform domain, mostly used in JPEG and MPEG compression standards.

For example, angiograms, which are images of blood vessels, contain primarily contrast enhanced blood vessels on an empty background, thus already look sparse in the pixel representation. They can be made even sparser by spatial finite-differencing. More complex imagery, such as brain images, can be sparsified in more sophisticated domains, such as the wavelet domain. Sparse representation is not limited to still imagery. Often videos can safely be compressed much more heavily. Dynamic MR images are highly compressible as well, since they are extremely sparse in the temporal dimension. For example, the quasi-periodicity of heart images has a sparse temporal Fourier transform [2].

The transform sparsity of MR images can be demonstrated by applying a sparsifying transform to a fully sampled image and reconstructing an approximation to the image from a subset of the largest transform coefficients. The sparsity of the image is then expressed by the percentage of transform coefficients sufficient for diagnostic-quality reconstruction. Of course, diagnostic quality is a subjective parameter that depends on the specific application. But we know from CS theory that the number of samples acquired depends on a quality parameter C that ensures faithful reconstruction of the image. Thus, rising its numerical value we can suit any accuracy requirement, at the cost of an increased number of sample to acquire.

To illustrate this, Michael Lustug, David Donoho and John M. Pauly [1] performed this experiment on two representative MR images: an angiogram of a leg and a brain image. The results show that a reconstruction involving $5\% - 10\%$ of the largest transform coefficients guarantees good consistency with the original images.

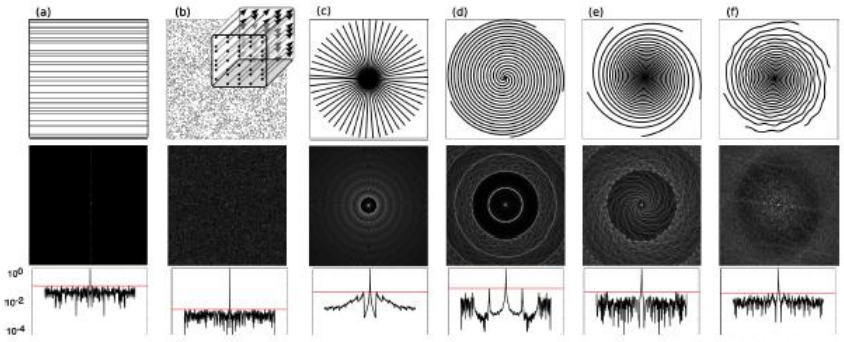

Figure 3.1: Various sampling trajectories and their PSF. From left to right: random 2D lines, random 2D points (also seen as cross-section of random 3D lines), radial, uniform spirals, variable density spirals, variable density perturbed spirals.

3.3 Incoherent Sampling and Point Spread Function

A central point to design a CS scheme for MRI is then to select a subset of the frequency domain which can be efficiently sampled and that leads to incoherent aliasing interference in the sparse transform domain.

In the original CS theory we assumed to sample a truly random subset of the k-space to guarantee a very high degree of incoherence. Neverthless complete random sampling in all dimensions is generally impractical, because of hardware and physiological constraints. Sampling trajectories have to follow smooth lines or curves to be implemented in practice and to be robust to non-ideal situations. Then, we aim to design a practical sampling scheme that emulates the interference results of complete random sampling but takes into account structural constraints.

We have to keep in mind that our goal is to allow rapid data collection, thus maintaining software and hardware implementation as simple as possible. Moreover, most MR images don't have a uniform energy distribution in k-space. Thus a uniform random distribution of samples wouldn't be as effective as expected. In fact, most energy of the MR images is concentrated close to the centre of the k-space, suggesting use of a variable-density sampling scheme. To match this property, we should sample randomly with sampling density scaling according to the distance from the k-space origin. Variable density acquisition trajectories such as Cartesian, radial and spiral imaging have been proved to have apparently incoherent aliasing, so that

the interference appears as white noise in the image domain. Cartesian grid undersampling is by far the simplest to implement: we should just modify the acquisition method of a traditional scanner so that it simply drops entire lines of phase encodes from an existing complete grid. In this way scan time reduction is exactly proportional to the degree of undersampling. However, it is suboptimal, since the achievable incoherence is significantly worse than with truly random sampling. On the other hand, radial and spiral imaging are more complex to implement, but could recreate a somewhat irregular acquisition that achieves high incoherence, yet allowing rapid collection of data. Fig.3.1 shows some of the most used sampling trajectories.

