DEPARTMENT OF INFORMATION ENGINEERING

PhD School in Information Engineering

Curriculum: Information Science and Technology

XXVII cycle

2012 - 2014

Compressive Sensing Applications in Measurement

Theoretical issues, algorithm characterization and implementation

Ph.D. School Head
Prof. Matteo Bertocco

Coordinator
Prof. Carlo Ferrari

Advisor
Prof. Claudio Narduzzi

Ph.D. Candidate
Guglielmo Frigo

ai miei genitori

a mio fratello

Compressive Sensing

Applications in Measurement

Theoretical issues, algorithm characterization and implementation

by
Guglielmo Frigo

Abstract

At its core, signal acquisition is concerned with efficient algorithms and protocols capable to capture and encode the signal information content. For over five decades, the indisputable theoretical benchmark has been represented by the well-known Shannon's sampling theorem, and the corresponding notion of information has been indissolubly related to signal spectral bandwidth.

The contemporary society is founded on almost instantaneous exchange of information, which is mainly conveyed in a digital format. Accordingly, modern communication devices are expected to cope with huge amounts of data, in a typical sequence of steps which comprise acquisition, processing and storage. Despite the continual technological progress, the conventional acquisition protocol has come under mounting pressure and requires a computational effort not related to the actual signal information content.

In recent years, a novel sensing paradigm, also known as Compressive Sensing, briefly CS, is quickly spreading among several branches of Information Theory. It relies on two main principles: signal sparsity and incoherent sampling, and employs them to acquire the signal directly in a condensed form. The sampling rate is related to signal information rate, rather than to signal spectral bandwidth. Given a sparse signal, its information content can be recovered even from what could appear to be an incomplete set of measurements, at the expense of a greater computational effort at reconstruction stage.

My Ph.D. thesis builds on the field of Compressive Sensing and illustrates how sparsity and incoherence properties can be exploited to design efficient sensing

strategies, or to intimately understand the sources of uncertainty that affect measurements. The research activity has dealt with both theoretical and practical issues, inferred from measurement application contexts, ranging from radio frequency communications to synchrophasor estimation and neurological activity investigation. The thesis is organised in four chapters whose key contributions include:

- definition of a general mathematical model for sparse signal acquisition systems, with particular focus on sparsity and incoherence implications;
- characterization of the main algorithmic families for recovering sparse signals from reduced set of measurements, with particular focus on the impact of additive noise;
- implementation and experimental validation of a CS-based algorithm for providing accurate preliminary information and suitably preprocessed data for a vector signal analyser or a cognitive radio application;
- design and characterization of a CS-based super-resolution technique for spectral analysis in the discrete Fourier transform (DFT) domain;
- definition of an overcomplete dictionary which explicitly account for spectral leakage effect;
- insight into the so-called off-the-grid estimation approach, by properly combining CS-based super-resolution and DFT coefficients polar interpolation;
- exploration and analysis of sparsity implications in quasi-stationary operative conditions, emphasizing the importance of time-varying sparse signal models;
- definition of an enhanced spectral content model for spectral analysis applications in dynamic conditions by means of Taylor-Fourier transform (TFT) approaches.

Contents

List of Figures

List of Tables

Chapter 1

Compressive Sensing Theory

1.1 Introduction

In 1948 Claude Elwood Shannon published his well-known sampling theorem [1], destined to influence the future development of modern digital era. Just one year later, he proposed a novel model of communication system [2], focusing the attention on a quantitative definition of information. No more a description from an epistemological point of view, but a thorough mathematical characterization. According to science historians, Shannon's discoveries mark the origin of modern Information Theory. In particular, his sampling theory has inspired countless designs for acquisition systems which need to convert analogue continuous signals into digital quantized data sets. Even after more than 65 years, the canonical approach to signal acquisition relies on Shannon's sampling theorem and the corresponding reconstruction formula. Over the years, alternative sampling protocols have been implemented, in order to overcome some practical limitations, primarily the constraint regarding the sampling rate. Nevertheless, the theoretical framework has not experienced significant variations: the signal information content has been intimately related with its spectral bandwidth.

In 2005 a totally novel sampling theory hits the headlines, promising to revolutionize the way signal information is intended, defined and captured. In 2008 the IEEE Signal Processing Magazine dedicates an entire issue to the so-called Compressive Sensing theory, briefly CS theory, which denies the common wisdom and

experimental custom, embodied by the Shannon-based acquisition protocols, and asserts that any signal could be accurately reconstructed even starting from an incomplete set of random measurements [3]. In other words, the signal is acquired directly in a compressed form, at the expense of a more complicated reconstruction procedure. At first sight, this could appear an unrealistic result, valid only under strict and unattainable assumptions. Conversely, CS theory relies on two theoretical principles, verified in most practical cases: sparsity, which pertains to the signals of interest, and incoherence, which pertains to the sensing modality [3].

In a short while, CS has brought a new slant to canonical approaches and has spread among any branch of Information Theory, from biomedicine to communication theory, from geology to military applications. Without an unequivocal theoretical benchmark, several publications have appeared in the literature, each providing a specific formulation of the sampling protocol or the reconstruction formula. A practitioner, who firstly approaches CS theory, is persuaded to cope with a disorganized mess, where similar results are associated on the base of common denominations, rather than common mathematical properties.

During my Ph.D. course, the first research effort has been concerned with compiling and editing a unified CS theoretical framework. A mathematical model for the sparse signal acquisition has been developed and thoroughly characterized. The role of sparsity and incoherence have been investigated, particularly their influences on the design of efficient CS sensing protocols. The proposed model can be easily tuned according to a priori knowledge or boundary conditions. On the other hand, its general formulation allows to cope even with blind operative conditions, i.e. ignoring any property or plausible prediction about the signal under investigation. In fact, the CS approach to the acquisition problem aims at capturing the sparse signal information content, without trying to understand its inner structure.

In this first chapter, the reader is introduced to the CS scenario and provided with mathematical tools which could be useful for a thorough understanding of this novel sensing theory. First of all, the notions of information, sparsity and incoherence are rigorously defined and briefly discussed by means of clarifying examples. Then, a mathematical model for sparse signal acquisition is presented. In this context, two fundamental approaches are briefly summarized and compared: on one side, the canonical approach based on the aforementioned Shannon's theorem; on

the other side, the CS approach which acquires the signal directly in a compressed form. Particular attention is devoted to the computational efforts required both at sensor and receiver side. Indeed, CS theory allows a drastic reduction of sampling frequency, at the expense of a more demanding reconstruction stage.
Another crucial point is represented by sampling protocol design. From a vector space perspective, two criteria for the most suitable measurement domain selection are introduced. For each one, an annotated dictionary of reconstruction theorems and sufficient conditions is presented. More precisely, the maximum achievable compression ratio is assessed as a function of signal sparsity and noise level.
After the sampling stage focus moves on the reconstruction stage. In fact, CS theory aims at recovering the original information content starting from what could appear to be an incomplete set of measurements. Accordingly, any CS acquisition protocol has to deal with an under-determined system, whose solution requires ad hoc algorithms. Two main algorithmic classes can be identified, relying respectively on proper Lebesgue norm approximation and greedy iterative search. For each class, three typical examples are presented and their pertinence and feasibility in sparse acquisition problems are briefly discussed. Performance analysis focuses on two specific algorithms, one from each class, which will be widely employed in the following chapters of this thesis. More precisely, their behaviour is characterized both in noiseless and noisy conditions, devoting particular attention to solution accuracy and computational complexity. Furthermore, an intuitive tool deduced from probabilistic theory, the phase transition graph allows to predict the success probability of a given acquisition protocol and immediately compare different behaviours in similar conditions. All things considered, a general guideline is proposed: if a blind approach is required, norm approximation represents the most suitable option; whereas greedy search exhibits lower computational complexity at the expense of lower solution accuracy.

In conclusion, this introductory chapter aims at collecting all those contributions, sometimes alternative, sometimes contrasting, and including them into an common framework. In this way, CS theory appears not only as an effective technology, but also as a novel paradigm for capturing sparse signal information. The mathematical model introduced, developed and characterized here, underlies the practical implementations to be discussed in the following chapters. In fact, ow-

ing to its signal-independent formulation, it can easily fit any sparse acquisition scenario. This thorough mathematical framework is expected to become a useful tool not only for properly designing the acquisition protocol, but also for rigorously validating considerations inferred by the results or intimately understanding the sources of uncertainty that affect measurements.

1.2 Lebesgue Norm Definition

The ℓ_p-norm, which proves to be extremely useful in CS theory, is formally introduced here:

Definition 1.2.1. Let $\mathscr{S}$ be a Lebesgue space of dimension n. For $\{p : p \in \mathbb{R}, p \geq 1\}$, the Lebesgue ℓ_p-norm of a vector $\mathbf{x} = (x_1, x_2, \cdots x_n) \in \mathscr{S}$ is defined as follows:

$$\|\mathbf{x}\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{\frac{1}{p}} \tag{1.1}$$

From a linear algebra point of view, the ℓ_p-norm can be interpreted as a measurement of the distance between the space origin and the point indicated by x coordinates. In the following, mostly ℓ_1 and ℓ_2 formulations are employed, corresponding respectively to the Manhattan and the Euclidean distance:

$$\|\mathbf{x}\|_1 = \sum_{i=1}^{n} |x_i| \qquad \|\mathbf{x}\|_2 = \sqrt{\sum_{i=1}^{n} x_i^2} \tag{1.2}$$

Computing the limit for $p \longrightarrow \infty$, the ℓ_∞-norm is defined as the maximum $\mathbf{x}$ value:

$$\|\mathbf{x}\|_\infty = \max(|x_1|, |x_2|, \cdots |x_n|) \tag{1.3}$$

With a slight formal abuse, it is possible to define also the ℓ_0-norm as the number of non-zero entries in $\mathbf{x}$:

$$\|\mathbf{x}\|_0 = \text{card}\,(i \in 1, 2, \cdots n : x_i \neq 0) \tag{1.4}$$

where card denotes the cardinality of the set passed as argument.

1.3 Signal Properties

Given a signal of interest, the acquisition process aims at capturing the actual information content and condensing it into a limited number of samples. In this

context, the effectiveness and appropriateness of an acquisition protocol are evaluated not only in terms of reconstruction accuracy, but also in terms of required computational effort and achieved compression ratio.

Evidently, the signal inner structure strongly affects the protocol performances. In the presence of reliable a priori knowledge, acquisition can be designed expressly to cope with the peculiar operative conditions under investigation. Otherwise, it is necessary to consider some general properties related to any kind of measurable signal, namely spectral bandwidth, innovation rate and sparsity.

1.3.1 Spectral Bandwidth

The modern digital society relies on the well known Shannon's theorem [2], where the sampling rate is required to be at least two times higher than the signal spectral bandwidth.

Definition 1.3.1. A signal s is said to be band-limited if its spectrum does not contain significant components whose frequency is higher than a finite threshold f_{max}, also known as the signal spectral bandwidth or Nyquist's rate.

1.3.2 Rate of Innovation

Shannon's theorem consists in a sufficient condition for perfect recovery of the original information content, but tends to overestimate the required sampling frequency. On the other hand, a suitable parametric signal model could assess the actual innovation rate. In fact, it is reasonable to expect that signal variations depend on a limited set of unknown parameters, i.e. that the signal exhibits a limited amount of degrees of freedom.

Definition 1.3.2. A signal s is said to have a finite innovation rate if it can be described by a finite number of parameters per time unit [4].

In more detail, a signal with finite innovation rate can be expressed as a linear combination of properly scaled and time-shifted functions belonging to a given dictionary $\mathscr{D}$:

$$s(t) = \sum_{i \in \mathbb{Z}} \sum_{k=0}^{K-1} \alpha_{i,k} d_k(t - t_i) \qquad \mathscr{D} = \{d_k, k = 0, 1, \dots K-1\} \tag{1.5}$$

Dictionary functions are assumed to be known a priori or by construction. Accordingly, the only unknown parameters are the scaling factors $\alpha_{i,k}$ and the time shifts t_i. Let us introduce a counting function $\mathscr{C}_s(t_1, t_2)$ which assesses the number of parameter variations over a given time interval $[t_1 \div t_2]$. The signal rate of innovation is thus defined as follows:

$$\rho = \lim_{\tau \to \infty} \frac{1}{\tau} \mathscr{C}_s\left(-\frac{\tau}{2}, \frac{\tau}{2}\right) \tag{1.6}$$

Given a signal with a finite rate of innovation ρ, it is reasonable to expect that a sampling rate equal to ρ is sufficient to correctly recover the original information content. Similarly, given a generic unbiased estimator in the presence of noise, the rate of innovation proves to be a lower bound for the ratio between the average estimation mean-squared-error and the noise variance, independently from the adopted sampling scheme [5].

1.3.3 Sparsity

From a mathematical point of view, the simplest way to model a sampled signal is represented by a vector $\mathbf{s}$ whose dimension n depends on the observation interval length and the adopted sampling scheme.

Definition 1.3.3. A vector $\mathbf{s} \in \mathbb{R}^n$ is said to be sparse if $\|\mathbf{s}\|_0 \ll n$ and, particularly, it is said to be k-sparse if $\|\mathbf{s}\|_0 \leq k \ll n$ [6].

From this two equally significant definitions follow immediately:

Definition 1.3.4. Given a generic vector $\mathbf{s}$, it is always possible to define its best k-sparse approximation $\mathbf{s}_k$ as:

$$\mathbf{s}_k = \min_{\mathbf{u}:\|\mathbf{u}\|_0 \leq k} \|\mathbf{s} - \mathbf{u}\|_p \tag{1.7}$$

The optimization problem returns the same vector $\mathbf{s}$ where the $n-k$ lower coefficients have been set equal to zero.

Definition 1.3.5. A vector $\mathbf{s}$ is said to be compressible if there exist two positive constants, namely $c > 0$ and $r > 1$, which verify the following inequality:

$$\|\mathbf{s} - \mathbf{s}_k\|_p \leq c k^{-r} \tag{1.8}$$

As long as k decreases, condition (1.8) becomes more and more strict and hard to satisfy. Indeed, the lower is the number of non-zero coefficients, the larger is the approximation error with respect to the original signal representation. This peculiar phenomenon shows why it is so difficult to achieve high compression ratios without compromising reconstruction accuracy.

It is worth noticing that in equations (1.7) and (1.8) the order p of Lebesgue norm is unspecified. In other words, both definitions hold independently from the adopted distance formulation. In accordance with the peculiar operative conditions, it is reasonable to tune this parameter in order to minimize the injection of spurious contributions or to avoid sparsity level misinterpretations.

1.4 A Mathematical Model for Sparse Signal Acquisition

A generic sparse acquisition problem could be modelled by the following linear matrix equation:

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{w} \tag{1.9}$$

where:

- $\mathbf{y}$ is the acquired samples vector of dimension $[m \times 1]$;
- $\mathbf{A}$ is the measurement matrix of dimension $[m \times n]$;
- $\mathbf{x}$ is the sparse signal representation of dimension $[n \times 1]$;
- $\mathbf{w}$ is the additive noise contribution of dimension $[n \times 1]$.

1.4.1 Sparse Signal Acquisition Protocols

The sparse acquisition problem (1.9) can be dealt with by two approaches. On the one side, the conventional protocol which implements the sampling theory introduced by Claude Shannon in 1948 [1]. On the other side, the compressive sensing protocol that exploits signal sparsity to accurately recover the original information content from a reduced set of measurements.

Shannon-Based Acquisition Protocol

After World War II, Claude Shannon worked at Bell Laboratories, focusing his studies on improving telegraphic transmission quality in the presence of noisy disturbances or spurious contributions. Shannon was particularly interested in the acquisition process and studied the relationship between sampling rate and distortion level in the recovered signal. For over five decades, his sampling theorem has been universally considered the benchmark for any signal sensing theory [2]:

Theorem 1.4.1. *If a signal* **s** *contains no frequencies higher than* f_{max} *Hz, it is completely determined by giving its ordinates at a series of points spaced* $t_s = \frac{1}{f_{max}}$ *seconds apart.*

The theorem authorship is actually a moot point. Indeed, similar formulations were published in a mathematical context under the name of Whittaker in 1929 [7], and Kotel'nikov in 1933 [8]. Shannon himself did never claim the result as a personal discovery, but rather as a common criterion gathered from experimental practice. It should be noticed that, also at Bell Laboratories, Harry Nyquist postulated an equivalent sufficient condition and showed its application to early telegraphic transmissions in 1928 [9]. Shannon always considered himself as an humble heir of this eminent tradition and decided to name the criterion threshold after his trailblazing predecessor. Henceforth, Theorem 1.4.1 is known as Shannon's theorem and f_{max} is called Nyquist's rate, ensuring eternal remembrance to both their authors.

The acquisition process consists of two main stages: sampling and reconstruction. Accordingly, after defining the anti-aliasing rate, Shannon provided also the corresponding reconstruction formula:

$$f(t) = \sum_{k \in \mathbb{Z}} f(kT)\, sinc\left(\frac{x}{T} - k\right) \tag{1.10}$$

where the acquired samples are interpolated by means of Dirichlet kernel functions. Once more, this is not an $\alpha\pi\alpha\xi\ \lambda\epsilon\gamma o\mu\epsilon\nu o\nu$: in 1915, Whittaker presented a totally equivalent formulation, claimed as Cardinal Series Expansion [10]; moreover, similar expressions could date back to further preceding publications [11].

Modern digital technology still relies on the sufficient condition stated by Theorem 1.4.1. Unfortunately, Shannon's formulation exhibits some significant discrepancies from the actual operative conditions. On one hand, the theorem applies only

on perfectly band-limited signals. On the other hand, the Nyquist constraint on sampling rate depends merely on the signal bandwidth, not considering how the information is distributed over the acquisition domain. Evidently, these problems affect the efficiency of the conventional protocol: a larger sampling rate corresponds to a greater amount of data to process, transfer or store. On the other hand, the conventional protocol is appreciated due to its simple and immediate implementation: in accordance with Occam's *lex parsimoniae*, in most cases, this straightforward solution proves to be the best compromise between costs and performances.

The conventional protocol relies on the time - frequency domain dualism. In this sense, its theoretical content descends immediately from common sense and experimental practice. The larger is the signal rate of change, the larger is the sampling rate required to reproduce it without distortions or omissions. Even from a mathematical point of view, the theoretical framework is intuitive and consists merely of a proper application of Fourier transform properties.

In the formulation considered in this thesis, the problem deals with a finite sequence of N consecutive and equally-spaced samples, acquired in the time domain at a uniform sampling rate equal to $f_s = 1/t_s$ Hz. In the frequency domain, the discrete Fourier transform of the sequence presents a periodic repetition of the signal spectral content centred around integer multiples of f_s. Accordingly, if f_s is compliant with Nyquist's constraint, i.e. $f_s \geq 2 \cdot f_{max}$, no mutual interference takes place and the signal content can be easily retrieved. On the contrary, if f_s is not correctly related to the signal spectral bandwidth, i.e. $f_s < 2 \cdot f_{max}$, the higher frequency components are partially superposed, causing the well-known *aliasing* effect, as shown in Fig. 1.1.

Block scheme: oversampling and then compression Shannon's theorem does not consider any signal property except for spectral bandwidth. As a consequence, the same acquisition scheme is employed in any operative condition, independently from the a priori information at disposal. For instance, even if the signal is known to be sparse in the frequency domain, the sampling frequency is set to be compliant with Nyquist's constraint. The sparsity assumption provides the only expedient to afterwards condense the signal in a compressed format, suitable for effective storage or processing applications.

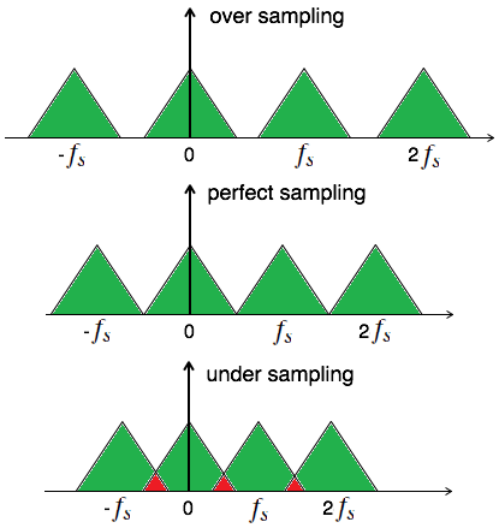


Figure 1.1: Consequences of sampling process in the DFT domain with reference to Nyquist's rate constraint. From top to down, respectively: over sampling, perfect sampling, and under sampling.

The block diagram in Fig. 1.2 summarizes the main acquisition stages, highlighting the respective functions of sensor and receiver side.
More precisely, at sensor side, the input vector $\mathbf{s}$ of dimension $[n \times 1]$ passes through a coherent sampling process at Nyquist's rate, resulting in $m \geq n$ samples equally spaced in time. The modern ADC technology is able to cope with sampling frequencies exceeding the order of GHz, and accordingly to account for very fast signal variations. On the other hand, it is reasonable to expect that the overall amount of acquired data is not scaled to the actual information content and often requires an ad hoc compression technique in order to be efficiently stored. In accordance with the peculiar operative condition, a proper compression stage reduces the measurement vector dimension up to the desired sparsity level $k \ll m$.
At receiver side, the inverse procedure is carried out. Before any further processing, the stored data are decompressed, returning to their original dimension $[n \times 1]$.

Final remarks: a universal but inefficient protocol Shannon's sufficient condition (1.4.1) ensures perfect recovery of the original information content. Furthermore, signal reconstruction requires only the straightforward computation of the cardinal

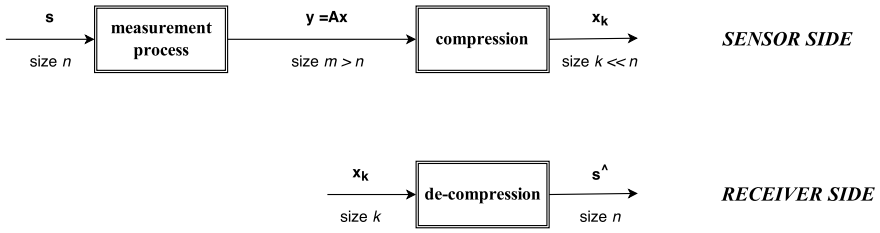


Figure 1.2: Traditional acquisition protocol block diagram.

series formula (1.10).

Nevertheless, it is worth observing that these protocol peculiarities do not provide any guarantee of performance enhancement. Computational complexity has to be assessed with respect to the actual amount of data and the employed compression algorithm.

As a matter of principle, the traditional protocol suggests to acquire the signal directly in the domain where it exhibits its sparsity properties. To this end, oversampling is not despised, as it makes more significant the ratio between the number of respectively, negligible and significant coefficients. For instance, in spectral analysis applications, oversampling is intentionally employed in order to reduce scalloping loss and mutual interference among superposed components.

CS-Based Acquisition Protocol

The actual innovation rate of a continuous time signal may be much smaller than suggested by its bandwidth. Similarly, a a discrete-time signal may depend on a number of degrees of freedom comparably much smaller than its finite length. More precisely, any signal, expressed in a proper domain, could be represented by a reduced set of parameters, i.e. can be called sparse.

Sparsity examples In the time domain, a pure sinusoidal tone exhibits its typical periodic trend, swinging from positive to negative values. Conversely, in the frequency domain, a single Dirac function is sufficient to convey the entire information content, as shown in Fig. 1.3.

A generic image can be seen as a two-dimensional matrix, whose entries repre-

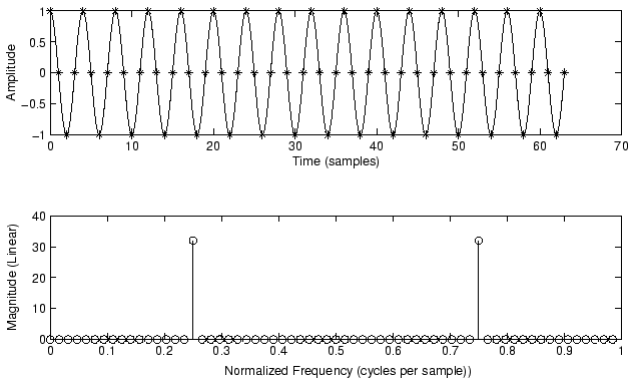


Figure 1.3: Given a cosine function at ω_f Hz, 64 samples are acquired at a sampling rate $\omega_s = 4 \cdot \omega_f$. The signal representation in both time (top) and frequency (bottom) domain are provided.

Figure 1.4: On the left, the well-known Lichtenstein image, considered a severe test bench for processing techniques. On the right, the result of a two level wavelet transform, comparable to the JPEG2000 compression format.

sent the gray levels of the corresponding pixels. A proper wavelet transform allows to detect which pixels account for smooth or constant regions and which pixels account for step or transient regions. In this way, the number of significant coefficients is drastically reduced, i.e. the image is sparse in the wavelet transform domain, as shown in Fig. 1.4.

A chemical compound mass spectrum can be estimated by means of induced electron ionization. For a given set of mass-to-charge ratio (m/Q) values, the abundance of ions is measured and the overall percentage is computed. Typically, the resulting trace is affected by large broadband noise with a signal-to-noise ratio (SNR)

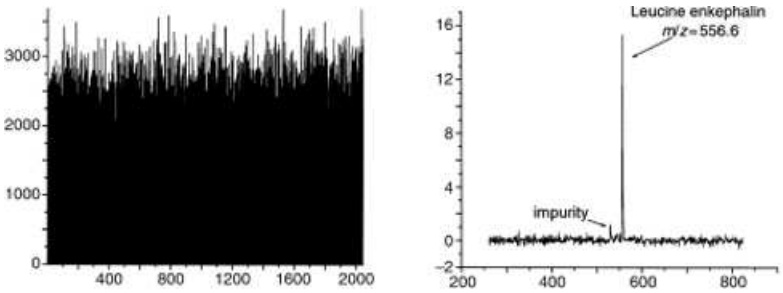


Figure 1.5: In mass spectroscopy context, comparison between a raw trace, on the left, and a resolution enhanced version obtained by Hadamard tomography technique, on the right.

at least equal to 1. Hadamard tomography decomposes the input signal into a linear combination of Walsh functions. Combination coefficients are distributed so that informative components are associated to larger weights, while noisy components are associated to smaller weights. In this way, the sparse structure of the acquired signal can be revealed, as shown in Fig. 1.5.

In order to develop a high-resolution map, LiDAR remote sensing technology measures distances by illuminating a target with a laser and analysing the reflected light. As shown in Fig. 1.6, the backscattered waveform accounts for obstacle heights. Over a wide region, the number of significant obstacles is expected to be reasonably small if compared with the area under investigation.

Finally, also radio-frequency (RF) acquisitions exhibit a double level of sparsity. In accordance to Nyquist's constraint, the RF spectrum should be acquired at a sampling rate on the order of some GHz. In order to avoid interference and superposition, each transmission is associated to a specific channel. Not necessarily, all transmitters are operating at the same time. In the frequency domain, the signal is said to be block-sparse [13], because the most significant coefficients condensate around the active channel carrier frequencies. It is reasonable to expect that, as a whole, significant blocks take up a reasonably small portion of the entire RF spectrum, as shown in Fig. 1.7. Subsequent acquisitions of the same channel may prove to be similarly sparse in the time domain. In fact, it is reasonable to expect a significant value only when the transmitter is operating, otherwise the acquired trace consists only of noise.

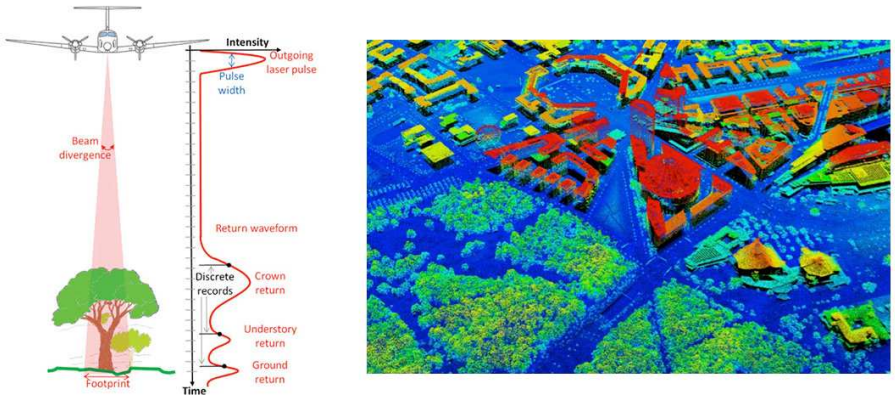


Figure 1.6: On the left, LiDAR acquisition protocol is illustrated; on the right, an example of LiDAR point elevation data rendered in 3D (Potsdamer Platz in Berlin, Germany) [12].

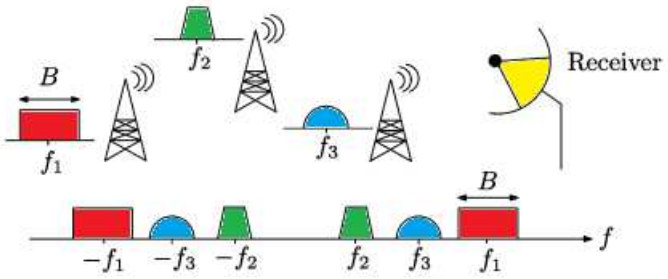


Figure 1.7: Multi-band signal model consisting of three different contributions, each located in a specific transmission channel [13].

From Shannon to CS theory Sparsity evidence has inspired many sub-Nyquist techniques, i.e. acquisition protocols adopting an actual sampling rate much lower than suggested by Nyquist's constraint. Mostly, these are ad hoc solutions, explicitly designed to exploit or cope with the operative conditions peculiar to a single class of signals. Shannon's general perspective towards the acquisition problem is replaced by a disjoint approach, consisting of specific and uncorrelated practical solutions. While the traditional protocol relies firmly on Shannon's theorem, the sub-Nyquist protocols lack a unified theoretical framework.

In this context the theory of compressive sensing, briefly CS, takes its origin and develops. It is not only a sampling protocol, but actually an acquisition paradigm that offers a totally innovative perspective: the signal can be acquired directly in

a compressed form, shifting most of the computation effort to the reconstruction stage. In principle, it is possible to recover the original information content even from what could appear as an incomplete set of measurements. To make this possible, CS relies on two principles: sparsity, which pertains to the signals of interest, and incoherence, which pertains to the sensing modality [3].

In a suitable basis Ψ, a sparse signal is supposed to exhibit a more concise representation than suggested by Nyquist's constraint. For instance, a discrete-time signal could be fully described by a number of degrees of freedom much smaller than its finite length. Similarly, a continuous-time signal could vary much more slowly than suggested by its finite spectral bandwidth. A novel concept, also known as information rate, goes beyond the definitions of Nyquist's and innovation rate: the acquisition protocol should focus on information content variations, rather than on signal variations. Independently from its representation in the observed domain, it is important to sample and characterize the signal whenever the conveyed information undergoes a significant change. For instance, according to CS theory, a stationary signal, no matter how large its Nyquist's or innovation rate is, can be recovered from a severely under-sampled acquisition.

From a linear algebra point of view, the incoherence property can be seen as a theoretical generalization of the well known duality between time and frequency domain, or similarly between Dirac and Dirichlet basis functions. A thorough mathematical characterization is beyond the scope of this introductory section and will be rigorously addressed later. However, it is worth noting here that a signal should be sampled in a basis Φ incoherent with Ψ, i.e. where its information content is not condensed into few coefficients, but almost uniformly spread out over the entire domain. In this way, no matter which is the sampling rate or scheme, any sample is reasonably expected to contain a small portion of the original information content. At first sight, this "incoherent" sampling strategy could appear to be in contrast with the sparsity assumption. Conversely, in practice, this proves to be the best suited solution, not requiring any a priori knowledge about the localization of significant coefficients. Challenging common sense and canonical theory intuition, CS theory could be summarized by the following motto.

Given a generic signal, a reduced set of samples, acquired at a sampling rate lower than Nyquist's and according to a random scheme, is sufficient to correctly recover

the original information content. Keep calm and randomize!

Block scheme: an under-determined problem CS theory completely revolutionizes the perspective towards the acquisition problem. In the traditional approach, the measurement process depends exclusively on signal spectral bandwidth: in the absence of detailed a priori knowledge, it is necessary to employ an anti-aliasing filtering stage, which may erase some information contributions. In the CS approach, instead, the underlying theoretical model relies only on the sparsity assumption, which almost any signal verifies in a proper domain. Moreover, in the matter of sampling scheme, no prior constraints have to be satisfied if carried out in a sufficiently incoherent domain.

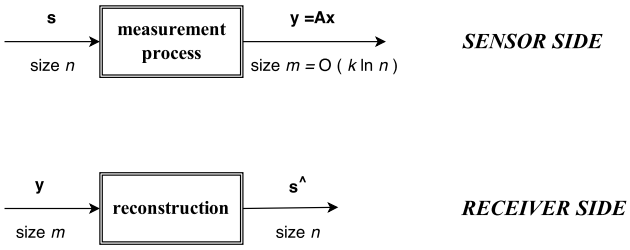


Figure 1.8: Compressive sensing acquisition protocol block diagram.

The block diagram in Fig. 1.8 summarizes the main acquisition stages, highlighting the respective functions of sensor and receiver side.
More precisely, at sensor side, the traditional sequence of measurement and compression is replaced by a single stage where the information content is directly acquired in a compressed form. Independently from its Nyquist's or innovation rate, the sufficient number of samples to be acquired is given by:

$$m = O(k \ln n) \tag{1.11}$$

where n is the input signal size and k is the expected or approximated sparsity level. In particular, the compression ratio provided by the CS protocol behaves as a logarithmic function of the number of samples. Indeed, the traditional compression stage could achieve much better performances, but it should be noticed that the CS protocol condenses acquisition and compression into a single stage and adopts

a totally signal-independent approach. On the one hand, compressed acquisition avoids the computationally expensive sequence of oversampling and compression: according to the traditional protocol, a huge amount of samples have to be acquired in order only to enhance the sparsity property and immediately disregarded as negligible coefficient, resulting in a worthless waste of computational resources. On the other hand, the compression stage does not require any a priori knowledge about how the information content is conveyed or distributed over the sparsity domain, resulting in a much better portability of the CS protocol to any branch of Information Theory.
At receiver side, the de-compression stage is replaced by a more demanding reconstruction stage. The acquired set of measurements is not consistent with the desired solution size and the acquisition inverse problem appears to be under-determined and not to provide a unique solution. Nevertheless, properly exploiting the sparsity assumption, it is possible to determine the most suitable sparse approximation of the input signal, at the expense of a larger computational complexity. The mathematical details of this reconstruction stage are beyond the scope of this introductory section and will be thoroughly addressed later in the following. However, it is worth noticing that the computational costs at receiver stage are largely balanced or even exceeded by the compression ratios achieved at sensor side.

Final remarks: a novel unified sub-Nyquist sampling theory According to CS theory, it is possible to design efficient acquisition protocols that capture the useful information content embedded in a sparse signal and condense it into a small amount of data. These sensing paradigms prove to be non-adaptive, rely on an undisputed sparsity assumption, and simply correlate the signal with a reduced set of basis functions which are only required to be sufficiently incoherent with the sparsity domain. What is most remarkable about this novel sampling theory is that CS-based sensors are capable to efficiently acquire the underlying information content without trying to reflect the inherent structure of the signal [3].

1.5 Most Suitable Acquisition Domain Selection

The acquisition model (1.9) is pretty generic and could be easily adapted to many different operative conditions. What is most remarkable about its mathematical for-

mulation is that the sampling process is modelled by a linear operator, namely the rectangular matrix **A**. In this context, it is worth noticing that the matrix elements do not depend on the signal under investigation **s** nor on its sparse representation **x**. More precisely, the matrix columns determine in which domain and according to which scheme the signal is going to be sampled. For this reason, **A** is also often called measurement matrix, because its knowledge yields the complete representation of the measurement process.

A proper measurement matrix could improve the acquisition performances significantly, in terms not only of reconstruction accuracy, but also compression ratio. In CS literature, many criteria have been developed in order to design **A** in accordance with the peculiar operative conditions and available a priori knowledge. In the following subsections, some of these criteria are briefly introduced and summarized, highlighting the theoretical generality, on one side, and the actual feasibility in practice, on the other side.

1.5.1 Incoherence

Given a generic signal **s**, knowing a priori the domain where it exhibits a sparse behaviour **x**, the linear model (1.9) can be decomposed as follows:

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{w} = \Phi\Psi\mathbf{x} + \mathbf{w} = \Phi\mathbf{s} + \mathbf{w} \tag{1.12}$$

where:

- Φ of dimension $[m \times n]$ is called sensing matrix, because its rows ϕ_i determine the sampling scheme:

$$y_i = \langle \mathbf{s}, \phi_\mathbf{i} \rangle \quad i = 1, 2, \dots m \tag{1.13}$$

- Ψ of dimension $[n \times n]$ is called sparsifying matrix, because its columns $\psi_\mathbf{j}$ form a basis for the domain in which **s** is sparse:

$$x_i = \langle \mathbf{s}, \psi_\mathbf{j} \rangle \quad j = 1, 2, \dots n \tag{1.14}$$

From a linear algebra point of view, Φ and Ψ can be seen as bases spanning specific vector subspaces over which the signal **s** has to be projected. Similarly, the inner product formulations in (1.13) and (1.14) can be seen as realizations of the well-known Hilbert space projection theorem [14].

The sparsifying matrix Ψ is a fixed signal feature; no degree of freedom is possible. On the contrary, the sensing matrix Φ can be tuned according to specific application requirements or user desiderata. Thus, a question plays a crucial role: how should the sampling scheme be implemented?

- uniform and coherent, as suggested by the traditional acquisition protocol?
- in accordance with a given statistical distribution, e.g. a Gaussian or a Bernoulli distribution?
- totally random, letting a pseudo-random generator to decide when and how a sample has to be acquired?

Many different alternatives are possible, but no one is directly related to sparsity.

According to CS theory, the most preferable option is represented by a sensing matrix Φ incoherent with the sparsifying matrix Ψ. From a mathematical point of view, a quantitative index for coherence could be defined as follows [3]:

$$\mu(\Phi,\Psi) = \sqrt{n}\max_{i,j}|\langle\phi_{\mathbf{i}},\psi_{\mathbf{j}}\rangle| \qquad i = 1,2,\dots m \quad j = 1,2,\dots n \tag{1.15}$$

This index measures the maximum correlation between any two vectors of Φ and Ψ, i.e. it is a sort of measure of similarity between Φ and Ψ. Given two orthonormal bases F and P, μ ranges:

- from 1, which corresponds to minimal coherence condition, optimal for CS-based applications;
- to $\sqrt{n}$, which corresponds to maximal coherence, unsuitable for CS-based applications.

The upper bound descends directly from the inner product properties: given two vectors of unitary norm, their product exhibits a module not larger than 1. The lower bound, instead, descends from the Parseval relationship:

$$\sum_{i=1}^{m}|\langle\phi_{\mathbf{i}},\psi_{\mathbf{j}}\rangle|^2 = \|\psi_{\mathbf{j}}\|_2^2 = 1 \qquad j = 1,2,\dots n \tag{1.16}$$

An Example of Incoherent Basis

In Fig. 1.9 a typical example of incoherent basis is presented, namely a sinusoidal tone is uniformly sampled both in the time and in the frequency domain:

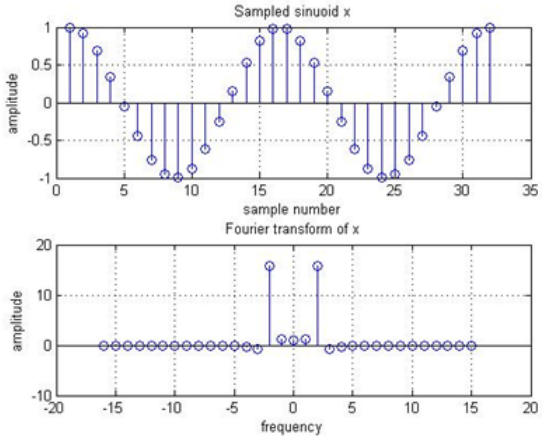


Figure 1.9: Sinusoidal tone uniformly sampled in the time (top) and frequency (bottom) domain.

- Φ is the canonical basis, such that its column vectors consist of suitably shifted Dirac functions around the integer multiples of the sampling period:

$$\phi_{\mathbf{i}}(t) = \delta(t - t_i) \tag{1.17}$$

- Ψ is the Discrete Fourier Transform (DFT) basis, such that its column vectors consist of DFT kernels suitably centered around the bins of the frequency axis:

$$\psi_{\mathbf{i}}(t) = \frac{1}{\sqrt{n}} e^{\frac{i2\pi jt}{n}} \tag{1.18}$$

Owing to the ideal duality between time and frequency domain, maximal incoherence is achieved, i.e. $\mu(\Phi, \Psi) = 1$. This condition between pulses and sinusoids is universally verified and holds independently from the actually dimensions of the considered space vector.

In practice, this corresponds to sampling in the time domain a signal that is sparse in the frequency domain. Is this the best way to exploit sparsity?

At first sight, intuition would suggest to sample in the frequency domain where the entire information is contained in only two coefficients. However, it is required to know a priori their location, otherwise a uniform sampling is necessary, resulting in lots of zero coefficients and the well-known spectral estimation problems (e.g., spectral leakage, scalloping loss, short and long range interference, etc.).
Conversely, CS theory suggests to sample in the time domain where the same infor-

mation content is almost uniformly spread out among all the representation coefficients. The greater is the incoherence between Φ and Ψ, the greater is the probability that any sample tells something about the signal.

Random Solution as Effective Expedient

In the literature, some frequent examples of incoherent basis couples can be found, e.g.: canonical / DFT, noiselet / Haar, wavelets / higher dimensional related functions (a few of them are listed in [15]). Noiselets prove to be particularly useful: in fact, they are maximally incoherent with both spikes and Fourier basis complex exponentials. Their increasing implementation in many different branches of Information Theory is easily motivated:

1. incoherence with most common sparse representation of image or whatever structured data type;

2. availability of fast algorithms for transform coefficient computation: state of the art solutions run in $O(n)$ time, and just like the Fourier transform, the noiselet matrix does not need to be stored to be applied to a vector.

However, sometimes the sparsifying matrix Ψ does not provide an obvious sensing matrix Φ. For instance, given a signal sparse in Hough transform (digital image processing) domain or in $Y-\Delta$ transform (electrical circuit analysis) domain, where should it be sampled?

An unexpected solution for this class of problems is represented by the statistical properties of random matrices. Given sparsifying basis Ψ, select an orthobasis Φ uniformly at random, e.g. by orthonormalizing n vectors sampled independently and uniformly on the unit sphere. Then, with high probability, the coherence between Φ and Ψ is about $\sqrt{2\log n}$ [3]. This is not the lower limit, but still offers good performances.

By extension, any random sampling scheme with independent identically distributed (i.i.d.) entries, e.g., Gaussian or ± 1 binary entries, is expected to be highly incoherent with any fixed sparsifying domain. It should be noticed that this result conceals a pretty surprising implication [3]: if sampling stage is required to be incoherent, then efficient sensing devices should acquire correlations with random waveforms, just like the white noise!

Incoherence-Based Criteria and Sufficient Conditions

Let us consider a signal $\mathbf{s} \in \mathbb{R}^n$ which admits a k-sparse representation $\mathbf{x}$ in the proper sparsifying basis Ψ.

Theorem 1.5.1. *Given a sensing basis Φ, select m measurements uniformly at random. If the measurement number m satisfies the following inequality for some positive constant C:*

$$m \geq C \cdot \mu^2(\Phi, \Psi) \cdot k \cdot \log n \tag{1.19}$$

then the acquisition problem 1.9 can be inverted and provides the exact unique solution with overwhelming probability [16].

More precisely, the probability of success is shown to exceed $1 - \delta$ if:

$$m \geq C \cdot \mu^2(\Phi, \Psi) \cdot k \cdot \log\left(\frac{n}{\delta}\right) \tag{1.20}$$

and the result is guaranteed for all k-sparse representations of $\mathbf{s}$.

Another interesting result investigates the relationship between the coherence index μ and the highest achievable compression ratio and is formalized in the following theorem.

Theorem 1.5.2. *Given the measurement matrix $\mathbf{A} = \Phi\Psi$, a mutual coherence index $\mu(\mathbf{A})$ is defined as the maximum product of two different columns $\mathbf{a}_i$ and $\mathbf{a}_j$:*

$$\mu(\mathbf{A}) = \max \mathbf{a}_i^T \mathbf{a}_j \,, \qquad i \neq j \tag{1.21}$$

Then, the sparsity level of the acquisition problem solution is lower bounded by the following mutual coherence function:

$$\|\hat{\mathbf{x}}\|_0 \geq \lfloor \frac{\mu(\mathbf{A}) + 1}{2\mu(\mathbf{A})} \rfloor \tag{1.22}$$

Three important implications descend from theorems 1.5.1 and 1.5.2 and are worth to be highlighted:

- the sufficient condition depends on the coherence index according to a quadratic function behaviour: the smaller is the coherence, the fewer samples are needed;
- even starting from an incomplete set of measurements, no significant information loss is expected;

- the reconstruction stage requires ad hoc algorithms, which are capable to cope with under-determined systems, and do not require any a priori knowledge with the exception of the desired or expected sparsity level.

1.5.2 Restricted Isometry Property

The incoherence criterion relies on the assumption that the signal $\mathbf{s}$ admits a k-sparse representation $\mathbf{x}$ in a given sparsifying domain Ψ. However, in practice this knowledge is unavailable a priori. The aforementioned decomposition into sensing and sparsifying matrix is no more feasible and the measurement matrix $\mathbf{A}$ has to be taken into account as a whole:

$$\mathbf{y} = \Phi\Psi\mathbf{x} + \mathbf{w} = \mathbf{A}\mathbf{x} + \mathbf{w} \tag{1.23}$$

In this context, a novel criterion, perfectly blind and independent from the inherent structure of the signal under investigation, is introduced. More precisely, measurement matrix $\mathbf{A}$ is required to satisfied the Restricted Isometry Property, briefly RIP, which can be defined as follows [3]:

Definition 1.5.1. Given a positive integer k, the isometry constant δ_k of the measurement matrix $\mathbf{A}$ is the minimum number which satisfies the following equation for any k–sparse vector $\mathbf{x}$:

$$(1-\delta_k)\|\mathbf{x}\|_2^2 \leq \|\mathbf{A}\mathbf{x}\|_2^2 \leq (1+\delta_k)\|\mathbf{x}\|_2^2 \tag{1.24}$$

As a consequence, A is said to exhibit the RIP of order k if $\delta_k < 1$.

In mathematics, the term isometry is commonly related to any function, transformation or, more generally, map which preserves the distance measure even passing from a vector space to another. Similarly, a measurement matrix $\mathbf{A}$ which satisfies the RIP of order k ensures to preserve the Euclidean length of any k-sparse signal $\mathbf{x}$. In other words, there exists no k-sparse vector which belongs to the null space of $\mathbf{A}$, a necessary condition to aim at a correct reconstruction of the original sparse representation.

The adjective restricted gives the property a peculiar connotation which may appear clearer in the following interpretation. If a matrix $\mathbf{A}$ of dimensions $m \times n$ with $m < n$ satisfies the RIP of order k, then all the vector subspaces spanned by

subsets of k columns taken from $\mathbf{A}$ are almost orthogonal. The exact orthogonality is feasible only in the case of square matrices, whereas in the present case the linear independence between two different subsets is not guaranteed because the number of columns exceeds the number of rows.

The relationship between RIP and CS protocol is immediately evident in the following example. Let $\mathbf{A}$ be the measurement matrix employed to acquire k-sparse signals from a given raw data flow. Suppose that $\mathbf{A}$ satisfies the RIP of order $2k$, i.e. twice the expected sparsity level, with an isometry constant δ_{2k} sufficiently lower than 1. Then, all pairwise distances between k-sparse signals are expected to be well preserved in the measurement space, with a relative approximation error not larger than $\pm\delta_{2k}$. It is equivalent to say that the following condition:

$$(1-\delta_{2k})\|\mathbf{x_1}-\mathbf{x_2}\|_2^2 \leq \|\mathbf{Ax_1}-\mathbf{Ax_2}\|_2^2 \leq (1+\delta_{2k})\|\mathbf{x_1}-\mathbf{x_2}\|_2^2 \tag{1.25}$$

holds for any pair of k-sparse vectors (x_1, x_2).

Once the RIP is verified for a given sparsity level, the following theorem allows to easily extend the property also to larger cardinality sets.

Theorem 1.5.3. *Suppose that* $\mathbf{A}$ *satisfies the RIP of order* k *with isometry constant* δ_k. *Let* c *be a positive integer. Then* $\mathbf{A}$ *satisfies the RIP of order* $k' = c\lfloor\frac{k}{2}\rfloor$ *with isometry constant* $\delta_{k'} < c \cdot \delta_k$.

The statement could appear to be trivial for $c = 1$ or 2, but for $c \geq 3$ and $k \geq 4$ it actually allows to extend the RIP (and thus the capability to invert the corresponding sparse acquisition system) to higher order formulations. On the other hand, δ_k is required to be sufficiently small in order for the resulting bound to be useful. In fact, theorem 1.5.3 holds only if $\delta_k < \frac{1}{c}$ and accordingly $\delta_{k'} < 1$.

In conclusion, two observations have to be highlighted:

- the Restricted Isometry Property does not concern the composition of rows or columns of $\mathbf{A}$: no constraints are introduced with reference to sensing domain or sampling scheme;

- the Restricted Isometry Property can be interpreted as a sort of measure of the numerical conditioning of $\mathbf{A}$: the distortion of input signal distances does not exceed a fixed threshold, related to the expected sparsity level.

RIP-based Criteria and Theorems

In order to understand the RIP implications, let us return to the aforementioned sparse acquisition problem:

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{w} \tag{1.26}$$

- AIM: to recover the k-sparse vector $\mathbf{x}$ which better approximates the measurement vector $\mathbf{y}$;
- PROBLEM: the measurement matrix $\mathbf{A}$ has m rows and n columns, with $m \ll n$, thus the problem is highly under-determined;
- SOLUTION: if $\mathbf{A}$ satisfies the RIP of order $2k$, i.e. twice the expected sparsity level, the following three theorems can be employed to achieve a reliable solution $\hat{\mathbf{x}}$.

Theorem 1.5.4. *If* $\mathbf{A}$ *satisfies the RIP of order* $2k$*, it is always possible to recover the unique k–sparse solution of the linear system* $\mathbf{y} = \mathbf{A}\mathbf{x}$ *[3].*

It should be noticed that this theorem guarantees the existence and unicity of the solution, but does not specify how it could be estimated. Furthermore, theorem 1.5.4 is valid only in noiseless conditions. Nevertheless, the fact that the RIP is a sort of sufficient condition for the feasibility of highly under-determined systems should not be underrated.

Theorem 1.5.5. *If* $\mathbf{A}$ *satisfies the RIP of order* $2k$ *with an isometry constant* $\delta_{2k} < \sqrt{2}-1$*, there always exist two positive constants* C_1 *and* C_2 *such that the following inequality is verified:*

$$\|\hat{\mathbf{x}} - \mathbf{x}\|_2 \leq C_0 \frac{\|\mathbf{x} - \mathbf{x_k}\|_1}{\sqrt{k}} + C_1 \varepsilon \tag{1.27}$$

where $\hat{\mathbf{x}}$ *is the recovered solution and* ε *quantifies the variability related to the additive noise term* $\mathbf{w}$ *[3].*

This theorem shows that the reconstruction error is limited by the sum of two terms. The first one quantifies how likely it is to obtain a k-sparse approximation $\mathbf{x_k}$ of the actual sparse signal representation $\mathbf{x}$, whose sparsity level is supposed to be unknown. The second term assesses how much the recovered signal $\hat{\mathbf{x}}$ is affected by additive noise. It is worth noticing that the first term is related to the adopted signal model, precisely to the expected sparsity level, whereas the second one is related

to a perturbation of the measurement system, which is independent from the user choices.

Just to give an idea of the expected weight of each term in summation (1.27), if $\delta_{2k}0.25$ it is sufficient to take $C_1 < 5.5$ and $C_2 < 6$. In other words, model inappropriateness and noisy disturbance provide nearly comparable contributions for a sufficiently small isometry constant. As long as δ_{2k} increases, conversely, also the model discrepancy is expected to grow, while the noise effect is expected to remain unvaried.

Theorem 1.5.6. *If* $\mathbf{A}$ *satisfies the RIP of order* $2k$*, there always exists a positive constant* C *such that the following inequality is verified [3]:*

$$m \geq Ck\log\left(\frac{n}{k}\right) \tag{1.28}$$

This theorem introduces a necessary condition for correct recovery regarding the minimum number of measurements to be acquired. At the same time, it allows to assess in advance the maximum compression ratio that could be achieved in the specific case under investigation. It is worth noticing that the number of measurements m depends linearly on the sparsity level k and depends on the input signal length n according to a logarithmic relationship. For instance, if $n = 1024$ and $k = 10$, then a plausible value for the number of measurements is $m \simeq 20$, resulting in a compression ratio nearly equal to 50.

Stability Implications of the RIP

Looking for a k-sparse solution of the acquisition problem (1.26), the measurement matrix $\mathbf{A}$ is required to satisfy the RIP of order $2k$ or higher. Otherwise, there always exists at least one $2k$-sparse vector $\mathbf{z}$ which belongs to the null space of $\mathbf{A}$, i.e.:

$$\mathbf{Az} = 0 \qquad \rightleftharpoons \qquad \mathbf{z} \in \mathcal{N}(\mathbf{A}) \tag{1.29}$$

Let $\mathbf{z}$ be decomposed as the sum of two k-sparse vectors x_1 and x_2:

$$\mathbf{z} = \mathbf{x_1} + \mathbf{x_2} \qquad \|\mathbf{z}\|_0 = card\left(i = 1,2,\dots n : \mathbf{x_1}_i \neq 0 \wedge \mathbf{x_2}_i \neq 0\right) \tag{1.30}$$

In the absence of the RIP of proper order, it is reasonable to expect that:

$$\mathbf{Ax_1} = \mathbf{Ax_2} \tag{1.31}$$

which denies any possibility to achieve a unique k-sparse solution.

In this context, it is crucial to accurately estimate or predict the signal sparsity level. Unfortunately, additive noise could easily affect this prediction and it is useful to introduce a specific notion of stability for sparse acquisition problems [17].

Definition 1.5.2. Let $\mathbf{A}$ be a generic measurement matrix and $\mathscr{F}:\mathbb{R}^m \rightarrow \mathbb{R}^n$ be an ad hoc sparse recovery algorithm. Given a positive constant c, the pair $(\mathbf{A},\mathscr{F})$ is said to be c-stable if the following inequality is satisfied:

$$\|\mathscr{F}(\mathbf{A}\mathbf{x}+\mathbf{w})-\mathbf{x}\|_2 = \|\hat{\mathbf{x}}-\mathbf{x}\|_2 \leq c\|\mathbf{w}\|_2 \tag{1.32}$$

for any k-sparse vector $\mathbf{x}$ and any measurement noise vector $\mathbf{w}$.

RIP stability limits the effect of additive noise on sparse prediction error. It guarantees that if a small amount of noise is added to the measurements, its impact on the recovered signal should not be arbitrarily large. In particular, it can be shown that the sensing matrix $\mathbf{A}$ is required to satisfy the RIP of order $2k$ with an isometry constant related to the stability constant c according to the following theorem.

Theorem 1.5.7. *If the pair $(\mathbf{A},\mathscr{F})$ is c-stable for k-sparse vectors, then any $2k$-sparse vector is shown to satisfy the following inequality [17]:*

$$\frac{\|\mathbf{x}\|_2}{c} \leq \|\mathbf{A}\mathbf{x}\|_2 \tag{1.33}$$

The comparison between (1.33) and (1.24) shows how the isometry constant is directly proportional to the corresponding stability constant. In other words, the larger is the capability to preserve distances and norms, the larger is the capability to cope with additive uncorrelated noise [18].

1.6 Under-Determined Solving Algorithms

The sparse acquisition problem (1.26) yields an under-determined systems of equation to solve. Any combinatorial approach exhibits a non polynomial complexity, i.e. it is NP-hard. Without a sufficiently performing algorithm, the benefits achieved at sensor side would be wasted at receiver side. In the literature, many ad hoc solutions have been implemented and characterized. Generally speaking, the algorithms could be classified according to two main paradigms:

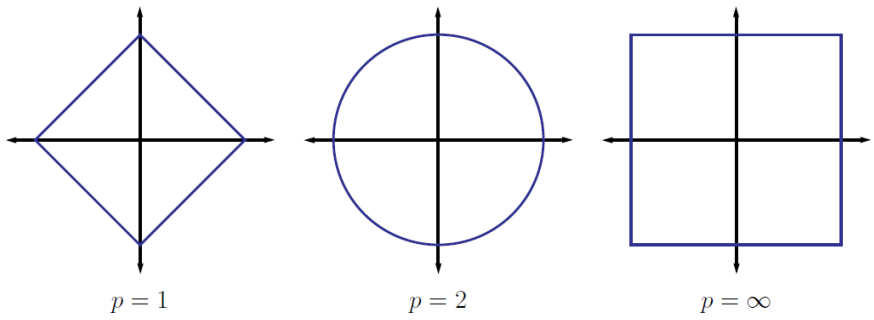


Figure 1.10: Representation of the unit ball in a Cartesian space reproducing the $\mathbb{R}^2$ domain, for three different ℓ_p-norm formulations: from left to right, respectively for $p = 1, 2$ and ∞ [17].

- convex optimization;
- greedy search.

The first one relies on a convex cost function minimization, whereas the second paradigm iteratively updates the solution including the current local approximation.

In this Ph.D. thesis, both these approaches have been taken into account and applied to specific operative conditions: particularly, the so-called LASSO minimization problem from the convex optimization class and the orthogonal matching pursuit (OMP) algorithm from the greedy search class. For the sake of completeness, these algorithms will be briefly introduced and their performances will be synthetically compared.

1.6.1 Norm Approximation

From a functional analysis point of view, a generic norm approximation problem is an optimization problem where the objective function to be minimized is a proper norm of the residuals.

A suitable norm approximation for the sparse acquisition problem (1.26) is provided by:

$$\hat{\mathbf{x}} = \min_{\mathbf{x}:\|\mathbf{x}\|_0=\mathbf{k}} \|\mathbf{A}\mathbf{x}-\mathbf{y}\|_\mathbf{p} \tag{1.34}$$

where the solution is obtained as the k-sparse vector $\mathbf{x}$ which minimizes the residuals ℓ_p-norm.

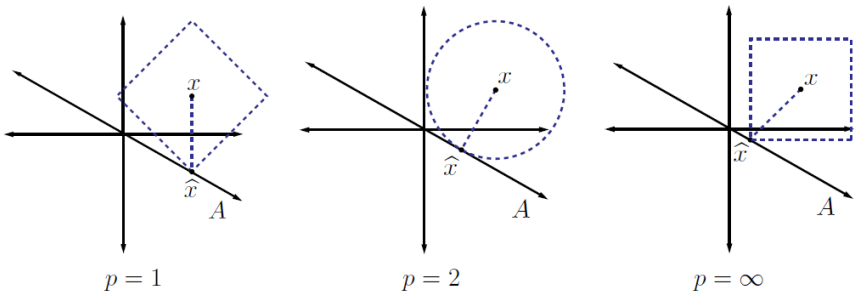


Figure 1.11: Norm approximation of a fixed point according to the same measurement matrix but different distance definitions. In a Cartesian space representation of the $\mathbb{R}^2$ domain, the solution is computed minimizing the residual ℓ_p-norm, respectively for $p = 1, 2$ and ∞ from left to right [17].

According to the peculiar norm adopted, a different penalty function and a different residuals distribution are obtained. In a vector space the notion of distance and, accordingly, the notion of measure is strictly related to the norm definition. In Fig. 1.10, the unitary ball in the $\mathbb{R}^2$ domain is represented for three different definition of ℓ_p-norm, specifically for $p = 1$, 2 and ∞. To each norm definition corresponds a peculiar distance measure. More precisely, in a generic n-dimensional vector space, given two points $\mathbf{x}$ and $\mathbf{y}$, their distance can be computed as follows:

$$d_1 = \sum_{i=1}^{n} |x_i - y_i| \qquad d_2 = \sqrt{\sum_{i=1}^{n} |x_i - y_i|^2} \qquad d_\infty = \max\{|x_i - y_i|, i = 1, \dots n\} \quad (1.35)$$

Owing to their massive employment in many Information Theory branches, these three distance measures are also known as, respectively, Manhattan, Euclidean and Chebyshev distance.

As a consequence, the best norm approximation of a given function in a given point changes on varying the order of the adopted norm. In Fig. 1.11 a linear function in $\mathbb{R}^2$ is employed to approximate a given measure $\mathbf{x}$. In this context, the solution is represented by the line point which exhibits the minimum distance with respect to the acquired measure. Evidently, the approximation error depends on the distance definition: each solution minimizes a specific cost function and, thus, optimizes a specific signal feature. According to the desired solution properties, it is possible to tune the objective function in order to endorse some features, e.g. measurement tracking rather than smoothing regularization, to the detriment of other

ones, not necessarily less informative, but less significant with respect to the case under investigation.

Application to Sparse Problems

Minimum ℓ_0-norm solution In the CS context, the solution is explicitly required to be sparse. A natural formulation for the solution recovery might be:

$$\hat{\mathbf{x}} = \min \|\mathbf{x}\|_0 \qquad s.t. \quad \|\mathbf{A}\hat{\mathbf{x}} - \mathbf{y}\|_2 \leq \varepsilon \tag{1.36}$$

where ε is related to the expected noise level in the acquired measurement vector $\mathbf{y}$.

Unfortunately, this approach is known to be non-tractable in terms of computational complexity, independently from the RIP order of $\mathbf{A}$. In general, ℓ_0-norm minimization problem requires searching among all the feasible subsets of $\mathbf{A}$ columns the one which simultaneously fits the measurements and exhibits the lowest cardinality. More precisely, this would require to implement a combinatorial search procedure among $\binom{m}{k}$ subsets. The computational complexity is expected to grow exponentially with the column number, i.e. with the solution vector cardinality: the larger is the desired resolution, the larger is the computational effort to achieve it. In conclusion, the ℓ_0-norm minimization ensures a sparse solution, but is nearly non-feasible in practice owing to its excessive computational load.

Basis Pursuit solution In machine learning literature, it is well-known that, given a measurement matrix $\mathbf{A}$ satisfying the RIP, highly sparse solutions can be obtained by convex optimization, namely by minimizing the solution ℓ_1-norm [19]:

$$\hat{\mathbf{x}} = \min_{\mathbf{x}} \|\mathbf{x}\|_1 \qquad s.t. \quad \|\mathbf{A}\hat{\mathbf{x}} - \mathbf{y}\|_2 \leq \varepsilon \tag{1.37}$$

In this way, the sparse recovery problem can be interpreted as a convex optimization problem, and can be efficiently solved via linear programming techniques based on the canonical simplex method or the more recent interior point method [20].

This formulation was initially proposed as an alternative approach to decompose the measurement vector $\mathbf{y}$ as a linear combination of $\mathbf{A}$ columns. In literature, this algorithm became famous under the denomination of Basis Pursuit (BP) [21].

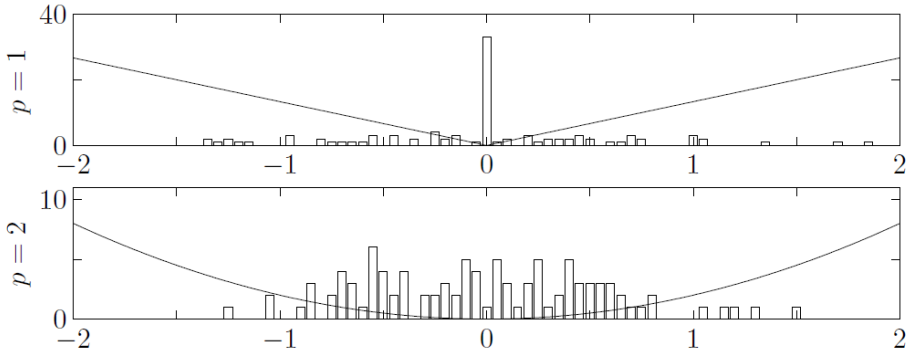


Figure 1.12: Histogram of residual amplitudes for $\|\mathbf{x}\|_p$ minimization with $p = 1$ and 2. The corresponding cost function is superposed for reference [22].

LASSO solution In terms of reconstruction accuracy, a further improvement is provided by the following regularized formulation:

$$\hat{\mathbf{x}} = \min_{\mathbf{x}} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 + \gamma \|\mathbf{x}\|_1 \tag{1.38}$$

where the cost function depends on three quantities:

- the ℓ_2-norm term, also known as least squares (LS) solution, minimizes the residual energy, thus ensuring de-noising and measurement data fitting;
- the ℓ_1-norm term, also known as basis pursuit (BP) solution, limits the number of non-zero coefficient, i.e. forces the solution to be sparse;
- the regularization parameter γ determines the respective weight of the preceding terms in the comprehensive cost function.

The different behaviour between LS and BP solution is explained in Fig. 1.12 where two specific histograms display the residuals amplitude distribution. For the sake of completeness, the corresponding objective functions are displayed, properly scaled and centred around the zero-valued residual amplitude.
The ℓ_1-norm function provides larger penalties in the surrounding of the origin, but grows linearly with the residual amplitude weights. Accordingly, the BP solution produces a large portion of residuals with an amplitude equal to zero or approximately negligible, but also some residuals whose amplitude is pretty high. In other words, the solution is sparse because it has many zero-valued residual, but affected by large

outliers or noisy distortions.
On the other hand, the ℓ_2-norm is characterized by a quadratic relationship between residual amplitude and penalty function. With respect to the previous case, the number of zero-valued residuals is expected to decrease, but the amplitude distribution is more uniform and covers a narrower interval. In other words, the lasso solution guarantees the solution to fit the measurement data with an approximation error driven from a nearly uniform distribution.

The regularization of LS and BP solution is also known under the denomination of LASSO and represents a good compromise between de-noising and sparsity.

Elastic Net solution Finally, a slightly more complex version of sparse norm approximations is provided by the so-called elastic net (EN) approach:

$$\hat{\mathbf{x}} = \min_{\mathbf{x}} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 + \beta\|\mathbf{T}\mathbf{x}\|_2^2 + \gamma\|\mathbf{x}\|_1 \tag{1.39}$$

where an additional ℓ_2-norm term is included, accompanied by the corresponding regularization parameter β. More precisely, this quadratic penalty function consists of a Tikhonov matrix $\mathbf{T}$ which multiplies the unknown vector $\mathbf{x}$. Mostly, the Tikhonov matrix is implemented as a weighted identity matrix ($\mathbf{T} = c\mathbf{I}, c \neq 0$), thus endorsing solution with smaller energy. Otherwise, another widely employed formulation defines $\mathbf{T}$ as a low-pass operator, e.g. a finite difference operator, thus endorsing smoother solutions.

In any case, the EN approach improves the numerical conditioning of the problem and guarantees better stability in the presence of noise. On the other hand, each term of the objective function corresponds to a specific assumption on the solution inherent structure. Looking for a totally blind approach, not requiring any a priori knowledge, the elastic net should be better considered as a performance reference, to be applied only when the signal under investigation has been previously studied and characterized.

1.6.2 Greedy Search

Convex optimization techniques prove to be a powerful tool for computing sparse representations from a given set of measurements. On the other hand, a large variety of greedy search techniques promise to yield similar performances at the expense of

a lower computational complexity. Generally speaking, an algorithm is said to be greedy if it solves an heuristic problem by iteratively selecting the locally optimal solution.

A vocabulary shows that the adjective greedy is correlated with words like hungry, craving, grasping, attributes which can hardly be associated with a sparse recovery algorithm. However, this odd denomination could help point out the fundamental difference between convex optimization and greedy search:

- the convex optimization approach refines a set of possible solution progressively optimizing a global cost function;
- the greedy approach considers a starting solution and extends it till a predefined stop criterion is satisfied: at each iteration only the maximal extension is considered and included into the current solution.

Typically, greedy solutions allow a drastic reduction of computational complexity, but they are practically feasible only on a specific class of problems, characterized by two peculiar properties: greedy choice and optimal substructure.
More precisely, a problem is said to admit greedy choices if its solution can be computed iteratively, progressively extending the current solution but never reconsidering the previous choices. Similarly, a problem is said to exhibit optimal substructure if an optimal solution to the problem is also an optimal solutions to its subproblems. Fortunately, the sparse acquisition problem is proven to satisfy both these properties and appears to be an ideal candidate for greedy search application.

Application to Sparse Problems

Matching Pursuit The forefather of greedy search techniques is undoubtedly the so-called Matching Pursuit (MP) algorithm [23], which provides the basic structure for all of the greedy algorithms to follow.

Once more, just like in BP case, the basic idea is to approximate the measurement vector $\mathbf{y}$ as a linear combination of the measurement matrix $\mathbf{A}$ columns. In this context, the problem is the correct identification of the most suitable coefficient vector $\mathbf{x}$.

The algorithmic main steps are summarized in Algorithm 1, where the local op-

timal selection procedure is defined as follows:

$$[hard(\mathbf{h},k)]_i = \begin{cases} h_i, & \text{if } |h_i| \text{ is among the } k \text{ largest coefficients of } \mathbf{h} \\ 0, & \text{otherwise} \end{cases} \tag{1.40}$$

where $\mathbf{h}$ is the vector to be approximated and k is the number of representation coefficients, i.e. its sparsity level. In the present case, the $hard$ procedure is applied directly to the current proxy vector $\mathbf{h}^\ell$ in order to identify only its largest coefficient.

Evidently, MP result is strongly dependent on the stopping criterion choice. In this context, many alternatives are feasible and different criteria could be properly combined in order to simultaneously control different solution features. For instance, a maximum iteration number is equivalent to a constraint on the maximum sparsity level. Conversely, a condition on the residual energy could be employed to enhance data fitting, but requires a preliminary assessment of the noise level which affects the measurements. Typically, there is no golden standard and stop criterion ought to be tuned according to the peculiar operative conditions.

Algorithm 1 Matching Pursuit algorithm

input: measurement matrix $\mathbf{A}$, measurement vector $\mathbf{y}$, stopping criterion

1: **initialize:**
   residual vector $\mathbf{r^0} = \mathbf{y}$, current solution $\mathbf{x^0} = []$, iteration $\ell = 0$
2: **while** stop criterion not satisfied **do**
3: **proxy:** $\mathbf{h}^\ell = \mathbf{A^H r}^\ell$
4: **update:** $\mathbf{x}^{\ell+\mathbf{1}} = \mathbf{x}^\ell + hard(\mathbf{h}^\ell, 1)$
5: $\mathbf{r}^{\ell+\mathbf{1}} = \mathbf{y} - \mathbf{A x}^{\ell+\mathbf{1}}$
6: $\ell = \ell + 1$

output: sparse solution $\hat{\mathbf{x}} = \mathbf{x}^\ell$, sparsity level ℓ

Iterative Hard Thresholding Some years later, a novel greedy algorithm also known as Iterative Hard Thresholding (IHT) was inspired by MP [24]. The algorithmic routine is nearly the same, but the current solution update is replaced by the following formulation:

$$\mathbf{x}^{\ell+\mathbf{1}} = hard(\mathbf{x}^\ell + \mathbf{h}^\ell, \mathbf{k}) \tag{1.41}$$

In other words, the selection procedure is applied not only on the proxy, but on its summation with the current solution. Furthermore, the selection could consider a

set of coefficients whose cardinality is equal to k, i.e. directly the desired sparsity level.

Owing to this peculiar update formula, IHT proves to be greedier than MP, still keeping the solution sparsity under control. Adopting such a sharp coefficient selection procedure, it is reasonable to argue that IHT could be more prone to excessive simplification of the signal under investigation, leading to an information content loss or misinterpretation. However, experimental practice shows how IHT represents a good alternative to the MP approach, especially in those problems where signal sparsity is evident and pronounced.

Orthogonal Matching Pursuit Even if IHT keeps spreading more and more among CS literature, another algorithm is worth to be considered as the actual legitimate successor of MP, namely the Orthogonal Matching Pursuit (OMP) algorithm [25]. As can be seen in Algorithm 2, the main structure draws inspiration directly from Algorithm 1, even though some differences are noticeable.

Firs of all, the solution support is introduced and defined as the set of non-zero solution coefficient locations:

$$S_{\mathbf{x}} = \{i \in \mathbb{Z}^+ \quad s.t. \quad x_i \neq 0\} \tag{1.42}$$

At each iteration, the support of the selection procedure result is included into the comprehensive signal support.

Furthermore, also the current solution update is renewed. According to MP algorithm, the current solution is provided by the summation of previous iteration result with $hard$ selection result. Conversely, OMP algorithm defines the current solution as the LS solution chosen among all the ℓ-sparse vectors whose support is given by $S^{\ell}_{\mathbf{x}}$. Though computing ℓ_2-norm approximation, OMP is still an iterative greedy algorithm: once a coefficient location has been included into the support, it will be never discarded nor could be selected again. On the other hand, the coefficient value could change on varying the iteration number. Indeed, OMP tries to approximate the measurement vector $\mathbf{y}$ as a linear combination of measurement matrix $\mathbf{A}$ columns: it is reasonable to expect that the larger is the number of employed columns, the more accurate will be their combination coefficients.

Looking at MP and OMP, a difference stands out immediately: the term orthogonal. According to Algorithm 1, MP coefficients are estimated by projecting the cur-

rent residual over the set of **A** columns selected in the previous iterations. This approach is proven to preserve the entire information content, but does not ensure asymptotic convergence to the actual global solution. In order to solve this sub-optimal behaviour, the orthogonality constraint is introduced in Algorithm 2. Indeed, from a linear algebra point of view, the LS update formula ensures the residual to be projected over the vector subspace $\mathbb{P}^{\ell}$ spanned by the so far selected **A** columns. At each iteration, a novel subspace is defined and it can be shown that $\mathbb{P}^{\ell}$ is orthogonal by construction with all its preceding versions $\mathbb{P}^{\ell-1}, \mathbb{P}^{\ell-2}, \ldots$.

The orthogonal constraint, provided by the LS approximation, guarantees the asymptotic convergence to the actual global solution. Moreover, it allows the algorithm to converge in a fewer number of iterations. In conclusion, OMP should be generally preferred over MP, because its larger computational effort results in more accurate estimations and fewer iterations.

Algorithm 2 Orthogonal Matching Pursuit algorithm

input: measurement matrix **A**, measurement vector **y**, stopping criterion

1: **initialize:**

residual vector $\mathbf{r^0} = \mathbf{y}$, current solution $\mathbf{x^0} = []$,

solution support $S_x^0 = \varnothing$, iteration $\ell = 0$

2: **while** stop criterion not satisfied **do**

3: **proxy:** $\mathbf{h}^{\ell} = \mathbf{A^H r}^{\ell}$

4: **identify:** $S_x^{\ell+1} = S_x^{\ell} \cup supp\left(hard(\mathbf{h}^{\ell}, 1)\right)$

5: **update:** $\mathbf{x}^{\ell+1} = \min_{\mathbf{x}} \|\mathbf{y} - \mathbf{Ax}\|_2 \qquad \forall \mathbf{x} : supp(\mathbf{x}) \subseteq S_x^{\ell}$

6: $\mathbf{r}^{\ell+1} = \mathbf{y} - \mathbf{Ax}^{\ell+1}$

7: $\ell = \ell + 1$

output: sparse solution $\hat{\mathbf{x}} = \mathbf{x}^{\ell}$, sparsity level ℓ, solution support $S_{\hat{\mathbf{x}}} = S_{\mathbf{x}}^{\ell}$

1.6.3 Performances Comparison

The previous sections have briefly summarized the main techniques for solving highly under-determined acquisition systems, properly exploiting the signal sparsity assumption and the measurement matrix RIP. Roughly speaking, these can be classified according to algorithmic families, namely convex optimization and greedy search, each one exhibiting its own advantages and disadvantages. However, a question still remains unresolved: how can the user choose the most suitable solving

technique, which better copes with the present operative condition? In the following, some useful theoretical results are provided in order to characterize the behaviour of each family in terms of required measurement number, noise sensitivity, and reconstruction accuracy.

Noiseless operative conditions Let us consider a noise-free measurement vector. In order to estimate its better sparse approximation, the norm approximation approach suggests to employ a simple BP algorithm. In fact, there is no need to include additive terms to enforce data-fitting or solution smoothing.

In order to recover a k-sparse vector, the measurement matrix $\mathbf{A}$ is required to exhibit a mutual coherence compliant with the following inequality [18]:

$$\mu(\mathbf{A}) \leq \frac{1}{\sqrt{n}} \tag{1.43}$$

where n is the cardinality of the sparse representation.

A similar condition can be expressed in terms of RIP whose isometry constant is required to satisfy the following inequality [26]:

$$\delta_k < 0.307 \quad \wedge \quad \delta_{2k} < 0.472 \tag{1.44}$$

It has been shown that this bound can not be substantially improved. In this context, it is worth noticing that any condition regarding the isometry constant for a given matrix is computationally very hard to verify. Instead of checking the RIP directly over a statistically significant set of k-sparse vectors, it is advisable to randomly define the measurement matrix and show that it satisfies the RIP with reasonably high probability [27]. In this way, it is possible to focus on RIP conditions which involve directly the isometry constant δ and are totally independent from the measurement matrix inherent structure.

More precisely, the Johnson-Lindenstrauss lemma ensures that a reduced set of high-dimensional vectors can be projected onto a lower-dimensional space, preserving the distances with a limited uncertainty level. In particular, given two k-sparse vectors $\mathbf{x_1}$ and $\mathbf{x_2}$ and a projection matrix $\mathbf{A}$ of dimensions $[m \times n]$, with $m \ll n$, it can be shown that:

$$(1-\delta_{JL})\|\mathbf{x_1}-\mathbf{x_2}\|_2^2 \leq \|\mathbf{Ax_1}-\mathbf{Ax_2}\|_2^2 \leq (1+\delta_{JL})\|\mathbf{x_1}-\mathbf{x_2}\|_2^2 \tag{1.45}$$

where δ_{JL} quantifies the uncertainty introduced by the dimensionality loss.
It is immediate to ascertain the similarities between (1.24) and (1.45). If a random

matrix satisfies the Johnson-Lindenstrauss Lemma with sufficiently high probability, it is reasonable to expect that the same matrix satisfies the RIP of order k with an isometry constant $\delta_k \simeq \delta_{JS}$.

In terms of measurement number, the BP approach requires a sufficient number of measurements which is nearly provided by a logarithmic function of the ratio between solution cardinality and sparsity level:

$$m \simeq 2k \log\left(\frac{n}{k}\right) \tag{1.46}$$

In terms of computational complexity, the BP approach generally exhibits a computational complexity equal to $\mathcal{O}(1000 \cdot m \log m)$ for iteration, even if performance enhancement is possible in some peculiar cases, where ad hoc optimization procedures are feasible. It should be noticed that the global solution achievement requires a number of iterations which is not directly related to the expected sparsity level, but depends on the cost function evolution that cannot be easily foreseen a priori.

In the matter of the OMP approach, the constraint over the mutual coherence becomes:

$$\mu(\mathbf{A}) < \frac{1}{2k-1} \tag{1.47}$$

much easier to be satisfied than the corresponding BP inequality (1.43).
Similarly, the measurement matrix $\mathbf{A}$ is required to verify the RIP of order $k+1$ with an isometry constant compliant with the following inequality:

$$\delta_{k+1} < \frac{1}{3\sqrt{k}} \tag{1.48}$$

In this condition, it can be shown that the OMP approach yields the best k-sparse approximation exactly in k iteration, resulting in a comprehensive computational complexity equal to $\mathcal{O}(kn \log n)$.

OMP performances appear to exceed BP performances, but it is important not to neglect the sufficient condition regarding the minimum number of measurement:

$$m \geq 2k \log(n-k) \tag{1.49}$$

Accordingly, it is worth noticing that in the noiseless case the BP approach requires fewer measurements, that is allows a larger compression ratio.

Noisy operative conditions Let us consider a measurement vector affected by an additive uncorrelated white noise. For the sake of simplicity, let us assume the noisy

contribution vector $\mathbf{w}$ can be modelled by a Gaussian distribution whose variance is equal to $\sigma^2_{\mathbf{w}}$. In this context, it is worth noticing that an excessive noise level could cancel the sparsity assumption. Accordingly, the following considerations are related to largely positive signal-to-noise ratio values. In this way, the signal sparse representation still exhibits a clear distinction between informative and noisy coefficients.

In the presence of noise, the BP approach becomes severely unstable and it is advisable to include a regularization term, switching to the LASSO configuration.

The additive noise contribution can be seen as an uncertainty source which degrades both norm approximation and greedy search results. In these operative conditions, LASSO and OMP exhibit similar performances, even if they mostly differ by the computational complexity.

The measurement number is the same for both the approaches:

$$m \geq 2k\log(n-k) \tag{1.50}$$

Comparing this value with the noiseless case, the OMP approach proves to be almost noise insensitive, while the norm approximation technique requires a significant increase of the measurement number.

Similarly, both the approaches requires the measurement matrix $\mathbf{A}$ to satisfy the RIP of order $2k$, namely twice the expected sparsity level, with an isometry constant equal to $\delta_{2k} \leq \sqrt{2}-1$, just like in (1.44).

The main difference between norm approximation and greedy search is thus the computational complexity. In this context, it is noticeable that OMP requires the same effort as in the noiseless case, while LASSO further increases the yet larger requirements of BP. In other words, the noise disturbance enlarges the performance gap, once more emphasizing that the greedy approach is particularly prone to capture the essence of sparse problem: fewer measurements imply also fewer iterations, otherwise the CS paradigm becomes totally ineffective.

Phase Transition

The performance indices provided in the previous paragraphs could be useful to foresee the behaviour of a given CS acquisition protocol in a given operative context. However, these indices are not always unanimous and the reconstruction algorithm

ought to be chosen, making a proper compromise between different, and often contrasting, requirements. In this context the need for an unequivocal and immediate criterion arises. To the best of authors' knowledge, a plausible answer is represented by the so-called phase transition graph, which intuitively shows if the implemented algorithm is near to the theoretically optimal performances or should be replaced by more effective solutions.

In geometry, phenomena like abrupt phase transitions can be observed in any kind of measurement space, independently from its dimension. Typically, a phase transition is defined as a rapid shift in the likelihood of a property's occurrence when a dimension parameter crosses a critical threshold [28]. Actually phase transitions become more and more significant in higher-dimension geometric probability, as they represent a powerful tool to predict the development of a given occurrence even in contexts where an immediate intuition is precluded. Indeed, phase transitions are expected to influence the model selection for the statistical analysis of large datasets, but also the assessment of the most suitable sampling scheme, capable to correctly acquire a digital data stream.

In CS theory context, phase transition graph becomes a useful tool to comprehend the actual potentiality of a reconstruction algorithm. Exempli gratia, let us consider a generic sparse acquisition problem where a k-sparse representation of cardinality n has to be retrieved from a measurement set of cardinality m with $m \ll n$. As reconstruction algorithm, let us compare the aforementioned BP and OMP approach. The reconstruction is said to be successful if the solution support actually contains the k most significant coefficient locations. For each considered algorithm, the success probability exhibits a multi-dimensional distribution, which depends on the compression ratio m/n and the sparsity level k.

In practice, a thorough mathematical characterization is infeasible, due to the amount of variables to be taken into account. Nevertheless, some interesting cases represent a theoretical benchmark in mathematics and their phase transition function can be inferred in a closed form. For instance, given a mixture of k independent Gaussian n-dimensional distributions, the probability of retrieving the exact original formulation starting from an incomplete set of $m < n$ measurements is fully characterized and its phase transition function can be expressed as a non-linear function of k, n and m. A recent paper has demonstrated how noiseless under-

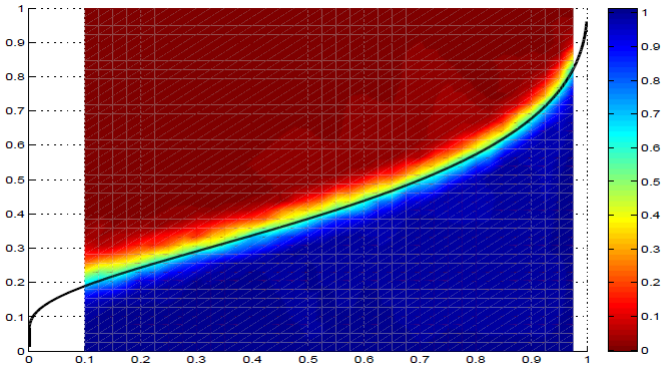


Figure 1.13: Compressed Sensing from random Fourier measurements. Shaded attribute: fraction of realizations in which l1 minimization (1.2) reconstructs an image accurate to within six digits. Horizontal axis: under-sampling fraction. Vertical axis: sparsity fraction [28].

determined problems behave like noisy well-determined problems, i.e. coping with incompleteness of the measurement data is (for random sensing matrices) similar to coping with Gaussian noise in complete measurements. In other words, it is possible to change the underlying distribution from Gaussian to any one of a variety of non-Gaussian choices, and the phase transitions are expected to occur at the same locations, i.e. in correspondence to the same acquisition settings.

Fig. 1.13 depicts the phase transition graph of BP success probability in the presence of compressed random Fourier measurements. The probability function is observed on varying the compression ratio m/n on the y-axis, and the sparsity fraction k/n on the x-axis. The black bold curve represent the theoretical phase transition function as inferred by the corresponding mixture of Gaussian distributions. It is worth noticing that this curve is a demarcation line which divides the success area below from the failure area above.

Given an expected sparsity level k, it is possible to set the acquisition parameters m and n in order to get into the success area and optimize the acquisition protocol performances. For instance, let us consider an experimental setting which corresponds to position 1 in Fig. 1.14. By increasing the problem size, i.e. by increasing n, the position inside the graph moves leftward from 1 to 2, still inside the failure area. By increasing both the problem size n and the measurement number m, it is possible to reach position 3, where high success probability is guaranteed. In particular,

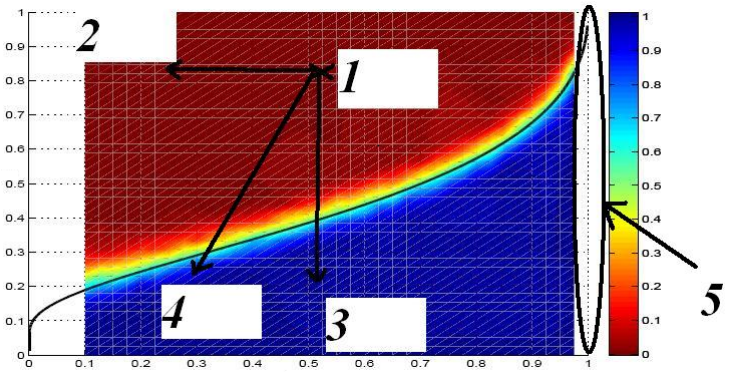


Figure 1.14: Different behaviours according to different position in the phase transition graph [29].

the passage from 1 to 3 requires that the compression ratio stays constant, while the sparsity fraction decreases. Alternatively, the position 4 could be reached with a path between 2 and 3, i.e. by decreasing both the compression ratio and the sparsity fraction. The lower is the distance from the theoretical limit, the higher is the probability that unexpected disturbances or malfunctions make the setting to get into the failure area. Finally, the area labelled as position 5 corresponds to the canonical acquisition protocol where $m = n$, where reconstruction probability is unitary, independently from the signal sparsity properties.

Phase transition graphs are also an immediate and intuitive tool to compare algorithm performances. For instance, Fig. 1.15 depicts the phase transitions graphs of four different under-determined solving approaches. Blue and red curves correspond respectively to BP and IHT approaches. Conversely, magenta and black curves represent two different implementations of the OMP approach, respectively the Subspace Pursuit, briefly SP [31],and the Compressed Sampling Matching Pursuit, briefly CoSaMP [32]. From the graph it is immediate to ascertain that the success area related to greedy search approaches is much smaller than the success are related to norm approximation approaches. Accordingly, an actual computational complexity reduction is feasible only in accordance with strict constraints on the experimental settings.

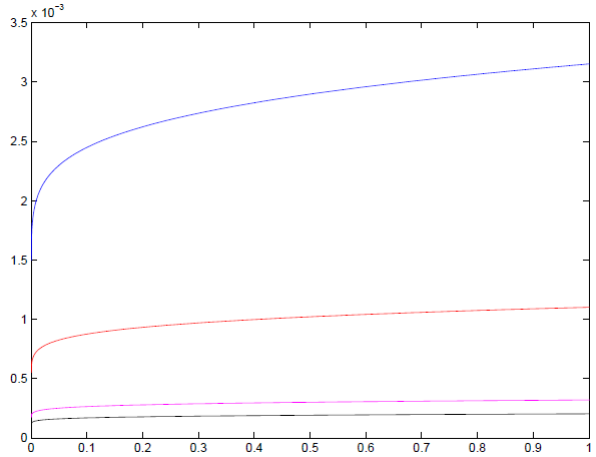


Figure 1.15: Phase transition graphs of BP approach (blue), IHT (red), SP (magenta), and CoSaMP (black) [30].

1.6.4 Guidelines for Practical Implementation

In CS literature, many solutions have been proposed in order to solve the under-determined acquisition system. More precisely, two algorithmic approaches could be identified, namely the norm approximation and the greedy search approach. By properly exploiting the sparsity assumption, these algorithms are able to estimate an accurate sparse representation of the observed signal, even starting from an incomplete set of measurements. As mentioned in the previous sections, there exist several theoretical results and criteria which allow to foresee the actual feasibility of a sparse acquisition protocol, in terms of expected reconstruction error and required measurement number.