To compare the efficiency of the scheme proposed, we need a quantitative measure of incoherence that allows us to put their performances in comparison. The *Point Spread Function* (PSF) is a natural tool to mesure incoherence and a practical implementation of the mutual coherence we introduced in the previous chapter.

Definition 3.1 (Point Spread Function)**.** Suppose that we sample a subset S of the k-space, and that F_S is the Fourier transform evaluated on the frequency subset S. If F_S^* denotes the adjoint operation (that we can imagine as a simple zero-filling followed by the inverse Fourier transform) we define the *Point Spread Function* (PSF) as the matrix with elements

$$PSF(i,j) = (F_S^* F_S)(i,j). \tag{3.2}$$

Under complete sampling, the PSF becomes the identity matrix with all the off-diagonal terms equal to zero. Undersampling the k-space induces nonzero off-diagonal terms, which shows that linear reconstruction of pixel i suffers interference by a unit impulse at pixel $j \neq i$.

Under undersampling conditions, let e_i be the i-th vector of the natural basis (having "1" at the ith position and zero elsewhere). Then $PSF(i,j) = {e^*}_j F_S^* F_S e_i$ measures the contribution of a unit-intensity pixel at the ith position to a pixel at the jth postion. In other words, the PSF measures how zero-filling linear reconstruction produces aliasing, leaking energy from pixel to pixel. This energy then shows up in the image as aliasing artifacts and blurring (see also Fig.3.2). In this context we define *coherence* to be the maximum off-diagonal value in the PSF matrix that describes the image of interest.

As we saw earlier, MR images are at most sparse in the transform domain rather than in the usual pixel domain. According to this, we generalize the notion of PSF to *Transform Point Spread Function* (TPSF), which measures how a single transform coefficient of the underlying object influences other transform coefficients of the undersampled image.

In order to do this, we recall the orthogonal sparsifying transform Ψ we introduced in the previous chapter. Then TPSF is given by

$$TPSF(i,j) = (\Psi^* F_S^* F_S \Psi)(i,j). \tag{3.3}$$

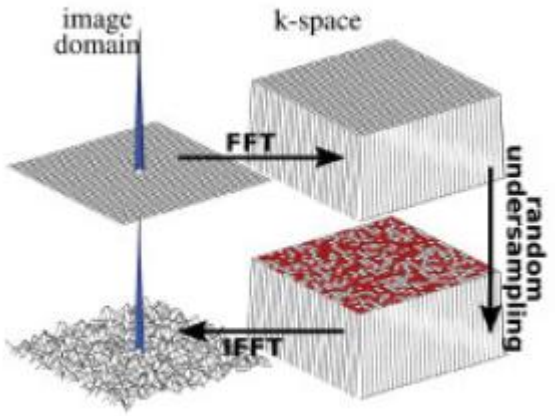

Figure 3.2: The PSF of random 2D k-space undersampling. We note how undersampling results in incoherent interference in the reconstructed image domain.

With this notation, coherence is measured as the maximum off-diagonal value of the TPSF matrix as well. Of course, we are looking for small coherence, i.e. incoherence.