The choice between norm approximation and greedy search should consider many different aspects. In fact, the practical implementation should be tuned in accordance with the specific operative conditions. Even adopting a signal independent perspective, with a totally random measurement matrix, some boundary conditions could suggest to employ a norm approximation technique rather than a greedy search technique, and vice versa.

In the following, a brief summary of pros and cons is presented.
Norm approximation copes with any kinfd of sparse problems, independently from the actual distribution of the significant coefficients. By minimizing a proper ℓ_1-

norm cost function, the provided solution takes a global meaning and accounts for the sparsest approximation of the measurement vector. By properly setting the acquisition parameters, this approach yields overwhelming success probability and noticeable compression ratio. On the other hand, the computational effort is enormous and requires suitable processing capabilities to be carried out in reasonably short time. In the presence of noise, it is necessary to enrich the cost function by introducing one or more ℓ_2-norm terms. In this way, the algorithm proves to be more robust, but also more computationally demanding.
Greedy search adopts an iterative procedure where a local solution is iteratively updated and extended till a pre-defined stop criterion is satisfied. Typically, greedy search techniques allow a drastic reduction of the computational effort, requiring no more than $k+1$ iterations to yield their k-sparse approximation. The noise robustness is such that no precautions are required in the presence of additive uncorrelated noise, provided that the sparsity assumption is preserved. On the other hand, this local perspective does not guarantee the convergence to the actual global solution. Greedy search is feasible only in those problems which exhibit optimal substructure and admit greedy choice solutions. Furthermore, in terms of success probability, greedy acquisition settings must be compliant with much more severe constraints than in norm approximation case.

The previous considerations could be condensed into the following guidelines:

- if a blind and general approach is required, the most suitable option is represented by norm approximation techniques, particularly BP in the noiseless case or LASSO in the noisy case;

- if problem substructure is proven to be suboptimal and the main focus concerns computational effort and implementation complexity, the most suitable option is represented by greedy search techniques, particularly OMP in any of its various implemented formulations.

Chapter 2

CS-Based Pre-Processor for Vector Signal Analysis and Cognitive Radio Applications

2.1 Introduction

Nowadays, radio frequency (RF) spectrum represents a much precious resource. In order to ensure its rational, fair and economical use, many committed authorities have been established and many norms and standards have been promulgated.

In this context, a still not completely resolved problem is represented by the co-existence of two kinds of users: primary users with a license for a specific transmission channel, and secondary users who occupy spectrum holes, i.e. temporarily unemployed spectrum portions. In order to avoid interference or superposition, secondary users have to monitor a large bandwidth and set their transmission parameters according to their acquisitions results. Such a dynamical technology is called cognitive radio, briefly CR [33]. Basically, CR could be considered a peculiar application of software defined radio (SDR) to wideband communications. From a practical implementation point of view, a cognitive radio is a smart transceiver, i.e. a device which combines both transmitting and receiving functionalities, but, at the same time, is provided with an autonomous processing and control unit, that can monitor

surrounding conditions and accordingly tune its operative parameters. Typically, a state-of-the-art CR includes five main functionalities: geographic localization, signals encryption and decryption, user authority identification, real-time spectrum sensing, and prompt adjustment of transmission parameters, like throughput, symbol rate, modulation format, etc.

This chapter focuses its attention particularly on the spectrum sensing issue. Restricting the analysis to this peculiar aspect, the problem is remarkably similar to the more general problem of detecting RF sources in monitoring applications using a vector signal analyser, briefly VSA. It requires to deal with RF signals in an extremely large spectrum bandwidth and with nearly complete absence of a priori knowledge. In fact, in most spectrum sensing and wireless communications analysis, signals of interest typically occupy only a few among several possible bands, and do so for short time bursts within a given observation interval. The frequency and time location of these signals may be known only approximately *a priori* (e.g., the nominal frequency of a wireless channel) or, in general, not accurately enough to set up more detailed measurements. Such a scenario suggests that a CS approach could be extremely effective in reducing the sampling frequency and the computational effort necessary to span the radio frequency spectrum and identify unemployed channels.

In this chapter a novel CS algorithm is developed and employed to provide accurate preliminary information and suitably preprocessed data for a vector signal analyser or a cognitive radio application.

The CS paradigm exploits signals sparsity not only to efficiently capture the signal information content, but also to accurately localise where the active contributions lie over the observed spectrum. The proposed algorithm is inspired by a novel protocol for analog-to-digital data conversion, called Modulated Wideband Converter [13]. It is able to extract specific signal bursts from a record of samples covering a longer time interval and a broader bandwidth. The accuracy of the extraction process is analysed and experimental results referring to vector analysis are presented. This provides spectrum-blind sensing and signal extraction capabilities that can effectively simplify the time consuming process of setting up a spectrum analyser.

Furthermore, the proposed algorithm is able to single out active sources which could be partially or totally superposed both in time and frequency domain. The ob-

served spectrum is divided into consecutive non-overlapping sub-bands. By properly analysing the eigenvalue distribution of the measurement autocorrelation matrix, active sub-bands are identified and demodulated at baseband. In this way, also a de-noising effect is achieved, as thoroughly characterized in the following sections. Finally, a histogram analysis stage considers the in-phase and quadrature components of each singled out baseband trend to predict the transmission modulation format. A simple threshold process performed on the histogram distributions allows to estimate a rather rough rendering of the original constellation graph, providing also the areas where it is more plausible to find a symbol. Application of pattern matching algorithms, either to such plots or directly to histogram data, can ultimately detect the modulation format [34].

2.2 Problem Scenario

Vector signal analysis extends the capabilities of spectrum analysers to the domain of digital modulations. It can be seen as an enhancement that capitalizes on the availability, in modern digital architectures, of high-resolution, high-speed digitising coupled with powerful signal processing resources.

A vector signal analyser (VSA) yields essential information for the characterisation of digitally modulated signals, such as constellation and vector diagrams, error vector magnitude (EVM) plots, etc. [35]. However, VSA measurements require accurate advance information about carrier frequencies and modulation formats of interest. Unless these are exactly known beforehand, knowledge has to be gained by preliminary measurements.

This can make the process rather cumbersome and time consuming in multi-source wireless environments, such as in spectrum monitoring, surveillance applications and cognitive radio [33], feeding the instrument with the required advance information becomes challenging.

In fact, correctly measuring intertwined and possibly elusive communication channels requires an accurate analysis of the mutual timing, frequency and levels of a set of signals employing different modulation formats. Signals of interest might occupy only a few among a number of possible bands and do so only for the short time required to send a radio-frequency (RF) burst, so that taking preliminary mea-

surements becomes challenging.

Such signals can be intuitively associated to the notion of *sparsity*, that implies their structure is recoverable from a comparatively small number of measurements defined in a suitable domain. This important feature is exploited in the signal processing paradigm called *compressive sensing* (CS), or compressed sampling, to achieve greater efficiency in data acquisition and measurement.

In this chapter we propose a CS-based measurement pre-processor, that significantly enhances the potential of VSA applications by providing the instrument with the ability to detect signals from multiple sources and autonomously identify their basic parameters.

The literature about potential CS applications is very extensive. Interested readers are referred in particular to published journals [36], [37], textbook [38] and the extensive references therein. Applications related to signal processing and measurement have been discussed in the literature [39], [40], including sensing and signal detection (e.g., [41]) which is relevant to this work. Design of hardware data acquisition devices based on CS principles is also an active research area [42], [43].

In this work, the focus on the algorithmic side of CS is motivated by the following considerations:

- data acquisition hardware in a vector/spectrum analyser is already a highly sophisticated and optimised system;

- as the CS algorithm is meant to implement a pre-processing stage, solutions impacting instrument architecture would be far beyond the aims of the work;

- sample records are usually several thousand samples in size and the envisaged pre-processor functions could be computation-intensive. Hence, a processing paradigm that allows data compression without sacrificing measurement accuracy can significantly improve efficiency.

So far, few works have addressed a measurement application discussing in detail its CS aspects. This chapter motivates the choice of a CS algorithm, discusses underlying assumptions and presents an analysis of potential measurement accuracy.

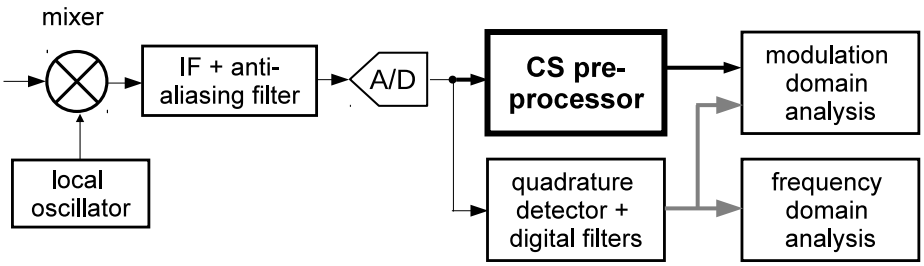


Figure 2.1: Simplified block diagram showing integration of the CS-based pre-processor into VSA firmware architecture.

2.3 CS algorithm for VSA measurement

The relationship of the proposed CS-based pre-processor to the general VSA instrument architecture is shown in Fig. 2.1. To simultaneously monitor different frequency channels of interest, the VSA analog intemediate-frequency (IF) filter would be employed as a pre-selector, using a wide-band setting and the pre-processor would become part of the digital intermediate-frequency software/firmware. A single record of N samples from the analog-to-digital (A/D) converter might then contain multiple RF signal bursts generated by different sources, whose time separation, length, carrier frequency, bandwidth and modulation format are not known.

An example of a time domain trace at the IF filter output is presented in Fig. 2.2, where bursts at three distinct carrier frequencies have been given different amplitudes to help distinguish them visually (a shorter peak is the result of partial superposition in time between the first two bursts). The CS algorithm is tasked with the detection of "active" sub-bands where RF bursts are found, followed by the extraction of relevant signal components, where appropriate. Its use guarantees a twofold advantage since, on the one hand, it allows the adoption of sub-Nyquist schemes that limit the computational load involved in spectrum sensing. On the other hand, a sharp distinction is made between informative and spurious components, extracting the former while providing some degree of noise reduction.

The algorithm operates on the sample record in three stages:

- *mixing and sub-Nyquist down-sampling*: this is obtained by a software implementation of the *modulated wideband converter* (MWC) concept [13], [44], which allows compressive acquisition of sparse wide-band signals and resam-

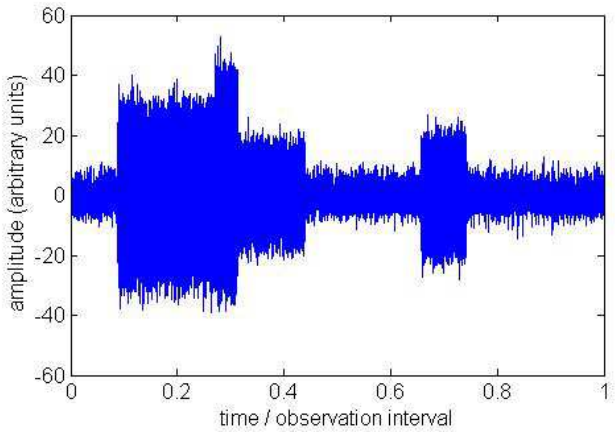


Figure 2.2: Example of a time domain trace comprising three bursts. Carrier frequencies are all different, but two bursts are partially superposed in time.

pling at sub-Nyquist rates;

- *finite support recovery*, by means of an OMP-based procedure based on a development of the method presented in [45];

- signal *reconstruction*, that simply requires inverting a linear measurement equation restricted to "active" sub-bands.

2.4 MWC Analysis and Characterization

A block diagram of the MWC is presented in Fig. 2.3. The sequence of Nyquist-rate samples from the A/D converter, indicated as $x[nT]$ with $0 \le n < N$, is sent in parallel to m MWC mixing channels, being multiplied in each channel by a different periodically repeating pattern of M random equiprobable sign values. We indicate each pattern by a sign vector: $\mathbf{a_i} = [\alpha_{i,0}, \alpha_{i,1}, \ldots, \alpha_{i,M-1}]^T$, with $\alpha_{i,h} \in \{-1,1\}$. The index $i = 1, \ldots m$ identifies the mixing channel, which is uniquely associated to a pattern as long as the pattern length is $M < 2^m - 1$.

Sign vectors are assumed to be mutually uncorrelated, that is, $\mathrm{E}_\alpha[\mathbf{a}_i^T \mathbf{a}_l] = 0$ for $i \neq l$, the appended subscript to symbol $\mathrm{E}_\alpha[\cdot]$ denoting that in this case the statistical expectation operator is referred to the probability of sign values. For the sake of simplicity, it is also assumed that the length of $x[nT]$ corresponds to an integer

number of P periods of the modulating sign pattern, so that the signal length can be factored as $N = P \cdot M$.

The spectra of periodical sign patterns are composed of non-zero terms at integer multiples of their common fundamental frequency $\frac{1}{MT}$. Given the discrete Fourier transform (DFT) of the Nyquist-rate sequence, $X(\frac{k}{NT})$, each mixer output is the superposition of different frequency-shifted replicas of the signal spectrum, obtained according to a channel-specific aliasing pattern $A_i(\cdot)$:

$$\sum_{l=0}^{M-1} A_i\left(\frac{l}{MT}\right) X\left(\frac{k}{NT} + \frac{l}{MT}\right), \tag{2.1}$$

where $A_i(\cdot)$ is the DFT of the i-th sign pattern. In this way, a "controlled" aliasing is introduced [44].

After mixing, signals are low-pass filtered and down-sampled by a decimation factor M. The digital low-pass filter is the same for all channels and its frequency response $H\left(\frac{k}{NT}\right)$ is designed to approximate an ideal low-pass response with $H\left(\frac{k}{NT}\right) = 0$ for $|k| > \frac{P}{2}$.

The amount of data at the MWC output is thus reduced from $N = MP$ samples in a single vector to a measurement matrix having size $m \times P$. The ratio $\frac{M}{m}$ can be called the *compression factor*.

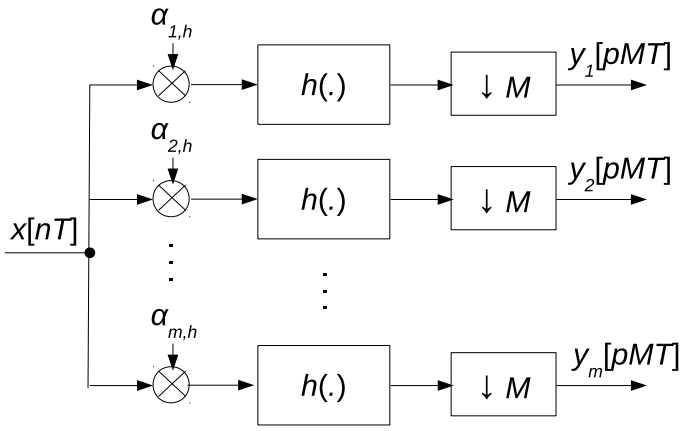


Figure 2.3: Block diagram of the modulated wideband converter [44].

2.4.1 MWC Measurement Equation

The output of the i-th MWC channel is a lower-rate sequence $y_i[pMT]$ of finite length P, that can be described by the time-domain equation:

$$y_i[pMT] = \sum_{l=0}^{M-1} A_i\left(\frac{l}{MT}\right) x_l[pMT] \qquad 0 \le p < P \tag{2.2}$$

Accordingly, an equivalent matrix equation is derived as follows:

$$\begin{bmatrix} y_1[0] & \dots & y_1[P-1] \\ y_2[0] & \dots & y_2[P-1] \\ \vdots & \ddots & \vdots \\ y_m[0] & \dots & y_m[P-1] \end{bmatrix} = \begin{bmatrix} A_1[0] & \dots & A_1[M-1] \\ A_2[0] & \dots & A_2[M-1] \\ \vdots & \ddots & \vdots \\ A_m[0] & \dots & A_m[M-1] \end{bmatrix} \begin{bmatrix} x_0[0] & \dots & x_0[P-1] \\ x_1[0] & \dots & x_1[P-1] \\ \vdots & \ddots & \vdots \\ x_{M-1}[0] & \dots & x_{M-1}[P-1] \end{bmatrix} \tag{2.3}$$

Each sequence $x_l[pMT]$, with $0 \le l < M$, represents the contribution to the signal $x[nT]$ from a spectrum slice of width $\frac{1}{MT}$, baseband-shifted from its original center frequency $l \cdot \frac{1}{MT}$. It is a complex-valued sequence that can be expressed as:

$$x_l[pMT] = \sum_{n=0}^{N-1} x[nT] e^{j\frac{2\pi}{M}nl} h[(n-pM)T]. \tag{2.4}$$

The MWC thus implicitly creates a grid on the frequency axis that determines a partition of the IF bandwidth into M equal-width sub-bands.

In a multi-source wireless environment, each spectrum slice could be related to a source employing a different frequency channel. Any time-multiplexed signal within a channel can be singled out directly from the samples $x_l[pMT]$. Furthermore, if the carrier frequency of an active source equals the center frequency $\frac{l}{MT}$, then $x_l[pMT]$ directly provides the demodulated signal, provided that the corresponding spectral contribution is entirely contained into the selected sub-band.

In each down-sampled sequence $y_i[pMT]$ sub-band contributions are mixed up by channel-specific aliasing patterns, as shown in (2.2) and (2.3). In the following it will be useful to refer to (2.3), using the compact form:

$$\mathbf{Y} = \mathbf{A} \cdot \mathbf{X}, \tag{2.5}$$

where $\mathbf{Y}$ contains the MWC outputs, that is, time-domain sequences $y_i[pMT]$ from MWC channels $i = 1, \dots, m$, arranged into rows. It is useful to define column vectors $\mathbf{y}_i = \left[y_i[0]\ y_i[MT]\ \dots\ y_i[(P-1)MT]\right]^T$, which allow to write $\mathbf{Y} = \left[\mathbf{y}_1\ \mathbf{y}_2\ \dots\ \mathbf{y}_m\right]^T$. Likewise, each row of $\mathbf{X}$ contains a baseband time-domain sequence $x_l[pMT]$ associated to the l-th spectral sub-band, $l = 0, 1, \dots, M-1$, of width $\frac{1}{MT}$. These rows, too,

can be arranged into column vectors $\mathbf{x}_l = [x_l[0]\ x_l[MT]\ \ldots\ x_l[(P-1)MT]]^T$ with $0 \le l < M$, yielding $\mathbf{X} = [\mathbf{x}_0\ \mathbf{x}_1\ \ldots\ \mathbf{x}_{M-1}]^T$.

Finally, matrix $\mathbf{A}$ can be factored as:

$$\mathbf{A}_\alpha \cdot \mathbf{W}, \tag{2.6}$$

where $\mathbf{W}$ is a DFT matrix of size $M \times M$, with elements $[w]_{h,l} = \frac{1}{\sqrt{M}} e^{-j\frac{2\pi}{m}hl}$, called the *sparsifying matrix* in CS parlance. The *sensing matrix* $\mathbf{A}_\alpha$ is a $m \times M$ Bernoulli matrix that contains the sign vectors as its row elements:

$$\mathbf{A}_\alpha = \frac{1}{\sqrt{M}} \begin{bmatrix} \alpha_{1,0} & \alpha_{1,1} & \ldots & \alpha_{1,M-1} \\ \alpha_{2,0} & \alpha_{2,1} & \ldots & \alpha_{2,M-1} \\ \vdots & & \ddots & \vdots \\ \alpha_{m,0} & \alpha_{m,1} & \ldots & \alpha_{m,M-1} \end{bmatrix} = \begin{bmatrix} \mathbf{a_1}^T \\ \mathbf{a_2}^T \\ \vdots \\ \mathbf{a_m}^T \end{bmatrix}. \tag{2.7}$$

For the sake of mathematical convenience both matrices have been suitably normalized, so that $\mathbf{W}$ is orthonormal and rows of $\mathbf{A}_\alpha$ have unit norm. The assumed incorrelation of the sign vectors implies that, "on average", rows of the sensing matrix are orthogonal, i.e.: $\mathrm{E}_\alpha\left[\mathbf{A}_\alpha \mathbf{A}_\alpha^T\right] = \mathbf{I}$, where $\mathbf{I}$ is an $m \times m$ identity matrix.

2.4.2 MWC Input Noise

In the conditions for which the CS-based pre-processor is intended, the VSA would operate with a comparatively wide-band IF filter setting (typically, in the 10-100 MHz range). As the displayed average noise level (DANL) of current VSA/spectrum analysers is typically in the region of -140 dBm for a 1–Hz bandwidth, noise power could be in the range -60 to -70 dBm, depending on the instrument specifications and setting.

To analyse how noise and uncertainty come into play, it is commonly assumed the measurement matrix $\mathbf{Y}$ is affected by additive *measurement* noise. This is not applicable to the MWC case, as additive noise affects instead the acquired sample record $x[nT]$. Thus, one should write: $x[nT] = s[nT] + v[nT]$, where $v[nt]$ is a sequence of zero-mean, uncorrelated *signal* noise samples with finite variance σ_v^2. A similar expression holds for each baseband sequence and can be written as vector relationship:

$$\mathbf{x}_l = \mathbf{s}_l + \mathbf{n}_l, \tag{2.8}$$

where $\mathbf{s}_l$ refers to the "clean" signal component from the l-th sub-band and $\mathbf{n}_l$ to the corresponding signal noise sequence, both baseband-shifted and filtered according to (2.4).

Collecting these vectors into two separate matrices, respectively **S** and **N**, yields the equation that correctly describes the MWC input-output relationship:

$$\mathbf{Y} = \mathbf{AX} = \mathbf{A}(\mathbf{S} + \mathbf{N}), \tag{2.9}$$

with $\mathbf{S} = [\mathbf{s}_0\ \mathbf{s}_1\ \ldots\ \mathbf{s}_{M-1}]^T$ and $\mathbf{N} = [\mathbf{n}_0\ \mathbf{n}_1\ \ldots\ \mathbf{n}_{M-1}]^T$.

A vector $\mathbf{n}_l$ is a segment of a noise process whose variance has been reduced to $\frac{\sigma_v^2}{M}$ by low-pass filtering. Noise vectors are uncorrelated zero-mean, and their mutual correlation is also made negligible by low-pass filtering within the MWC, hence: $\mathrm{E}[\mathbf{n}_l^T \mathbf{n}_{l'}] \simeq 0$ for $l \neq l'$, the symbol $\mathrm{E}[\cdot]$ indicating the expected value. Therefore, one has: $\mathrm{E}\left[\frac{1}{P}\mathbf{N}\mathbf{N}^T\right] = \frac{\sigma_v^2}{M} \cdot \mathbf{I}$, with **I** an $M \times M$ identity matrix.

One may also consider a sub-band signal-to-noise ratio (SNR), defined as:

$$SNR_l = \frac{\|\mathbf{s}_l\|_2^2}{\mathrm{E}\left[\|\mathbf{n}_l\|_2^2\right]} = \frac{\frac{1}{P}\mathbf{s}_l^T \mathbf{s}_l}{\frac{\sigma_v^2}{M}}. \tag{2.10}$$

It has to be remembered this quantity is a "virtual" SNR, since sub-bands at the MWC input are actually not accessible. It will however be useful as a reference later on, to better understand the behaviour of the CS algorithm with regards to noise.

2.4.3 Sparsity Assumption and MWC Design Trade-Off

The measurement matrix **A** has size $m \times M$ with $m < M$, consequently, the set of equations (2.9) is underdetermined. The problem is overcome by the *a priori* assumption that **X** is *sparse*, that is, only a few among its rows contain non-zero signal components. In the VSA measurement context, sparsity thus refers to sub-bands within the monitored frequency span. As most RF sources employ signal bursts, the hypothesis that active signal sources are found only in a few sub-bands is representative of the actual situation, provided the observation interval (i.e., the total measurement time NT) is short enough. It has to be remarked that the sparsity hypothesis refers to the fact that only few among the *signal* vectors $\mathbf{s}_l$ are expected to be non-zero. On the contrary, since noise components $\mathbf{n}_l$ will *all* have non-zero variance, $\mathbf{X} = \mathbf{S} + \mathbf{N}$ could never be strictly sparse, unlike **S**. In its application to (2.9) the

definition has to be modified accordingly, so that only row vectors of **X** significantly above the noise level are considered "non-zero".

The practical implication of this revised criterion is that a signal contribution in the l-th sub-band is deemed significant (i.e., $l \in S$) when its *total* power, averaged over the whole observation interval, exceeds a suitably defined threshold, ideally: $SNR_l > 1$. From this viewpoint, short high-power bursts and longer, low-power bursts can be equally detectable.

With these assumptions, (2.9) can be referred to a class of CS problems called *multiple measurement vectors* (MMV) [46].

Finding a sparse solution for the unknown **X** in (2.9) allows to extract time domain sequences associated to specific individual sub-bands, so that each can be processed to obtain the preliminary information required by the VSA, and subsequently analysed as an individual entity by the instrument modulation domain algorithms .

Formally, the *support* of the matrix **X** is defined as the set S of indices such that, if $l \in S \subset \{0,\ 1,\ \ldots,\ M-1\}$ vector $\mathbf{x}_l$ is non-zero (i.e., $|x_l[pMT]| \neq 0$ for at least one value of p). In practice, an assumption on the *maximum* signal sparsity level can generally be made *a priori*, limiting the cardinality of the support to at most $|S|_{max}$.

Since the sparsity pattern is considered independent of the column index p, the columns of matrix **X** are said to be *jointly* sparse. Provided it can be reduced to just $|S|$ unknowns, (2.9) could be solved as soon as $m = |S|$, allowing to de-alias measurements without loss of information.

Design parameters for the MWC are then the two integers m and M, that also determine the size of the measurement matrix **A**. From what has been said above, M determines the partitioning of the frequency axis, while the number m of MWC output channels should be chosen to reflect the sparsity level of the analysed signals. However, elements of S are not known in advance and have to be found by optimization algorithms that typically require $m \geq 2 \cdot |S|$ to solve the problem, if **A** has full rank m [47, 18]. Accordingly, it suffices for m to be not less than $2 \cdot |S|_{max}$. Making M as large as possible, within the bound $M < 2^m - 1$ mentioned above, could be attractive in the interest of increasing the compression factor. However, this may not always be the best choice, as it should be reminded that increasing M results in narrower spectrum slices, making subsequent steps of the CS algorithm potentially

harder to manage. As in any measurement problem, a suitable tradeoff has to be found, considering as well signal bandwidths, SNR, etc.

2.5 Blind Sensing and Finite Support Recovery

Finite support recovery is aimed at the determination of the support of $\mathbf{X}$, according to the notion of sparsity introduced in Section III. In a VSA application this stage achieves blind sensing of active spectrum sub-bands, by indicating the index values of the vectors $\mathbf{x}_l$ that contribute significantly to the MWC outputs.

The process is based on the singular-value decomposition (SVD) of the matrix $\mathbf{YY}^T$, with the aim of reducing the MMV problem dimension. As this matrix is symmetric, SVD yields:

$$\mathbf{YY}^T = \mathbf{V}_Y \cdot \mathrm{diag}[\lambda_i^2] \cdot \mathbf{V}_Y^H, \tag{2.11}$$

where λ_i, $i = 1, \ldots m$ are the eigenvalues of $\mathbf{YY}^T$, with $\lambda_1 > \lambda_2, \ldots, \lambda_m$. The columns of $\mathbf{V_Y}$ form the corresponding set of eigenvectors: $\mathbf{V_Y} = [\mathbf{v}_Y[1]\, \mathbf{v}_Y[2]\, \ldots\, \mathbf{v}_Y[m]]$.

It was shown in [45] that, when noise is negligible, $\mathbf{X}$ has the same support as the matrix $\mathbf{U}$ obtained by solving the matrix equation $\mathbf{V_Y} = \mathbf{AU}$ under a joint sparsity constraint. For the noiseless case, then, information about the sparse support is recoverable from almost any eigenvector of $\mathbf{YY}^T$. In the more realistic situation considered in (2.9), eigenvalue-based subspace decomposition can help convey better information content.

It is first necessary to determine whether signals are actually present within the observation interval. A detection criterion can be based on the ratio λ_1/λ_m of the largest to smallest eigenvalue. When only noise is present, elements of $\mathbf{Y}$ can be described by independent and identically distributed (i.i.d.) complex Gaussian random variables and $\frac{1}{P}\mathbf{YY}^H$ is a complex Wishart matrix. Its statistical properties have been widely studied, in particular the signal detection threshold:

$$\frac{\lambda_1}{\lambda_m} > \frac{(\sqrt{M}+\sqrt{P})^2}{(\sqrt{M}-\sqrt{P})^2}\left[1+\frac{1}{\sqrt[3]{\left(\sqrt{M}+\sqrt{P}\right)^2\sqrt{MP}}}\right], \tag{2.12}$$

depends only on the dimensions of matrix $\mathbf{Y}$ and can be shown to result in a probability of false alarm not larger than 5% [48].

For our purposes, this leads in practice to the condition that a useful signal is detected when the eigenvalue spread is at least an order of magnitude, that is, the de-

tection threshold is $\lambda_1/\lambda_m \leq 10$, for which the probability distribution of the largest eigenvalue was obtained in [49]. Based on results provided there, a suitable threshold for detecting signal components could be determined, keeping the effect of statistical variability into account. For instance, a probability of false alarm not larger than 5% is guaranteed by the following threshold value:

$$\frac{\sigma_v^2}{M}\cdot\left(1+\sqrt{\frac{M}{P}}\right)^2\cdot\left[1+\sqrt[3]{\frac{1}{\sqrt{M}}+\frac{1}{\sqrt{P}}}\right] \tag{2.13}$$

However, in practice the criterion requires a preliminary estimate of noise variance σ_v^2/M. The threshold is considered reliable till the signal-to-noise ratio is not lower than -10 dB. For instance, let us consider an experimental setting where the sampling frequency F_s is equal to 1 MHz, the number of sub-bands M is equal to 195, the number of mixer channels is equal to 50, and the number of measurements per channel P is equal to 121. Fig. 2.4(a) depicts the eigenvalues distribution relative to three equal amplitude sinusoidal tones, whose frequency is given by:

$$f_i = l_i\cdot\frac{1}{MT} \qquad i=\{1,2,3\} \quad l_i=\{2,37,45\} \tag{2.14}$$

A similar condition is represented in Fig. 2.4(b) where the only difference is that the three sinusoidal tones are still contained in a single sub-band block, but their frequency no longer coincides exactly with the sub-band centre:

$$f_i = l_i\cdot\frac{1}{MT} \qquad i=\{1,2,3\} \quad l_i=\{2,37.2,45.6\} \tag{2.15}$$

Looking at Fig. 2.4, it is worth noticing that, on the left side, there exists a single eigenvalue, the first one, which clearly emerges from the others; whereas, on the right side, three eigenvalues stand above the others. This phenomenon is motivated by the peculiar MWC acquisition protocol. Once demodulated in baseband, the three components of eq. (2.14) provide the same contribution and cannot be discriminated. Conversely, the three components of eq. (2.15) present a specific remaining modulation contribution, resulting into three different eigenvalues. Accordingly, this eigenvalue-based criterion is not a measure of signal sparsity level, but rather a measure of similarity between informative components.

In order to thoroughly understand which factors and parameters affect the eigenvalue distribution, we consider first the properties of the sample correlation matrix $\frac{1}{P}\mathbf{X}\mathbf{X}^H$, whose expected value is:

$$\mathrm{E}\left[\frac{1}{P}\mathbf{X}\mathbf{X}^H\right]=\frac{1}{P}\mathbf{S}\mathbf{S}^H+\frac{\sigma_v^2}{M}\cdot\mathbf{I}. \tag{2.16}$$

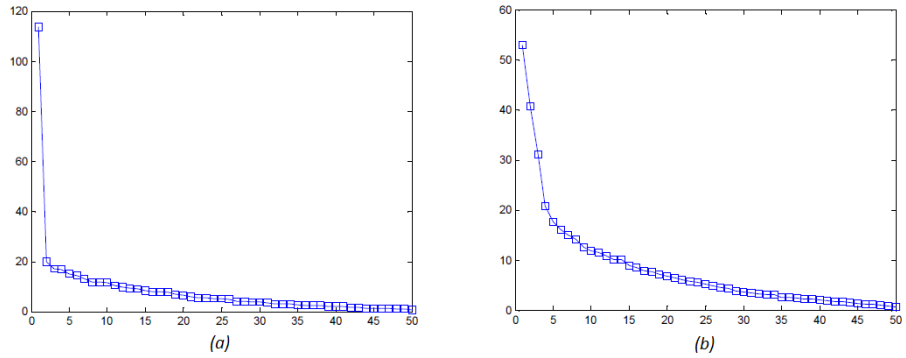


Figure 2.4: Eigenvalues distribution of the measurement autocorrelation matrix: the eigenvalue magnitudes and corresponding indices are presented respectively on the ordinates and coordinates axis. In (a) all the frequencies are centred around the corresponding sub-band; in (b) each frequency has a specific shift with respect to the corresponding sub-band centre.

For the signal component, SVD of the Hermitian matrix $\frac{1}{P}\mathbf{S}\mathbf{S}^H$ yields:

$$\frac{1}{P}\mathbf{S}\mathbf{S}^H = \mathbf{V}_S \cdot \mathrm{diag}[\xi_i] \cdot \mathbf{V}_S^H, \tag{2.17}$$

with ordered eigenvalues ξ_i, $i = 1, \dots M$. Columns of the unitary matrix $\mathbf{V_S}$ form the corresponding ordered set of eigenvectors. Given the sparsity assumption, the rank of $\mathbf{S}\mathbf{S}^H$ cannot be greater than $|S|$, therefore $\xi_i = 0$ for $|S| < i \leq M$. On the other hand, the presence of a noise component ensures the matrix has full rank, since the noise contribution is spread over the whole space spanned by the eigenvectors of matrix $\mathrm{E}\left[\frac{1}{P}\mathbf{X}\mathbf{X}^H\right]$. Therefore ,its eigenvalues are equal to $\frac{\sigma_v^2}{M}$ for $|S| < i \leq M$ and approximately equal to $\xi_i + \frac{\sigma_v^2}{M}$ for $1 < i \leq |S|$.

This space can be decomposed into orthogonal signal and noise subspaces, based on the assumption that eigenvectors whose associated eigenvalues are larger than $\frac{\sigma_v^2}{M}$ span the signal subspace and the remaining ones are related to the noise subspace. The former are (approximately) sparse, whereas the latter contain no useful information for support recovery.

Of course, $\mathbf{X}\mathbf{X}^H$ is not known in practice, and attention has to be turned to the $m \times m$ MWC output sample correlation matrix $\frac{1}{P}\mathbf{Y}\mathbf{Y}^T$. Since, recalling (2.9), its expected value can be written as:

$$\mathrm{E}\left[\frac{1}{P}\mathbf{Y}\mathbf{Y}^T\right] = \mathbf{A}\mathrm{E}\left[\frac{1}{P}\mathbf{X}\mathbf{X}^H\right]\mathbf{A}^H, \tag{2.18}$$

we need to analyze how signal and noise subspaces are mapped through **A** to see whether subspace decomposition can still be considered.

Let **v** be a generic vector belonging to the noise subspace. From the inequality: $\|\mathbf{Av}\| \leq \|\mathbf{A}\|\|\mathbf{v}\|$ it follows that worst case mapping occurs when: $\|\mathbf{Av}\|_2^2 = \lambda_{A\_max}\|\mathbf{v}\|_2^2$, where λ_{A_max} is the largest eigenvalue of **A**. As this is a random matrix whose elements satisfy suitable bounds on their moments, its first eigenvalue converges to:

$$\lambda_{A\_max} \longrightarrow \left(1+\sqrt{\frac{m}{M}}\right)^2 \tag{2.19}$$

where $m < M$ holds for the measurement matrix [50]. This means norms in the noise subspace are not increased significantly. This limit function has been experimentally demonstrated in a practical case where $m = 10$, and $M = 51$: the computed maximum eigenvalue and its theoretical expected value are, respectively, equal to 1.781 and 1.876.

Vectors in the signal subspace are approximately sparse, therefore in their case we need to refer to the *restricted isometry property* (RIP) in relation to matrix **A**. This is said to have the RIP of order K if the following inequalities hold [51]:

$$(1-\delta_K)\frac{m}{M}\|\mathbf{v}\|_2^2 \leq \|\mathbf{Av}\|_2^2 \leq (1+\delta_K)\frac{m}{M}\|\mathbf{v}\|_2^2, \tag{2.20}$$

with $0 \leq \delta_K < 1$, for any vector **v** having at most K non-zero elements (i.e., *K-sparse*). The worst case mapping for a generic $|S|$-sparse vector **v** in the signal subspace is obtained by taking the lower bound in (2.20) with $\delta_K = \delta_{|S|}$. To ensure (2.9) has a unique $|S|$-sparse solution, the RIP of order $2|S|$ is required, which measurement matrix **A** is known to satisfy as long as $m \geq |S|\log N$ [52]. Of course, this implies RIP of order $|S|$, hence a value $0 \leq \delta_{|S|} < 1$ does exist.

The ratio of the noise upper bound to the signal lower bound:

$$\frac{\left(1+\sqrt{\frac{m}{M}}\right)^2}{(1-\delta_{|S|})\frac{m}{M}} = \left[\frac{\left(1+\sqrt{\frac{m}{M}}\right)^2}{(1-\delta_{|S|})}\right]\cdot\frac{M}{m} \tag{2.21}$$

shows that the MWC output SNR is degraded, with respect to the "virtual" input SNR, by a factor which is slightly larger than the compression factor $\frac{M}{m}$. The phenomenon is known as *noise folding* [53] and ratio (2.21) accounts for loss of sensitivity by eigenvalue-based detection. Nevertheless, the result confirms that the approach is applicable to the MWC output sample correlation matrix as long as SNR is large enough.

Contributions from different active sub-bands are thoroughly mixed through the measurement matrix **A**. Therefore, information about the sparsity pattern of **X** can be recovered from a single eigenvector of $\frac{1}{P}\mathbf{Y}\mathbf{Y}^T$, provided it is selected from among those associated with the signal subspace.

This would allow to:

- determine whether signals are present within the observation interval;
- possibly perform decomposition into orthogonal signal and noise subspaces.

On this basis, when a signal is detected within the observation interval we shall employ for support recovery the *principal component* of the MWC output sample correlation matrix, which arguably ensures the best SNR.

Simulation analysis with the kind of signals typical of VSA measurements confirmed that the *principal eigenvector* $\mathbf{v}_Y[1]$ contains enough information to allow recovery of the sparse support S in most cases. The joint support of **X** is then determined as the set of locations of the non-zero elements of the vector **u** that, by an orthogonal matching pursuit (OMP) algorithm [54], is found to be the sparsest solution of equation $\mathbf{v}_Y[1] = \mathbf{A}\mathbf{u}$.

The soundness of the approach was confirmed by several simulation tests. It should be remembered that, with m being seldom much larger than order of 10 in practice, the $m \times m$ matrix $\frac{1}{P}\mathbf{Y}\mathbf{Y}^T$ has a comparatively small size and computational effort is acceptable. It is also important to note that the process is entirely driven by the acquired data, which gives the CS algorithm its spectrum-blind sensing capabilities.

2.6 Signal Reconstruction

Once a finite support, represented by a (small) set of indices $S \subset \{0, 1, \ldots, M-1\}$ is determined, a correspondingly *restricted* measurement matrix $\mathbf{A_S}$ can be obtained by keeping only the columns of **A** whose column indexes l satisfy the condition: $l \in S$. While solving the underdetermined matrix equation (2.5) is not possible in general, the sparsity assumption here comes into play and makes extraction of significant contributions feasible [44].

The signal components of interest are reconstructed by means of the pseudo-inverse:

$$\hat{\mathbf{S}}_S = \mathbf{A}_S^{\dagger}\mathbf{Y} = \left(\mathbf{A}_S^H\mathbf{A}_S\right)^{-1}\mathbf{A}_S^H\mathbf{Y}, \tag{2.22}$$

where the rows of matrix $\hat{\mathbf{S}}_\mathbf{S}$ contain the estimates of the subsequences $s_l[p]$ for $0 \le p < P$, with $l \in S$. Thus, individual contributions from each wireless source corresponding to a sub-band, within the observed IF band and time interval determined by the VSA, can be returned separately.

To analyze reconstructed signals accuracy, (2.22) can be rewritten as:

$$\hat{\mathbf{S}}_S = \mathbf{S}_S + \mathbf{A}_S^{\dagger}\mathbf{A}\mathbf{N}, \tag{2.23}$$

where the rows of matrix $\mathbf{S}_\mathbf{S}$ contain the extracted components, while the second term on the right-hand side is the noise contribution.

A feature of particular interest is the de-noising effect, which is inherent in any procedure where only the main components of a signal are picked out. To understand the noise relationship between the acquired record $x[nT]$ and the reconstructed signal components, We consider again how a generic vector $\mathbf{v}$ belonging to the noise subspace would be mapped through (2.23), using the norm inequality: $\|\mathbf{A}_S^{\dagger}\mathbf{A}\mathbf{v}\| \le \|\mathbf{A}_S^{\dagger}\|\|\mathbf{A}\|\|\mathbf{v}\|$. The same inequalities given in (2.20), with $\delta_K = \delta_{|S|}$, also provide bounds for the eigenvalues of the restricted matrix $\mathbf{A}_\mathbf{S}$. Furthermore, it was shown in [17] that the singular values of $\mathbf{A}_S^{\dagger}$ are reciprocals of the singular values of $\mathbf{A}_\mathbf{S}$. It follows that:

$$\|\mathbf{A}_S^{\dagger}\mathbf{A}\mathbf{v}\|_2^2 \le \frac{M}{m}\cdot\frac{\left(1+\sqrt{\frac{m}{M}}\right)^2}{(1-\delta_{|S|})}\|\mathbf{v}\|_2^2. \tag{2.24}$$

Further simulation analyses confirmed that noise superposed on the extracted signal components can be described by an i.i.d. vector random process with approximately uncorrelated components, and the expected value of its sample correlation matrix is:

$$\mathrm{E}\left[\frac{1}{P}\mathbf{A}_S^{\dagger}\mathbf{A}\mathbf{N}\mathbf{N}^H\mathbf{A}^H(\mathbf{A}_S^{\dagger})^H\right] \simeq \frac{\sigma_v^2}{m}\cdot\frac{\left(1+\sqrt{\frac{m}{M}}\right)^2}{(1-\delta_{|S|})}\mathbf{I} \tag{2.25}$$

It should be remarked that element values of $\mathbf{A}_\alpha$ are actually fixed, once the MWC channel patterns have been chosen. However, considering average properties of a Bernoulli matrix will be useful later on in the discussion of signal reconstruction uncertainty, when analysis of estimation noise statistical properties will be conditioned on the value of $\mathbf{A}_\alpha$.

Thus, as a rough rule of thumb, SNR for a recovered signal components satisfies:

$$\hat{SNR}_l \leq m \cdot \frac{\frac{1}{P}\mathbf{s}_l^T \mathbf{s}_l}{\sigma_v^2}, \qquad l \in S \tag{2.26}$$

It should be noted that, depending on the monitored signals and the outcome of the support recovery step, different column subsets of matrix **A** can be selected. RIP bounds account for the eigenvalue spread due to this fact. Tighter bounds could be obtained by introducing assumptions on the probability distribution of index values $l \in S$. It should also be observed that the RIP constant $\delta_{|S|}$ is referred to the whole set of $m \times M$ random matrices **A** . The analysis of possible dependence on specific sign patterns in (2.7) is still an open research issue.

2.6.1 Noise Folding

In CS literature, noise effects over a reduced set of measurements are widely discussed. In particular, the impact of additive Gaussian noise in sparse signal reconstruction is also known as *noise folding* phenomenon, which is dealt with by the following theorem.

Theorem 2.6.1. *Suppose that* **A** *satisfies the RIP of order k with isometry constant* δ_k*. Furthermore, suppose that the rows of* **A** *are orthogonal and that each row of* **A** *has equal norm. If* **N** *is a zero-mean, white random vector with* $\mathbb{E}(\mathbf{N}\mathbf{N}^T) = \sigma^2\mathbf{I}$*, then* **AN** *is also a zero-mean, white random vector with* $\mathbb{E}(\mathbf{A}\mathbf{N}\mathbf{N}^T\mathbf{A}^T) = \tilde{\sigma}^2\mathbf{I}$*, where:*

$$\frac{n_s}{n_m}\sigma^2(1-\delta_k) \leq \tilde{\sigma}^2 \leq \frac{n_s}{n_m}\sigma^2(1+\delta_k) \tag{2.27}$$

where n_s *is the sparse representation cardinality and* n_m *is the number of measurement collected [17].*

Theorem 2.6.1 implies two main considerations. On the one hand, the CS acquisition process does not affect the additive noise structure, which remains white and Gaussian. On the other hand, the noise level could be highly amplified if the compression ratio is excessive.

Intuitively, this result makes sense and provides an immediate motivation from a linear algebra point of view. The measurement matrix **A** projects the entire noise (originally spread out over n_s coefficients) down into n_m measured coefficients (with

$n_m \ll n_s$, by definition). The orthogonal projection operation is expected to preserve the noise power. Accordingly, the average noise level affecting the measured coefficients is expected to increase with respect to the original conditions.

In the present case, the noise folding phenomenon is partially balanced by the sub-bands decomposition. In other words, the CS acquisition considers each sub-band separately: accordingly, the input noise level is the result of baseband demodulation and subsequent low-pass filtering stage.

The following theorem allows to quantify the SNR loss due to noise folding:

Theorem 2.6.2. *Suppose that* $\mathbf{A}$ *satisfies the RIP of order* k *with isometry constant* δ_k. *Furthermore, suppose that the rows of* $\mathbf{A}$ *are orthogonal and that each row of* $\mathbf{A}$ *has equal norm. If* $\mathbf{N}$ *is a zero-mean, white random vector with* $\mathbb{E}(\mathbf{N}\mathbf{N}^T) = \sigma^2\mathbf{I}$, *then the expected SNR loss is bounded by [17]:*

$$\frac{n_s}{n_m}\sigma^2(1-\delta_k) \leq \text{SNR loss} \leq \frac{n_s}{n_m}\sigma^2(1+\delta_k) \tag{2.28}$$

From its proof it immediately descends the loss assessment in decibels:

$$\text{SNR loss [dB]} \simeq 10\log_{10}\left(\frac{n_s}{n_m}\right) \tag{2.29}$$

As a consequence, if n_m is halved (or correspondingly the under-sampling ratio is increased by one octave), the expected SNR loss increases by 3 dB.

2.7 Algorithm Characterization

2.7.1 Synthetic measurements - Single burst

Pre-processing for VSA measurements can be seen as a spectrum-blind sensing application. The structure of the algorithm outlined in the previous section appears to be ideally suited to this kind of problem and, in fact, motivates this investigation.

The CS algorithm was tested first with synthetic measurement data that reproduce, in simplified form, a signal digitized either by a high-speed baseband sampling device or at the IF filter output of a VSA. This provides a controlled test environment with exactly known carrier frequencies and RF burst positions.

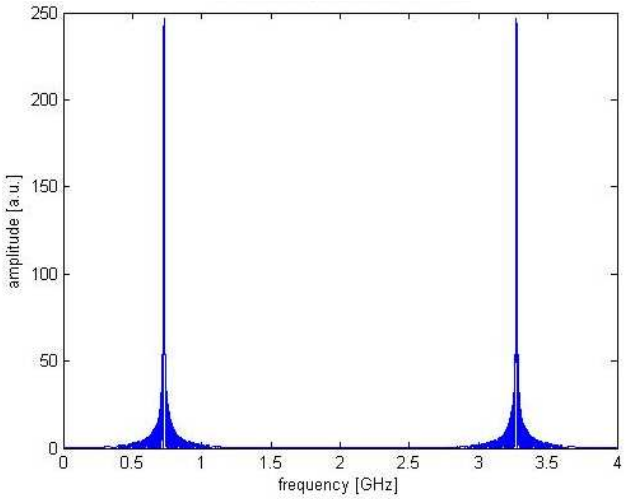


Figure 2.5: Spectrum of a short RF burst. Carrier frequency: $f_B = 727.405$ MHz.

2.7.2 Signal Features

We consider first a single burst described by the equation:

$$b(t) = A \cdot p_{T_B}(t - \tau_B) \cdot \cos\left[2\pi f_B (t - \tau_B) + \phi\right], \tag{2.30}$$

where f_B, A and ϕ are, respectively, the carrier frequency, amplitude and instantaneous phase (referred to the mid-point) of the burst. The function $p_{T_B}(t)$ limits the burst length to a duration T_B and is centred on time τ_B. In a modulated signal A and ϕ are time varying functions of time.

To model typical transmitter turn-on/off transients, $p_{T_B}(t)$ has been defined as a cosine tapered window, i.e., a rectangle whose edges are smoothed by convolution with a raised cosine function $w_C(t)$ of length T_C:

$$p_{T_B}(t) = w_C(t) * \mathrm{rect}\left(\frac{t}{T_B}\right). \tag{2.31}$$

Total signal length is then $T_B + T_C$, where T_C is the length of $w_C(t)$. The duration T_B is correctly measured between the mid-point levels of the on/off transients. Taper is typically a short fraction of the overall length, the assumed range of values of the ratio $\alpha = T_C/(T_B + T_C)$ being between 0.01 and 0.05.

In the following we consider a record of $N \simeq 30000$ samples at the baseband sampling rate of 4 GHz. A burst of length $T_B = 100$ ns (i.e., 400 samples), with tapering parameter $\alpha \simeq 0.025$ and carrier frequency $f_B = 727.405$ MHz is positioned in the

middle of a 6.5 μs observation interval. Hence, the test signal in this first case is $s(t) = b(t)$, with A and ϕ constant. To understand its relevance to MWC characterization it should be noticed that, in terms of the bandwidth to center frequency ratio, it is representative of an ultra-wide band (UWB) signal. Cosine tapering ensures that its spectral content does not practically extend beyond $\pm(2/T_B)\cdot(1/\alpha)$ from the center frequency f_B. Fig. 2.5 shows a bandwidth of approximately 800 MHz, as expected. Because of the small value of α, the burst spectrum within that band approximates the $(\sin x/x)$ behaviour of an non-tapered rectangle window.

2.7.3 MWC Detection and Recovery

We refer to a MWC where the aliasing rate is set to $M = 195$, so that the width of each spectral sub-band is about 20.5 MHz. This bandwidth can account for approximately 80% of the power of $s(t)$, which is still acceptable for the eigenvalue-based detection discussed in Section IV. The MWC mixing stage is composed of $m = 50$ channels, which results in matrix $\mathbf{Y}$ having size 50×151 and in a reduction of the overall amount of data by almost 75%.

It has to be noticed that assumed burst length is close to the lower time resolution limit of this MWC, as it corresponds to at most 3 sample intervals at the output sample rate. This analysis exemplifies the short high-power burst case since, with $T_B = 100$ ns, signal length is less than 2% of the observation interval, resulting in an actual peak power to noise ratio nearly 20 dB better than the given SNR values, where the total signal power within the observation interval is considered.

Another peculiarity is that carrier frequency f_B has been intentionally located almost exactly in between two MWC sub-bands. The test signal support is $S = \{35, 36, 160, 161\}$, inclusive of image components. Recalling (2.21), any residual contribution outside the four sub-bands is unlikely to be detected. In any practical setting this situation needs to be addressed, since the carrier position with regards to the frequency grid induced by the MWC is not known in advance. Here, any two adjacent sub-bands with $l \in S$ are considered as parts of the same spectral component and, during recovery, their contributions are recombined to describe the entire signal information[1].

[1] The issue can be suitably formalized by introducing the notion of *block* sparsity [55]. We leave this to future research developments.

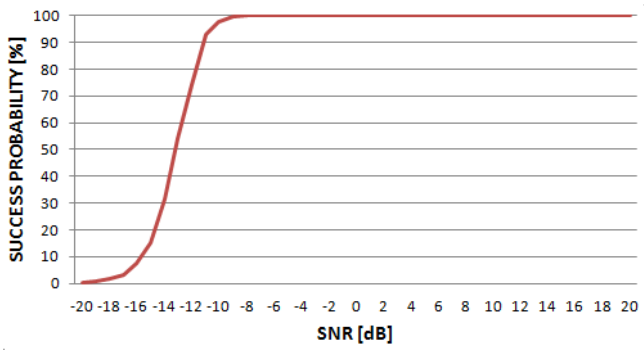


Figure 2.6: Support recovery: success probability versus SNR.

SNR [dB]	**Success** [%]
−5	100
−10	97.6
−15	15.2
−20	0.3

Table 2.1: Support recovery success probability for negative SNR values.

2.7.4 Support Recovery Success Probability

We investigate algorithm reliability as a function of SNR. For this purpose the OMP algorithm was run and support recovery performed without applying the detection threshold given in (2.12). Simulations were repeated until a set of 200 successful trials were obtained, success meaning that the recovered support coincides with the nominal one.

White Gaussian noise, uncorrelated with the burst, was added to $s(t)$ and its power varied to obtain SNR values ranging from −20 to +20 dB. The plot of Fig. 2.6 relates measured *success probability* to specific SNR values and shows that support recovery becomes unreliable when SNR gets lower than −10 dB. Success probability quickly falls off to almost negligible values below this threshold. In the following, reconstructed signals from successful recoveries with SNR > −10 dB are analysed to evaluate measurement uncertainty.

A too large noise level conceals the signal spectral content under a uniform trace

of uncorrelated fluctuations. In similar conditions, also the auto-correlation matrix hardly exhibits significant contributions. The support recovery necessarily fails and returns different values at each simulation, related with the noise and no more with the signal. The scarce success rate in Tab. 2.1 can be considered a consequence of this phenomenon. Under the threshold, any successful simulation is more likely due to random circumstances than to effective algorithm capabilities.

2.8 Single-Carrier Digital Modulation

Preliminary knowledge of the carrier frequency (and modulation format) is essential for VSA measurements on single-carrier modulations. The minimum requirement is to provide an initial estimate that allows VSA algorithms to lock on the carrier frequency. For instance, with the Agilent 89600 VSA software toolset the instrument center frequency would be set to the *estimated* carrier frequency which, as a general rule, is allowed to differ from the actual carrier frequency by no more than ±3% of the symbol rate employed by the modulation under analysis [56].

2.8.1 Carrier Frequency Estimation

The CS-based pre-processor allows direct estimation of carrier frequency from the extracted signal component samples $\hat{s}_l[pMT]$, $l \in S$. Sequences returned by (2.23) are downshifted into the baseband spectrum slice, which is helpful as a first demodulation step for vector measurements. However, unless the carrier frequency of the wireless source represented by $s_l[p]$ is exactly in the middle of the associated spectrum slice (i.e., $f_B = l/MT$), a second step is required to estimate f_B.

Since selected sub-bands indexed by the support S can be dealt with individually[2], it is possible to assume in advance that a single spectral peak is present within a sub-band. Therefore, considering the magnitude of the discrete Fourier transform (DFT) $|\hat{S}_l[k/NT]|$, the carrier frequency corresponds (with the notable exception of frequency-shift keying modulation) to the location of the one spectral peak, provided the modulation spectrum is symmetric. Frequency resolution is proportional to the grid step $(1/NT)$, since the observation interval length remains equal to NT also at the MWC output.

[2] At most, two adjacent sub-bands are conisidered as a pair and recombined, as mentioned previously.

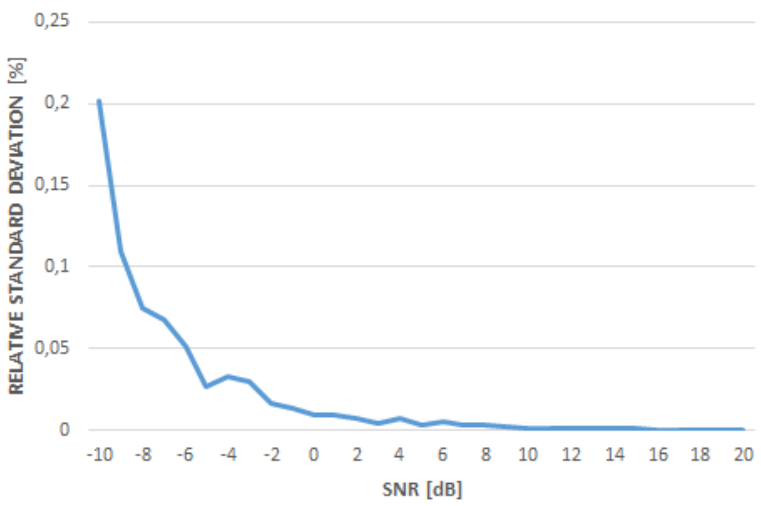


Figure 2.7: Carrier frequency estimation standard deviation. The spectrum main lobe has been pre-smoothed by a Savitzky-Golay digital filter.

Carrier frequency can then be determined to within $\pm(1/2NT)$ of its actual value. With the observation interval considered in section 2.7.1, this value is approximately ± 75 kHz. This result can be enhanced by about one order of magnitude with two- or three-point interpolation [57], using DFT coefficients adjacent to the peak.

Applying interpolation turns out to be comparatively simple for an unmodulated UWB-like burst, less so for a typical digital modulation. In the latter case, in fact, the signal spectrum is characterized by a broader and flatter main lobe, which makes the preliminary peak search step vulnerable to even moderate amounts of noise. To counter this, we pre-filter $|\hat{S}_l[k/NT]|$ by a digital Savitzky-Golay filter [58] to smooth the magnitude spectrum. In Fig. 2.7 the relative standard deviation of carrier frequency estimation is presented. As expected, the estimation variability uncertainty decreases as SNR increases.

2.8.2 Time-Domain Accuracy: Burst Length and Position

In the time domain, primary signal parameters are the length T_B and position of the burst. Using the selected MWC output channels, baseband contributions are recovered according to (2.22), then interpolated and up-shifted back to the original carrier frequency. Burst length T_B is estimated as $N_B T$, where N_B is the number

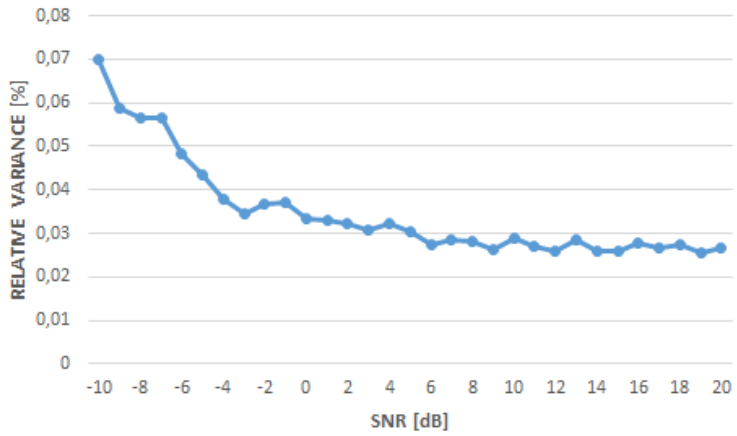


Figure 2.8: Burst length estimation variance. The burst length has been assessed by means of a threshold process.

of samples whose amplitude exceeds a given threshold. Position τ_B is defined as the mean between the burst endpoints determined by this threshold. In this case, the amplitude threshold is set to half the nominal rms value defined in the signal generation stage.

Accurate time information allows VSA algorithms to correctly define the time window required to demodulate and analyse specific signal bursts, even when multiple sources are simultaneously present. Fig. 2.8 shows that variance decreases significantly as SNR improves. For SNR ≥ 0 dB the two-sigma confidence level is nearly equal to ± 10 samples, i.e. ± 2.5 ns. In the present configuration, the confidence level corresponds nearly to two periods of the observed carrier frequency. This estimation error is consistent with any detection method based on threshold crossing.

2.8.3 Waveform Accuracy

For the purposes of this work the accuracy of waveform recovery is characterised with regards to typical VSA measurements. The aims are:

- ensure that the modulation format associated to the extracted component can be identified;
- determine whether signal reconstruction is accurate enough for VSA measurements.

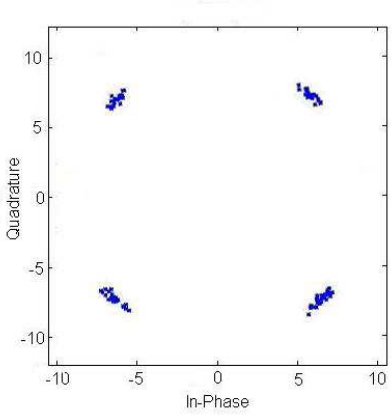


Figure 2.9: Constellation for recovered burst - no trace noise.

To summarize these features into a readily understandable result, we analysed constellation diagram measurements obtained from the CS-recovered signal burst, and compared them with the reference one.

Argument-modulation was considered for the test, since this is common to widely employed formats in RF wireless communications. To represent the signal, (2.30) can be formally modified by replacing the constant ϕ with the function $\phi(t)$ that represents the argument modulation function associated to burst $b(t)$.

We report the analysis for a modulation format employing four symbols (which can be, equivalently, a 4-QAM or a QPSK modulation, since in both cases amplitude remains constant). As a suitably large number of recovered symbols is needed to allow reliable constellation measurement, typical burst length in this analysis is about 30% of the whole trace. Accordingly, the observation interval lenght is set equal to 2.5 ms, whereas the symbol rate is fixed at 625 kHz.

Only a single burst is contained in the sequence $x[nT]$, so that no adverse effects can be caused by either time-domain or frequency-domain interference among multiple bursts.

To establish a baseline reference, a “clean” signal coming directly from an Agilent E4433B RF signal generator was measured by an Agilent N9010A vector signal analyzer with continuous modulation (no burst). This yielded 35.8 dB modulation error ratio (MER), corresponding to 1.6% relative EVM, given as a root-mean-square (RMS) value.

A time domain trace with no additive noise was acquired first. The diagram in Fig. 2.9, obtained from the signal recovered by the CS algorithm, shows that symbols are closely clustered around their nominal positions. The residual dispersion provides an assessment of how the CS algorithm affects accuracy, since any difference from the ideal constellation can be attributed to the processing algorithm. MER was estimated to be 24 dB, which, for this kind of modulation, corresponds to a 6% EVM value. In the present case, the main factors which affect and degrade the constellation estimation are phase noise (error) and carrier frequency error. These two phenomena cause the constellation to spread around the expected symbol points and be slightly rotating around the constellation centre. Evidently, with a greater number of symbols, a greater MER is required to correctly detect the constellation. The proposed algorithm reduces the noise approximately at the same level, independently on the number of symbols. The resolution gain is almost the same for any kind of modulation format: in some cases it would be sufficient, in other it would not.

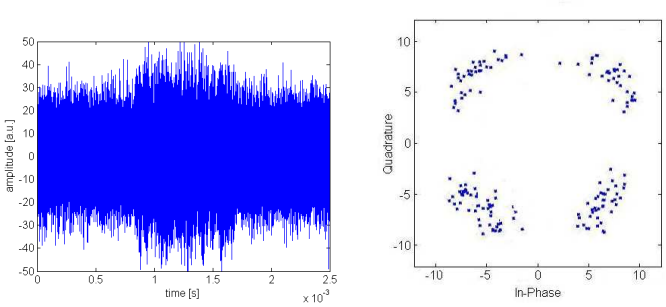
The corresponding bound on relative amplitude uncertainty for the constellation of Fig. 2.9 is approximately ±5%, while angle uncertainty is below ±10 degrees. These values reflect the basic accuracy level attainable in the current CS algorithm implementation.

It should be noticed that these results were obtained by the use of $m = 50$ parallel MWC channels, which may appear oversized if one considers the likely number of simultaneously present components in a trace. However, according to (2.26) using a larger number of channels improves reconstructed signal SNR at the price of a loss in compression. As such, this MWC size represents a trade-off between reconstruction accuracy and compression, that may need to be investigated more closely.

Tests were repeated with additive noise of increasingly higher levels, Fig. 2.10 showing the constellation obtained with a -5 dB SNR (noise power approximately three times greater than signal power).

Even with such poor signal-to-noise ratio, recovery by the CS algorithm allows to identify the constellation. Relative amplitude uncertainty has risen to about ±30% and angle uncertainty is in the region of ±30 degrees but, as this effect depends on SNR, a similar degradation in accuracy would occur with any algorithm.

The limiting SNR for the recovery of a single digitally modulated burst was found to be approximately -15 dB. At this point a burst is barely discernible from the noise



(a) original time domain trace; (b) measured constellation diagram.

Figure 2.10: Constellation diagram for recovered burst; -5 dB signal-to-noise ratio.

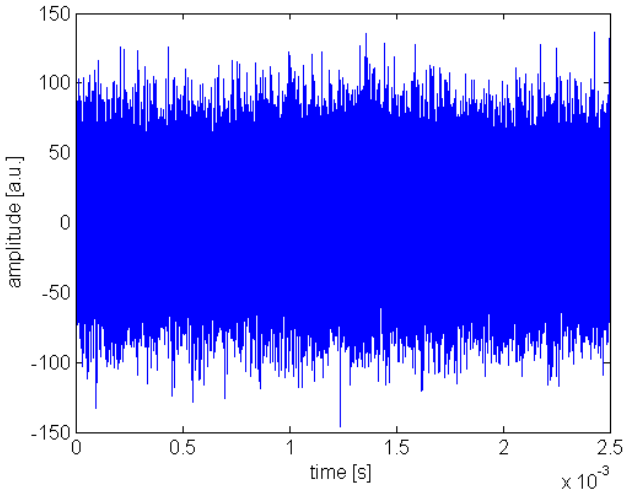


Figure 2.11: Time domain trace of burst at limiting SNR value (-15 dB).

background in the time domain trace, as shown in Fig. 2.11, although the modulation format can still be inferred from the analysis of the recovered signal. The plots of Fig. 2.12 summarise the results obtained for the variability of the measured constellation, indicated by maximum relative amplitude and phase uncertainty as a function of different SNR values.

2.9 Experimental Analysis – Multiple Bursts

The case where multiple sources using different carrier frequencies come into play, represents the typical spectral monitoring situation and a most significant test

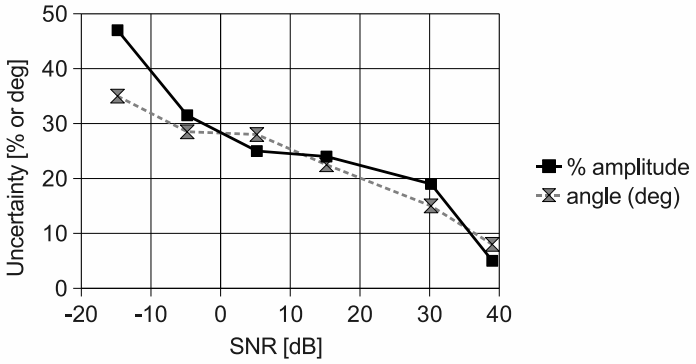


Figure 2.12: Amplitude and phase variability in constellation diagram measurements for different SNR values.

for a CS-based algorithm. Even in this case, discontinuous transmission means active sources are hardly more than a few at a time. Therefore a mathematical model for the observed signal (e.g., the one shown in Fig. 2.2) can be given as the sum of J bursts and an additive white noise term $v(t)$:

$$s(t) = \sum_{j=1}^{J} b_j(t) + v(t). \tag{2.32}$$

In terms of the CS algorithm, the sparsity level is equal to $2J$, which is the total number of bandwidths occupied by the signal at some time during the observation interval.

2.9.1 Design of Test Signals

To generate input signals that realistically reproduce typical conditions of spectrum monitoring in a wireless bandwidth. We relied on a RF arbitrary waveform generator, which allows great flexibility in the design of a variety of complex signals while ensuring controlled and repeatable test conditions.

The instruments employed in this work were an E4433B signal generator and a N9010A vector signal analyzer from Agilent Technologies. Purposely computed waveform samples were downloaded into the two signal generator RAM waveform memories (dedicated, respectively, to the in-phase (I) and quadrature (Q) components).

A preliminary check on the correct implementation of the basic waveform design was achieved by comparing the synthesized digitally modulated signals with

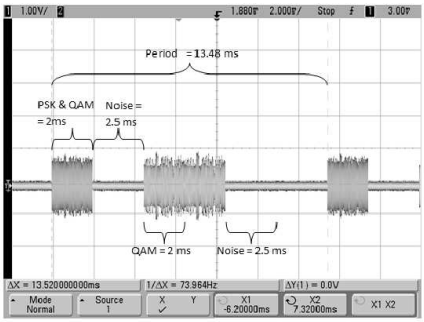


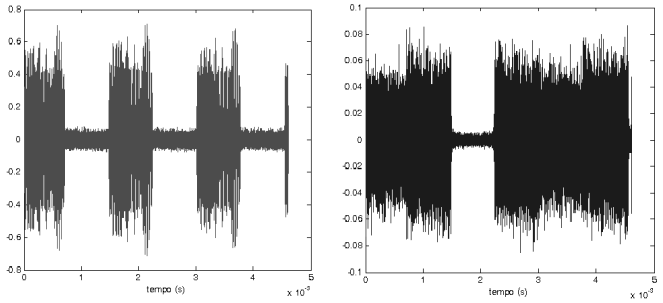
Figure 2.13: Sequence obtained from the E4433B signal generator: baseband in-phase component auxiliary output, as measured by a digital oscilloscope.

those obtained by the E4433B built-in internal signal generation options. EVM measurements obtained by the N9010A VSA gave comparable values in the two cases.

Memory in the Agilent E4433B can be partitioned into a number of individual segments, each containing a specific waveform. This allows the creation of waveform sequences by the concatenation of segments and also provides the possibility to specify for each segment the number of iterations before progressing to the next one. Sequence control represents a powerful tool for the generation of very realistic test signals, which reproduce the superposition and interference of different RF sources in a controlled way.

Each segment can be defined as the sum of up to four components, such as bursts of the form (2.30) with individually defined modulation formats. To better reproduce real-life conditions, we built segments with up to three bursts, to which arbitrary white gaussian noise (AWGN) is added. Although segments can be sequenced, the generator does not allow to change the carrier frequency concurrently. Therefore, individual bursts are digitally pre-modulated with an offset frequency, so that the combination of the two modulations yields the desired result.

The generator sample clock has a maximum frequency of 40 MHz and the output reconstruction filter bandwidth was set to 8 MHz, so that comparatively broad band wireless channels could be reproduced. Fig. 2.13 shows, as an example, the baseband in-phase component of a composite signal, where different modulated carriers alternate and superpose during a 13.5 ms period. It can be seen that the trace is obtained from three individual segments, some of which are repeated to



(a) test signal with 16-QAM modulated bursts; (b) CS algorithm reconstruction error for the test signal.

Figure 2.14: CS-based recovery of 16-QAM modulated bursts.

give a total of six. One segment contains noise only, while another presents the time superposition of two different modulated signals.

2.9.2 Signal Extraction for VSA Analysis

Results presented next refer to the monitoring and recovery of RF bursts employing a 16-QAM modulation format. Accurate vector measurements are more difficult to achieve with these signals than with four-symbol modulations, because of the closer spacing of the constellation symbols. Although denser schemes exist, these are far less common in wireless links, therefore characterisation for the 16-QAM case is considered to provide a good indication of the actual usefulness of the proposed CS-based preprocessor.

To characterise the CS algorithm, VSA measurements were taken using the widest allowable IF bandwidth for the N9010A, i.e., 10 MHZ. The instrument operates at a 12.8 MHz sampling frequency, each acquired data record numbering over $10e^5$ samples.

Set-up parameters of the CS algorithm are the decimation factor L and the number m of MWC channels. The decimation factor was chosen so that each channel produced a down-sampled sequence shorter than $10e^3$ samples, which means $L > 100$. Consequently, in the experiments that follow the 10 MHz IF bandwidth is ideally divided into sub-bands whose width is approximately 70 kHz.

The original trace, shown in Fig. 2.14(a), has been designed as the repetition of four segments, alternating between RF bursts and noise. Only one of out two

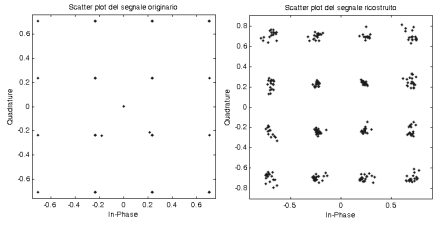
consecutive 16-QAM bursts provides a clean signal, while the other includes AWGN, whose amplitude is approximately 10% of the signal amplitude. To better understand how the CS algorithm operates, it is interesting to look at the signal reconstruction error over the whole time interval, shown in Fig. 2.14(b). In all but one of the segments the reconstruction error has the same magnitude as the AWGN component in the original trace. This means the algorithm has correctly recovered a signal component, at the same time reducing the incidence of noise. The part of the plot where the reconstruction error is one order of magnitude lower refers to the single segment containing a clean signal and shows that contributions due to the algorithm itself are limited. It is also important to note that the position and length of each burst are always determined correctly.

Tests have also been carried out with composite signals where some of the segments contain time superposed RF bursts with different carrier frequencies. In this case, the algorithm has shown the capability to correctly separate each component, allowing the user to choose between measuring each of them or discard some as interferences.

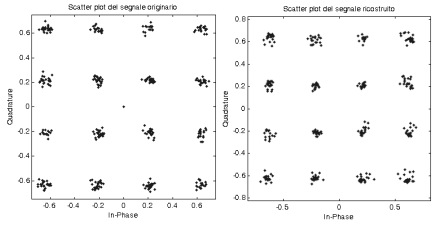
Fig. 2.15 shows that the constellation diagram of a 16-QAM modulated burst can actually be improved, after extraction by the CS algorithm, if the trace is moderately noisy. For this kind of test a 16-QAM modulation with inceasing levels of AWGN was generated and the signal was the recovered by the CS-based algorithm. In this case, the parameter that best summarises quality is MER, which was measured for the test signal at each different noise level. Constellation diagrams of the original and recovered signal are presented side by side for decreasing MER values of, 32 dB, 28 dB and 26 dB, respectively.

It is interesting to note that Fig. 2.15(b) shows very similar behaviour in the two diagrams and that the diagram for the recovered signal in Fig. 2.15(a) is also almost equal. This seems to suggest that for this kind of modulation 28 dB is approximately the limiting MER value that can be obtained with CS-based recovery. This assessment is further supported by the fact that, if one takes into account the difference in power distribution between the symbols of the two modulations, the 28 dB limit for 16-QAM turns almost exactly into the 24 dB limit already found for the QPSK modulation shown in Fig. 2.9.

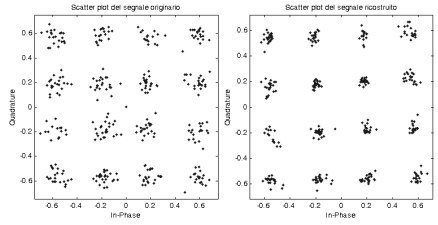
It should also be observed that CS-based recovery has impoved the constella-



(a) test signal with 32 dB MER (left); recovered signal (right);



(b) test signal with 28 dB MER (left); recovered signal (right);



(c) test signal with 26 dB MER (left); recovered signal (right).

Figure 2.15: CS-based recovery of 16-QAM modulated bursts.

tion diagram for the case of 26 dB MER, suggesting that in some cases the denoising properties of the algorithm might also be beneficial.

2.10 Histogram Analysis

In cognitive radio (CR) research and development, the aim becomes the characterization of legitimately intertwined communication signals. This involves the ability to determine the mutual coordination and timing of primary and secondary sources, verify appropriate power levels and generally help troubleshoot CR systems under development. In the latter case, it has to be further assumed that unwanted interference might have occurred.

Enhancements to basic VSA capabilities are thus desirable to effectively support CR system development. To this aim, in the section we propose the use of multimodal amplitude histograms to implement a modulation detector with spectrum-blind capabilities. We discuss its performances and show that it can be integrated with VSA measurement algorithms, providing the instrument with the ability to autonomously detect sources under analysis and identify their basic parameters.

In the previous sections, the proposed CS algorithm has been proven to be able to extract significant components from the acquired data. Hereinafter, the integration of the sensing part with a simple modulation detection procedure based on multimodal amplitude histograms is discussed and characterized by means of experimental results.

To this end, specific benchmarks have been implemented both in a purely simulation context and with synthesized signals obtained from an RF arbitrary waveform generator. A well defined and controlled environment, where impairments and sources of uncertainty are being progressively introduced, is necessary to retain the ability to manage the complex interplay of factors affecting measurement performances.

Two modulation formats have been tested, 16-QAM and 8-PSK, these being representative of amplitude and argument modulation, respectively. In the QAM case symbols would be located on an orthogonal grid along the I and Q axes.

Histograms in Fig. 2.16 in fact both present four clear peaks distributed according to a symmetric and uniform pattern. Using a comparatively small number of

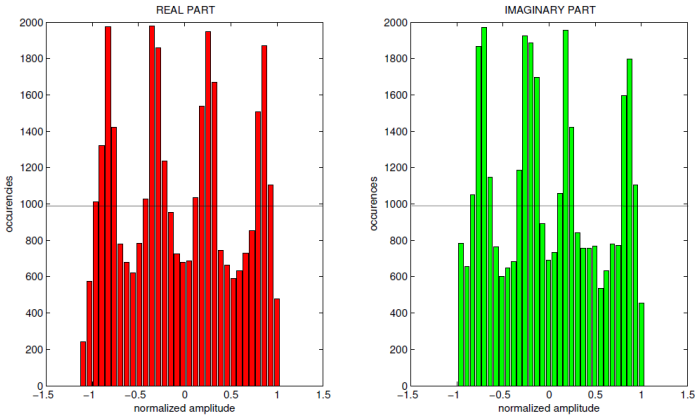


Figure 2.16: Analysis of the 16-QAM case: histograms of signal I and Q components.

bins in histograms is a simple way to reduce irregularity, clustering results into fewer large peaks. This is appropriate for modulation detection, since only an approximate constellation shape is sought. In Fig. 2.16 64 bins were employed.

The PSK case provides an example of an argument modulated signal, for which a representation in polar coordinates is better suited, as shown in Fig. 2.17. As expected, the amplitude histogram presents a single peak centred around the burst envelope amplitude, hence the phase histogram is the main source of information for argument modulation. In this example it exhibits eight clear peaks, each corresponding to one of the phase values. As peak density is highest in phase histograms, peak detection can be more difficult.

Traces in Fig. 2.18 demonstrate the multi-source detection capability of the algorithm. In this case two different signal bursts fall within the same observation interval and SNR in the acquired sample sequence is 0 dB. The two measured sources are separated in frequency and employ different modulation formats (respectively, 16-QAM and 8-PSK). Reconstructed bursts are superposed on the original trace, showing correctly reproduced waveforms with an enhanced SNR.

Histograms in Fig. 2.18 were constructed using just 32 bins, since the symbol density of the two modulations allows a reduced amplitude resolution for better clustering of samples. Nevertheless, the resulting peaks allow correct identification only for the 16-QAM modulation format, whereas inspection of the 8-PSK phase histogram does not give a clear suggestion.

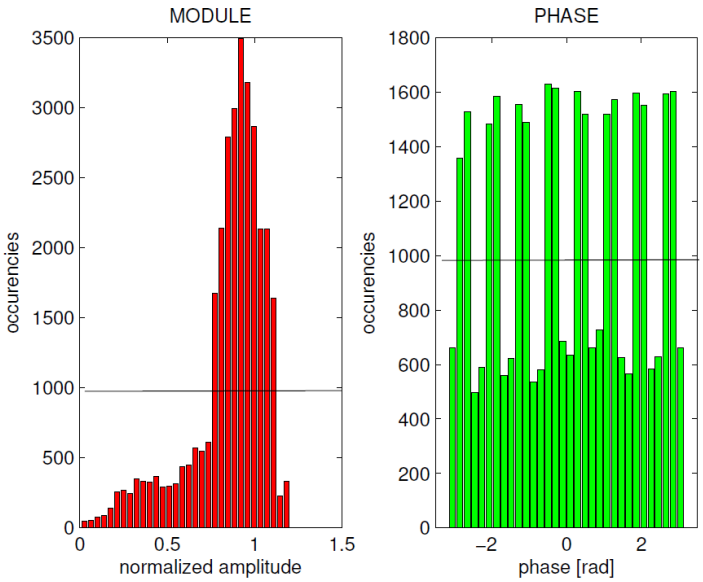
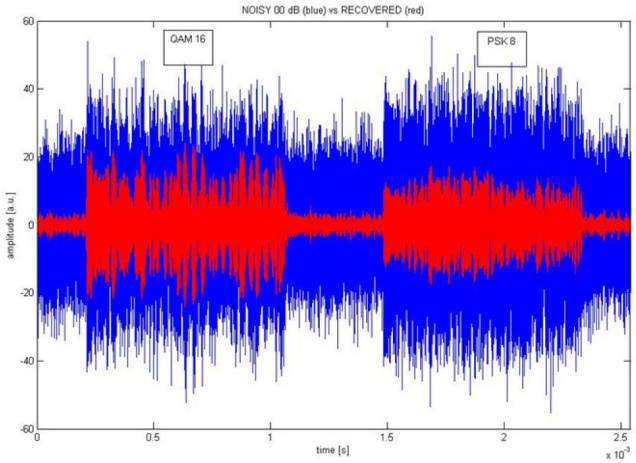


Figure 2.17: Analysis of the PSK case: histograms of module and phase.

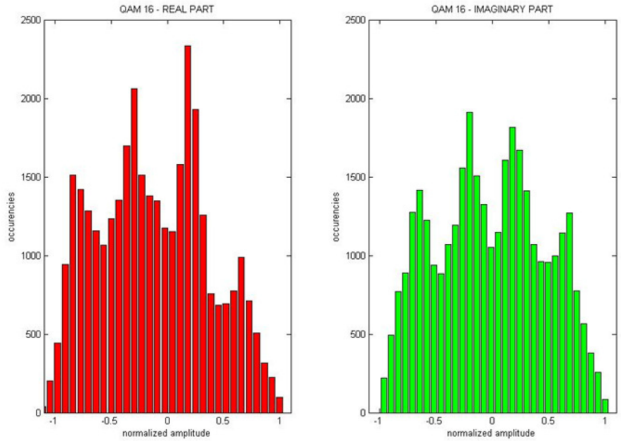
This result points to a critical issue in the use of histograms, that is, the need to ensure that the extracted modulation waveform contains a sufficiently uniform set of constellation symbols.

The situations shown in Figs. 2.16 and 2.17 are in fact particularly favourable, since histograms were built from a set of about ten consecutive observations of 850 μs bursts from the same source. With clearly defined peaks, detection thresholds are easily defined as the expected value of the relevant quantity, indicated by horizontal lines in the figures. Areas where peaks are located are thus delimited and plots providing a rough rendering of an ideal constellation can be drawn. Application of pattern matching algorithms, either to such plots or directly to histogram data, can ultimately detect the modulation format.

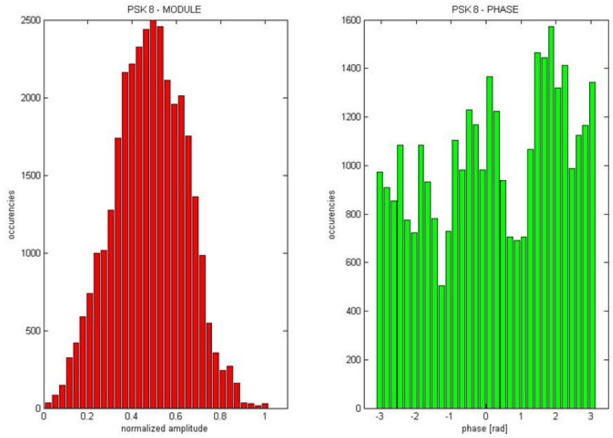
In practice, measured histograms may not always correspond to a large enough symbol population. Shorter observation intervals introduce an additional cause of variability, since the condition on equal probability of symbols may not be satisfied. Thus, even though the number of amplitude samples can be large enough for the histogram estimate to be stable, peak heights may vary significantly from one symbol to another as well as in successive observation intervals, making a simple threshold detection much more difficult. This aspect is totally unrelated to the features of



(a) signal burst recovery;



(b) 16-QAM histogram of I and Q components;



(c) 8-PSK histogram of module and phase.

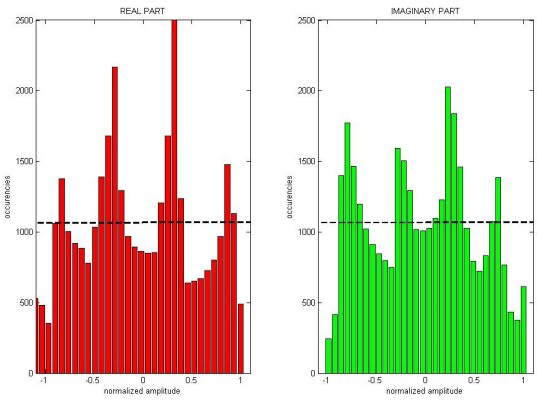
Figure 2.18: Multi-source detection from a single 2.5 ms sample sequence.

the CS algorithm, but has to be accounted for in experimental characterization.

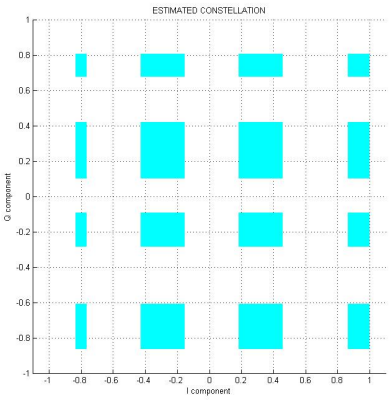
In the case of Fig. 2.18, both bursts have the same length which corresponds to 64 symbols. However, it is clear from the combined inspection of the I and Q histograms that in this particular observation the 16 constellation symbols do not have the same number of occurrences.

The worst-case effect of a short measuring interval is shown in the phase histogram of Fig. 2.19.c, that refers to another example of a 8-PSK modulation and was constructed from the observation of a single burst composed of just 64 symbols. Histogram sensitivity to statistical fluctuations with regards to the expected equal probability of symbols is quite high and less frequent symbols in this case are missed, resulting in a partial constellation plot. It should be emphasized this is not a shortcoming of the algorithm, but the effect of statistical instability caused by an excessively short observation interval.

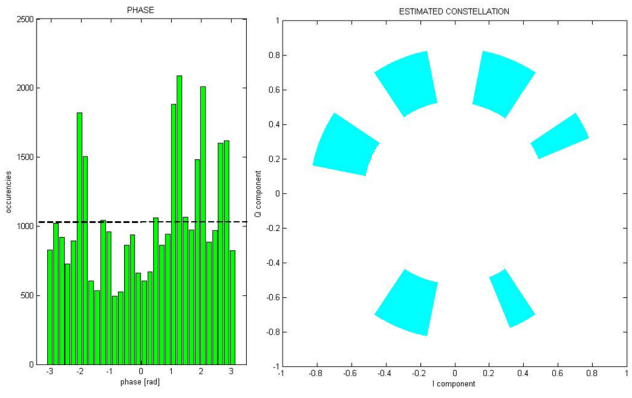
Another source of uncertainty in constellation estimation is presented in Fig. 2.19.b, that refers to a single burst of 16-QAM consisting of 64 symbols. In this case, the denial of equal probability assumption implies that all the peaks exceed the threshold but their width is not uniform. As a consequence, narrower peaks in the Q component histogram could result in an asymmetric constellation plot, i.e. the expected position of some symbols could be erroneously predicted.



(a) shaded area: 16-QAM histogram of I and Q components, dotted line: peak detection threshold;



(b) 16-QAM estimated constellation;



(c) shaded area: 8-PSK phase histogram, dotted line: peak detection threshold (left); corresponding estimated constellation (right).

Figure 2.19: Effect of deviation from symbol equal probability for short observation intervals. In QAM case, narrower peaks in the Q component histogram result in an asymmetric constellation plot. In PSK case, undetected peaks in the phase histogram result in a partial constellation plot.

Chapter 3

CS-Based Spectral Analysis

3.1 Introduction

A classical problem for spectral analysis is the accurate measurement of a multi-tone waveform by the estimation of the individual sinusoidal component parameters. With algorithms based on discrete Fourier transform (DFT), spectral leakage is known to affect the accuracy of amplitude estimation as well as the ability to resolve closely spaced frequency components [59].

As long as the frequency separation of waveform components is large enough, several forms of frequency interpolation of DFT coefficients allow to significantly improve estimation accuracy, approaching the theoretical limit set by the Cramér-Rao bound [60, 61, 62]. However, if sinusoid frequencies are close to each other, interpolation is likely to fail and different approaches are required, among which are a variety of super-resolution methods [63]. In this chapter, the super-resolution problem in the DFT domain is dealt with by a novel approach, relying on CS theory principles. The chapter is divided into three main parts.

In the first part, the frequency super-resolution problem is contextualized into the CS context and properly formulated as a sparse acquisition problem. From a spectral analysis point of view, the generic signal under investigation is modelled by a weighted linear combination of cisoids, also referred to as multi-tone signal. In the DFT domain, this model is expected to be sparse or at least compressible. Thus, a CS approach is not only feasible, but also recommended. In fact, the literature

contains many CS-based solutions, but their performances are still unsatisfactory if compared with the results obtained by similar CS approaches in other research fields. To overcome this limitation, this chapter introduces a novel super-resolution algorithm, in the following referred to as CS-DFT, and characterizes it in terms of resolution threshold, estimation accuracy and noise sensitivity.
The DFT coefficients are defined over a discrete frequency grid, whose step is the theoretical limit for any super-resolution technique. The proposed algorithm employs an over-complete dictionary which explicitly accounts for spectral leakage effects. In this way, CS-DFT achieves a resolution threshold equal to 1.5 times the DFT grid step, whereas the current CS benchmark techniques require a minimum separation not lower than 3 times the DFT grid step. By properly exploiting the sparsity assumption, the proposed algorithm ensures a resolution enhancement of one order of magnitude, independently from the observation interval length. Support recovery is carried out by means of an OMP procedure, thus allowing a computationally tractable implementation. This stage is particularly sensitive to noise and requires that SNR be not lower than 15 dB. CS-DFT estimations are defined over a much denser discrete grid than the original DFT grid. Both long-range interference and scalloping loss effect are accordingly reduced. Therefore, inaccuracy in amplitude and phase estimates is mainly due to additive noise variability.