In conclusion, we aimed to find an optimal sampling scheme that maximizes the incoherence for a given number of samples. However, we have to keep in mind that this problem is combinatorial and might be considered intractable. Since we know that choosig samples at random usually results in a good, incoherent solution, practical procedures build a probability density function (that takes into account what we saw about variable density trajectory and the image's energy density distribution) upon a complete Cartesian sampling grid. Then, simply drawing indices at random from the pdf, we undersample the grid. Repeating this procedure iteratively and choosing the pattern that exibits the lowest interference might seem wasteful, but turns out to be optimal since the sampling pattern so obtained can be used again for future scans.

3.4 Non-linear Reconstruction and Thresholding Method

As we know from the CS theory, the recovery method is complex and computatively heavy. To get an intuition about how it can be realistically imple-

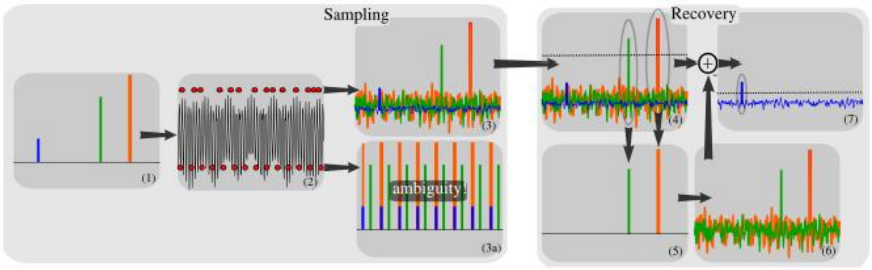

Figure 3.3: Heuristic procedure for reconstruction from undersampled data. (1) A sparse signal and its k-space domain representation (2). The signal is undersampled in its k-space either through pseudo-random undersampling (higher dots line) and equispaced undersampling (lower dots line). The latter results in signal aliasing preventing recovery (3a), while the former (3) can be recovered by use of iterative thresholding: some components rise above the noise level (4) and can be recovered (5). Their contribution to interference is calculated (6) and can be subtracted from the whole signal(7), thus lowering the total interference level.

mented and to show how incoherence is fundamental for the feasibility of CS in MRI, we will first develop an intuitive exampe of 1D signal reconstruction (Fig. 3.3).

A sparse signal is undersampled; since it is already sparse in the transform domain, we can think its sparsifying transform to be simply the identity. The common procedure of zero-filling the missing values and inverting the Fourier transform generates artifacts that depend upon the sampling pattern used. We can now distinguish two different cases, depending on whether we use equispaced undersampling or random undersampling.

Equispaced k-space undersampling and reconstruction generates coherent aliasing. In fact, the reconstructed signal shows a superposition of shifted signal replicas that results in inherent ambiguity. Since every signal copy is equally likely, it is impossible to determine which signal is the original one, leading to the impossibility of signal reconstruction.

On the other hand, random undersampling results in a complete different situation. The Fourier reconstruction shows incoherent artifcts that behave much like additive random noise to the original signal's transform. Of course, the aliasing artifacts aren't noise, rather interferences that can be explained with an energy-distribution argument. In fact, random undersampling reconstruction causes leakage of energy away from the individual nonzero coefficients of the original signal that tends to spread randomly to other reconstructed signal coefficients, including those wich had been zero

in the original signal.

If we have a certain knowledge of the underlying original signal and of the sampling pattern, it is possible to calculate this leakage analytically. Then a plausible recovery method relies on a nonlinear iterative procedure (also shown in Fig. 3.3), based on *thresholding*. If we analyze the reconstructed image and pick up only the strongest component of the signal, we are able to calculate the interference they caused in the overall reconstruction. Therefore, subtracting this interference from the entire signal, we can reduce the total interference level, thus allowing smaller components, previously submerged, to stand out and be recovered. Iterative repetition of this procedure permits to recover the rest of the sigal components.