In the second part of this chapter, the super-resolution algorithm is applied to power quality and phasor measurements in power systems. The test signals considered for its characterization reproduce a typical smart grid scenario and its challenging operative conditions. Typically, a fundamental frequency component is affected by harmonic and interharmonic disturbances and undergoes dynamic changes of its frequency, amplitude and instantaneous phase. Experimental results show how the CS-DFT approach yields power quality indices and fundamental phasor estimates, which are mostly compliant with current regulations and standards.

In the third part, a two stage procedure is implemented in order to further enhance the super-resolution accuracy. In fact, CS-DFT results are still defined over a discrete grid. A component frequency not lying on a grid bin is approximated to the nearest grid point. This means that a sort of quantization error is committed and its size depends on the discrete grid step. To overcome this limitation, a polar interpolation technique is applied to CS-DFT estimation. This procedure relies on a

Basis Pursuit approach, which is more demanding in terms of computational complexity, but provides estimates defined over a continuous set of values. This research field is also known as CS off-the-grid and is applicable to all those contexts where a discrete approximation is infeasible or unsatisfactory. This two-stage approach does not affect the the CS-DFT resolution threshold, but significantly enhances frequency estimation accuracy. The idea of computing polar interpolation only at the CS-DFT estimated frequencies and not over the entire denser grid allows a strong reduction of the overall computational complexity.

3.2 Part I: CS Super-Resolution in DFT-Based Spectral Analysis

3.2.1 Spectral Super-Resolution Scenario

Frequency-domain waveform analysis is a problem for which a variety of well-known solutions have been proposed in the literature [64, 65]. A multi-tone waveform can be expressed as a sum of cisoids, whose spectrum after sampling is given by the following complex-weighted spike train:

$$X(\lambda) = \sum_h A_h e^{j\phi_h} \delta(\lambda - \lambda_h). \tag{3.1}$$

Here, as in the following, frequency $\lambda \in [0,1]$ is normalized with respect to the sampling rate $f_s = \frac{1}{T_s}$ (i.e., $\lambda = f T_s$).

Non-parametric approaches based on the discrete Fourier transform have been a workhorse of waveform analysis for decades [59]. Given a sample sequence of length N, its discrete Fourier transform (DFT) coefficients are located on a frequency grid with step $\Delta_\lambda = \frac{1}{N}$. Therefore, they represent (3.1) exactly only when spike locations λ_h fall on this grid.

Reconstruction by DFT alone yields poor results for off-grid frequencies, as the capability to resolve signal components at closely spaced frequencies is limited by spectral leakage. Interpolation of DFT coefficients [60] can provide much more accurate estimates of the weights $A_h e^{j\phi_h}$ and the off-grid normalized frequencies λ_h, with variances approaching the relevant Cramér-Rao bounds [61]. However, required minimum separation between adjacent frequencies is increased by a factor

k_R, that depends on the kind of interpolation algorithm and on the weighting (if any) applied to the time-domain samples.

Since $k_R > 1$, the minimum allowed distance $k_R\Delta_\lambda$ is always significantly larger than Δ_λ. Parametric approaches, e.g., Pisarenko harmonic decomposition, MUSIC and ESPRIT, can allow much better resolution, but do so at the price of greater computational complexity.

In this chapter the problem is addressed by introducing a finer frequency grid, with smaller step size Δ'_λ, and relating the set of N samples to $N' = P \cdot N$ coefficients of the DFT defined on the finer grid, that is associated to an integer super-resolution or *refinement* factor $P = \frac{\Delta_\lambda}{\Delta'_\lambda}$.

In the literature on compressive sensing (CS), super-resolution algorithms based on this idea have been proposed. Random waveform samples can be related to DFT coefficients by way of the measurement equation:

$$\mathbf{x} = \mathbf{W}^H_{(N\times N')}\,\mathbf{a} + \mathbf{z}, \tag{3.2}$$

where $\mathbf{x} \in \mathbb{R}^N$ contains N time-domain samples $x(n)$ and elements of $\mathbf{a} \in \mathbb{C}^{N'}$ are DFT coefficients on the finer grid. Vector $\mathbf{z}$ represents additive zero-mean white noise with variance σ_z^2. In the standard approach $\mathbf{W}^H_{(N\times N')}$ is a compressive random measurement matrix defined as a $N \times N'$ partial inverse Fourier transform, whose rows are randomly drawn from the full $N' \times N'$ matrix. A fine-grid solution can be found by convex ℓ_1 minimization, exploiting an *a-priori* sparsity constraint on vector $\mathbf{a}$, provided a minimum distance $4\Delta_\lambda$ exists between adjacent components [66], [67]. The issue of poor numerical conditioning caused by highly coherent column vectors in $\mathbf{W}^H_{(N\times N')}$ is addressed by algorithms based on *frequency inhibition* [68] or coherence *band exclusion* [69], the latter achieving a minimum separation of $3\Delta_\lambda$. Components whose frequency separation is closer to the Rayleigh threshold Δ_λ seemingly remain unresolvable by CS methods.

We show in this chapter that a CS approach can indeed resolve closely spaced frequency components, when $\mathbf{x}$ contains *sequentially* sampled values, if the measurement equation is suitably defined to reflect this. Rather than adapt matrix $\mathbf{W}^H_{(N\times N')}$, as in [70], we exploit the feature by explicitly introducing information about spectral leakage into the measurement equation. With our approach this becomes:

$$\mathbf{x} = \mathbf{W}^H\mathbf{D}\mathbf{a} + \mathbf{z}, \tag{3.3}$$

where $\mathbf{W}^H$ is a full inverse DFT of smaller size $N \times N$ and spectral leakage is modelled by matrix $\mathbf{D}$, of size $N \times N'$. The latter is in fact an overcomplete convolutional dictionary [71].

The proposed algorithm has been characterized by means of numerical simulations, which show that components as close as $1.5\Delta_\lambda$ can be successfully resolved. As a further advantage, total acquisition time for $\mathbf{x}$ in (3.3) is strictly NT_S, whereas random sampling acquisition of $\mathbf{x}$ in (3.2) approximately requires the time $N'T_S$.

The improved lower bound on frequency separation is of interest in monitoring applications, where it reduces the need to acquire longer sample sequences for a given resolution, thereby avoiding possible adverse effects on the capability to track waveform variations.

3.2.2 Frequency Interpolation as a CS Problem

In the frequency domain, the measurement process is accurately modelled by a generic sparse acquisition problem:

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n} \tag{3.4}$$

where $\mathbf{y}$ is the measurement column vector of dimensions $[N \times 1]$, $\mathbf{A}$ is the measurement matrix of dimensions $[N \times N']$, and $\mathbf{x}$ is the unknown column vector of dimensions $[N' \times 1]$, with $N' > N$ by definition. The vector $\mathbf{n}$ accounts for all those spurious disturbances which are expected to affect and degrade the measurements. For the sake of simplicity, in the following $\mathbf{n}$ is modelled as an additive white Gaussian noise vector, statistically uncorrelated from the informative vector $\mathbf{x}$.

It is worth noticing that the sparse acquisition problem is formulated in the frequency domain, unlike its previous version 3.3 in the time domain. In this context, the measurement matrix $\mathbf{A}$ accounts for both the DFT and the compression stage. The unknown vector $\mathbf{x}$ is assumed to be sparse, i.e. it is assumed to exhibit only a reduced set of non-negligible coefficients, if expressed in the proper domain. In the present case, it is immediate to argue that $\mathbf{x}$ is sparse precisely in the DFT domain. The choice of relating a time domain signal with a frequency domain set of measurements exactly fits with the incoherence criterion presented in chapter 1.

By construction, the problem 3.4 requires to solve an under-determined systems of equations. Any combinatorial approach would result in non-polynomial compu-

tation times. Nevertheless, by properly exploiting the sparsity assumption, it is possible to obtain a sparse signal approximation, directly inferred from the apparently insufficient set of measurements.

Owing to the presence of noise, no exact solution is feasible and, according to BP approach, the best estimate coincides with the sparsest vector $\hat{\mathbf{x}}$ such that:

$$\hat{\mathbf{x}} = \min_{\mathbf{x}} \|\mathbf{x}\|_1 \qquad s.t. \quad \|\mathbf{A}\hat{\mathbf{x}} - \mathbf{y}\|_2 \leq \varepsilon \tag{3.5}$$

A multi-tone waveform is the sum of a number of complex exponential components (with suitable symmetries if the waveform is assumed to be real). Indicating by $S_{N'}$ the set of values of the index h corresponding to a waveform component, the signal can be written in sampled form as:

$$x[n] = \sum_{h \in S_{N'}} A_h e^{j(\phi_h + 2\pi\lambda_h n)}, \qquad -\infty < n < +\infty \tag{3.6}$$

where component frequencies have been expressed in the normalised form $\lambda_h = f_h T_s$, and T_s is the sampling interval. The starting point for the problem at hand is the equation that defines the DFT of a multi-tone waveform, as calculated from N time-domain samples acquired with sampling interval T_s:

$$X\left(\frac{k}{N}\right) = \sum_{h \in S_{N'}} A_h e^{j\phi_h} \left[\frac{\sin \pi (k - \lambda_h N)}{N \sin \frac{\pi}{N} (k - \lambda_h N)} e^{-j\pi \frac{N-1}{N}(k - \lambda_h N)} \right] e^{-j\frac{2\pi}{N}(k - \lambda_h N) n_0} \tag{3.7}$$

where n_0 is the starting index of the time-domain sequence.

The normalised frequency granularity of a N-point DFT is $1/N$ and the target of this chapter is to estimate frequency components whose separation may be close to, or even lower than the granularity. Accordingly, a denser frequency grid is defined, whose minimum step is $\Delta_\lambda = \frac{1}{N'}$, with $N' \gg N$ (one can assume $N' = NP$, where P is the interpolation factor), which allows to express normalised component frequencies as integer multiples of the step Δ_λ, yielding:

$$\lambda_h = \frac{h}{N'}, \qquad h \in S_{N'} \subset \{0, 1, \ldots N' - 1\} \tag{3.8}$$

Of course, the calculation of a N'-point DFT could provide the desired granularity, however this would be achieved at the cost of a longer observation interval spent acquiring the samples.

The idea discussed in this chapter is to process a small set of M data records, each having length N and with different known starting indexes n_m defined with

respect to a common time reference, or trigger point. These are arranged into a *compressive* measurement matrix **Y** composed of M complex column vectors, each containing the DFT coefficients of a single record. Varying n_m allows to exploit the corresponding changes induced on the phases of individual DFT coefficients in different records.

By properly combining (3.7) and (3.8), if no additive noise is present, each element of the measurement matrix **Y**, indicated as $y_{k,m}$, can be written as:

$$y_{k,m} = \sum_{h \in S_{N'}} A_h e^{j\left(\phi_h + 2\pi \frac{h}{N'} n_m\right)} \left[\frac{\sin \pi \left(k - \frac{h}{N'} N\right)}{N \sin \frac{\pi}{N} \left(k - \frac{h}{N'} N\right)} e^{-j\pi \frac{N-1}{N} (k - \frac{h}{N'} N)} \right] \tag{3.9}$$

It should be noticed that the complex exponential term $e^{-j\frac{2\pi}{N}(k-\lambda_h N)n_0}$, that appears in the general expression (3.7), has been split in two terms:

- $e^{j2\pi \frac{h}{N'} n_m}$ has been added as a constant known phase term to the corresponding signal component;
- $e^{-j\frac{2\pi}{N} k n_m}$ is simply set to zero by always using $n_0 = 0$ as the starting index for DFT computation, regardless of the actual start time.

Equation (3.9) shows that the frequency interpolation problem can be seen as the problem of solving the matrix equation:

$$\mathbf{Y} = \mathbf{D}\mathbf{X} + \mathbf{N} \tag{3.10}$$

where the columns of matrix **D** contain the complex samples of the Dirichlet kernel (within square brackets in (3.9)), each progressively shifted by fractional frequency amounts Δ_λ. Thus, the row index is k while the column index is h. The matrix **N** accounts for measurement noise and uncertainty.

In Fig. 3.1 the peculiar structure of matrix **D** is presented: a three-dimensional representation is combined with a row and column trend.

The amplitude and phase of sinusoidal components forming the multi-tone waveform are represented by the non-zero complex elements of a matrix **X**, composed of M vectors x_m whose size N' agrees with the desired, denser, frequency grid. It should be noticed that each vector $\mathbf{x}_m$ is associated to a specific value of n_m, since from (3.9) one has:

$$x_{h,m} = A_h e^{j\left(\phi_h + 2\pi \frac{h}{N'} n_m\right)} \tag{3.11}$$

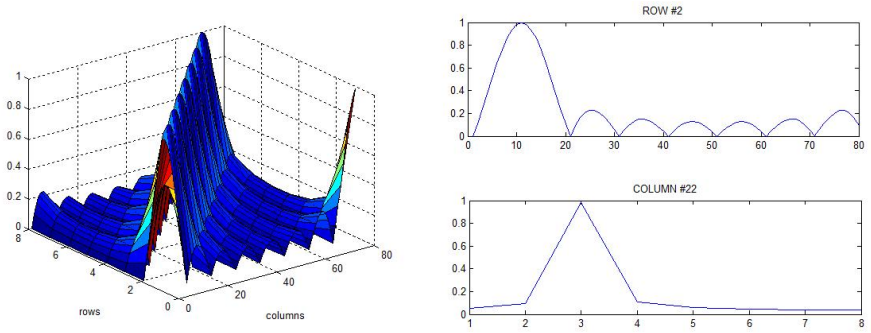


Figure 3.1: On the left, three-dimensional representation of matrix **D** with $N = 8$ and $P = 10$; on the right, typical row (top) and column (down) trends.

Thus, the matrix **X** can be factored as:

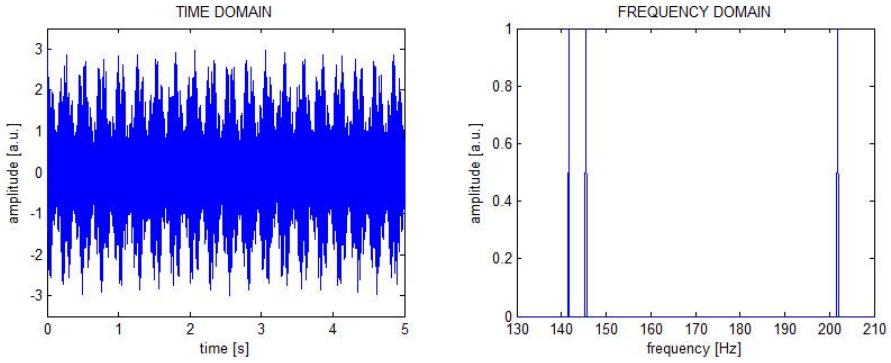
$$\mathbf{X} = \begin{bmatrix} A_1 e^{j\phi_1} & \dots & 0 \\ 0 & \ddots & 0 \\ 0 & \dots & A_{N'-1} e^{j\phi_{N'-1}} \end{bmatrix} \begin{bmatrix} e^{j2\pi \frac{0}{N'} n_0} & \dots & e^{j2\pi \frac{0}{N'} n_{M-1}} \\ \vdots & \ddots & \vdots \\ e^{j2\pi \frac{N'-1}{N'} n_0} & \dots & e^{j2\pi \frac{N'-1}{N'} n_{M-1}} \end{bmatrix} \tag{3.12}$$

The fundamental assumption that allows to formulate frequency interpolation as a CS problem and achieve super-resolution is that, whatever the individual value of n_m, the frequencies of the non-zero components are the *same* for all measurements. Hence, vectors $\mathbf{x}_m$ are sparse and, additionally, they are known to share the same set $S_{N'}$ of frequency indexes [45], corresponding to the non-zero waveform components. In CS terms, this is called a multiple measurement vectors problem.

3.2.3 Compressed Sensing Based Fourier Analysis

From a measurement point of view, the proposed CS-based data acquisition allows to reduce the length of the acquisition interval (and of the total sample record), since the sets of samples from which each DFT vector in **Y** is computed can be broadly overlapped. If the average difference between consecutive starting indexes is Δ_n, the total number of samples required is $N+\Delta_n(M-1)$, with $\Delta_n \ll N$. Assuming the CS algorithm allows an interpolation factor P, this should be compared with the length $N' = NP$ required to achieve the same frequency granularity by a standard DFT algorithm.

The first step in solving the compressed sensing problem is support recovery, which has the purpose of determining which elements in $\mathbf{x}_m$ are non-zero. In prac-



Parameters	Values
Sampling frequency, F	500 Hz
Length of sequence, N	256 samples
Interpolation factor, P	10
Components frequencies	141.7969, 145.7031, 201.9531 Hz
Components amplitudes	1, 1, 1 a.u.
Components phases	0, 0, 0 rad
Original support	726, 740, 1034

Figure 3.2: TOP: on the left, signal real part in the time domain; on the right, detail of signal Fourier transform amplitude. BOTTOM: multi-tone signal parameters.

tice, the aim is to determine the set $S_{N'} \subset \{0, \dots N'-1\}$, as defined above. As far as Fourier analysis is concerned, this is in fact the key part of the algorithm, since $S_{N'}$ contains the frequency locations of the signal components.

Following the approach proposed in [45], this step is carried out via a singular-value decomposition of the matrix $\mathbf{Y}\mathbf{Y}^H$, where the superscript denotes transposition and complex conjugation. Decomposition allows to obtain the matrix $\mathbf{V}$, whose columns are the eigenvalues of $\mathbf{Y}\mathbf{Y}^H$ multiplied by the square roots of the corresponding eigenvectors. Collapsing the columns of $\mathbf{V}$ into a single vector $\mathbf{v}$, and solving the equation $\mathbf{v} = \mathbf{D}\mathbf{u}$ by an Orthogonal Matching Pursuit (OMP) algorithm allows to find the support for $\mathbf{u}$, which can be shown to be the same as $S_{N'}$. By construction, $\mathbf{V}$ and $\mathbf{D}$ are known: the first is obtained from the measurements result, the second defines the measuring scheme.

To test the approach, a signal formed by three known complex exponential components, according to (3.6), was considered and different sets of M records of length $N = 256$ samples were generated. In the multi-tone waveform given in Fig. 3.2 spectral components have normalised frequencies $\lambda_1 = 72.6$, $\lambda_2 = 74$, $\lambda_3 = 103.4$, so that the second component contributes no spectral leakage, but is affected by leakage from the first one. On the top right figure, sampling rate is set to 500 Hz, while com-

ponent frequencies on the top left figure correspond to 141.80, 145.70 and 201.95 Hz. The third component is employed as a control element, being sufficiently far away from the first two to be only affected by scalloping loss.

In the test configuration $M = 64$ and the starting points of each sample record are separated by 3 samples. Component amplitudes were set to 1, their initial phases to 0. In the following, amplitude values are expressed in arbitrary units, while phase values are expressed in radians. Simulations were iterated 100 times for each test condition. Trials were first carried out with no noise. The desired interpolation factor is $P = 10$, ensuring that frequencies fall exactly on points of the denser grid (i.e., they are integer multiples of Δ_λ). With these values the support recovery step always provided the correct component frequencies.

The subsequent algorithm step involves the construction of a restricted matrix $\mathbf{D}_S$ by deleting columns whose index h is not in $S_{N'}$. This turns (3.10) into an overdetermined matrix equation, that can be solved by computing the Moore-Penrose pseudo-inverse:

$$\hat{\mathbf{X}}_S = \left(\mathbf{D}_S^\dagger \mathbf{D}_S\right)^{-1} \mathbf{D}_S^\dagger \mathbf{Y} \tag{3.13}$$

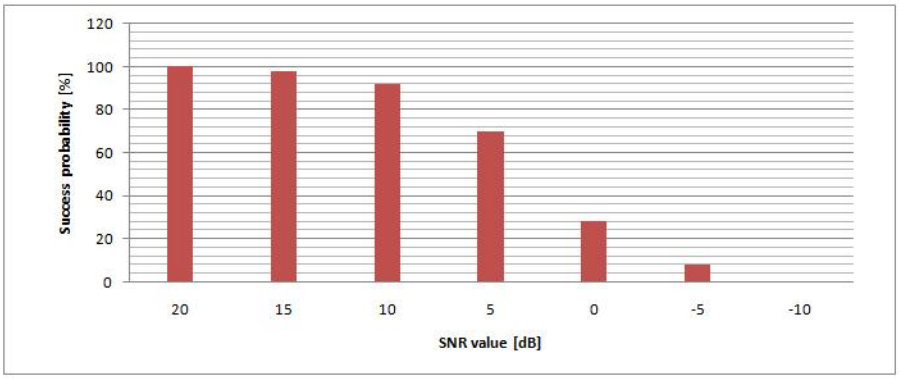
which only contains the non-zero rows of $\mathbf{X}$. The symbol † identifies the Hermitian transpose operation.

The final step is to invert the relevant rows according to (3.12); this accurately reconstructs component amplitudes and phases (to within machine error), provided the number of measurement vectors M is greater than the number of signal components.

When measurement noise is present, replacing (3.10) into (3.13) shows the estimate to be:

$$\hat{\mathbf{X}}_S = \mathbf{X} + \left(\mathbf{D}_S^\dagger \mathbf{D}_S\right)^{-1} \mathbf{D}_S^\dagger \mathbf{N} \tag{3.14}$$

provided the support recovery step can still be successfully completed. However, noise variability affects the support recovery algorithm, whose results are no longer deterministic, but susceptible to changes: each iteration could provide a different support, with obvious consequences also on amplitude and initial phase estimates. This phenomenon becomes more and more significant as the signal-to-noise ratio decreases. Then, it is reasonable to assume there exists a limiting value of the signal-to-noise ratio (SNR), below which the support is wrongly identified and accurate component estimation becomes impossible.



SNR value [dB]	Success Probability [%]
20	100
15	98
10	92
5	70
0	28
-5	8
-10	0

Figure 3.3: Success probability with progressively reduced SNR.

To find this limiting value, tests were carried out by repeating the CS algorithm 100 times for a number of different SNR values, recording the number of successful support recoveries. In this case, to avoid any interference contribution the normalised frequencies were set to $\lambda_1 = 72.6$, $\lambda_2 = 98$, $\lambda_3 = 103.4$, so that the second component, which still causes no leakage, is also far enough to be unaffected by leakage from any of the other two components.

Results are shown in Fig. 3.3, which evidences that successful support recovery cannot be guaranteed below 20 dB SNR. It should be noted that only an exact coincidence means a successful result.

3.2.4 Super-Resolution Experimental Results

So far, it has been shown that the proposed CS algorithm can accurately estimate the frequency position of signal components on a dense (interpolated) grid, thus allowing to compensate for the effects of spectral leakage. It remains to show that components that are placed very close in frequency can be picked out and correctly estimated. Then, achievable performances should be characterized.

Given the initial test configuration described above, in the absence of noise the second component was shifted in frequency, by steps of 0.1 bins, towards the first

Number of measurements	Amplitude mean error component #1	Amplitude mean error component #2	Amplitude mean error component #3
64	1,11E-016	4,44E-016	8,88E-016
32	6,66E-016	4,44E-016	4,44E-016
16	1,11E-015	4,44E-016	2,22E-016
8	8,88E-016	0,00E+000	1,22E-015
4	4,44E-016	2,33E-015	3,33E-016
2	4,84E-001	5,13E-001	3,13E-003
1	6,67E-001	6,67E-001	6,67E-001

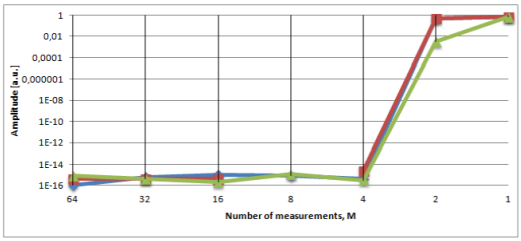


Figure 3.4: On the left, amplitude means errors according to different number of measurements; on the right, logarithmic plot of mean error trends.

component. Results were found to depend on signal parameter settings, in particular on the initial phases. If the phase shift between the first and second components is equal to 0, $\pi/2$ or π the resolution threshold at which support recovery is successful is equal to 1.1 bins. By raising the threshold to just 1.2 bins, the two equal amplitude components are correctly resolved with any phase configuration. For any random set of phase values, reconstruction error is negligible.

Spectral interference becomes worse if one component is larger than the other. Keeping the first component amplitude constant, we progressively reduced the second one, up to a ratio of 1 : 1000. In this case, a minimum separation of 1.4 bins is required to ensure a correct resolution for any possible setting. This condition was verified by simultaneously varying the amplitudes and phases of both components with a random pattern.

It is also important to find out which is the minimum required number of measurements. While correct support recovery is possible even for $M = 1$, at least $M = 4$ is necessary to provide good accuracy in waveform reconstruction, as can be seen in Fig. 3.4.

On the contrary, the relative spacing between the initial points n_m does not affect significantly the results, even though by reducing the superposition of two sequences, their statistical correlation is less significant. On the other hand, by a proper choice of the number and position of the measurement initial points the total observation interval can be markedly reduced. A good compromise was found to be $M = 4$ and $n_m = [\,0, 3, 6, 9\,]$. In this way, only 256 samples are acquired to obtain an estimate over a grid of 2560 frequency points.

In the presence of noise, the method's effectiveness depends critically on sup-

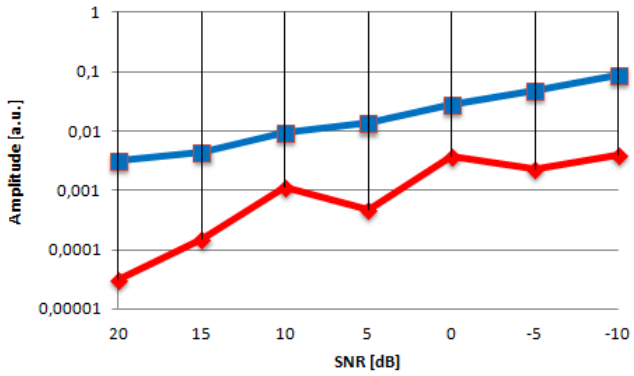


Figure 3.5: Mean (diamonds) and standard deviation (squares) of amplitude estimation error vs. SNR on second signal component: $A_2 = A_1 = 1$ a.u.

port recovery. Adopting a SNR equal to 20 dB, method's performances are comparable with the noiseless case. A distance of 1.4 bins represents a resolution threshold, once more not dependent on component amplitudes and initial phases. For components with different amplitudes, the minimum ratio has to be limited to $1:10$ (i.e., -20 dB), otherwise the second component is hidden by noise fluctuations.

To understand the impact of support recovery in comparison to noise effects on component parameter estimation with known support, two sets of simulations were carried out, where SNR varied but support was known *a priori*. Results are presented in Fig. 3.5 for $A_1 = A_2 = 1$ a.u. and in Fig. 3.6 for $A_1 = 1$ and $A_2 = 0.1$ a.u. In both cases, the mean deviation of the amplitude estimation error for A_2 is almost one order of magnitude below the standard deviation. It is important to note that, when the support is known, estimation variance is proportional to SNR, as implied by Eq. (3.14).

In conclusion, obtained results suggest that a resolution threshold equal to 1.4 bins can be considered valid until SNR value is not lower than 20 dB. Within these constraints, the proposed method recovers the original support and provides accurate estimates of component amplitudes and initial phases.

3.2.5 Leakage Dictionary Advantages

Till now, the problem of resolving frequency components close to Rayleigh threshold has been address by introducing a super-resolution technique which exploits the

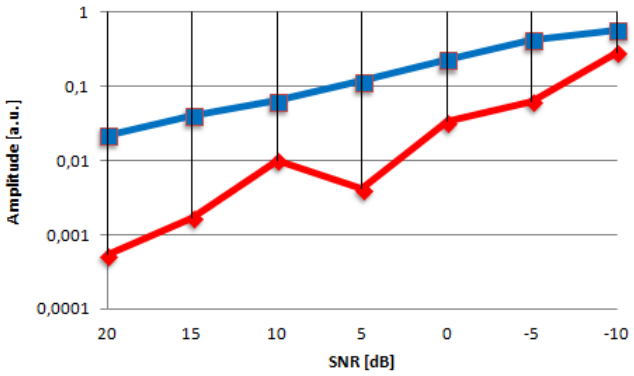


Figure 3.6: Mean (diamonds) and standard deviation (squares) of amplitude estimation error vs. SNR on second signal component: $A_2 = \frac{A_1}{10} = 0.1$ a.u.

components sparse representation in the DFT domain. In Section 3.2.4, the experimental results show that the proposed algorithm achieves good estimation accuracy, without significantly extending the overall observation interval. In this regard, it should be noticed that challenging operative conditions have been implemented in order to determine the actual resolution threshold and its dependency from the component parameters.

In this section, the proposed CS approach is compared with some of the most renowned and widely employed super-resolution techniques. Particular attention is devoted to the alternative CS approaches, in order to highlight the performance enhancement provided by the proposed algorithm.

Approaches Not Relying on CS Theory

Let us consider a generic signal consisiting of a weighted sum of cisoids:

$$X(\lambda) = \sum_h A_h e^{j\phi_h} \delta(\lambda - \lambda_h) \qquad \lambda = f T_s \in [0,1] \tag{3.15}$$

For the sake of simplicity, in the following the frequencies are assumed to be normalised by the sampling rate f_s. Let us consider a conventional uniform sampling protocol compliant with Shannon's sampling theorem, omitting for the moment the sparsity assumption.

In this context, the literature offers three main approaches:

- the canonical DFT approach provides exact estimations only for those components whose frequency lies exactly on a discrete grid bin, i.e. is an integer multiple of $\Delta_\lambda = 1/N$;
- the DFT coefficients interpolation techniques ensure a spectral resolution equal to $k_R\Delta_\lambda$ with $k_R > 1$; within this threshold, also off-grid estimates are feasible with variances approaching Cramér-Rao bounds;
- the parametric methods, e.g., Pisarenko, MUSIC and ESPRIT, achieve much better resolution, but at the expense of greater computational complexity.

CS Norm Approximation Approaches

Even starting from N time domain samples, the basic idea is to obtain an estimate resolution proportional to $\Delta'_\lambda = 1/N'$, with $N' = NP$. Typically, the integer P is called also super-resolution or refinement factor $P = \Delta_\lambda/\Delta'_\lambda$.

Let us consider the typical sparse acquisition protocol in the time domain, modelled by Eq. 3.3. The canonical CS approach provides a convex norm relaxation approach, which could be BP or LASSO according to the expected level of additive noise. Generally, this estimation technique provides rather unsatisfactory performances:

- the achieved resolution is scarce, with $k_R = 4$ [66][67];
- the poor numerical conditioning of the measurement matrix $\mathbb{W}$ has to be properly taken into account: for instance, frequency inhibition [68] or coherence band exclusion [69] techniques ensure a resolution enhancement up to $k_R = 3$;
- the Rayleigh threshold is seemingly unresolvable, or even unapproachable by CS super-resolution approaches.

Over-Complete Dictionary of Leakage

Let the vector $\mathbf{x}$ contain sequentially sampled values. More precisely, let us consider a set of M subsequent N-points data records, and let the measurement equation be suitably defined to reflect this.

Accordingly, the super-resolution problem can be interpreted as a sparse acquisition problem, simply with a further dimension related to the number of acquired

records. It is worth noticing that the M data records are expected to be jointly sparse, i.e. different realizations of the same aleatory process which share the same support. As aforementioned, this kind of sparsity could be employed to enhance the CS support recovery performances. In the following, two plausible solutions are briefly discussed and compared.

Inverse matrix truncation Fig. 3.7(a) shows a typical representation of CS acquisition protocol. Let us consider a full rank system of equations:

$$\mathbf{X} = \mathbf{W}_{\mathbf{N}' \times \mathbf{N}'} \mathbf{A} + \mathbf{Z} \tag{3.16}$$

where both the measurement and the unknown vector have cardinality N', and the measurement matrix is square. Evidently, this system is determined and admits a unique solution to be computed by a simple matrix inverse:

$$\hat{\mathbf{A}} = \mathbf{W}^{\dagger}_{\mathbf{N}' \times \mathbf{N}'} (\mathbf{X} - \mathbf{Z}) \tag{3.17}$$

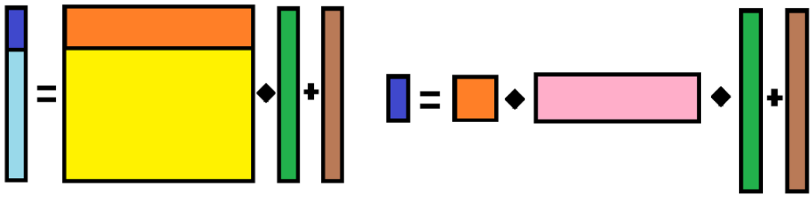
In this context, the CS acquisition protocol is equivalent to selecting a set of $N = N'/P$ rows of the measurement matrix $\mathbf{W}$ and accordingly obtaining the typical underdetermined sparse problem:

$$\mathbf{X} = \mathbf{W}_{\mathbf{N} \times \mathbf{N}'} \mathbf{A} + \mathbf{Z} \tag{3.18}$$

As shown in Fig. 3.7(a), the novel matrix linear system is characterized by a rectangular measurement matrix, which is only required to satisfy the RIP of a suitable order.

In the literature, this straightforward CS approach to super-resolution problem is also known as inverse matrix truncation [70]. As aforementioned, this is a completely blind approach which exploits the only knowledge a priori about the signal sparsity in the DFT domain. The resolution threshold depends essentially from the row selection stage. However, even implementing the coherence band exclusion criterion, the resulting truncated $\mathbf{W}$ matrix is expected to exhibit scarce conditioning number and isometry constant. Consequently, components distance too close to Rayleigh threshold are difficult to be correctly recovered.

Spectral leakage contribution The CS algorithm presented in this chapter, instead, enhances frequency resolution by introducing a dictionary which explicitly accounts



(a) inverse matrix truncation; (b) spectral leakage dictionary.

Figure 3.7: Comparison between different CS approaches to super-resolution problem in DFT spectral analysis.

for spectral leakage on a fine grid. As shown in From Fig. 3.7(b), some significant differences are noticeable with respect to the previous paragraph approach. First of all, the measurement equation is formulated as follows:

$$\mathbf{X} = \mathbf{W}^{\mathbf{H}}_{\mathbf{N}\times\mathbf{N}}\mathbf{D}_{\mathbf{n}\times\mathbf{N}'}\mathbf{A} + \mathbf{Z} \tag{3.19}$$

where the inverse DFT matrix assumes its canonical square form, with dimensions $[N \times N]$. The compression stage is committed to the over-complete dictionary $\mathbf{D}$ of dimensions $[N \times N']$, which accounts for leakage contribution inherent in any DFT measurement protocol. In other words, signal sparsity is expressed not in terms of an orthonormal basis but in terms of an over-complete dictionary [71].

In CS literature, the adoption of this kind of dictionary is currently spreading, owing essentially to two motivation. On the one hand, it is not always possible to define a sparsyfing basis which is exactly orthonormal, while an over-complete dictionary is required only to have more columns than rows. On the other hand, this approach offers more flexibility and adaptability: in linear matrix inverse problem it is reasonable to expect that suitable over-complete representations could be helpful in reducing approximation errors and making the CS procedure more robust in the presence of noise or artefacts. In the present case, for instance, the over-complete dictionary allows to achieve $k_R \simeq 1.5$ even starting from reasonably short data records, that is nearly half the best resolution threshold provided by any other CS approach in the literature.

Table 3.1: Minimum distance – equal-amplitude components.

P	3	5	7	9	11	13
Δl	5	8	10	13	16	22
$\lvert\lambda_2-\lambda_1\rvert\cdot N$	**1.67**	**1.6**	**1.43**	**1.44**	**1.45**	**1.69**
P	15	17	19	21	23	
Δl	22	26	28	30	34	
$\lvert\lambda_2-\lambda_1\rvert\cdot N$	**1.47**	**1.53**	**1.47**	**1.43**	**1.48**	

3.2.6 Support Recovery Stage

Fine-grid DFT coefficients are the non-zero elements of **a**, whose values are determined by computing the pseudo-inverse:

$$\hat{\mathbf{a}}_S = \frac{1}{N}\left(\mathbf{D}_S^H\mathbf{D}_S\right)^{-1}\mathbf{D}_S^H\mathbf{W}\mathbf{x}. \tag{3.20}$$

In this equation, $\mathbf{D}_S$ is a *restricted* dictionary matrix obtained by keeping only the columns of **D** with index belonging to the recovered support S, and $\hat{\mathbf{a}}_S$ is the correspondingly restricted vector.

Support recovery is the critical part of the algorithm, as its task is to identify waveform components. Its function is the equivalent, on the finer frequency grid, of peak search in traditional DFT-based spectral analysis and, likewise, a signal-to-noise ratio (SNR) threshold holds.

To find out the lower bound for frequency separation we considered a signal with $|S_a| = 3$, i.e., three cisoidal components, each having unit amplitude and initial phase randomly taken from a uniform distribution between 0 and 2π. In the frequency domain two components, at frequencies λ_1 and λ_2, had their distance progressively reduced, while the third, located further away, was employed as a control element.

The limiting distance between the two close components was determined when support recovery success rate dropped below 100%. Results obtained in noiseless conditions are reported in Table 3.1, where Δl is the minimum difference between the indexes of the corresponding two non-zero vector elements a_{l_1} and a_{l_2}, and $|\lambda_2-\lambda_1|\cdot N = \frac{\Delta l}{P}$.

Considering a sequence length $N = 256$, trials were repeated for different super-

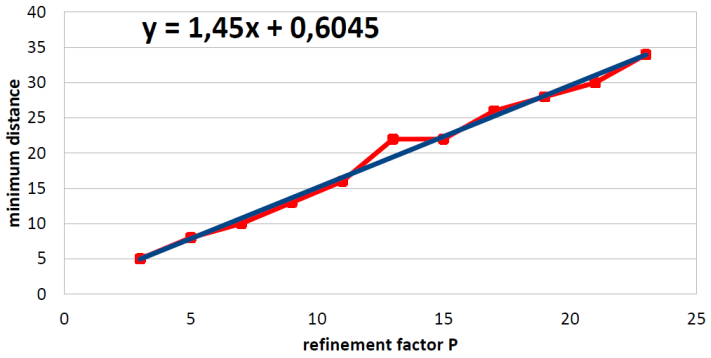


Figure 3.8: Minimum distance between equal amplitude components in the presence of different values of refinement factor P (red squares); least-squares trend line whose gradient is approximately equal to 1.5 (blue).

resolution factors. We assigned to P integer values between 3 and 23 (a range of non-critical values for the numerical conditioning of $\mathbf{D}$), selecting only coprimes with N to prevent $\mathbf{D}$ from becoming singular. Fine grid step size is thus approximately one order of magnitude smaller than Δ_λ. Within the given range for P the threshold lies at approximately $1.5\Delta_\lambda$, that is, half the value required by coherence band exclusion [69]. Variability in the computed value of $\Delta_\lambda \cdot N$ is due to the fact that only frequencies falling exactly on the finer grid have been considered, to avoid interaction with finite-grid errors that are the object of separate tests discussed in the next section. In Fig. 3.8, the minimum resolved distance between equal amplitude components, in noiseless condition, is presented as function of the refinement factor P. Furthermore, also the least-squares trend line is superposed and confirms that the resolution threshold is equal approximately to $1.5\Delta_\lambda$ and can be considered independent from P.

Results obtained for different values of the signal-to-noise ratio (SNR) show that, for a sequence of $N = 256$ samples, S_a is correctly recovered down to approximately SNR = 15 dB. This means 100% success rate in finding components whose frequencies λ_h lie on the finer grid.

Success probability versus SNR, for $N = 256$ and different values of P, is presented in Fig. 3.9. Performance degradation is evident as P gets larger, since coherence among the columns of $\mathbf{D}$ is increased as well. Nevertheless, OMP can still achieve a success rate in excess of 90% with SNR = 10 dB and $P < 15$.

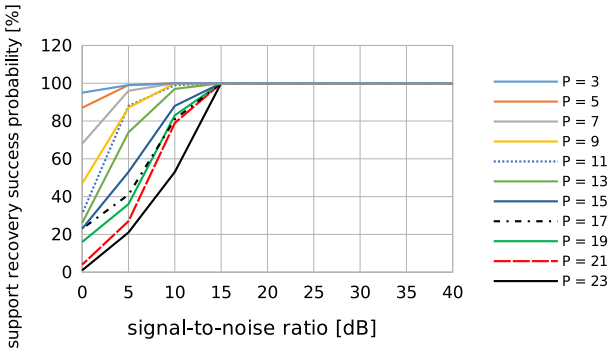


Figure 3.9: Probability of successful support recovery for different values of SNR and super-resolution factor P.

3.2.7 Finite-Grid Error and Noise

Amplitude estimation by (3.20) can yield accurate estimates, whose covariance with noisy data is:

$$\mathrm{cov}[\hat{\mathbf{a}}_S] = \frac{\sigma_z^2}{N}\left(\mathbf{D}_S^H \mathbf{D}_S\right)^{-1}. \tag{3.21}$$

Matrix $\mathbf{D}_S^H \mathbf{D}_S$ approximates the identity for component separations greater than Δ_λ, so that amplitude estimates are, for practical purposes, uncorrelated and their variance is close to the single-component Cramèr-Rao bound [72]. This was confirmed by a set of 100 simulations, repeated for different values of SNR, considering again the three-component signal introduced in the previous section. A super-resolution factor $P = 11$ was selected as it is a prime integer and approximately corresponds to an order-of-magnitude improvement of the frequency grid step size, that is, equivalent to what can safely be achieved in practice by the interpolated DFT approach [60]. The first two waveform components were placed quite close to each other, at $(\lambda_2 - \lambda_1)\cdot N = \frac{14}{11} = 1.27$, while the third, included as a far-distance reference, was at $(\lambda_3 - \lambda_2)\cdot N = \frac{292}{11} = 26.5$. Over a wide range of SNR values, estimated amplitude variances differed very little among them, their values being only marginally larger than the quantity $\frac{\sigma_z^2}{N}$.

Component frequencies may not actually coincide with grid points, therefore to a more limited extent leakage can still be present. To show how amplitude estimates are affected, the plots of Fig. 3.10 were obtained by varying the frequency

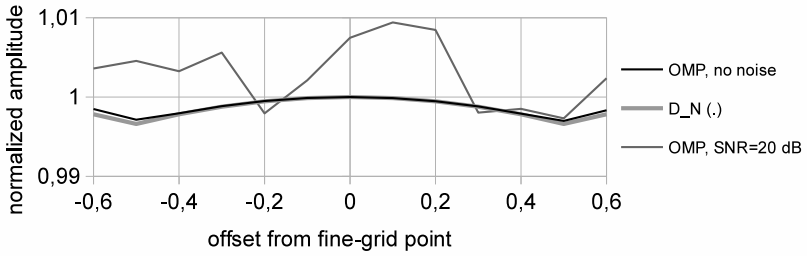


Figure 3.10: Amplitude estimation for a single off-grid component.

of a single sinusoidal component in a $\pm 0.5\Delta'_\lambda$ neighborhood of a fine-grid point $\frac{k_h}{N'}$. In the noiseless case, relative error in amplitude estimation is almost exactly $1 - D_N(\Delta'_\lambda)$ and depends on the scalloping loss associated to the Dirichlet kernel [73]. The largest error occurs for $|\lambda| = 0.5\Delta'_\lambda$, but attenuation is much smaller than the Dirichlet kernel worst-case scalloping loss, since $D_N(0.5\Delta'_\lambda) \gg D_N(0.5\Delta_\lambda)$ even with the moderate super-resolution factor $P = 11$ employed for the plots.

Fig. 3.10 also presents an example where signal samples are affected by zero-mean random white noise with SNR = 20 dB. The plot shows that estimate variations caused by noise can be considerably larger than scalloping loss effects. To further analyze the latter aspect, a set of 100 simulations with a single sinusoidal component and SNR = 20 dB was repeated at a number of frequencies within $\pm 0.5\Delta'_\lambda$ of a fine-grid point. In this case the total root-mean-square error is a more useful performance indicator than pure variance, since the resulting mean deviation from reference values is significant. This is plotted in Fig. 3.11 and, when compared with Fig. 3.10, shows the considerable impact of noise.

3.2.8 OMP Noise Sensitivity

The success rate of the non-linear support recovery stage included in the CS algorithm drops significantly when SNR gets below a certain limiting value, as shown in Fig. 3.9. This performance can be enhanced, at the cost of a moderate increase in measurement time, by jointly processing a set of time-shifted but strongly overlapped N-sample sequences.

For this purpose we consider a set of M measurement vectors $\mathbf{x}_m$, with $0 \le m < M$, each containing N sequentially acquired waveform samples. The index of the

Figure 3.11: Total root-mean-square error for amplitude estimation of a single off-grid component, with SNR = 20 dB.

first sample in each sequence is known, since one may start from 0 for the first acquisition, and simply record the index difference at the start of subsequent records. With this convention, the start index of a sequence will be indicated by n_m and $n_0 = 0$. Setting $\Delta n = n_{m+1} - n_m$, the total acquisition length becomes: $N + (M-1)\cdot\Delta n$.

To understand the effect of an index shift on the measurement equation, it is useful to consider DFT expression (3.9). Given a sample record starting at n_m, each complex term in the summation on the right-hand side has the form:

$$A_h e^{j(\phi_h - 2\pi\lambda_h n_m)} D_N\left(\frac{k}{N} - \lambda_h\right) e^{-j2\pi\frac{k}{N}n_m}, \tag{3.22}$$

where the exponential term at the end of the expression is independent of the frequency λ_h. This can actually be set to zero assuming that, whenever a DFT is computed, the time index runs from 0 to $N-1$, rather than from n_m to $n_m + N - 1$.

Each vector $\mathbf{x}_m$ can therefore be associated to a measurement equation of the form (3.3): $\mathbf{x}_m = \mathbf{W}^H \mathbf{D}\mathbf{a}_m + \mathbf{z}_m$, where vectors $\mathbf{z}_m$ are uncorrelated, being time-shifted with respect to each other. On the other hand, the net effect of an index shift on the waveform-related term is a phase rotation of the complex weight associated to the h-th waveform component, by the quantity $-2\pi\lambda_h n_m$. Neither magnitude A_h nor, more importantly, frequency location λ_h are affected. Consequently, the support set S_a is common to all vectors $\mathbf{a}_m$.

After arranging the set of measurement vectors into an $N \times M$ matrix $\mathbf{X} = [\mathbf{x}_0, \mathbf{x}_1 \ldots, \mathbf{x}_M - 1]$, our problem can be formulated by means of a *multiple measurement vector* (MMV) equation with *jointly sparse* support [74]:

$$\mathbf{X} = \mathbf{W}^H \mathbf{D}\mathbf{A} + \mathbf{Z}, \tag{3.23}$$

with $\mathbf{Z} = [\mathbf{z}_0, \mathbf{z}_1 \ldots, \mathbf{z}_M - 1]$. Matrix $\mathbf{A}$, with size $N' \times M$, can be factorized as:

$$\mathbf{A} = \text{diag}\{\mathbf{a}\}\mathbf{R}. \tag{3.24}$$

where diag{$\mathbf{a}$} is a diagonal $N' \times N'$ matrix whose non-zero entries are the elements of vector $\mathbf{a}$. The M columns of $\mathbf{R}$ contain the phase rotation terms defined at each point of the fine frequency grid for each time shift n_m.

For the measurement correlation matrix $\mathbf{X}\mathbf{X}^T$ the following equality holds:

$$\mathbf{X}\mathbf{X}^T = \mathbf{W}^H\mathbf{D}\left(\sum_{i=0}^{M-1} \mathbf{a}_i\mathbf{a}_i^H\right)\mathbf{D}^H\mathbf{W} + \sum_{i=0}^{M-1} \mathbf{z}_i\mathbf{z}_i^T \tag{3.25}$$

where, as already noted, white noise vectors are uncorrelated. Therefore, we can apply singular-value decomposition (SVD): $\mathbf{X}\mathbf{X}^T = \mathbf{V}_X \cdot \text{diag}[\lambda_i^2] \cdot \mathbf{V}_X^H$, with $\lambda_1 > \lambda_2, \ldots, \lambda_N$ to separate signal and noise subspaces. Since the signal subspace rank is equal to the order of the signal model (3.1), a condition involving the cardinality of S_a, namely: $M > |S_a|$, must be satisfied, which determines the minimum size of the measurement correlation matrix.

Interpretation of SVD as a Karhunen-Loève expansion for $\mathbf{X}\mathbf{X}^T$ suggests that, by increasing M, at low SNR values the larger eigenvalues will increase their energy without increasing in number, whereas noise-related eigenvalues will increase in number, but not in amplitude. This is shown in the plots of Fig. 3.12, where the eigenvalue profile of the measurement correlation matrix for a waveform with three cisoidal components is presented for different values of SNR. The value $M = 10$ was chosen so that a suitable number of noise-related eigenvalues could also be calculated, evidencing that, by setting a suitable threshold, signal-related eigenvalues can be singled out from noise. This holds approximately up to SNR = -8 dB, which is the threshold for the maximum likelihood estimator [75].

In the low SNR case, therefore, SVD decomposition of the measurement correlation matrix allows to carry out joint support recovery for the MMV equation (3.23) by searching for the sparsest solution, $\mathbf{u}$, to the equation:

$$\mathbf{v}_{X(th)} = \mathbf{W}^H\mathbf{D}\mathbf{A}\mathbf{u}, \tag{3.26}$$

which, compared to the approach proposed in [45], has been modified so that $\mathbf{v}_{X(th)}$ is formed as a suitable linear combination only of the columns of $\mathbf{V}_X$ whose corresponding eigenvalues are above the threshold. As the support of $\mathbf{u}$ is equal to S_a,

estimates of waveform components can be obtained again by (3.20), using any of the measurement vectors $\mathbf{x}_i$.

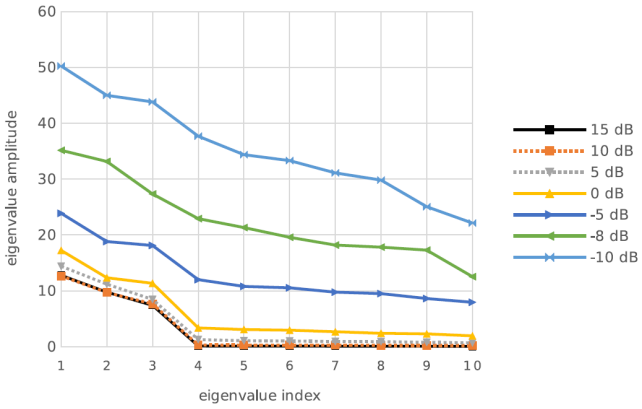


Figure 3.12: Eigenvalue distribution of the measurement autocorrelation matrix, with $M = 10$, $P = 11$.

3.2.9 CS-DFT Approach Overview

Spectral estimation of multi-tone waveforms is an extensively studied problem. In this chapter we addressed it by a CS-based approach, which in the following will be addressed also as CS-DFT, showing that fine-grid frequency estimates can be obtained and component amplitudes and phases reconstructed exactly when normalized component frequencies are given by l_h / N'. Referring to a finite grid may appear as a limitation compared to interpolation and parametric methods where, in principle, frequency can be considered a continuous variable. However, finite signal-to-noise ratio (SNR) places a lower-bound as well, which is equivalent for all practical purposes to just considering a discrete fine grid.

While application of the CS algorithm does not eliminate spectral leakage, by reducing granularity on the frequency axis it increases, in normalised terms, the frequency separation between adjacent components. This virtually eliminates long-range frequency interference effects, leaving only scalloping loss to be dealt with. Of course, computational complexity of the algorithm is higher, but still within manageable limits for the processing capabilities of current equipment. The end result is that a similar order-of-magnitude improvement can be achieved in frequency resolution, with a bare 5% increase in measurement time.

In principle, interpolation allows to consider frequency as a continuous variable. However, finite signal-to-noise ratio (SNR) places a lower-bound, which is equivalent for all practical purposes to just considering a finer discrete grid with normalized step $\Delta'_\lambda < \Delta_\lambda$. Accordingly, any enhancement can be assessed by considering a finite *super-resolution factor*: $SRF = \frac{\Delta_\lambda}{\Delta'_\lambda}$.

So far, performance of OMP in support recovery has been found adequate for applications in electrical engineering, allowing to correctly detect all components of analysed multi-tone waveforms. Investigation into CS alternatives with better noise robustness could allow to better exploit in the future the potential of some of the features discussed in the chapter.

For lower SNR values, locating components lying on the finer grid requires preliminary subspace decomposition of a measurement correlation matrix. As this is obtained from M time-shifted sample sequences, each of length N, just a moderate increase in measurement time is required [76].

3.3 Part II: CS-DFT Application to Power Quality and Phasor Measurement

3.3.1 Power System Measurements Scenario

Measurement in power systems, particularly smart grids and microgrids, is often concerned with the analysis of voltage and current waveforms [77, 78], in which Fourier analysis plays a leading role. For instance, power quality analyzers (PQAs) can determine the harmonic content of current or voltage waveforms at selected points in a network [79, 80]. Phasor measurement units (PMUs) can also make use of Fourier-based algorithms to carry out time-synchronized measurements of phasors at the fundamental line frequency [81], [82], although the basic estimation model provided in IEEE Standard C37.118-2011 refers to a digital quadrature demodulator [83].

Any measurement algorithm based on the discrete Fourier transform (DFT) is characterized by a fundamental time-vs-frequency tradeoff. DFT coefficients obtained from a sequence of N samples, taken at uniform intervals T_s, are determined on a discrete frequency grid whose step size Δ_f is the reciprocal of the observation

time NT_s. This determines the minimum theoretical frequency separation at which two waveform components can be discerned. To resolve harmonics of the power-line frequency, observation time would have to be equal to at least one waveform cycle, T_0.

Resolution and accuracy in DFT-based algorithms are affected by spectral leakage, which can be countered either by windowing of the sample sequence followed by frequency-domain interpolation of DFT coefficients, or by synchronization of the sampling rate to the line frequency. The former approach has been considered for phasor measurement [84]. The need to allow for window mainlobe width means that, in principle, a larger component separation would be necessary, resulting in longer observation times, of between two and four waveform cycles [85]. Reporting latency is also affected since, for any DFT-based algorithm, this cannot be shorter than half the observation time.

One-cycle phasor estimators can be realized as well, as long as a very limited range of variation is allowed for the powerline frequency, so that inaccuracies due to residual leakage effects in off-nominal situations can be made small enough to be negligible [86]. This does not take into account the possibile presence of interharmonics, whose interference might affect phasor estimates.

Line frequency synchronization is considered, for the measurement of harmonics and interharmonics, in IEC Standard 61000-4-7, where interharmonic components are dealt with by defining interharmonic groups [87]. As a 5-Hz frequency resolution is required, suggested observation time is 200 ms (equal to $10T_0$ for 50 Hz systems or $12T_0$ for 60 Hz systems), again showing that any resolution improvement has to be paid for by an increase in measurement time.

These well-known conditions determine basic limitations of the Fourier approach, particularly when dynamic conditions are considered. To overcome the shortcomings of the static signal model underlying Fourier analysis different methods have been proposed, allowing phasor measurement [88], [89] and harmonic analysis [90] under dynamic conditions. These approaches were shown to be able to meet most or even all requirements for dynamic synchrophasor measurements as specified in IEEE Standard C37.118-2011, but require longer measurement times [91], [92]. Recently, Prony's method was shown to have the potential for accurate synchrophasor estimation in a one-cycle observation interval [93].

Phasor and harmonic measurements are commonly related to HV and MV distribution grids. However, the emerging scenario of smart microgrids [94, 95] points to the pervasive use of intelligent meters in LV distribution grids, providing continuously updated measurements of phasors and power quality indexes for use, e.g., in distributed control. Low-cost measuring equipment are expected to provide accuracy performance indexes similar to those already achieved by PQAs and standard PMUs, but under more distorted and disturbed operating conditions, as experienced in LV distribution grids. At the same time, the need for fast reporting rates is emphasized.

In this context, CS-DFT represents a promising candidate as accurate estimation algorithm for harmonic analysis, interharmonic and phasor measurement. The performance assessment in [40] employs specifically implemented benchmark signals, extending on the work already presented in [96].

Experimental results, carried out in non-ideal conditions, validate what has been previously demonstrated in theory. The proposed CS-DFT algorithm can achieve an order-of-magnitude improvement in frequency grid resolution, compared to the DFT step size Δ_f, without significantly extending total observation time.

While its application does not eliminate spectral leakage, the algorithm significantly reduces scalloping loss effects, while long-range frequency interference are dealt with by jointly processing frequency domain information when estimating waveform components [76]. Although still relying on a static signal model, the finer frequency grid of the CS-DFT algorithm allows to reduce the effects of dynamic conditions on measurement accuracy, so that requirements of IEEE Standard C37.118-2011 [83] can mostly be met. As the algorithm is non-parametric, it can be made adaptable enough to provide either PQA or PMU measurements by a single device as the need arises.

3.3.2 CS-DFT for Real-Valued Periodic Signals

Consider a multisine waveform expressed by the sum of a number of complex exponential components (with suitable symmetries):

$$s(t) = \sum_h \left[\frac{A_h}{2} e^{j(\phi_h + 2\pi f_h t)} + \frac{A_h}{2} e^{-j(\phi_h + 2\pi f_h t)} \right] \tag{3.27}$$

This equation can be referred to harmonic analysis, in which case $f_h = h \cdot f_0$ and $f_0 = \frac{1}{T_0}$ is the fundamental frequency, as well as to a more general situation where no relationship among frequencies may exist, e.g., in the analysis of power system interharmonics.

Given a sequence of N waveform samples, the DFT algorithm provides a corresponding set of N frequency-domain coefficients, that can be expressed by the equation:

$$S\left(\frac{k}{N}\right) = \sum_{h \in S_H} \frac{A_h}{2}\, e^{j\phi_h} D_N\left(\frac{k}{N} - \nu_h\right) e^{-j2\pi\left(\frac{k}{N} - \nu_h\right) n_0}\,, \tag{3.28}$$

where $0 \le k < N$, $\nu_h = f_h T_s$ is the frequency of the h-th waveform component normalized by the sampling rate $\frac{1}{T_s}$, and $D_N(\cdot)$ is the Dirichlet kernel:

$$D_N(\nu) = \frac{\sin \pi N \nu}{N \sin \pi \nu} e^{-j\pi(N-1)\nu}\,. \tag{3.29}$$

Index n_0 refers to the time position of the first sample in the sequence and will henceforth be set to 0.

DFT coefficients in (3.28) are given at discrete normalized frequency values $\nu = k/N$, that is, for integer multiples of the frequency step $1/NT_s$. Hermitian symmetry implies the sum in (3.28) extends to image components at normalized frequencies $1 - \nu_h$. In short, this is denoted by $h \in S_H$, where S_H is the set that includes all contributing complex exponential terms.

We now define a finer frequency grid, with a smaller step $\Delta'_f = \Delta_f / P$, so that the total number of grid points is $N' = P \cdot N$. The integer P can be called the *interpolation factor*. Referring waveform frequencies to this finer grid allows to express them as:

$$\nu_h = \frac{l_h + \delta'_h}{N'} = \frac{l_h + \delta'_h}{P} \cdot \frac{1}{N} \tag{3.30}$$

for some integer $l_h \in [0, 1, \ldots, N'-1]$, with $|\delta'_h| \le \frac{1}{2}$. The closest approximation to ν_h on this new grid is:

$$\hat{\nu}_h = \frac{l_h}{NP}. \tag{3.31}$$

Let $S_h \subset \{0,\ 1,\ \ldots,\ N'-1\}$ be the subset of integers corresponding to waveform components, i.e., $l \in S_h$ implies $l = l_h$ for some value of the index h. Replacing ν_h by $\hat{\nu}_h$ allows to approximate (3.28) as:

$$S\left(\frac{k}{N}\right) \simeq \sum_{\substack{l \in S_h \\ 0 \le l < N'}} \frac{A_h}{2}\, e^{j\phi_h} D_N\left(\frac{k}{N} - \frac{l}{N'}\right), \quad 0 \le k < N\,. \tag{3.32}$$

Of course, under our assumptions the number of elements in S_h, i.e., its cardinality $|S_h|$, is the same as $|S_H|$. Equation (3.32) can be written in a more compact form as a matrix relationship between the frequency grid indexed by k (the DFT index) and the finer grid indexed by l:

$$\mathbf{s} \simeq \mathbf{D}\mathbf{a}\,. \tag{3.33}$$

DFT coefficients are contained in vector $\mathbf{s}$ and elements of the $N \times N'$ matrix $\mathbf{D}$ are defined by: $d_{k,l} = D_N\left(\frac{k}{N} - \frac{l}{N'}\right)$. The unknown vector $\mathbf{a}$ contains complex amplitude values that can be associated to waveform components.

The actual measurement equation:

$$\mathbf{x} = \mathbf{s} + \mathbf{n} = \mathbf{D}\mathbf{a} + \mathbf{e}\,. \tag{3.34}$$

also accounts for measurement noise and uncertainty. Here $\mathbf{x}$ is the vector of computed DFT coefficients and it should be reminded that, if the additive random contribution affecting the N time domain samples has variance σ_n^2, then $\mathbf{n}$ is a zero-mean, uncorrelated random complex vector, with covariance $(\sigma_n^2/N)\mathbf{I}$, which turns out to be approximately Gaussian [97]. The vector $\mathbf{e}$ in the rightmost term summarizes both $\mathbf{n}$ and the approximation effect introduced by (3.31).

Given the vector $\mathbf{x}$, estimates of the components of $s(t)$ are the non-zero elements a_l in vector $\mathbf{a}$ contributing terms to summation (3.32), whose index values satisfy $l \in S_h$. Since $|S_h| \ll N'$, vector $\mathbf{a}$ is considered *sparse*, all its other elements being zero. S_h is called the *support* of $\mathbf{a}$.

In a nutshell, this is the problem feature that leads to considering a CS approach for Fourier analysis of the waveform measurement. The CS solution to (3.34) can be formulated as follows:

$$\hat{\mathbf{a}} = \arg\min_{\mathbf{a}} \|a\|_0 \quad \text{subject to: } \|\mathbf{x} - \mathbf{D}\mathbf{a}\|_2 \leq \epsilon\,, \tag{3.35}$$

where $\|\mathbf{a}\|_0$ indicates the ℓ^0 pseudo-norm, that is, the number of non-zero elements of $\mathbf{a}$.

As problem (3.35) is known to have combinatorial complexity, in practice $\hat{\mathbf{a}}$ is found either by convex relaxation (i.e., constrained minimization of the ℓ^1 norm $\|\mathbf{a}\|_1$ in place of $\|\mathbf{a}\|_0$), or by employing a so-called *greedy algorithm*. We shall follow the latter approach, which is computationally less demanding, using orthogonal matching pursuit (OMP) [25, 54] to recover the vector estimate $\hat{\mathbf{a}}$.

OMP is an iterative approximation algorithm whose steps are given as follows:

Step 1 the initial approximation residual $\mathbf{r}_0$ is equal to the measurement vector: $\mathbf{r}_0 = \mathbf{x}$, the initial estimated support set is empty: $\hat{S}_0 = \varnothing$. We also introduce the matrix $\mathbf{D}_{S_t}$, that will be composed only of the columns of $\mathbf{D}$ selected through the steps of the algorithm. Initially, $\mathbf{D}_{S_0}$ has zero columns. The iteration counter is set to $t = 1$;

Step 2 find the index l_t as:

$$l_t = \arg \max_{0 \le l < N'} \|\mathbf{D}^H \mathbf{r}_{t-1}\|_2, \tag{3.36}$$

where the superscript H denotes a complex conjugate transpose matrix. Iterations of this step cover the *support recovery* part of the algorithm;

Step 3 augment the estimated support set: $\hat{S}_t = \hat{S}_{t-1} \cup \{l_t\}$. The column of $\mathbf{D}$ with the corresponding index is appended to the matrix: $\mathbf{D}_{S_t} = \left[\mathbf{D}_{S_{t-1}}\ \mathbf{d}_{l_t}\right]$;

Step 4 calculate a new vector:

$$\hat{\mathbf{a}}_{l_t} = \arg\min_{\mathbf{a}} \|\mathbf{x} - \mathbf{D}_{S_t}\mathbf{a}\|_2 = \left(\mathbf{D}_{S_t}^H \mathbf{D}_{S_t}\right)^{-1} \mathbf{D}_{S_t}^H \mathbf{x}, \tag{3.37}$$

which has t non-zero elements. Iterations of this step provide a progressively refined estimate;

Step 5 calculate the new residual: $\mathbf{r}_t = \mathbf{x} - \mathbf{D}_{S_t}\hat{\mathbf{a}}_{l_t}$, then increment the iteration counter: $t = t + 1$ and return to Step 2.

The OMP algorithm can have two stopping criteria:

- if the number of non-zero elements of $\mathbf{a}$ is known in advance, the number of iterations t_{MAX} can be predefined: $t_{MAX} = |S_h|$;
- otherwise, the algorithm can be stopped when the approximation residual drops below a given threshold ϵ, as introduced by (3.35), in which case t_{MAX} is the first value of t that satisfies: $\|\mathbf{r}_t\|_2 < \epsilon$ and the *estimated* cardinality of the support of $\mathbf{a}$ is: $|\hat{S}_h| = t_{MAX}$.

The resulting sparse vector $\hat{\mathbf{a}}_{t_{MAX}}$ is the desired estimate of the components of (3.27). The important point is that, by defining a suitable approximation threshold, the algorithm can solve (3.35) without advance knowledge of the support cardinality $|S_h|$, although it may be useful to set a limiting number of OMP iterations as a safety criterion.

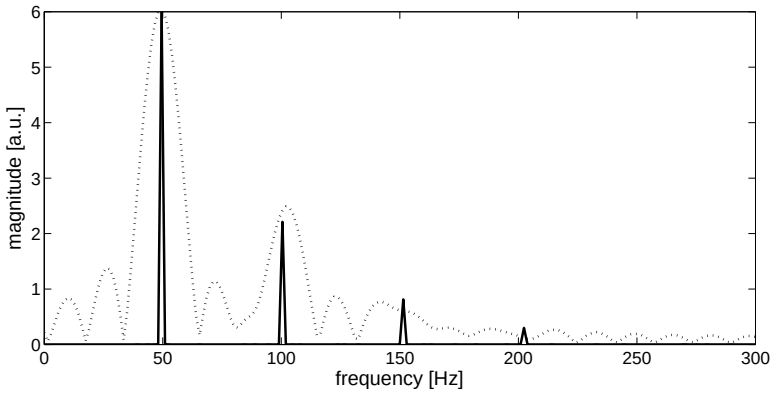


Figure 3.13: Magnitude estimates of a four-harmonic waveform obtained by the CS-DFT algorithm. A zero-padded 256-point DFT (dotted) is plotted for comparison.

As discussed in [76], the required minimum frequency separation between two waveform components lying on the fine grid is approximately $1.5 \cdot \Delta_f = 1.5 \cdot (\Delta_l / P)$, where Δ_l is the integer difference between the fine-grid indexes of the two components. When this condition is satisfied, OMP correctly recovers the support S_h and amplitude estimates are accurate. Furthermore, the finer grid allows to estimate component frequencies to within $\pm(\Delta'_f/2)$.

It should be remembered, however, that the latter quantity cannot be made infinitely small by the expedient of increasing P at will. Matrix **D** represents the effect of spectral leakage, through values of the Dirichlet kernel computed on a suitable two-dimensional grid. When large values of the interpolation factor are considered, the numerical conditioning of **D** tends to get worse. Thus, it is advisable to aim at values of P not much larger than order of 10.

3.3.3 Measuring a Simple Harmonic Waveform

The proposed CS-DFT approach promises to achieve enhanced frequency accuracy without requiring an extension of the observation time, allowing to overcome the traditional time-vs-frequency tradeoff of DFT-based methods. The CS framework provides a rigorous and effective procedure to determine from experimental data the support S_h on a fine frequency grid and, subsequently, reconstruct the composition of a waveform.

The example shown in Fig. 3.13 refers to the analysis of a periodic waveform

Table 3.2: CS-DFT estimation of a four-term harmonic waveform with SNR = 40 dB

Harmonic no.	**1**	**2**	**3**	**4**
Frequency [Hz]	49.50	99	148.51	198.01
Magnitude [a.u.]	6	2.21	0.812	0.299
Mean dev. (relative)	-4×10^{-5}	3×10^{-4}	3×10^{-4}	2×10^{-3}
Std. dev. (relative)	6×10^{-4}	1.5×10^{-3}	4×10^{-3}	1×10^{-2}
Phase [rad]	3.96	4.85	4.75	3.11
Mean dev. [rad]	1.2×10^{-4}	3×10^{-4}	-7×10^{-4}	-8×10^{-3}
Std. dev. [rad]	1.7×10^{-2}	2×10^{-2}	2×10^{-2}	2×10^{-2}

composed of four harmonic terms. The initial phase of each component is independently drawn from a random uniform distribution in the interval $[0,2\pi)$. Data record size is N = 256 samples and the assumed sampling frequency is 4100 Hz. This results in a DFT frequency grid step Δ_f = 16.02 Hz, while the finer frequency grid with an interpolation factor P = 11 is Δ'_f = 1.456 Hz. With these parameters, the fundamental frequency f_1 = 49.5 Hz is an exact multiple of the fine grid step: $f_1 = l_1\Delta_f$, with $f_1 = 34$.

The minimum separation condition given in the previous section is satisfied, allowing OMP to correctly locate on the fine grid the frequencies of the four components, with the corresponding magnitude estimates shown by the four sharp peaks in Fig. 3.13. For comparison, the figure also presents a plot of the magnitude of the DFT computed from a 256-point sample sequence, zero-padded with $(P-1)N$ zeroes. Since the latter is defined on the same finer grid as the CS-DFT, comparison readily evidences discrepancies and the adverse effects of interference among the waveform components, that would affect any purely DFT-based estimate. In particular, for the second harmonic the significant discrepancy in magnitude from the CS-DFT can be attributed to interference from the fundamental.

An assessment of CS-DFT measurement uncertainty for this example is presented in Table 3.2, which summarizes the outcomes from a set of 100 simulation runs where random white noise was added to the reference harmonic waveform. The assumed SNR = 40 dB can be considered representative of practical situations where

the equivalent number of bits (ENOB) of the waveform digitizer is low (e.g., an 8-bit analog-to-digital converter).

It can be seen that estimation bias is negligible when component frequencies fall on the fine grid. For all four harmonic components, the relative mean deviation from the reference values of magnitude and phase is almost one order of magnitude smaller than the relative standard deviation due to the presence of noise.

At SNR = 40 dB and with a coverage factor equal to 3, uncertainty on the phasor magnitude of the fundamental component would be better than 0.2%. For phase estimates, estimation uncertainty is about 0.06 radians.

It has to be mentioned that, for the smaller components, noise may sometimes affect the outcome of the support recovery procedure. In a single case, out of 100 simulations, the frequency of the third harmonic component was not estimated correctly. Results reported in Table 3.2 for this component do not include the outcome of the simulation run where support recovery partly failed.

Harmonic frequencies in Table 3.2 were intentionally chosen to fall on the CS-DFT fine grid. However, the equality $\nu_1 = (l_1/NP)$ denotes a coherent sampling situation, where l_1 periods of the fundamental frequency are *exactly* contained in an interval of length NP. In practice, it can be more realistically assumed that the measuring device will be designed so that the observation interval *approximately* matches the condition given above, that is, the CS-DFT algorithm will operate in

Table 3.3: CS-DFT estimation of a four-harmonic waveform with SNR = 40 dB off-grid components

Harmonic no.	**1**	**2**	**3**	**4**
Frequency [Hz]	50	100	150	200
Magnitude [a.u.]	6	2.21	0.812	0.299
Mean dev. (relative)	5×10^{-4}	-5×10^{-3}	1×10^{-2}	9×10^{-2}
Std. dev. (relative)	1×10^{-2}	2×10^{-2}	2×10^{-2}	2.5×10^{-2}
Phase [rad]	6.01	5.54	0.16	3.91
Mean dev. [rad]	5×10^{-2}	9×10^{-2}	-1×10^{-1}	-1×10^{-1}
Std. dev. [rad]	2×10^{-2}	1.3×10^{-2}	1×10^{-2}	8×10^{-2}

quasi-coherent sampling conditions. Results presented in Table 3.3 reflect this assumption, as the fundamental frequency $f_1 = 50$ Hz differs from the previous case by just 1%. Of course, in this case spectral leakage affects component estimates to some extent, introducing bias in magnitude due to scalloping loss and a limited amount of interference. To counter this, record length was set to $N = 256$ rather than the theoretical minimum $N = 128$, so that component frequency separation Δ_l/P is twice the limiting value previously defined. Comparison with Table 3.2 shows that both the estimate mean deviation and the standard deviation are increased.

3.3.4 Application to Power Systems

To show that power systems applications are within the range of feasibility of the proposed CS-DFT algorithm, in the following we shall consider in greater depth two suitable fields, power quality analysis and phasor measurement.

PQA – Harmonic and interharmonic measurement

The synchronized measurement approach of IEC Standard 61000-4-7 is well suited to test the CS approach, as far as harmonic and interharmonic measurement is concerned. In fact, a frequency grid step corresponding to the required resolution of 5 Hz could be obtained by the acquisition of even a single fundamental cycle of the waveform, using interpolation factor $P = 10$ (or $P = 12$). Accounting for the minimum components separation constraint, this would have to be increased to one and a half cycle.

It should be remembered that the IEC Standard assumes an observation interval equal to an integer number of power-line cycles. Therefore, sampling frequency needs to be adjusted accordingly (e.g., by phase-lock techniques) if the fundamental frequency varies. For the purposes of this analysis, it is equivalent and simpler to refer to a static condition with exactly coherent sampling. It is then possible to consider a degree of offset resulting in quasi-coherent sampling, although it seems reasonable to assume that in this case the measuring device would be designed to keep as close as possible to the ideal condition.

It should be noticed that interharmonics usually have far smaller magnitudes than some harmonic components, therefore spectral interference due to the quasi-coherent sampling condition might have more significance. To counter this more

Table 3.4: CS-DFT estimation of interharmonics in a four-harmonic waveform

Harmonic no.	**1**	**2**	**3**	**4**
Frequency [Hz]	50.23	100.46	150.69	200.92
Magnitude [a.u.]	6	5.16	4.44	3.83
Interharmonic no.	**0.5**	**1.5**	**2.5**	**3.5**
Frequency [Hz]	34.58	58.60	121.94	167.80
Magnitude [a.u.]	0.51	0.39	0.59	0.31
Deviation (relative)	1.4×10^{-2}–0.1×10^{-2}		-2×10^{-2}	0.1×10^{-2}
Phase [rad]	6.01	5.54	0.16	3.91
Deviation [rad]	0.16	0.1	-0.1	-0.2

effectively, the observation interval can be further lengthened.

Throughout this section, data record size is $N = 512$ samples and sampling frequency is 4100 Hz, which yields $NT_s \simeq 125$ ms, still almost 40% shorter than the interval suggested in the IEC Standard. The DFT frequency grid step is $\Delta_f = 8$ Hz, while the finer frequency grid with an interpolation factor $P = 11$ is $\Delta'_f = 0.728$ Hz.

The test case adopted for this study considers a waveform with 8 components, namely, four harmonics and four interharmonics. The latter are between 5% and 10% of the magnitude of the fundamental. The initial phase of each component is independently drawn from a random uniform distribution between 0 e 2π. In this controlled environment, the degree of sparsity is known in advance, namely, $|S_h| = 16$ including image terms. In practice, with suitable information about the maximum harmonic order and the extent of interharmonic distortion, a reasonable guess can be arrived at.

As far as the algorithm is concerned, results are independent of the nature, either current or voltage, of the analyzed waveform. Distorion levels considered in this example are more consistent with an experimental situation involving current waveforms, whereas smaller values would be involved in voltage harmonic and interharmonic analysis.

For this set of simulations no additive broad-band noise was included, as its effect is similar to what is already shown in previous sections and further discussed

in [76]. The analysis that follows is focused on the ability to detect and measure interharmonic components.

It should be noticed that harmonic magnitudes are larger than in the example of the previous section. Although possibly less realistic, this choice was made to stress-test support recovery performances for interharmonic components.

Table 3.4 refers to the case where all harmonic frequencies are located on the fine CS-DFT grid, thereby avoiding any interference with the much smaller interharmonic components. As the latter cannot instead be assumed to fall upon fine-grid points, the worst case has been considered, with frequencies falling exactly amidst two fine grid points. Hence, the OMP support recovery step will point to approximate frequency locations for interharmonic components (within ±0.364 Hz).

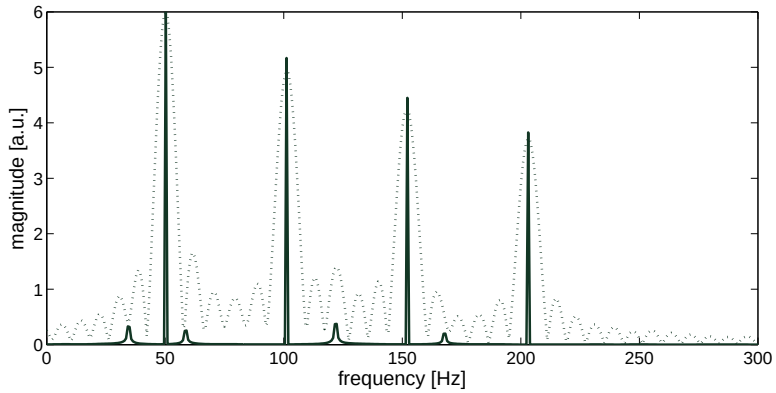


Figure 3.14: CS-DFT analysis of a signal with harmonic and inter-harmonic content. A zero-padded 256-point DFT (dotted) is plotted for comparison.

For the sake of simplicity, Table 3.4 does not report performance results for the harmonic components of the test waveform. In fact, their positioning on the fine frequency grid ensures they are exactly recovered, therefore the same results presented in Table 3.2 apply to them. Furthermore, they do not interfere with the estimation of interharmonics.

The algorithm correctly detects all interharmonic components achieving a satisfactory degree of accuracy, as shown by reported deviations from reference values in magnitude and phase. It should be remarked that scalloping loss effects with the CS-DFT algorithm are much smaller than in the standard DFT case. The resulting magnitude estimates are plotted in Fig. 3.14.

The quasi-coherent case can be analyzed by assuming 50 Hz as the fundamental frequency, which means a 0.5% deviation from the value given in Table 3.4. As far as harmonic components are concerned, results are similar to those already reported in Table 3.3 (actually slightly better, on account of the smaller off-grid offset) and will not be mentioned again.

For interharmonics, it should be reminded that the resolution required by the IEC Standard is 5 Hz. It seems reasonable, therefore, to assume this as a minimum frequency separation, in particular, between an interharmonic and a neighboring harmonic component. In terms of the fine grid step size given above, this distance is greater than $6 \cdot \Delta'_f$. In the "very nearly" coherent sampling conditions assumed for this test, the distance is large enough to ensure that interference caused by leakage from slightly off-grid harmonics is negligible. Therefore, results given for interharmonics in Table 3.4 also apply in this case. In some instances, DFT record size N could be adjusted to help meet the "large separation" condition.

PMU – Phasor measurement in steady state conditions

With reference to PMUs, compliance trials for phasor measurements have been carried out reproducing all test conditions specified in IEEE Standard C37.118.1 [83]. Steady-state results are reported here, in terms of TVE, while results for dynamic conditions are reported in the next section.

The total vector error (TVE) of a phasor measurement is defined as the relative difference between the phasor estimate and the theoretical phasor value of the signal being measured. Steady state compliance requires that TVE should not exceed 1%. For the purposes of this work, the only relevant difference between P-class and M-class devices concerns off-nominal frequency conditions, where a variability range of ±2 Hz has to be allowed in the former case, while a range of ±5 Hz is required for the M class.

Results discussed in this section refer to a data record size $N = 256$ samples. Sampling frequency is 5000 Hz, which corresponds to an observation interval of 51.2 ms. The DFT frequency grid step is $\Delta_f = 19.53$ Hz, while the finer frequency grid with an interpolation factor $P = 11$ is $\Delta'_f = 1.776$ Hz. It should be noted that phasor measurement refers to the fundamental frequency only, therefore the OMP algorithm can be terminated after just one iteration. Referring to the first of the two

stopping criteria aforementioned, this means setting $t_{MAX} = 1$.

As far as the effects of noise are concerned, support recovery is the critical part of the algorithm, as its task is to identify waveform components. Table 3.5 reports the maximum and root-mean-square (RMS) values of TVE at different SNR levels. It is assumed the fundamental frequency lies on the fine grid, so that only the effect of noise can be evidenced.

Table 3.5: Maximum and RMS values of TVE for different SNR levels

SNR	**MAX**	**RMS**
60 dB	2.7×10^{-4}	1.4×10^{-4}
50 dB	8×10^{-4}	4×10^{-4}
40 dB	3×10^{-3}	1×10^{-3}
30 dB	1×10^{-2}	5×10^{-3}
20 dB	3×10^{-2}	1×10^{-2}
10 dB	1.2×10^{-1}	6×10^{-2}

In the current OMP-based implementation, support detection becomes more difficult as the signal-to-noise ratio (SNR) drops below 20 dB [76]. This can be considered an acceptable performance limitation in power system measurement, with the possible exception of noise temporarily induced by transient phenomena such as lightning strikes or the opening of circuit breakers. In a realistic phasor measurement context, SNR could be expected to remain around 40 dB or better.

TVE variability caused by the addition to waveform samples of white noise with 30 dB SNR is summarized in the histogram of Fig. 3.15. This shows an approximately Rayleigh distribution, in agreement with the fact that, as noted above, DFT coefficients are affected by complex Gaussian white noise.

Accuracy analysis follows straightforwardly by considering (3.37) where, since $t_{MAX} = 1$, $\mathbf{D}_S$ is simply the vector $\mathbf{d}_{l_1}$ of the elements of $\mathbf{D}$ having the column index l_1 associated with the estimated fundamental frequency. Remembering (3.34), the estimate is:

$$\hat{a}_{l_1} = \left(\mathbf{d}_{l_1}^H \mathbf{d}_{l_1}\right)^{-1} \mathbf{d}_{l_1}^H (\mathbf{s} + \mathbf{n}), \tag{3.38}$$

where the noise contribution from $\mathbf{n}$ has zero mean and variance $(\mathbf{d}_{l_1}^H \mathbf{d}_{l_1}) \cdot \sigma_n^2 / N =$

σ_n^2/N, the equality following from the fact that $\mathbf{d}_{l_1}^H \mathbf{d}_{l_1}$ is the equivalent noise bandwidth of the Dirichlet kernel, that is equal to 1.

Assuming negligible interference from residual waveform components, in the neighborhood of l_1 the sum in (3.28) can be restricted to the fundamental, allowing to use for the DFT coefficients the simplified expression:

$$S\left(\frac{k}{N}\right) = \frac{A_1}{2}\, e^{j\phi_1} D_N\left(\frac{k}{N} - \nu_1\right). \tag{3.39}$$

The deterministic part of (3.38) is then:

$$\mathrm{E}\left[\hat{a}_{l_1}\right] = \frac{A_1}{2}\, e^{j\phi_1} \frac{\sum_{k=0}^{N-1} D_N^*\left(\frac{k}{N} - \frac{l_1}{N'}\right) D_N\left(\frac{k}{N} - \nu_1\right)}{\sum_{k=0}^{N-1} \left|D_N\left(\frac{k}{N} - \frac{l_1}{N'}\right)\right|^2}, \tag{3.40}$$

where $D_N^*()$ is the complex conjugate of $D_N()$.

It follows from (3.40) that TVE depends on the value of δ_1', which expresses the distance between ν_1 and l_1/NP as a fraction of the fine grid step. The plot showing the maximum value of TVE versus δ_1', for SNR = 40 dB, is given in Fig. 3.16. It is clear that the possibility to better approximate frequency, thanks to the fine CS-DFT grid, allows to keep TVE comfortably below the 1% limit.

Tests were also carried out with regards to the effects of harmonic distorion and out-of-band interference. Since the proposed algorithm has very good frequency selectivity, these effects are easily bypassed. The resulting maximum TVE is below 0.03% for out-of-band interference and better than 0.015% for harmonic distortion.

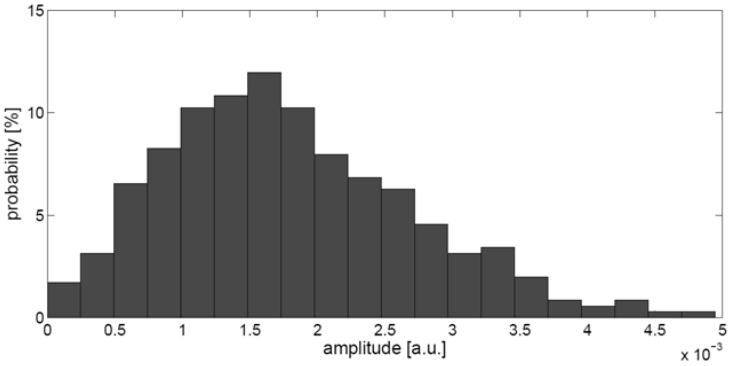


Figure 3.15: Histogram of TVE for the case of Table 3.5 and SNR = 30 dB.

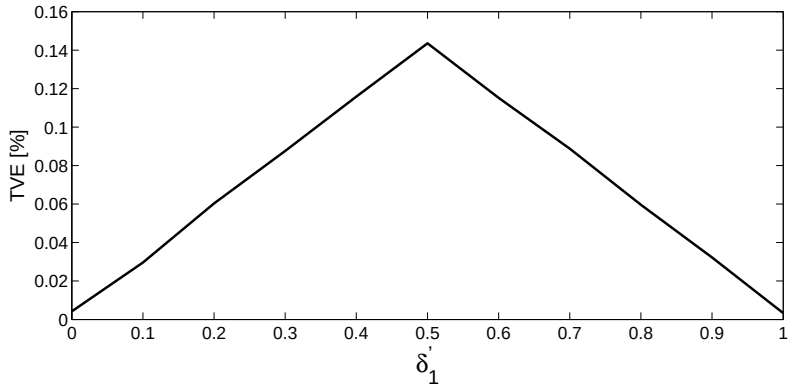


Figure 3.16: Variation of maximum TVE, calculated for SNR = 40 dB, as a function of the normalized offset from the fine grid, δ_1', of phasor frequency. The range of variation spans a full fine grid step.

3.3.5 Phasor Measurement – Dynamic Test Conditions

Dynamic compliance requirements specified by IEEE Standard C37.118.1 include an extensive set of test conditions. In the following, unless otherwise specified, the same measurement configuration employed in Section 3.3.4 is assumed. It should be remembered that, in this case, the reference phasor changes with time and, therefore, TVE will also be a time-varying quantity, whose value can be determined at reporting times t_r.

Assuming the CS-DFT algorithm can process a continuous stream of samples in real time, so that no dead time occurs between consecutive records (this will depend on the computing power actually available in the algorithm implementation), the maximum rate at which phasor measurements can be made available is $1/(NT_s)$. Since estimates obtained by a DFT algorithm are referenced to the centre of the relevant sample record, the CS-DFT estimate at reporting time $t_r = r \cdot NT_s$ has to be compared with the reference phasor at time $\left(r - \frac{1}{2}\right) NT_s$.

With these conventions, a number of TVE plots, referring to the different test conditions specified in the IEEE Standard, have been determined to characterize the algorithm.

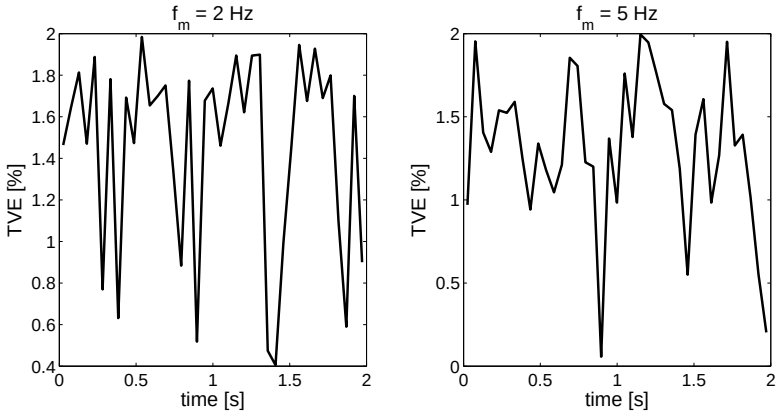


Figure 3.17: TVE plots for dynamic compliance measurement bandwidth test.

Amplitude and Phase Modulation

A test input with simultaneous amplitude and phase modulation is required to determine synchrophasor measurement bandwidth, according to the standard [83]. For single-phase measurement, the test signal is defined by the equation:

$$s(t) = A\big[1+k_x\cos(2\pi f_m t)\big] \times \cos\big[2\pi f_0 t + k_a\cos(2\pi f_m t - \pi)\big] \tag{3.41}$$

with $k_x = 0.1$ and $k_a = 0.1$ rad.

With the assumptions made above, the maximum reporting rate would be approximately 20 Hz. Modulation frequency f_m should be varied between 0.1 Hz and the maximum value of 2 Hz for P class PMUs, while for M class devices the maximum could be just one-fifth of the maximum reporting rate, i.e., 4 Hz. Nevertheless, simulation tests were still carried out up to 5 Hz and results are reported for this higher frequency.

TVE requirements for measurement bandwidth are somewhat mitigated, an error not greater than 3% being allowed. Fig. 3.17 shows plots of TVE for two different values of modulation frequency, $f_m = 2$ Hz and $f_m = 5$ Hz, calculated over a simulation time of 2 s. Although accuracy was found to progressively decrease as the rate of change of amplitude and phase gets higher, the algorithm complies with specifications for both M and P classes, with TVE never exceeding 2%.