We will now formalize the image reconstruction strategy by slightly modifying equation (2.12) in order to make it more robust to non-ideal situations, since we have always to take into account some errors, noise and imperfection of the system. Recalling that the image of interest is a vector X, that Ψ denotes the linear transformation from pixel to an appopriate sparsifying domain, that Φ_S represents the undersampled Fourier transform and that Y stands for the reconstructed image vector, we can write

$$\begin{aligned} \text{minimize} \quad & ||\Psi X||_1 \\ \text{so that} \quad & ||\Phi_S X - Y||_2 < \epsilon. \end{aligned} \tag{3.4}$$

As we saw in the CS theory, this constrained optimization problem exploits l_1 minimization to enforce sparsity. The second equation in (3.4) introduces a new bond, since ϵ represents roughly the expected noise level, thus controlling the fidelity of the reconstruction to the measured data.

In other words, among all the solutions consistent with the acquired data, eq.(3.4) finds the solution that is compressible by the transform Ψ [1]. Since iterative algorithms used to solve such optimization problem in effect perform thresholding and interference cancellation at each iteration, it is now explicit how such formal approach relates to the informal idea we introduced earlier.

Of course, eq.(3.4) can be further refined in many practical cases. For example, when finite differences are used as sparsifying transform, the objective in this equation is usually referred to as *Total Variation* (TV), since it computes the sum of the absolute variations in the image. Even when other sparsifing transform are used, often a TV penalty is included as well, requiring the image to be sparse by both specific transform and finite-differences at the same time. In the latter case, the problem is modified as follows

$$\begin{aligned} \text{minimize} \quad & ||\Psi X||_1 + \alpha TV(X) \\ \text{so that} \quad & ||\Phi_S X - Y||_2 < \epsilon, \end{aligned} \tag{3.5}$$

where α is a parameter that balances Ψ sparsity with finite-differences sparsity.

Lastly, eq.(3.4) should take into account that instrumental sources of phase error can cause low-order phase variation in MRI. These variations don't provide any physical information, but create artificial variation in the image that makes it more difficult to sparsify. To restrain this unwanted drawback, we introduce a low-resolution estimation of each pixel's phase, obtained by sampled k-space information. Then, this phase information is incorporated by slightly modifying eq.(3.4) so that $||\Phi_S P X - Y||_2 < \epsilon$, where P is a diagonal matrix whose entries give the estimated phase of each pixel.

Development of fast algorithms to solve eq.(3.4) accurately is an increasingly popular research topic. As we know, the iterative reconstruction is more computationally intensive than linear reconstruction. However, some of the methods proposed in the last years show great potential to reduce the overall complexity [1].

Experiments of reconstructions from undersampled accelerated acquistion have shown that the l_1 reconstruction tends to slightly shrink the magnitude of the reconstructed sparse coefficients. Therefore the resulting reconstructed coefficients are often smaller than the original ones. We can conclude that, in CS, images with high contrast can be easier undersampled and reconstructed, since high contrast often results in large distinct sparse coefficients. As a consequence, these coefficients can be recovered even at high acquisition accelerations, while features with lower contrast will be completely submerged in the noise level, i.e. are irrecoverable. We want to stress this last umpteenth CS peculiarity: while, with traditional acquisition methods, increased acceleration usually causes loss of resolution or increase interference (that results in blurring), CS accelerated acquisition looses low-contrast features in the images. Therefore, CS is particularly attractive for applications where images exhibit high resolution, high contrast features and rapid imaging is desiderable.

Chapter 4

Application of Compressed Sensing to MRI

In this section we will describe several potential applications of CS in MRI, all inspired to the work of M. Lustig, D. Donoho, M. Santos, J. Pauly [1], [2]. The main goal of these simulations was to test CS performances in image reconstruction, with increased undersampling, compared to traditional methods, such as low-resolution sampling or linear reconstruction method. The second aim was instead to demonstrate the effectiveness of variable density random undersampling over uniform density random undersampling.

It is worth mentioning that different applications have to face different constraints, imposed by MRI scanning hardware or by patient considerations. Therefore, the three requirements for successful CS reconstructions are differently matched in different applications. The inherent freedom to choose sampling trajectories and sparsifying transform in CS theory allows to offset these constraints.

All the experiments were performed on a 1.5T Signa Excite scanner, while all the CS reconstruction were implemented in Matlab.

4.1 Rapid 3D Angiography

Angiography is the most promising application for CS in MRI. In fact, the problem matches the CS requirements and angiography often needs to cover large field of view with high resolutions and crucial scan times.

Angiography is important for diagnosis of vascular diseases. Since important diagnostic informations are contained in the blood dynamics, often a contrast agent is injected to increase blood signal compared to the background tissue. Angiograms appear to be sparse already to the naked eye and can be further sparsified by finite-differencing. The need for high temporal and spatial resolution strongly encourages undersampling, and CS can improve current strategies reducing the resulting artifacts.

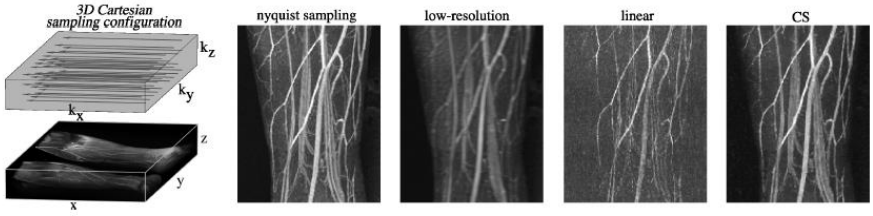

Figure 4.1: 3D contrast enhanced angiography. Left: random undersampling strategy. Right: performance comparison of different strategies. Even with 10-fold undersampling CS reconstruction is trustworthy. It achieves artifacts reduction over linear reconstruction and resolution improvement over low-resolution acquisiton.

The sampling strategy consists in the acquisition of a variable density, pseudo-random subset of equispaced parallel lines of the k-space. In this way undersampling is combined with incoherent acquisition (see also Fig. 4.1). CS is able to significatively accelerate MR angiography, as it recovers most of the information revealed by Nyquist sampling even with 10-fold undersampling (i.e. an acceleration factor of 10). Moreover, nonlinear reconstruction clearly outperforms linear reconstruction, since CS manages to avoid most of the artifacts that result from undersampling.

4.2 Whole Heart Coronary Imaging

MRI is emerging even as a non-invasive alternative to the X-ray coronary angiography, which is the standard test to evaluate coronary artery disease. This kind of diagnostic examination requires a very high resolution imaging, since coronary artery are always in motion. Precise synchronization, motion compensation and scan speed are fundamental requirements to offset the cardiac cycle and the respiratory motion blurring. Moreover, the standard practice to acquire the images during a short breath-held interval defines strict timing, which suggests the CS application.

A multi-slice acquisition method[1] with efficient spiral k-space trajectory, and finite-differences (to sparsify the piece-wise smooth coronary images) results in good CS reconstruction. In fact, undersampling artifacts are suppressed without degrading the image quality.

[1] In 3D MRI imaging it is possible to selectively excite thin slices through the whole volume. This method reduces the data collection to 2D k-space for each slice, thus simplifying the complete volumetric acquisition.

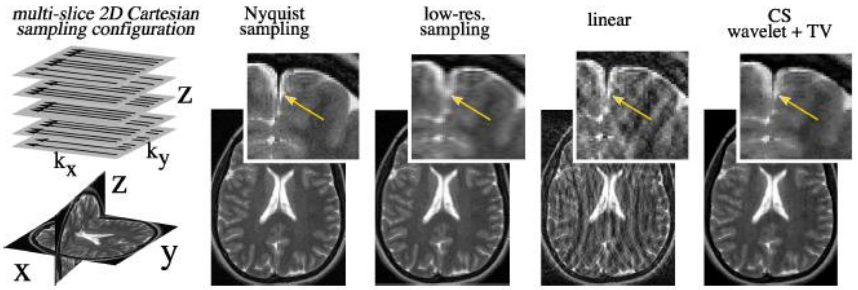

Figure 4.2: 3D brain imaging. Left: random undersampling strategy. Right: CS outperforms linear reconstruction in aliasing artifacts suppression, low-resolution acquisition in definition, and is comparable to a full Nyquist-sampled image even with reduced scan time.