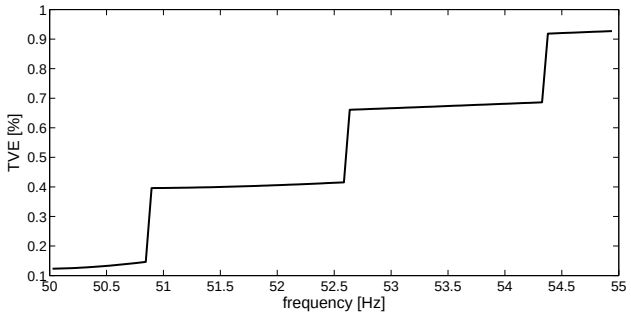


Figure 3.18: Plot of TVE versus instantaneous frequency $f(t)$ for the frequency ramp test. Frequency range: 50 to 55 Hz.

Frequency Ramp

In the frequency ramp test, signal frequency is ramped either up or down while keeping amplitude constant. Starting from the nominal frequency f_0, the frequency range is either ±2 or ±5 Hz respectively for P class and M class requirements. Instantaneous frequency is $f(t) = f_0 + R_f t$, with $R_f = 1$ Hz/s. The total frequency variation over an observation interval is $R_f \cdot NT_s$, which turns out to be rather limited, making this a quasi-stationary test input.

The TVE plot for an up-ramp from 50 to 55 Hz is shown in Fig. 3.18, similar results being obtained for the downward frequency ramp. Given the value of R_f the simulation covers a time interval of 5 s, but it is more useful to indicate the corresponding instantaneous frequency on the abscyssa. It has to be remembered that signal frequency falls exactly on the fine grid whenever $f(t) = l \cdot \Delta'_f$, with $\Delta'_f =$ 1.776 Hz.

It can be easily seen from Fig. 3.18 that TVE increases by steps whose corresponding width on the frequency abscyssa is exactly Δ'_f. By proper choice of the algorithm parameters, TVE in the neighborhood of the power-line frequency of interest can be kept below the required 1% limit.

Magnitude and Phase Step

In the step change test the signal consists of a pure sinusoidal tone, whose magnitude or phase is given a step variation for which the standard specifies, respectively, a ±10% change or a ±10 degrees ($\pi/18$ rad) phase shift. M and P classes re-

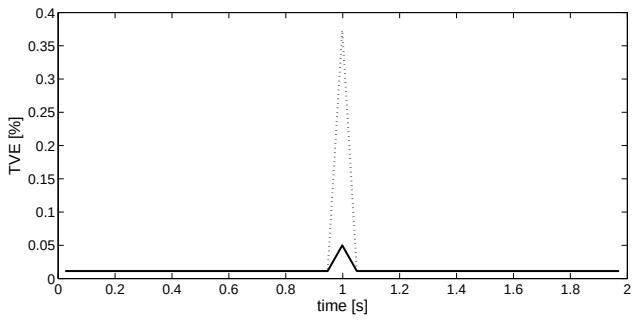


Figure 3.19: TVE plots for magnitude (+10%) and phase ($+\pi/18$) step changes. Continuous line indicates magnitude step TVE, dotted line phase step TVE.

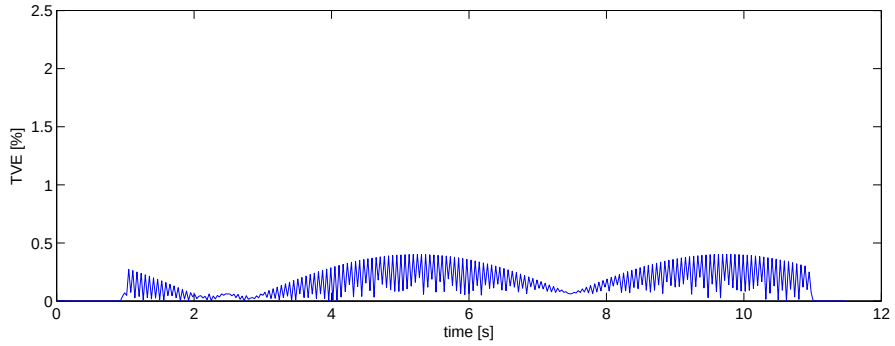
quire different response times, but are both required to keep TVE below 1%.
Fig. 3.19 shows that the algorithm complies with test requirements, remaining below the upper bound by almost an order of magnitude. Results refer to a +10% amplitude step and to a $+\pi/18$ phase step, plots for the corresponding negative steps being equal.

As can be expected with any DFT-based estimator, the effect of a step extends over twice the observation interval, that is, slightly more than 100 ms in this instance. It is apparent that a phase step has a much greater impact on TVE, yet this still remains well within specified bounds.

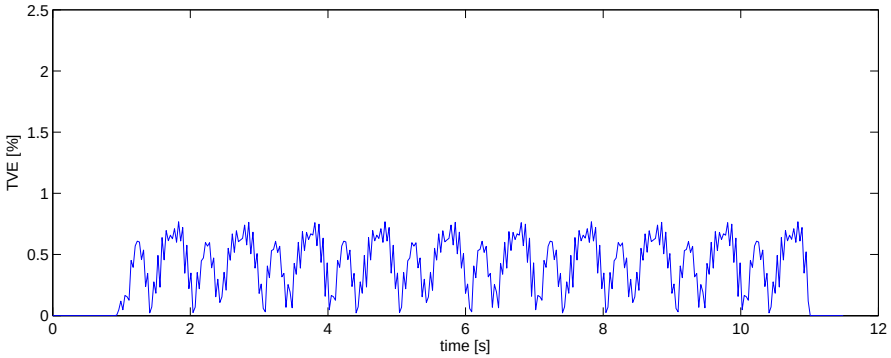
Experimental validation

Analysis by numerical simulation usually allows thorough characterization of a measurement algorithm. Nevertheless, further validation by means of some test bench may still be desirable. Results are found to be extremely close to those presented above, confirming the correctness of the simulation set-up employed in this work.

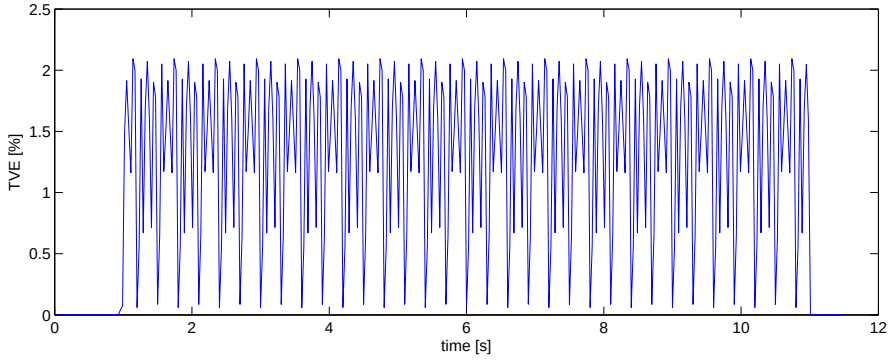
Simpler tests, notably for step changes, were performed by generating voltage waveforms with an Agilent 6812B AC Power Source/Power Analyzer having an adjustable output voltage up to 300 V. Signals are acquired with a National Instruments NI-cDAQ 9188 unit equipped with a voltage module (National Instruments NI 9225, 24-bit resolution, ±300 Vrms, sampling rate 50 kHz), and further processed by means of MatLAB code.



(a) $f_m = 0.1$ Hz;



(b) $f_m = 1$ Hz;



(c) $f_m = 5$ Hz;

Figure 3.20: TVE under combined amplitude and phase modulation at different modulation frequencies f_m.

The measurement bandwidth test has been implemented by a low-voltage test bench, which allows full characterization of measurement algorithms while neglecting the impact on PMU performances of HV or MV voltage and/or current transducer accuracies. Required voltage waveforms are obtained by means of a set of three laboratory voltage generators controlled by a LabVIEW software application. A Hewlett-Packard 8904A Signal Generator provides the output waveform and is connected to two Agilent 33220A Waveform Generators providing suitable modulation inputs. Samples are acquired through a National Instruments data acquisition module (NI USB-6211, 16-bit resolution, ± 10 V, sampling rate 250 kHz) and processed by means of MatLAB code, which yields TVE estimates at the desired reporting rate. Synchronization signals from the generators have also been recorded, for further processing and verification.

It should be noted that in both systems digitizer resolution is much higher than what was assumed in simulation analysis. Therefore, measurement variability can actually be lower than reported in the previous sections.

Results of the measurement bandwidth test are shown in Fig. 3.20. Each signal acquisition is 12 s long overall and consists of three portions. The first and last segment have a duration of 1 s and provide as a reference signal the unmodulated waveform in nominal conditions, that means, a pure sinusoidal tone at 50 Hz with initial phase equal to 0. The middle portion, instead, contains the test specific perturbation for a length of 10 s. Modulation frequency f_m can range from 0.1 to 5 Hz, while modulation parameters are the same considered in the numerical experiment in Section 3.3.5. The outcomes of simulation anaysis are fully confirmed.

3.3.6 Practical Feasibility and Computational Considerations

In the literature, DFT-based methods constitute a reference model for phasor measurement and power quality analysis. A thorough comparison is beyond the aim of this work, however it can be claimed that the combination of DFT and CS-based methods appears promising in achieving significant reductions of measurement time. Of course, computational complexity of the CS algorithm is higher than a simple DFT, but still within manageable limits for the processing capabilities of current equipment. In particular, computation of DFT coefficients by an FFT algorithm is a highly parallelizable task that can be implemented efficiently on a field-

programmable gate array (FPGA) architecture. This allows to quickly process sample records, while possibly leaving to a main processor in the measuring equipment the CS algorithm. For phasor measurement, however, the single OMP iteration could as well find room within an FPGA.

Discussion of different measurements in this chapter has shown that the CS-DFT algorithm is remarkably adaptable, allowing to easily trade speed for accuracy by simply changing the record length N. In PMU measurements, the moderate frequency separation requirement for a correct resolution of waveform components in any possible setting gives the possibility to keep the observation time well below two periods of the fundamental component. This in turn affects the choice of the number of samples in a record, N, for which the best solution should be a good compromise between peculiar application requirements and available computational power. The interpolation factor P can also be changed to some degree, adding a further possibility to optimize performance. By proper parameter settings, the algorithm could provide class M PMU accuracy at the faster reporting rate of class P devices.

The main features of the proposed algorithm make it of interest for application in smart microgrid projects. In fact, its basic structure can be adapted to a range of PQA and PMU measurements by a comparatively simple set of management and control rules, while the underlying hardware may remain unchanged. This goes towards the aim of realizing low-cost, high-performance measurement nodes in distribution-level smart microgrids, which are one of the targets of current research in the field.

From the viewpoint of complexity, the CS-DFT algorithm is more demanding than other DFT-based algorithms, but for PMU measurement is arguably on a similar level as weighted least squares and dynamic Fourier-based algorithms, with the advantage that shorter observation intervals are allowed.

Short-time high-resolution waveform analysis is usually associated with parametric, model-based methods, such as Pisarenko harmonic decomposition and its developments. As already noted, Prony's method, which has similar computational requirements to other model-based algorithms, has also been successfully applied to power system measurement [93]. While the CS-DFT algorithm is inferior in terms of accuracy, it may prove more adaptable for multi-function smart microgrid measurement.

Development of this work will follow along two lines, pointing to the study of efficient and cost-effective implementations, as well as to better performing algorithms for support recovery and signal reconstruction. Capabilities of the approach at lower levels of voltage harmonic and interharmonic distortion need to be further investigated, in connection with phasor and waveform analysis. Further aspects requiring consideration are performances beyond the scenario outlined by the requirements of IEEE Std C37.118.1-2011, particulalry in the more dynamic environment of distribution grids with distributed energy resources, for which the proposed CS-DFT algorithm should prove well suited.

3.4 Part III: From Finite Grid to Continuous Basis Pursuit Estimation

3.4.1 Off-the-Grid Estimation Scenario

Accurate measurement of a multisine waveform is a classic problem in spectral analysis, that also represents a benchmark for most signal processing algorithms. It is well-known that algorithms based on the discrete Fourier transform (DFT) have to contend with spectral leakage, which affects both amplitude estimation accuracy and frequency resolution [59]. It is also common knowledge that approaches based on a parametric signal model, such as Pisarenko harmonic decomposition, MUSIC and ESPRIT, can achieve much better frequency resolution, but this is obtained at the price of greater complexity [65].

A straight DFT-based estimate has an ultimate limit given by frequency granularity, that is consequent to the acquisition of a sample sequence having finite length N. For a sampling rate f_s, the resulting frequency step size is $\Delta_f = (f_s/N)$. The actual capability to resolve signal components at closely spaced frequencies is further limited, on account of spectral leakage, to a minimum separation $k_R\Delta_f$, with $k_R > 1$.

Interpolation of DFT coefficients [60] allows to overcome the granularity limit and consider frequency values defined on a continuous interval. In this way much more accurate estimates of waveform component amplitudes and frequencies, with variances approaching the relevant Cramér-Rao bounds, are possible [61]. However, the minimum separation between adjacent frequencies needs to be further in-

creased to avoid spectral interference, although this can be countered by weighting the time-domain samples by means of proper window functions [73].

A recent addition to the array of waveform analysis tools is the class of super-resolution algorithms based on compressive sensing (CS) [67], [68]. Basically, these algorithms allow to introduce a finer frequency grid with smaller step size $\Delta'_f = \Delta_f / P$, although continuous frequency values still cannot be obtained. However, minimum frequency separation limits are not always reduced accordingly [70].

In previous sections, it has been achieved what is so far the minimum separation limit for CS-based spectral analysis [76], fully analysed the features of a high-resolution CS algorithm employing an overcomplete dictionary [98] and successfully demonstrated the algorithm potential for power quality and phasor measurement [40]. Yet the granularity issue, although less significant, is not entirely overcome.

This led to consider a two-stage approach, whose features and performances are discussed in the following sections. A CS-based spectral analysis algorithm [98] provides the initial input to a second signal processing stage, which is inspired by an algorithm called *continuous basis pursuit* (CBP) [99]. As the name implies, CBP provides continuous-valued frequency estimates whose uncertainty is limited only by signal-to-noise ratio (SNR). Integration of the two steps into one effective algorithm requires some careful consideration of algorithm parameters, which is discussed in the following together with results obtained by simulation analysis.

3.4.2 Application of Continuous Basis Pursuit

Matrix $\mathbf{D}$ is also called a *finite dictionary* and its N' columns are termed the atoms of the dictionary. Let $\mathbf{d_l}$ be a generic column vector of $\mathbf{D}$, i.e., an element, indexed by l, defined as:

$$\mathbf{d_l} = [d_{0,l},\, d_{1,l} \,\ldots\, d_{N-1,l}]^T \quad 0 \le l < N'. \tag{3.42}$$

Support recovery in the CS-DFT algorithm associates each signal component to a single corresponding element of the finite dictionary $\mathbf{D}$. From (3.29) and the definition of $d_{k,l}$, it can be seen that elements in the dictionary are related by a frequency shift relationship, with the shift being an integer multiple of Δ'_λ. CBP allows to represent signal components by a linear combination of atoms, that is, to interpolate among neighbouring elements. This gives the ability to provide continuous-valued

frequency estimates.

Consider a vector $\mathbf{d}(\lambda)$ that depends on the continuous-valued quantity λ instead of l/N', and a generic shift $\Delta\lambda$. The set of all possible frequency-shifted and amplitude-scaled versions of this vector defines a non-linear two-dimensional manifold, $M_{\mathbf{d},\Delta\lambda}$. Setting: $\mathbf{d}_{\Delta\lambda}(\lambda) = \mathbf{d}(\lambda - \Delta\lambda)$, the manifold can be defined as:

$$M_{\mathbf{d},\Delta\lambda} \triangleq \{a \cdot \mathbf{d}_{\Delta\lambda}(\lambda) : a \geq 0, \Delta\lambda \in [0, 0.5]\}. \tag{3.43}$$

By construction, the sparse signal of interest belongs to $M_{\mathbf{d},\Delta\lambda}$. The discrete grid model, employed by the CS-DFT stage, constitutes a linear subspace approximation of the manifold. To obtain more accurate frequency estimates, it is necessary to explicitly account for the continuous dependency from frequency.

As discussed in [99], the ℓ_2-norm of $\mathbf{d}(\lambda)$ is unitary and translation invariant. Consequently, $M_{\mathbf{d},\Delta\lambda}$ lies on a hypersphere and has a constant curvature, so that it is reasonable to approximate it by an arc of a circle. For the problem considered in this work, the polar interpolation approach is considered.

Let λ_h be the frequency of a signal component, and $l_h \in S_a$ the element of the support providing its closest approximation on the dense CS-DFT grid. Generalizing [100], if a manifold segment defined in the interval $[\hat{\lambda}_h - m\Delta'_\lambda, \hat{\lambda}_h + m\Delta'_\lambda]$ is considered, an approximation can be defined by the circular arc which includes the three dictionary elements: $\mathbf{d}_{\mathbf{l_h}-\mathbf{m}}$, $\mathbf{d}_{\mathbf{l_h}}$ and $\mathbf{d}_{\mathbf{l_h}+\mathbf{m}}$.

In a proper system of polar coordinates, the "virtual" dictionary element at frequency λ_h is approximated by the use of a trigonometric spline, which yields:

$$\mathbf{d}(\lambda_h) \simeq \mathbf{c}_m(l_h) + \rho\cos(\delta_h\theta)\mathbf{u}_m(l_h) + \rho\sin(\delta_h\theta)\mathbf{v}_m(l_h) \tag{3.44}$$

where ρ and θ are, respectively, the radius and half the subtended angle of the arc.

Vectors $\mathbf{c}_m(l_h)$, $\mathbf{u}_m(l_h)$ and $\mathbf{v}_m(l_h)$ are obtained from:

$$\begin{bmatrix} (\mathbf{d}_{\mathbf{l_h}-\mathbf{m}})^T \\ (\mathbf{d}_{\mathbf{l_h}})^T \\ (\mathbf{d}_{\mathbf{l_h}+\mathbf{m}})^T \end{bmatrix} = \begin{bmatrix} 1 & \rho\cos\theta & -\rho\sin\theta \\ 1 & \rho & 0 \\ 1 & \rho\cos\theta & \rho\sin\theta \end{bmatrix} \begin{bmatrix} \mathbf{c}_m^T(l_h) \\ \mathbf{u}_m^T(l_h) \\ \mathbf{v}_m^T(l_h) \end{bmatrix} \tag{3.45}$$

The same approach can be easily extended to the entire set of component frequencies Λ identified during the CS-DFT stage. Let us introduce a more compact matrix formulation for the polar approximation:

$$\hat{\mathbf{x}} = \mathbf{C}_\mathbf{m}(\Lambda)\mathbf{p} + \mathbf{U}_\mathbf{m}(\Lambda)\mathbf{q} + \mathbf{V}_\mathbf{m}(\Lambda)\mathbf{r} \tag{3.46}$$

where $\mathbf{p} = \{p_h, \forall h : l_h \in S_a\}$ accounts for the component amplitudes, while $\mathbf{q} = \{q_h, \forall h : l_h \in S_a\}$ and $\mathbf{r} = \{r_h, \forall h : l_h \in S_a\}$ determine the frequency positions over the arc.

Equation (3.45) does not provide a closed form solution. The three coefficients have to be estimated by a proper optimization, where the cost function is explicitly designed to account for the polar space interpretation [99]. Peculiar constraints descending from the manifold geometry or the sparsity assumption prevent the problem from being ill-posed.

Once the coefficients are known, the component frequency can be better approximated by:

$$\tilde{\lambda}_h = \hat{\lambda}_h + \frac{\delta_h}{2\theta} \tan^{-1}\left(\frac{r_h}{q_h}\right) \tag{3.47}$$

From the computational point of view, the CBP stage exhibits a rather high complexity. In fact, the norm relaxation process involves at least some tens of iterations and operates with a set of three atoms for each component frequency. However, it is worth noticing that this procedure is computed only around the selected frequencies $\hat{\lambda_h} \in \Lambda$.

The higher computational effort is balanced by the capability to obtain off-grid estimation, even working on a reduced set of compressed measurements.

3.4.3 Spectral Analysis Results

The proposed method relies on two consecutive stages: the CS-DFT and the polar interpolation. The first one allows to increase the DFT grid granularity by one order of magnitude. The second one provides a further enhancement, explicitly accounting for continuous frequency dependency.

For performance validation, a set of tests has been designed to stress the algorithm in challenging situations, that combines different and concurrent phenomena.

The interpolation result is affected by the three dictionary elements selection, namely by the parameter m. In this context, the most suitable setting depends on the actual operative conditions, in particular, on the presence of interfering or partially superposed components. Accordingly, two different test conditions have been implemented in the Matlab programming environment. For the CS-DFT stage, the

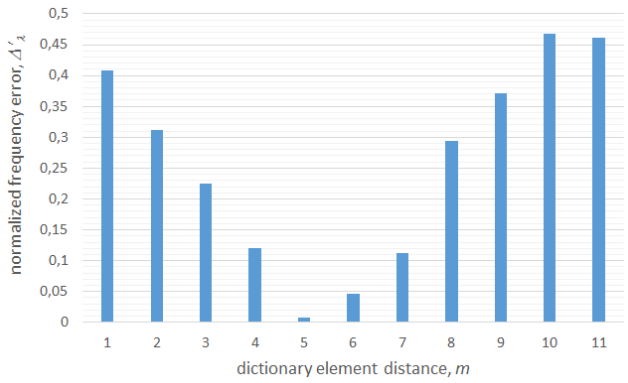


Figure 3.21: Normalized frequency estimation error for $m = 1,2,\dots P$.

refinement factor P is set equal to 11, corresponding to a resolution gain slightly larger than one order of magnitude.

In the first test, the signal under investigation consists of a single sinusoidal tone with normalized frequency $\lambda_0 = 155.45 \cdot \Delta'_\lambda$. It is worth noticing that the component frequency does not belong to the finite grid, independently from the observation interval length N, i.e. from the actual spectral resolution. In fact, the component frequency lies nearly halfway between two super-resolved finite grid points. Evidently, this represents the accuracy worst case for the CS-DFT estimate.

In the following, the frequency error is normalized with respect to the CS-DFT grid bin Δ'_λ. In Fig. 3.21, the estimation accuracy has been evaluated for $m = 1,2,\dots P$. The best performance is achieved for $m = 5$, i.e. $m \simeq P/2$. In other words, the best performance is achieved when the selected dictionary elements are separated by nearly $\Delta_f/2$.

The same test condition has been employed to assess also the dependency on the observation interval length N. In particular, given $m = 5$, the estimate accuracy has been evaluated in correspondence to different values of N. As depicted in Fig. 3.22, the frequency error decreases as N increases, and tends to settle around a $0.005 \cdot \Delta'_\lambda$ deviation. On the other hand, in the observed range the performance differences are limited and the dependency on N could be considered negligible.

The second test condition provides a multi-sine signal with two components of equal amplitude, partially superposed in the frequency domain. Once more, the worst case condition is taken into account. In fact, the signal components are sepa-

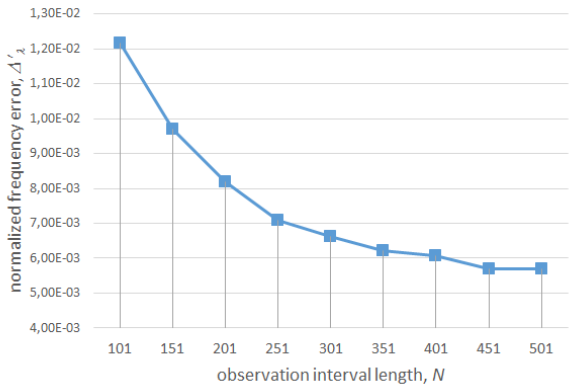


Figure 3.22: Normalized frequency estimation error for $N = 101, 151, \ldots 501$.

rated by $1.5 \cdot \Delta_f$, which is the minimum separation correctly resolved by the CS-DFT stage [98]. In particular, the signal consists of two sinusoidal tones whose frequencies are given respectively by $\lambda_0 = 155.45 \cdot \Delta'_\lambda$ and $\lambda_1 = 171.95 \cdot \Delta'_\lambda$. For the sake of simplicity, let us focus our attention to λ_0 estimation, but totally equivalent results can be obtained for λ_1.

With regard to the optimization problem, two different objective functions have been implemented and the corresponding dependencies on m are presented in Fig. 3.23. In more detail, the first function (bold) models only the λ_0 component, while the second one (dotted) accounts explicitly also for the λ_1 contribution. It is worth noticing that, in both cases, the best performance is achieved for $m = 3$. With respect to Fig. 3.21, at the starting point the trend is similar but rapidly increases for larger values of m. In fact, owing to the presence of an interfering component, the optimal distance between the selected dictionary elements should be lower than the single component case.

Moreover, the comparison between bold and dotted lines shows how the adoption of a more detailed objective function provides a significant reduction of estimation error and spectral interference effect.

Finally, a multi-sine signal comprising five components has been considered. The observation interval length has been set equal to 501 samples. The broadband disturbances have been modelled by an additive white Gaussian noise for a resulting SNR equal to 20 dB. Table 3.6 compares the component frequency actual values with the estimates provided respectively by the CS-DFT stage and the subsequent

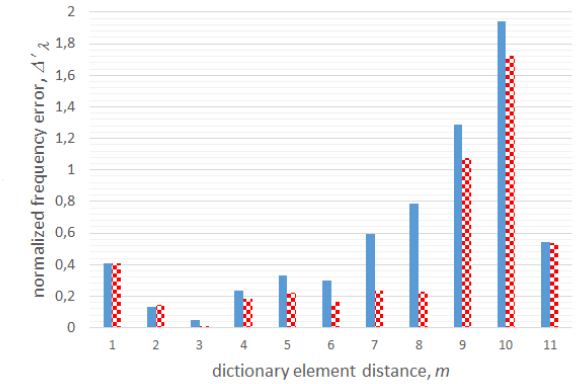


Figure 3.23: Normalized frequency estimation error for $m = 1,2,\ldots P$: bold line refers to a single component model, while dotted line refers to a more detailed signal model comprising also the interfering component.

Table 3.6: Normalized frequency estimation errors for a multi-sine signal in the Presence of 20dB AWGN.

COMPONENT INDEX	NOMINAL FREQ. [Hz]	CS-DFT NORM. ERR.	CBP NORM. ERR.
λ_0	150	−0.301	−0.188
λ_1	238	−0.294	0.047
λ_2	1456	0.178	0.099
λ_3	2561	0.241	0.262
λ_4	3925	0.122	−0.083

polar interpolation stage. The first ones are limited to $0.5 \cdot \Delta'_\lambda$, while the second ones provides an accuracy enhancement in almost any case.

3.4.4 Results Discussion and Improvement Issues

Results presented in these sections provide a full assessment of the enhancements in frequency estimation that the combination of CS-DFT and CBP algorithms can provide. The approach does not claim to be a universal tool for spectral analysis of multi-tone waveforms. Nevertheless, it shows interesting performances in the quest for overcoming the frequency limitations of DFT-based spectral analysis, in

particular, of the acquisition time vs frequency resolution trade-off.

Use of the combined algorithm allows to more accurately estimate components in a multi-tone waveform with short observation times, and could prove useful in applications where iterative refinement of waveform measurements is possible.

A still open question is represented by computational complexity. In particular, the CBP stage requires a noticeable effort to minimize the objective function. Similar convex relaxation techniques have to be considered and implemented, in order to minimize the computation time, preserving the off-the-grid estimation capability.

Chapter 4

Quasi-Stationary Sparse Signal Models in Dynamic Scenarios

4.1 Introduction

In the literature, conventional sparse signal models are usually related to stationary operative conditions. Stationariness of the sparse coefficients as well as the sparseness of their support, along with incoherence assumptions related to RIP, are fundamental to compressive sensing and sparse optimization. However, an always increasing number of sparse applications necessitate solving ill-conditioned optimization metrics and tracking rapidly fluctuating coefficients where such incoherence and stationariness assumptions are difficult to satisfy [101]. For instance, the DFT underlying signal model is static and periodic: in the presence of dynamic trends, any DFT-based method provides only an average estimation over the entire observation interval.

This gap between mathematical optimality and practical constraints can be closed by introducing more realistic sparse signal models which account also for time-variant conditions. Specifically, this chapter is concerned with quasi-stationary sparse signal models, where the sparsity and incoherence properties are assumed to be constant if evaluated over reasonably short observation intervals. In particular, a spectral analysis scenario has been taken into account and employed as benchmark for testing and validating the mathematical properties of the implemented algorith-

mic solutions.

Quasi-stationary spectral analysis consists mainly of two approaches.
The first one relies on assumptions like piecewise stationariness and conditional independence. Accordingly, the signal is divided into consecutive short data records, to be analysed singly. The spectral estimation is provided by DFT-based methods, explicitly designed to cope with reduced observation intervals. In this context, a typical example is represented by the so-called Short Time Fourier Transform, briefly STFT [102] which admits two different interpretations. On the one side, it can be seen as a DFT computed over short time records, in case weighted by proper windowing functions. On the other side, it can be seen as a bank of filters, properly designed to exhibit good estimation performances also on reduced observation intervals.
The second approach, instead, involves a Taylor expansion of the canonical DFT representation. Higher order derivative terms are included in the signal model and explicitly account for dynamic contributions. The computational effort is expected to increase with respect to canonical DFT-based methods. On the other hand, a significant enhancement in terms of estimation accuracy is expected. In this context, a typical example is represented by the so-called Taylor-Fourier Transform, briefly TFT [90]. Just like STFT, also TFT admits two plausible interpretation, the one driven from functional analysis, the other in terms of filtering stage.

Independently from the implementation details, both these approaches necessitate dealing with short time data records, which are expected to severely affect the actual resolution power in the frequency domain. On the other hand, it is worth noticing that sparsity or incoherence property have not been considered. As a matter of fact, CS theory can be coupled with quasi-stationary spectral analysis, in order to efficiently capture the original information content even from a reduced set of measurements. In quasi-stationary conditions, the signal support is expected not to vary significantly over the short observation interval. In other words, dynamic effects do not succeed in blurring or distorting the signal support, which could be recovered with sufficient sharpness.

In such a scenario, a two stage procedure is advisable. The CS acquisition protocol yields the informative component frequencies. Properly exploiting this a priori knowledge, the ad hoc quasi-stationary methods estimate component ampli-

tudes and initial phases. In this chapter, a plausible combination between CS-based super-resolution and TFT-based spectral analysis is discussed and validated by means of experimental results, coming from two different application contexts. The chapter is divided into three main parts.

In the first part, the notion of quasi-stationary signal model is introduced and its implications to spectral analysis are briefly discussed. Then, particular attention is devoted to TFT formulas and properties. The pros and cons of this novel approach are emphasized by means of intuitive interpretations as an orthogonal projection onto a specific vector subspace or as a filter bank. Finally, the most common measurement approaches related to TFT are summarized. In this context, the proposal of a joint CS-TFT approach is introduced as a novel signal-independent approach for spectral analysis in quasi-stationary operative conditions.

In the second part, this combined approach is applied to power measurements in smart grid scenario. More precisely, CS-TFT is proposed as a Phasor Measurement Unit (PMU) algorithm. The aim is to exploit, in a joint method, the properties of CS and TFT to identify the most relevant frequency components of the signal, even under dynamic conditions, and to model them in the estimation procedure, thus limiting the impact of harmonic and interhamonic interferences. The approach is verified using composite tests derived from the test conditions of the current synchrophasor standard and simulation results are presented to show its potentialities.

In the third part, the problem of gradient artefact removal in simultaneous recording of electroencephalography (EEG) and functional magnetic resonance imaging (fMRI) is addressed. A novel removal algorithm, based on CS-TFT approach, is proposed and validated on both simulation and experimental data. Experimental results show a significant reduction of spurious components in all the considered conditions. No significant distortions are introduced in spectral power distribution, allowing reliable clinical interpretation of the acquired trace.

4.2 Part I: Taylor-Fourier Transform for Quasi-Stationary Spectral Analysis

4.2.1 Quasi-Stationary Signal Models

Let us consider a generic signal acquired in consecutive partially overlapping records, each one related to a specific short time frame. The signal is said to be quasi-stationary if its statistics are locally static in each frame but vary from one frame to another. This signal model is particularly suitable for acquisition protocols and applications which operate at high sampling rates. For instance, speech and audio recordings are generally recognized as quasi-stationary signals.

In quasi-stationary conditions, spectral analysis must be performed by means of ad hoc optimized methods. Indeed, conventional DFT approaches imply two main limitations.

On the one hand, DFT-based estimates are defined over a discrete grid, which unifprmly span the observed spectral bandwidth. The larger is the number of transform coefficients, the larger is the grid granularity, and accordingly the lower is the estimation error. In quasi-stationary conditions, the observation interval length is reduced by definition. In order to increase granularity, it could be useful to increase the sampling frequency. However, this solution lacks of practical convenience, because it implies an inefficient oversampling, and even of theoretical motivation: an higher sampling frequency corresponds to a larger spectral bandwidth to span.

On the other hand, DFT-based methods rely on a static and periodic signal model. Accordingly, in the presence of dynamic conditions, DFT-based estimations consist in the average approximation computed over the entire observation interval. It should be noticed that this second limitation does not provide an immediate expedient. The proposed signal model is excessively ideal and provides a too scarce approximation of the actual operative conditions.

Many non-conventional approaches have been implemented for quasi-stationary signal spectral analysis. Independently from the algorithmic detail, a common requirement is the capability to provide accurate estimations from reduced set of measurements. To this end, CS principles could be employed to further improve the estimation performances. A CS approach could provide not only a resolution enhancement, as shown in the previous chapter, but also a more flexible sensing protocol,

scaled to the information rate, rather than the Nyquist's rate. In particular, this chapter is concerned with an extension of canonical DFT, also known as Taylor-Fourier Transform. In the following section, its fundamental properties are summarized and discussed, even if a thorough characterization is beyond the scope of this chapter.

4.2.2 Taylor-Fourier Transform

Let us consider a dynamic multi-tone sampled signal model:

$$x[n] = \sum_{h \in S_h} \frac{a_h[n]}{2} \cos(2\pi f_h n T_s + \phi_h[n]) \qquad n = 0, 1, \dots N \tag{4.1}$$

where S_h is the signal support, i.e. the set of the significant component indices. Typically, the first support index is related to the fundamental or most informative component, while the others represent harmonic and interharmonic components, enumerated according to a descending order of importance.

The same model could be expressed also adopting a phasor formulation:

$$x[n] = \sqrt{2} \left[\sum_{h \in S_h} p_h[n] e^{j2\pi f_h n T_s} \right] \qquad p_h[n] \triangleq \frac{a_h[n]}{\sqrt{2}} e^{j\phi_h[n]} \tag{4.2}$$

where the amplitude RMS value is extrapolated, $T_s = 1/f_s$ is the sampling period and N determines the length of the observation interval.

Taylor-Fourier transform, briefly TFT [90], consists in a K-th order Taylor series expansion of the canonical DFT. In particular, taking into account the transformation kernel of both DFT and TFT, it is immediate to notice their reciprocal connection:

$$\begin{aligned} &\text{DFT:} \quad \psi_h[n] = e^{-j2\pi f_h n T_s} \\ &\text{TFT:} \quad \psi_h^k[n] = \tfrac{(nT_s)^k}{k!} e^{-j2\pi f_h n T_s} \qquad k \in [0, 1, \dots K] \end{aligned} \tag{4.3}$$

where K represents the Taylor series expansion order. According to TFT approach, each component belonging to the support is assumed to be modulated by a time function, expressed by a Taylor polynomial expansion of a given order $k > 0$. Thus, TFT coefficients contain both the instantaneous values and the time derivatives of component amplitudes and phases, computed with respect to the observation interval midpoint:

$$\theta_h^k[n] = \frac{T_s^k}{2k!} \cdot \frac{d^k a_h(t) e^{j\phi_h(t)}}{dt^k} \tag{4.4}$$

In the TFT domain, the dynamic signal model (4.2) can be expressed as follows:

$$x[n] = \sqrt{2} \quad \Re\left[\sum_{h\in S_h}\left(\sum_{k=0}^{K_h}\theta_h^k[n]\right)\psi_h^k[n]\right] \tag{4.5}$$

Compared with the canonical DFT approach, the TFT-based model allows a more accurate and detailed signal representation. In fact, by introducing the higher order derivative terms, TFT is explicitly designed to cope with dynamic conditions. More precisely, the higher order coefficients account for time-varying contributions, while the static one accounts for stationary contributions.

It is worth noticing that TFT 0-th order coefficients are formally equivalent to DFT counterparts. On the other hand, TFT accuracy is reasonably expected to overcome DFT accuracy. Eventual dynamic contributions are entirely captured by higher order derivative terms and do not affect static term estimation.

Functional Analysis Interpretation

Let $\mathscr{H}_n(\mathbb{R})$ be a generic n-dimensional Hilbert space of real-valued functions. No constraints are introduced with reference to functions periodicity, stationary or bandwidth properties. In this context, DFT and TFT are defined over two specific sets of linearly independent basis functions, one for each signal component:

$$\begin{aligned} &\text{DFT basis:} \quad \Psi_D = \{e^{-j2\pi f_h n T_s}, h \in S_h\} \\ &\text{TFT basis:} \quad \Psi_T = \left\{\tfrac{(nT_s)^k}{k!}e^{-j2\pi f_h n T_s}, k \in [0,1,\dots K], h \in S_h\right\} \end{aligned} \tag{4.6}$$

Let $x \in \mathscr{H}_n(\mathbb{R})$ be a generic n-dimensional signal. From a linear algebra point of view, the signal DFT or TFT coefficients result from its orthogonal projection over the vector subspaces spanned respectively by the basis Ψ_D and Ψ_T. More precisely, the TFT coefficients can be computed in closed form as the following WLS solution:

$$\theta_h^k = \left(\Psi_T^H W^H W \Psi_T\right)^{-1} \Psi_T^H W^H x \tag{4.7}$$

where the superscript H denotes the Hermitian transpose matrix and W is a weighting matrix, which could be properly defined according to specific application requirements or boundary conditions.

It is worth noticing that TFT basis functions span a subspace which contains the DFT subspace. It is thus rigorously proven that TFT approximation error is less

than or equal to the DFT approximation error. As aforementioned, even estimates for $k = 0$ improve on the DFT counterpart affected by infiltration of the neglected derivative terms.

The most significant difference between DFT and TFT basis is, though, concerned with orthogonality property. DFT relies on an orthonormal complete basis, namely on a Hilbert basis. Conversely, TFT basis vectors form only a frame, lacking of orthogonality property.

Accordingly, the well-known Karhunen-Loève theorem holds in the DFT subspace, but not in the TFT subspace. In other words, in the DFT subspace, any signal can be decomposed as a linear combination of the orthonormal basis vectors. Conversely, in the TFT subspace, this is not guaranteed. In fact, basis vectors linear independence is sufficient for ensuring that a TFT solution exists [90], but particularly negative conditions may severely affect the conditioning number of Ψ_T. For instance, if the signal support contains much similar frequencies, the differences between respective vector sets are hardly noticeable, particularly if the higher order terms are considered. In such conditions, dynamic contributions related to a specific component may leak in the adjacent component expansion and vice versa. As a result, the condition number of Ψ_T degrades significantly and the corresponding coefficients (4.7) may be unreliable and inaccurate.

In conclusion, the functional analysis interpretation of TFT shows that:

- a properly designed TFT approach can cope with dynamic conditions and overcome the corresponding DFT approach accuracy;
- the TFT approach accuracy depends on the signal support likelihood: in the presence of spurious or erroneous frequencies, the basis matrix Ψ_T might be severely ill-conditioned.

Filter Bank Interpretation

The same TFT admits also a filter bank interpretation, as depicted in Fig. 4.1. For each component of interest, a specific bank of filters is centred around the component frequency. The bank consists of K filters, one for each derivative order. It should be noticed that filter response shape varies on varying the derivative order:

- 0-th order filter captures static contributions;

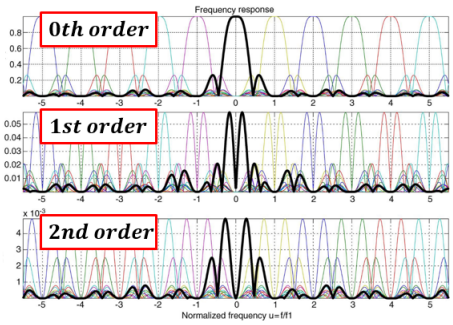


Figure 4.1: Filter bank interpretation of TFT [103].

- 1-st order filter senses eventual amplitude or phase variations;
- 2-nd order filter assesses variation gradients.

Each filter processes the acquired data record simultaneously and returns the respective TFT coefficient. Preliminary weighting functions are often employed to enhance filtering stage performances. For instance, proper data windowing results in narrower and nearly flat 0-th order bandpass response.

An accurate frequency support becomes crucial also in the filter bank interpretation. If the filters are centred around the desired components, the information content is entirely recovered, whereas the disturbance injections from adjacent components are limited as much as possible. Conversely, if the filters are inadequately located, it is reasonable to expect that the original information content undergoes partial losses or distortions.

As observed in the previous section, too similar component frequencies may lead to inaccurate results. In the filter bank approximation, the motivation appears clear and immediate. Even if the filter is correctly located, adjacent components lie within the bandpass and cause non-negligible disturbance injections. These mutual interference phenomena degrade the TFT estimation accuracy and confirm the require narrow bandpass filter responses to be mitigated.

4.2.3 Taylor-Fourier Expansion Advantages

As far as spectral analysis is concerned, the main advantages of the TFT approach can be summarized as follows:

- each component is projected on its own set of basis vectors, resulting in lower disturbance injection;
- dynamic effects are captured by higher order terms, ensuring a better static estimation.

On the other hand, TFT approach performances depend from the underlying spectral content model:

- the frequency support must be accurate and detailed: this knowledge a priori allows to design a specific TFT basis, thus preventing from information distortion or loss;
- the frequency support must include both informative and spurious components: in this way, the second ones are projected over specific vector sets, thus limiting their disturbance injection over the informative components.

4.2.4 TFT-Based Measurement Approaches

The component frequencies knowledge a priori plays a crucial role in any TFT-based measurement algorithm. Hereinafter, the case of a sinusoidal tone at 50 Hz, potentially affected by harmonic and interharmonic disturbances, along with additive spread spectrum uncorrelated noise. Furthermore, the signal is expected to undergo dynamic variations of frequency, amplitude and phase values. This peculiar operative condition is typical of power system scenario, particularly of distribution grid scenario, which is thoroughly discussed in the second part of this chapter. Nevertheless, similar considerations may be inferred in much different contexts, where a quasi-stationary spectral analysis by means of TFT is implemented.

As aforementioned, the TFT vector subspace is constructed according to the signal spectral content model at disposal.
Conventional TFT-based approaches employ the nominal frequency, without trying to infer any further knowledge from the measurement data:

1. TFT-based estimation computed around the only nominal fundamental frequency 50 Hz;
2. harmonic component modelling for a TFT expansion around all the nominal harmonic frequencies 50, 100, 150, … Hz.

Experimental results suggest to include all the significant components, both informative or spurious, within the spectral content model. To this end, the component frequencies are treated as unknown model parameters to be identified before computing the TFT-based estimation:

1. adaptive version of weighted least square TFT approximation, capable to detect fundamental transients or fast changes and accordingly update the corresponding spectral model;

2. signal spectral content model (comprising fundamental, harmonic and interharmonic components) identified by means of CS super-resolution technique and employed as reference to locate TFT basis vectors on a fine frequency grid.

In the following, this last approach is also referred to as CS-TFT, emphasizing that different theories can be coupled, without loosing their respective prerogatives.

4.3 PartII: CS-Based Enhanced Model for Synchrophasor Estimation by Taylor-Fourier Transform

4.3.1 Smart Grid Scenario

The contemporary society is developing an always-increasing consciousness and concern towards the environmental issues. In this context, the use of renewable energy is crucial to cope with emission reduction targets and ensure stability and feasibility of electrical supply. Unfortunately, these alternative energy sources are intermittent and unless the energy flows are measured and controlled, their exploitation could cause power quality degradation, even leading to wide-spread blackouts.

The so-called “smart grid” projects introduce a novel concept of electrical grid: no longer an inert distribution means, but rather an automated organism which gathers information about grid state and accordingly act to improve efficiency and sustainability. It is likely to expect that smart grids will allow the reliable and enduring use of large amounts of renewable energy. In this scenario, new measurement tools are essential for their stable operation. The GPS technology has opened up an efficient synchronization infrastructure over wide geographical areas: the knowledge of grid parameters and behaviour in different grid points and referred to the

same common time could give helpful advice to design, control and stabilize the smart grids of the future.

Currently, smart grids are primarily required to balance renewable supply with demand in real time, maintaining the grid stability and avoiding the power quality degradation. However, the penetration of distributed renewable generation causes high challenging dynamic conditions, prone to cascading failures or even power blackouts. Network operators need tools to accurately and promptly assess the grid state, the power quality and the supply stability.