4.3 Brain Imaging

Brain scans are actually the most common clinical application of MRI. Most of these scans are 2D Cartesian multi-slice acquisitions. Exploiting the brain images transform sparsity in the wavelet domain, CS application can reduce collection time while improving the resolution of current imagery.

Undersampling differently and randomly each slice promotes incoherence, obtaining perfect reconstruction of the desired image even with a speedup factor of 2.4. Moreover, if we include a TV penalty in the reconstruction algorithm (eq.(3.5)), the CS strategy perfectly recovers most of the image's details. Even with high undersampling, CS manages to reconstruct the images with quality comparable to a full Nyquist sampled set (see also Fig.4.2).

4.4 Random undersampling strategies in comparison

Lastly, we want to persuade the reader of the deep importance of which random undersampling strategy to use. In order to do so, we compare CS reconstruction performances with two linear schemes: low-resolution (LR) and zero-filling with density compensation (ZF-w/dc). The latter consists of a reconstruction by zero-filling the missing k-space data and applying a density compensation computed from the pdf from which the random samples were drawn. LR, instead, consists of reconstruction from a Nyquist sampled low-resolution acquisition.

The low-resolution reconstruction shows a decrease in resolution as the

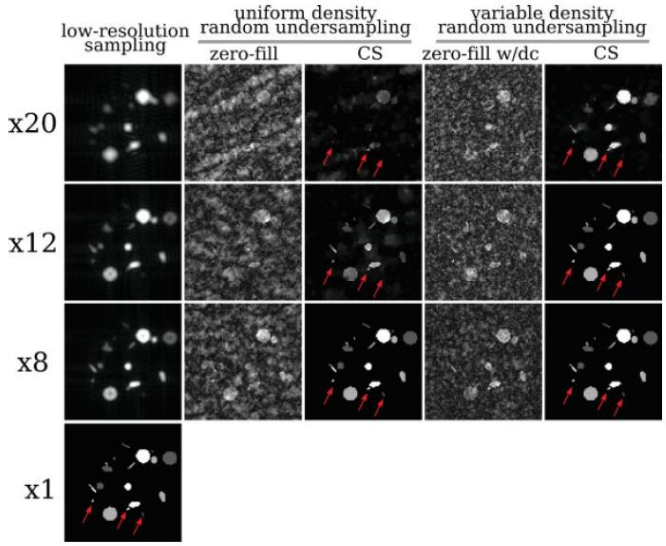

Figure 4.3: Reconstruction artifacts as a function of acceleration. CS with variable density undersampling significantly outperforms other reconstruction methods even with increased acceleration.

data acquisition accelerates, loosing small structures and showing diffused boundaries. The ZF-w/dc reconstruction exhibits a decrease in apparent SNR[2] because of the incoherent undersampling interference, which completely obscures dim and small features. The uniform density undersampling interfrence is significantly larger than in the variable density case.

With an 8-fold acceleration (i.e. acquiring approximatively 3 times more samples than the sparse coefficients) we get exact recovery from both uniform density and variable density undersampling. But, with increased acceleration (as with 12-fold and 20-fold acceleration) only the variable density undersampling gives us exact recovery, while uniform density random undersampling causes most of the features to disappear in the background.

Therefore, once again performance is traded-off with hardware/software complexity. For many applications, in which low accelerations are sufficient to meet the required specific, uniform density random undersampling is to prefer, since it is simpler to implement. For specific applications, in which high-resolution and short timing are fundamental restraints, it is worthwhile using variable-density random undersampling to achieve cutting-edge performances.

[2]Signal-to-noise ratio (often abbreviated SNR) is a measure used in science and engineering that compares the level of a desired signal to the level of background noise. It is defined as the ratio of signal power to the noise power. A ratio higher than 1:1 indicates more signal than noise.

Conclusions

In this thesis, we presented the CS theory and the details of its implementation for rapid MR imaging.

As we saw, Compressed Sensing is an hard mathematical theory that has developed since 2004. It is fascinating for its random approach, its completeness and robusteness, as well as for its wide range of applications in many different fields of science and engineering, that makes it astonishing. This theory, as we discussed deeply, could competely rewrite the Information Theory history. Indeed, it goes beyond the traditional Nyquist-Shannon limitations we encounter in signal acquisition and elaboration. In particular, its implementation in MRI could be fundamental in future Bioengineering applications. We are sure that, when technology, signal processing and optimization algorithms will be available for economical implementation and diffuse distribution, CS based MRI will be preferred to traditional imaging because of the advantages it offers for patients and health care economics.

We focused our attention on the natural fit between CS and MRI, and we reported experimental verification of several implementations for 2D and 3D imaging. Moreover, we underlined how MR images sparsity can be exploited to significantly reduce scan time, or alternatively, improve the resolution of MR imagery. However we have to keep in mind that CS-MRI is still in its infancy. Many crucial issues remain unsettled, including optimizing sampling trajectories, developing improved sparse transforms, studying reconstruction quality in terms of clinical significance, and improving the speed of reconstruction algorithms. Contemporary and future research has to face many tough challenges to turn this approach into great opportunities in improved medical care.

Bibliography

[1] Michael Lustig, David Donoho, John M. Pauly, *Sparse MRI: The Application of Compressed Sensing for Rapid MR Imaging*, *Magn Reson Med*, vol. 58, pp.1182-1195, 2007.

[2] Michael Lustig, David Donoho, Juan M. Santos, John M. Pauly, *Compressed Sensing MRI*, *IEEE Trans. Information Theory*, 2007.

[3] Emmanuel J. Candés, Justin Romberg, Terence Tao, *Robust Uncertainty Principles: Exact Signal Reconstruction From Highly Incomplete Frequency Information*, *IEEE Trans. Information Theory*, 52(2):498-509, 2006.

[4] Emmanuel Candés, Justin Romberg, *Sparsity and Incoherence in Compressive Sampling*, November 2006.

[5] Pablo A. Parrilo, *The Convex Algebraic Geometry of Rank Minimization*, *International Symposium on Mathematical Programming*, August 2009, Chicago.

[6] Justin Romberg, *Imaging via Compressive Sampling*, *IEEE Signal Processing Magazine*, March 2008, Chicago.

[7] Emmanuel Candés, *Compressive Sampling*, *International Congress of Mathematicians*, Madrid, Spain, 2006.

[8] D.L. Donoho, *Compressed Sensing*, *IEEE Trans. Information Theory*, 52(4):1289-1306, April 2006.

[9] G. Wright, *Magnetic Resonance Imaging*, *IEEE Signal Processing Magazine*, vol.14, no.1, pp.56-66, January 1997.

[10] AA. VV., *Risonanza Magnetica Nucleare ed applicazioni in campo medico*, *GDE Utet*, IV edizione, vol.XVII, pp.523-532.

[11] AA. VV., *Risonanza Magnetica Nucleare ed applicazioni*, *Enciclopedia tematica Garzanti*, vol.X, pp.1280-1281, 2006 Garzanti Libri s.p.a., Milano.

[12] Michael Levy (Editor), *Magnetic Resonance Imaging*, *Britannic Illustrated Science Laboratory*, vol. *Technology*, pp.54-55.

[13] Alan V. Oppenheim, Alan S. Willsky, *Signals and Systems*, *Prentice Hall International*, Second Edition, 1997