The Phasor Measurement Units (PMUs) technology represents a great opportunity to understand the complex dynamics of a smart grid. At a given grid point, they measure the electrical waves in terms of RMS magnitude, instantaneous phase, frequency deviation from the nominal value, and rate of change of frequency.

Typically, the PMUs have been employed to monitor the high voltage (HV) transmission lines. It is only recently, however, that PMUs information has been exploited to estimate the grid state and the network impedance. Moreover, it is widespread among scientific community the belief that the PMUs could be applied also to medium and low voltage (respectively, MV and LV) distribution lines. Such a context opens up both more severe requirements and more promising potentialities. On the one hand, the shorter lines and the non-negligible disturbances require an ad hoc measurement algorithm, which should be capable to provide standard performances also in non-ideal conditions. On the other hand, a capillary knowledge of grid variables allows more efficient control strategies and more precise fault localization.

In conclusion, the design of PMUs for application in electric distribution grids is an extremely challenging task. Similar units have been successfully deployed in high-voltage electric power transmission networks; a PMU design for use in distribution networks is both more demanding, in terms of cost/performance tradeoffs, and more promising in allowing the efficient inclusion of distributed renewable energy sources.

4.3.2 PMU Estimation Algorithms

Phasor Measurement Units (PMUs) are the most innovative measurement devices in power network monitoring and are expected to become a fundamental tool for managing and supervising both transmission and distribution networks. IEEE

Standard C37.118.1-2011 [83], along with its amendment [104], defines PMU outputs and applicable measurement accuracy limits. In particular, synchrophasor, frequency and rate of change of frequency (ROCOF) measurements are introduced with the specific aim of describing the behavior of power network signals under dynamic conditions. Accuracy requirements are presented for steady-state conditions, in the presence of off-nominal frequency, harmonic and interharmonic disturbances, as well for dynamic conditions, such as amplitude and phase modulations, linear frequency ramp and step changes for both amplitude and phase-angle. Two performance classes are defined by the standard, M-class and P-class. The former is intended for measurement applications, where accuracy is emphasized, while the latter is designed for protection, which requires fast responses to dynamic events that can be critical for network operation.

Several algorithms for PMU applications have been proposed and characterized in the literature (notably, [89, 105, 106, 107, 108, 109, 110, 40]) and rely on different signal processing techniques and models. For instance, the Interpolated Discrete Fourier Transform (IpDFT) approach is used for synchrophasor and frequency estimation in [107, 111], exploiting the capability to cope with static off-nominal conditions. In [108] a demodulation and filtering approach, with frequency tuning, is followed to fulfill P- and M-class requirements, as suggested also by Annex C of [83]. In [89, 105, 109, 106, 112], phasor estimation is improved by approximating slowly changing phasors with a complex Taylor series expansion around the estimation time point. In [89] and [105] better synchrophasor estimation performance is achieved by correcting the estimation errors of sequential phasor estimates computed with DFT and Short-Time Fourier Transform (STFT) by post-processing. In [109] the IpDFT is extended to compute the phasor derivatives of a second order expansion, thus also frequency and ROCOF, from three DFT components around the fundamental frequency.

An algorithm using a linear non-orthogonal transform, defined as Taylor weighted least squares (TWLS) is introduced in [106] and generalized to harmonics as Taylor-Fourier Transform (TFT), in [90]. It is based on a weighted least squares (WLS) approximation of an observation window with respect to a second order Taylor model, performed as a linear filter bank.

As investigated in [113], TFT-WLS synchrophasor estimation outperforms other

methods in all dynamic conditions except under step tests. To this end, an adaptive version of the TFT-WLS, which detects when the signal is undergoing fast changes and, then, refines phasor estimation, has been proposed in [112] and [114] to enhance performance under transient conditions. In [110] the adaptive algorithm and a suitable detector are used to define a single output PMU that complies with both P-class and M-class requirements for synchrophasor and frequency measurements, at the reporting rate of 50 frames/s. A detection approach, using the filtering paradigm, is also used in [115] to realize a hybrid M-class PMU with faster responses to transients and good measurement under steady state conditions.

The limits set by the standard and their physical meaning have been thoroughly discussed and this even led to the revision or suspension of some specific requirements [104]. Great attention is being paid to the design of estimation algorithms with the capability to follow the relevant current or voltage dynamics and to cope with the disturbances that can actually arise in the context of emerging power networks, where highly evolving scenarios are expected.

It has been highlighted that some of the disturbances, in particular those related to out-of-band (interharmonic) interfering signals, can be extremely difficult to deal with and can seriously affect synchrophasor, frequency and ROCOF estimations [115]. To limit the injection of such disturbances in the signal estimation at the fundamental frequency, the filtering properties of long measurement observation intervals are usually exploited. On the contrary, the ability to follow signal dynamics requires fast responsiveness to changing conditions and thus asks for shorter observation windows and wide frequency pass-bands around the relevant component of the measured signal.

With performance limits stemming from such fundamental time-frequency trade-off, requirements are seemingly impossible to reconcile. However, recent research into the application of Compressive Sensing (CS) has shown that it is in fact possible to push beyond those limits and achieve frequency super-resolution or, alternatively, enhanced short-term performance [76, 98].

The possibility to cope with such contrasting requirements, by means of the combined use of a CS approach for frequency support estimation and a TFT adapted to the CS stage outputs, has been investigated in [116] for a two-step estimation algorithm. The idea was to find the spectral components of the signal, in order to

model them and limit their impact on synchrophasor estimation, while keeping latency and response times reasonably low [40]. In the following sections, a unified compressive sensing Taylor-Fourier transform (CS-TFT) is introduced, where frequency support estimation also considers higher-order derivatives, and allows simultaneous estimation of the phasor. The proposed approach is tested under different conditions, even beyond strictly standard specifications, with particular attention to out-of-band interference and possible concurrent disturbances.

4.3.3 Modeling for Accurate Phasor Measurement

Synchrophasor Dynamic Model

According to [83], the time-varying synchrophasor representation of a sinusoid signal $x(t)$, whose amplitude and frequency can vary with time, is given at the time instant t (taking as a reference $t_0 = 0$) by:

$$\mathbf{X}(t) = \frac{X_m(t)}{\sqrt{2}} e^{j(2\pi \int g(t)\, dt + \phi_0)} \tag{4.8}$$

where ϕ_0 is the phase-angle at time t_0 and $g(t) = f(t) - f_0$ is the difference between the instantaneous frequency and the nominal power-line frequency f_0. With this notation the signal can be expressed as:

$$\begin{aligned} x(t) &= \mathrm{Re}\left[\sqrt{2}\mathbf{X}(t)e^{j2\pi f_0 t}\right] \\ &= \left[\frac{\mathbf{X}(t)}{\sqrt{2}} e^{j2\pi f_0 t} + \frac{\mathbf{X}^*(t)}{\sqrt{2}} e^{-j2\pi f_0 t}\right] \end{aligned} \tag{4.9}$$

Let $x[n]$ be a sequence of samples, having finite length N. For the sake of simplicity, but without loss of generality, in the following N is assumed to be even and $-N/2 \le n \le N/2 - 1$, so that the time reference for the synchrophasor computation is located at $n = 0$, in the center of the sample record.

The evolution of amplitude and phase-angle within the observation window (modeled in (4.8) by time-varying terms) can affect synchrophasor estimation. Measurements that rely on a static model, like those based on the DFT, provide a sort of "average" synchrophasor (see [113] for a more detailed discussion). The TWLS approach [106] considers instead the Taylor expansion of the phasor $\mathbf{X}(t)$ around the reference time t_0, so that time variations occurring within the observation interval can be represented, in addition to the zero order synchrophasor term. Using a K-th order expansion, the approximated model of the dynamic phasor becomes, in

discrete form:

$$\mathbf{X}(nT) = \sum_{k=0}^{K} p^k \frac{(nT)^k}{k!} \tag{4.10}$$

where the complex number p^k is the k-th order derivative of the dynamic phasor $\mathbf{X}(nT)$ at $n = 0$ and T is the sampling interval. Generally speaking, TWLS can be defined as the projection of the signal $x(t)$ on a set of linearly independent basis functions:

$$\psi_k(t) = t^k e^{\pm j2\pi f_0 t}, \qquad k = 0, \ldots K. \tag{4.11}$$

When a finite sample record is considered, basis functions are replaced by their basis vector counterparts, $\psi_k[n]$.

It can be seen from (4.11) that Taylor-Fourier basis functions/vectors are usually referred to the nominal powerline frequency whereas, in practice, fluctuations may occur. If the product ROCOF $\cdot\, NT/2$ is small enough in relation to f_0, the sinusoid frequency can be assumed to remain approximately constant within the observation interval, but its value f_1 will in general differ from the nominal value.

One may arguably expect that using f_1 in place of f_0 in (4.9) and (4.11) a more accurate TFT measurement could be obtained. In [110, 117] frequency feedback is used for the purpose of determining f_1, exploiting the estimated first phase derivative. Thus, the synchrophasor $\mathbf{X}_1(nT)$, where f_1 is taken as reference for the basis function, is considered instead of $\mathbf{X}(t)$. The resulting signal model becomes:

$$x[n] = \mathrm{Re}\left[\sqrt{2}\mathbf{X}_1(nT)e^{j2\pi f_1 nT}\right] \tag{4.12}$$

and, using the Taylor expansion, the signal can then be modelled as:

$$x[n] = \sum_{k=0}^{K} \frac{(nT)^k}{k!}\left[\frac{p_1^k}{\sqrt{2}}e^{j2\pi f_1 nT} + \frac{(p_1^k)^*}{\sqrt{2}}e^{-j2\pi f_1 nT}\right] \tag{4.13}$$

where, similar to (4.10), p_1^k represents the k-th order derivative of the dynamic phasor $\mathbf{X}_1(nT)$ at $n = 0$.

Synchrophasor Model with Harmonics and Interharmonics

In actual conditions a more complex signal model may be needed, because harmonic and interharmonic disturbances can be present. Thus, the generic electric signal is not a pure sinusoid and can be represented as:

$$x[n] = \sum_h a_h[n]\cos(2\pi f_h nT + \phi_h[n]) \tag{4.14}$$

where a_h and ϕ_h are the (time-varying) amplitude and phase angle of the h-th component, respectively. For harmonics, f_h would be an integer multiple of f_1, otherwise it may represent a generic interharmonic frequency.

The Taylor expansion represented by (4.10) was extended to harmonic estimation in [90] which, however, still refers to the nominal fundamental frequency. Defining a generic phasor:

$$\mathbf{X}_h[n] \triangleq \frac{a_h[n]}{\sqrt{2}} e^{j\phi_h[n]} \tag{4.15}$$

allows to further generalize the approach to the generic signal (4.14), considering the expansion of each harmonic/interharmonic phasor $\mathbf{X}_h$ around the reference time, up to order K.

In the general situation the signal can thus be approximated as:

$$x[n] = \sum_h \sum_{k=0}^{K} \frac{(nT)^k}{k!} \left[\frac{p_h^k}{\sqrt{2}} e^{j2\pi f_h nT} + \frac{(p_h^k)^*}{\sqrt{2}} e^{-j2\pi f_h nT} \right] \tag{4.16}$$

where p_h^k is the k-th derivative of the synchrophasor at frequency f_h (p_h^0 is the harmonic/interharmonic phasor). A different order of expansion K_h may be used for each component h, as in [118]. If the frequencies f_h are known, from (4.16) it is possible to compute the phasor, along with its derivatives, by a weighted least squares approach [90] (less relevant in this context, harmonic phasors of interest can be computed as well).

In synchrophasor measurement the number of fundamental frequency cycles for each observation interval should be kept as low as possible, in order to better follow the dynamics of the input quantities. This requirement conflicts with the aim of accurately estimating f_h, since accurate estimation is, generally speaking, impractical without using very long observation intervals. However, the use of CS allows to circumvent some limitations and to estimate signal components on a sufficiently fine frequency grid, while keeping the observation interval reasonably short [76].

4.3.4 CS-TFT Estimation Algorithm

Taylor Fourier Transform by Compressive Sensing

Let $\nu_h = f_h T$ be the signal component frequencies normalized by the sampling rate $1/T$. Given a sample record of size N, the DFT algorithm projects them onto the orthogonal basis referring to the set of N normalized frequencies $\nu_m = m/N$, with

$m = 0, \ldots, N-1$ whose elements are:

$$\xi_m[n] = e^{j2\pi\nu_m n}, \quad n = -N/2 \ldots N/2-1, \tag{4.17}$$

The corresponding set of N coefficients is defined on a uniform frequency grid with step $\Delta_f = 1/NT$ and can be related to a static synchrophasor model. Defining each basis element as a column vector, the DFT basis set can be described by matrix $\mathbf{W}^H$, whose columns are the basis vectors. The superscript denotes conjugate transposition, as it should be reminded that $\frac{1}{N}\mathbf{W}$ defines the DFT operation in matrix form.

Assuming a finer frequency grid with a smaller step $\Delta'_f = \Delta_f / P$ (P a suitable integer), so that the total number of grid points is $N' = P \cdot N$, a generic signal component frequency ν can be expressed as:

$$\nu = \frac{\hat{h} + \delta'_{\hat{h}}}{N'} = \frac{h + \delta'_{\hat{h}}}{P} \cdot \frac{1}{N} \tag{4.18}$$

for some integer $\hat{h} \in [0, 1, \ldots, N'-1]$, with $|\delta'_{\hat{h}}| \leq \frac{1}{2}$. The closest approximation to ν on the new grid is: $\hat{\nu}_{\hat{h}} = \hat{h}/N'$. The set S_h of the index values $\hat{h}$ providing the frequency estimates (including image components) is called the signal *support*.

Defining a full set of TFT basis vectors at each point on this finer grid yields the generic basis vector expression:

$$\psi_k^h[n] = (nT)^k e^{j2\pi\frac{h}{N'}n}, \quad n = -N/2 \ldots N/2-1 \tag{4.19}$$

where $h = 0, \ldots, N'-1$ and the discrete normalized frequencies are now h/N'. The integer $k = 0, \ldots, K_h$ is the order of derivative of the Taylor expansion. Considering the same expansion order K for every h, (4.16) can be translated into matrix form, considering all basis vectors in (4.19), as follows:

$$\mathbf{x} = \mathbf{B}\mathbf{p} + \mathbf{e} \tag{4.20}$$

where $\mathbf{x} = [x[-N/2], \ldots, x[+N/2-1]]^{\mathrm{T}}$ is the vector of signal samples, $k \in \{0, \ldots, K\}$ and $\mathbf{e}$ is a vector representing noise and uncertainty that can arise, for instance, in the data acquisition system (e. g. because of analog-to-digital converter (ADC) quantization error).

Vector $\mathbf{p}$ includes all the complex parameters p_h^k for $h \in \{0, \ldots, N'-1\}$ and can be seen as the concatenation of the N' vectors $\mathbf{p}_h = \left[p_h^0, \ldots p_h^K\right]^{\mathrm{T}}$, whose length is $(K+1)N'$:

$$\mathbf{p} = \left[\mathbf{p}_0^{\mathrm{T}}\, \mathbf{p}_1^{\mathrm{T}} \cdots \mathbf{p}_{N'-1}^{\mathrm{T}}\right]^{\mathrm{T}} \tag{4.21}$$

The $N \times (K+1)N'$ matrix $\mathbf{B}$ is the generalization of the static model basis matrix to the finer grid and to Taylor-Fourier derivatives. Its column vectors are defined by (4.19).

The elements p_h^0, which represent the phasors at all frequencies of the finer grid, are different from zero in correspondence of the fundamental component and in the presence of harmonic or interharmonic terms. The same may be true of elements p_h^k associated with the dynamic extension of the phasor model. Significant contributions can therefore be associated with the condition $\|\mathbf{p}_h\|_2 \gg 0$, i.e., when static as well as, possibly, dynamic terms point to a significant contribution at the frequency index h. Since $\mathbf{p}$ is the concatenation of blocks of size K and most of these will still be approximately zero, the vector can be considered extremely *sparse* or, more precisely, *block sparse*. This feature justifies considering a CS approach for solving (4.20), which can be formulated as:

$$\hat{\mathbf{p}} = \arg\min_{\mathbf{p}} \|\mathbf{p}\|_0 \quad \text{subject to: } \|\mathbf{x} - \mathbf{B}\,\mathbf{p}\|_2 \leq \zeta\,, \tag{4.22}$$

where the pseudo-norm $\|\mathbf{p}\|_0$ indicates the number of non-zero elements of $\mathbf{p}$, $\|\cdot\|_2$ is the Euclidean norm of the vector and ζ is a given threshold. The block-sparse feature can be exploited to speed-up the solution of (4.22).

Enhanced Resolution by Compressive Sensing Approach

Expanding the set of TFT basis vectors, as shown in the previous sub-section, is the most straightforward way of creating a finer frequency grid. However, improved granularity does not imply a corresponding enhancement in frequency resolution. As discussed in [98], an observation length of at least 3-4 cycles is still required to resolve off-grid equal-amplitude frequency components. This would be the case, for instance, with an off-nominal power-line frequency and its image.

To actually enhance frequency resolution or, equivalently, to achieve the same resolution with a shorter measurement interval, the use of an overcomplete dictionary has been proposed in [76], [98] to model truncation effects related to the finite observation length. Hereinafter, the same approach is applied to TFT-based phasor measurement and the effectiveness of the resulting CS-TFT algorithm is shown.

To obtain the required overcomplete dictionary, the N-point DFT of the sample sequence $x[n]$ is considered. From (4.16), using standard properties of Fourier

transforms, one has the Fourier components:

$$X\left[\frac{m}{N}\right] = \sum_{h \in S_h} \left\{ \sum_{k=0}^{K_h} \frac{p_h^k}{\sqrt{2}} \left[\frac{1}{k!} \left(\frac{j}{2\pi} \right)^k D^{(k)} \left(\frac{m}{N} - \nu_h \right) \right] \right\}, \tag{4.23}$$

where $0 \leq m < N$ and $D^{(k)}(\nu)$ is the k-th order derivative of the Dirichlet kernel:

$$D(\nu) = \frac{\sin \pi N \nu}{N \sin \pi \nu} \tag{4.24}$$

Replacing ν_h in (4.23) by its finite-grid approximation h/N' allows to reformulate the relationship in matrix form with TFT coefficients defined on the finer frequency grid. In the time domain the resulting equation is:

$$\mathbf{x} = \mathbf{W}^H \sum_{k=0}^{K_h} \mathbf{D}^{(k)} \mathbf{p}^k \tag{4.25}$$

where $(\mathbf{p}^k)^{\mathrm{T}} = [p_0^k \; p_1^k \; \dots \; p_{N'-1}^k]$ and $[\mathbf{D}^{(k)}]_{m,h} = D^{(k)}(\frac{m}{N} - \frac{h}{N'})$. By suitably recombining matrices and vectors, (4.25) can be rewritten in a form equivalent to (4.20):

$$\mathbf{x} = \mathbf{W}^H \mathbf{D} \mathbf{p} + \mathbf{e} \tag{4.26}$$

It should be noticed that $\mathbf{W}^H$ is a matrix of size $N \times N$, while $\mathbf{D}$ has size $N \times (K+1)N'$. The latter is an overcomplete dictionary which explicitly model spectral leakage, accounting for both sparsification and Taylor expansion. The CS-TFT problem can be formulated, accordingly, as:

$$\hat{\mathbf{p}} = \arg\min_{\mathbf{p}} \|\mathbf{p}\|_0 \quad \text{subject to: } \|\mathbf{x} - \mathbf{W}^H \mathbf{D} \mathbf{p}\|_2 \leq \zeta \,. \tag{4.27}$$

It might be noted that zero-padding of the time-domain samples could achieve a similar improvement in frequency grid granularity. However this simple solution does not reduce leakage, therefore actual resolution would remain poor.

OMP-Based Support Recovery for CS-TFT Solution

The solution $\hat{\mathbf{p}}$ can be found by employing a greedy algorithm, such as Orthogonal Matching Pursuit (OMP) [54]. The algorithm is iterative, starting from the given measurement vector $\mathbf{x}$ defined above. At each iteration i the largest frequency component, approximated to the nearest bin in the N'-point grid, is found and its index $\hat{h}_i$ is included in the estimated support, which becomes: $\hat{S}_h(i) = \{\hat{h}_1, \dots, \hat{h}_i\}$. Accordingly, a reduced matrix of dictionary elements $\mathbf{D}_{\hat{S}_h(i)}$ is obtained as a submatrix of $\mathbf{D}$ by keeping only the columns whose indices belong to $\hat{S}_h(i)$.

Model coefficients are computed by solving the reduced system $\mathbf{x} = \mathbf{W}^H \mathbf{D}_{\hat{S}_h(i)} \hat{\mathbf{p}}(i)$. This can be achieved by means of WLS approximation, which yields the solution of (4.20) when the support is known.

The current solution $\hat{\mathbf{p}}(i)$ is updated and the corresponding signal reconstruction $\hat{\mathbf{x}}_{i-1}$ is subtracted from the measurement vector.

It should be noticed that the ith index is determined as:

$$\hat{h}_i = \arg \max_{0 \le h < N'} \|(\mathbf{D}^{(0)})^{\mathrm{H}} \mathbf{W}(\mathbf{x} - \hat{\mathbf{x}}_{i-1})\|_2, \tag{4.28}$$

where $\mathbf{D}^{(0)}$ is the submatrix of $\mathbf{D}$ corresponding to 0-th order derivatives and $\hat{\mathbf{x}}_{i-1}$ denotes the solution at previous iteration $i-1$. The search range for h can be obviously limited to the range $[0, N'/2)$ for real signals and further restricted to limited frequency regions, depending on specific cases.

The iteration number defines the support cardinality. A proper stop criterion allows the exclusion of all negligible contributions from signal model (4.16). The overall number of components is unknown a priori, but iterations can be stopped when different termination criteria, or a combination of them, are met: e.g., the latest component found is below a desired threshold, current residual norm is lower than a threshold that accounts for the noise level affecting measurements, or the maximum iteration number (maximum assumed support cardinality) has been reached.

As discussed in [98], the CS approach requires a minimum frequency separation of approximately $1.5 \cdot \Delta_f$ between two equal-magnitude waveform components lying on the fine grid. When this condition is satisfied, the indices h can be determined correctly to within $\pm(\Delta_f / 2P)$, that is, with the enhanced resolution provided by the finer grid. It should be remembered this cannot be made arbitrarily small, as the numerical conditioning of $\mathbf{D}$ tends to get worse for large values of P. Still, improvement by about one order of magnitude is achievable.

4.3.5 Algorithm Validation in Realistic Conditions

In this section the tests performed to evaluate the performance of the proposed algorithm under different test conditions are reported. Adopted test signals are inspired by those indicated in [83] for individual operating conditions (in particular for the maximum reporting rate of 50 frame/s). However, tests are designed to stress the algorithm in more problematic situations, that include different phenomena,

to highlight its peculiarities. In fact, the aim is to create test conditions that combine two or more scenarios and can thus be representative of harder conditions in a changing environment. Thus, the severe conditions indicated by the standard for both classes are mixed to create representative test signals. In particular, for each steady state and dynamic test condition presented by the standard, the additional presence of interharmonics is considered. Besides, additive white noise at a signal-to-noise ratio of 60 dB is superimposed on every test signal.

Several configurations have been reproduced in the MatLab programming environment. As an accuracy index, the total vector error (TVE), which measures the relative deviation between the theoretical phasor and the estimated one, is chosen. This will be the main quantity discussed in the following, while reporting the results obtained using different test conditions. In particular, because of the presence of random noise in the test signals, results are reported in terms of average and standard deviation of percent TVE.

The sampling rate is assumed equal to 5000 sample/s and the powerline frequency, in nominal conditions, is $f_0 = 50$ Hz. The observation interval corresponds to five powerline cycles, with $N = 500$ samples, where not otherwise stated, while the interpolation factor is $P = 10$. Accordingly, interference among spectral components is negligible up to a minimum frequency separation of 15 Hz and frequency resolution better than ± 0.5 Hz is achieved through the CS approach.

In the following, two classes of tests will be shown: tests in presence of steady state conditions, both for the fundamental frequency component and the disturbances, and tests in presence of dynamic evolution of the signal.

Tests under steady state conditions

With regards to interharmonics, very few assumptions can be made a priori. In the most general case, these spurious components do not fall on any pre-determined point of the frequency grid and do not exhibit any degree of regularity. In terms of frequency, the standard [83] indicates that the separation between the fundamental and the nearest interharmonic is at least equal to half the reporting rate. As a preliminary test, a sinusoidal signal at nominal frequency, with an additive interarmonic disturbance at $f_{ih} = 76$ Hz (with an amplitude of 10 % with respect to the fundamental component) plus noise has been used. Table 4.1 reports the average

Table 4.1: TVE estimation errors in the presence of interharmonic and 60 dB noise.

Test	**Method**	**Avg TVE [%]**	**Std TVE [%]**
$f_1 = 50$ Hz,	*CS-TFT*	0.011	0.005
$f_{ih} = 76$ Hz	*f-TWLS*	1.305	0.147
	TWLS	1.305	0.147

extimated TVE, in a 2-s test, for the proposed method CS-TFT, the TWLS based on the frequency estimation of only the first component (f-TWLS), and the classical TWLS [106].

When the full estimated frequency model, given by CS-TFT, is used, the rejection of interharmonic interference allows better estimation performance. Traditional and frequency tuned TWLS present, in this case, the same accuracy, because the fundamental frequency component is always correctly identified by the right frequency bin, corresponding to the nominal frequency.

Table 4.2 shows the impact on the TVE, under the same test conditions, of the window duration. In particular, three window lengths, corresponding to 3, 5 and 7 nominal cycles ($N = 300$, 500, 700 respectively), are used. It is clear how larger observation intervals give lower TVEs, because of the enhanced frequency resolution and of the narrower bandwidth that allows better noise rejection.

A first set of combined tests is then performed using a single interharmonic interfering signal at $f_{ih} = 76$ Hz or $f_{ih} = 21.5$ Hz) with amplitude 10 % with respect to the fundamental component, added to a severely frequency deviated sinusoidal signal at 45 Hz or 55 Hz (maximum deviations for off-nominal frequency tests for M-class in [83]). Such interharmonics were chosen in the out-of-band frequency region suggested by the standard [83] and the first one is off the standard N-point frequency grid, while the other is chosen to be even off the finer grid. Table 4.3 reports the corresponding results for the same three algorithms and $N = 500$ as in Table 4.1.

Interharmonic interference is strongly reduced by the inclusion of the found in-

Table 4.2: TVE estimation errors of CS-TFT in the presence of interharmonic and 60 dB noise with different observation window lengths.

Test	**Window duration [nominal cycles]**	**Avg TVE [%]**	**Std TVE [%]**
$f_1 = 50$ Hz, $f_{ih} = 76$ Hz	3	0.088	0.121
	5	0.011	0.005
	7	0.008	0.005

terharmonic in the model. While the TVE increases if the interharmonic is closer to the fundamental in the classic approach, the error is almost the same if the extended model is adopted. In addition, under off-nominal conditions, the super-resolution given by CS approach allows a suitable frequency tuning that leads to similar estimation accuracy with respect to previous test case at nominal frequency. The frequency tuning only is not sufficient to give better performance, since the interhamonic disturbance can have a predominant role and centering the filter towards the correct frequency can yield lower attenuation of the out-of-band signals.

A second composite test is performed using an interharmonic signal at $f_{ih} = 76$ Hz along with a 10 % second or third order harmonic ($f_h = 100$ Hz or $f_h = 150$ Hz, respectively), superimposed to a sinusoid at nominal frequency, plus noise. Table 4.4 reports the test results. TWLS are not reported because, as for Table 4.1, they are identical to those of f-TWLS, since the fundamental frequency is always correctly identified. Also in this case, the benefits of CS-TFT are patent.

Tests in the presence of dynamic conditions

According to IEEE Standard C37.118.1a-2014 [104], synchrophasor measurement bandwidth is determined by tests in the presence of amplitude modulation (AM) and phase modulation (PM) of the sinusoidal signal at nominal frequency. For the tests, while keeping the same configuration of noise and interhamonic disturbances,

Table 4.3: TVE estimation errors in the presence of off-nominal conditions, inter-harmonics and 60 dB noise.

Test	**Method**	**Avg TVE [%]**	**Std TVE [%]**
$f_1 = 45$ Hz,	*CS-TFT*	0.011	0.005
$f_{ih} = 76$ Hz	*f-TWLS*	1.184	0.196
	TWLS	2.331	1.106
$f_1 = 55$ Hz,	*CS-TFT*	0.013	0.006
$f_{ih} = 76$ Hz	*f-TWLS*	2.252	0.197
	TWLS	2.341	0.950
$f_1 = 45$ Hz,	*CS-TFT*	0.013	*0.007
$f_{ih} = 21.5$ Hz	*f-TWLS*	2.377	0.530
	TWLS	2.101	0.771
$f_1 = 55$ Hz,	*CS-TFT*	0.010	0.005
$f_{ih} = 21.5$ Hz	*f-TWLS*	1.593	0.417
	TWLS	2.092	0.632

Table 4.4: TVE estimation errors in the presence of harmonic, interharmonic and 60 dB noise.

Interfering Signals		**Method**	**Avg TVE**	**Std TVE [%]**
f_h	f_{ih}		**[%]**	**[%]**
$f_h = 100$ Hz	$f_{ih} = 76$ Hz	*CS-TFT*	0.011	0.006
		f-TWLS	1.325	0.260
$f_h = 150$ Hz	$f_{ih} = 76$ Hz	*CS-TFT*	0.011	0.005
		f-TWLS	1.307	0.158

two modulated signals are adopted: an AM with modulation level $k_x = 0.1$ and modulation frequency $f_m = 5$ Hz and a PM with modulation level $k_a = 0.1$ and modulation frequency $f_m = 5$ Hz (corresponding to the maximum values indicated by [104] for M-class).

Table 4.5 shows the results in terms of average TVE and its standard deviation for both the TWLS using only fundamental frequency and the proposed method. It is clear that, also under dynamic conditions, the rejection of the interharmonic, by its estimation and inclusion in the model, leads to lower measurement errors. Once the interharmonic is removed by the filtering process, the residual estimation error is due to the passband of the filters, affecting the modulated signals.

The algorithm performance has been characterized also during a linear ramp of the fundamental frequency. The ramp rate of change of frequency has been set to 1 Hz/s, as indicated by the standard, and a 10-s test has been performed, thus letting the frequency change in the range [45,55] Hz. Fig. 4.2 shows the percent TVE when noise and interhamonic (at 76 Hz) are present during the whole ramp duration. It is clear that in the f-TWLS, when only fundamental frequency is considered, the percent TVE is higher. Different steps in TVE are visible, due to the discrete grid of frequency bins. For CS-TFT, an average TVE of about 0.01 % (with a standard de-

Table 4.5: TVE estimation errors in the presence of modulations, interharmonic and 60 dB noise.

Test	**Method**	**Avg TVE [%]**	**Std TVE [%]**
AM $k_x = 0.1$, $f_m = 5$ Hz,	*CS-TFT*	0.119	0.046
$f_{ih} = 76$ Hz	*f-TWLS*	1.326	0.208
PM $k_a = 0.1$, $f_m = 5$ Hz,	*CS-TFT*	0.124	0.039
$f_{ih} = 76$ Hz	*f-TWLS*	1.299	0.273

viation of 0.016 %) can be reached also during the ramp and similar results can be obtained changing the interharmonic frequency. Obviously the different TVE behaviour depends on the relative shift between fundamental and interhamonic components, which changes during the frequency ramp duration. With the aim of testing all the possible dynamic conditions suggested by [83], the step response of the algorithm has been verified with both amplitude (±10 %) and phase-angle (±10°) steps applied to a sinusoidal signal at nominal frequency, corrupted, as before, by an interharmonic interfering component at 76 Hz and 60 dB random noise. Fig. 4.3 shows the results. CS-TFT, allowing interharmonic rejection, shows, as in Table 4.1, a TVE in steady-conditions before and after the step change below the 1 % TVE limit permitted by the standard and used to define the dynamic behaviour of the estimator. Thus the response time of the algorithm, using the same definition as in [83], corresponds to about 32 ms and 88 ms for amplitude and phase-angle steps, respectively. On the contrary, with f-TWLS the TVE is larger than 1 % even in the pre- and post-step steady-state conditions, thus impeding the calculation of the response time.

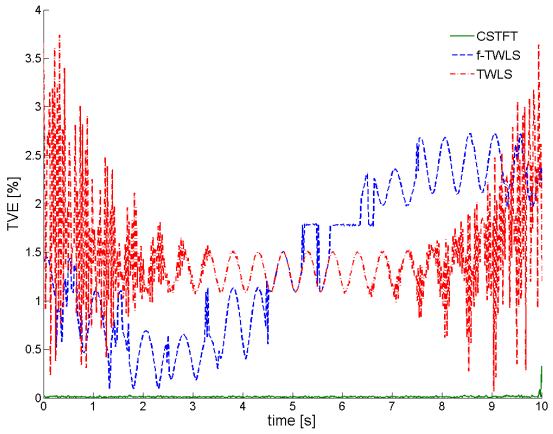


Figure 4.2: TVE trends during ramp frequency test (with interharmonic and noise).

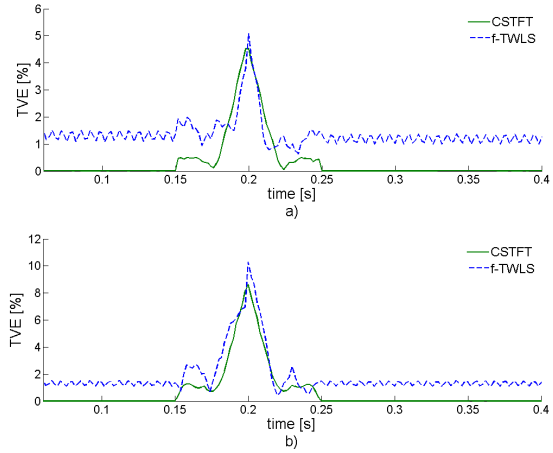


Figure 4.3: TVE trends in the presence of step changes (with interharmonic and noise): a) +10 % amplitude step; b) +10° phase-angle step.

4.4 Part III: CS-Based Template for EEG Signal Artefact Removal

4.4.1 Brain Activity Acquisition Scenario

Simultaneous recording of electroencephalography (EEG) and functional magnetic resonance imaging (fMRI) represents a powerful tool to thoroughly study brain activities.

Magnetic resonance imaging provides a mapping of the brain areas related to specific tasks or stimuli. Unfortunately, localization of activation events is accurate in the space domain, but still too poor in the time domain, where resolution is limited by the complex procedure required for measuring a magnetic moment relaxation. Moreover, fMRI does not evaluate actual neuronal activity, but an indirect effect, e.g. the blood oxygenation level dependent (BOLD) effect.

Electrophysiologycal analysis is able to detect the brain electrical activity directly. No system constraints limit the sampling frequency, i.e., the signal is acquired in real time. Conversely, EEG is affected by low spatial resolution: its activation mapping is blurred and relies only on a scalp surface model, i.e., three-dimensional information is not available.

By properly combining the two measuring techniques, it is possible to enhance event resolution and source localization in both time and space domains. Many applications of simultaneous EEG and fMRI analysis have been developed, ranging from the study of epilepsy and resting state conditions to the relationship between evoked potential and BOLD effect variations. Unfortunately, EEG data collected during concurrent fMRI are affected by very large artefacts related to both technical and physiological phenomena.

From the instrumentation point of view, such a large discrepancy between informative and spurious components requires a specific design of the measuring system. The amplifier stage needs a large dynamic range to ensure accurate EEG representation. Conversely, its bandwidth has to include at least all the significant bands, but a too large value makes the system prone to saturation. It is worth noticing that artefact correction cannot be partial or inaccurate, as residual components could still spoil any inference or result.

The main spurious contributions are named, respectively, gradient artefact (GRA)

and ballistocardiographic (BCG) artefact. The first one is related to the sawtooth profiles of magnetic flux inside the MRI scanner. The second is related to the electrodes pulsating motion due to scalp blood circulation. Both artefacts are generally orders of magnitude larger than the EEG trace, but can be reduced by post-processing if some system features (e.g., the gradient sequence or the ECG trace) are known. In both cases, the provoking causes exhibit inherent periodicity, whose timings are measurable or can be inferred.

This last chapter part is concerned with the gradient artefact. GRA magnitude depends directly on the gradient sequences imposed within the MRI scanner. The periodic inversion of magnetic flux produces a Faraday effect over the loops formed by the head tissues and the EEG leads. In the frequency domain, GRA contributes large equally spaced peaks over the entire spectrum. Experimental data acquired from a water phantom showed a clear relationship between artefact fundamental frequency and fMRI parameters, such as volume time and number of slices [119]. Unfortunately, affected frequencies include also the bands significant to neurological investigation. Low pass pre-filtering, e.g. with cut-off frequency at 250 Hz, could reduce the overall artefact magnitude, but would not resolve the problem.

Patient condition also affects artefact magnitude, e.g., proper positioning of the patient head inside the MRI scanner ensures a 30% GRA reduction [120]. Dynamic range can be accordingly reduced, enlarging the amplifier bandwidth and the frequency range under investigation. This acquisition setting does not affect functional measurement quality, as the head remains in homogeneous field conditions.

Hereinafter, a novel GRA removal algorithm is presented and validated over both simulation and experimental data. For each artefact occurrence, a specific template is computed and then subtracted in time domain.

The algorithm consists of two main stages. In the first one, the spectral content of the artefact is estimated. The GRA spectrum consists of a train of spikes, whose spacing depends on magnetic resonance parameters. This peculiar trend suggests the application of a harmonic estimation method. For this purpose, a super-resolution method, based on compressive sensing (CS) theory, provides accurate Fourier coefficients from a short set of EEG samples [76]. In the second stage, detected frequencies are used to form a basis in Taylor-Fourier transform (TFT) space [90]. The template for GRA removal is defined by projecting the sampled signal over the subspace

spanned by this basis. It is worth noting that the TFT has been developed exactly for dynamic harmonic analysis. Therefore, it is able to cope with the aforementioned inaccuracy causes, such as patient movement or misalignment of EEG channel timing.

4.4.2 Problem Statement

The signal under investigation consists of two main contributions: EEG and GRA. Information about neuronal activity is conveyed by the former, whose power content is mainly located in a low frequency range, not exceeding 250 Hz. In practice, the signal is not so strictly band-limited, but physiological studies have shown an inverse relationship between frequency and magnitude of brain electrical activity.

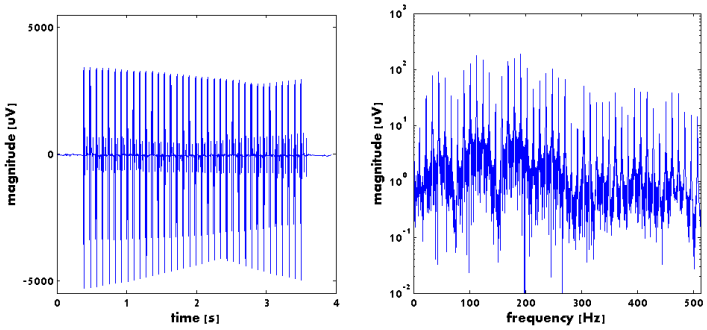


Figure 4.4: Characteristic GRA profile in the time (left) and frequency (right) domains.

An example of a gradient artefact superposed on an EEG trace in concurrent EEG-fMRI acquisition is shown in Fig 4.4. These large artefacts, related to the effect of alternating magnetic field gradients, are periodically superposed on the EEG trace. Every signal portion affected by GRA can be determined by a threshold process and artefact removal has to be applied to it.

In theory, the component frequencies of each GRA are exact multiples of a fundamental frequency, obtained from the fMRI acquisition parameters:

$$f_0 = \frac{N_v}{T_v} \tag{4.29}$$

where N_v is the number of slices in a volume, and T_v is the volume acquisition time [119]. In practice, a dynamic correction term has to be introduced for each har-

monic. Therefore, harmonic estimation cannot be simply carried out by means of the discrete Fourier transform (DFT), since this relies on a strictly periodic signal model, i.e., it is suitable only in static conditions. A suitable model for the GRA contribution is then:

$$x(t) = \sum_{h=1}^{H} A_h(t) cos(2\pi h f_0 t + \phi(t)) \tag{4.30}$$

where components are characterized by time varying amplitudes, $A_h(t)$ and instantaneous phases, $\phi_h(t)$ with $h = 1, \cdots H$.

In the presence of harmonic fluctuations in amplitude and/or phase, DFT can only provide an average value over the observation interval. This limitation can be partially solved by the adoption of shorter observation intervals, but this causes a reduction of the estimation grid granularity and an increase in long-range interference due to spectral leakage.

Given the sampling period T_s, a record of the signal under investigation can be expressed as follows:

$$y[n] = \sum_{h=1}^{H} A_h[n] cos(2\pi h f_0 n + \phi[n]) + e[n] = x[n] + e[n] \tag{4.31}$$

The useful EEG trace is actually represented by the *noise* contribution $e[n]$. The proposed algorithm is aimed at the estimation of GRA, which is the dominant part of $y[n]$.

In the literature, the most widely used algorithm is average artefact subtraction (AAS) [121]. Considering the artefact occurrences on all EEG channels, an average template is computed and then subtracted from each affected signal portion. AAS poses two main requirements. On one hand, it must be possible to accurately compare and superimpose different artefact occurrences over different times and/or channels. On the other hand, dynamic range should be larger than the artefact variability range. The former issue is addressed by synchronizing EEG channels with the fMRI scanner. The latter calls for high-resolution signal digitisation.

A further problem for AAS implementation is represented by patient movements, that are sufficient to change the magnetic field distribution. In this case, the adoption of sliding time windows to compute the average template is suggested in [121], but results are still unsatisfactory. The occurrence and time localization of movements can be inferred from MRI realignment parameters, so that specific templates can be computed for dynamic conditions [122]. Another approach classifies the oc-

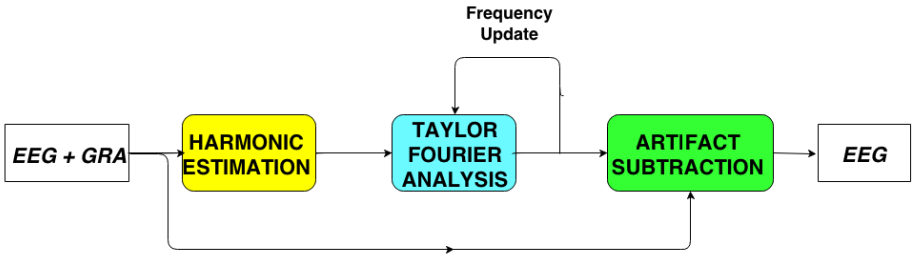


Figure 4.5: Block diagram of algorithm main steps.

currences according to their mutual similarity [123]. Both proposals reduce the cardinality of the set over which the average is computed, weakening the hypothesis of uncorrelated EEG traces.

4.4.3 CS-TFT Approach for GRA Removal

In the proposed algorithm, each artefact occurrence is treated singly. A GRA-affected portion is divided into consecutive non-overlapping segments of N samples. Since the portion length is not necessarily an integer multiple of N, zero-padding is employed for the last segment. Generally, N is assumed odd to preserve central symmetry in spectral analysis investigation.

As can be seen in Fig. 4.5, the data record passes through three steps. First, a super-resolution method determines GRA component frequencies. Secondly, the resulting frequencies are employed to form a basis in TFT subspace. The weighted least-squares (WLS) approximation provides the k-th order derivatives of each GRA component, for $k \in \{0,1,2\}$. The first derivative of the phase term allows an immediate correction of the estimated component frequency. The procedure is iterated until the first derivative term becomes negligible. Finally, the dynamic estimates of component amplitude and phase are employed to construct an artefact template, which is subtracted from the original record.

Spectral Content Model Identification

The super-resolution method employed in the first stage of the algorithm has been presented in [76]. It is summarized here for the sake of completeness.

The composition of $x[n]$ could be estimated by Fourier analysis. A difficulty of

this approach is that frequencies have to be determined very accurately, thus requiring very long sample sequences to achieve the necessary resolution. To overcome this limit, the super-resolution algorithm models the DFT coefficients of $y[n]$, arranged in the vector $\underline{Y}$, by the linear relationship:

$$\underline{Y} = \mathbf{D}\underline{A} \tag{4.32}$$

where $\underline{A}$ is a vector of complex Fourier amplitudes defined on a dense frequency grid with step $\Delta_f = 1/(P \times NT_s)$. The integer P is called the super-resolution factor and determines the resolution gain with respect to the standard DFT, whose frequency step is just $1/(NT_s)$.

Matrix $\mathbf{D}$ in (4.32) contains samples of the Dirichlet kernel representing spectral leakage at the different frequencies of the dense grid. Since $\dim[\underline{Y}] \ll \dim[\underline{A}]$, system (4.32) is under-determined. However, in the frequency domain information is sparse, i.e., only a reduced set of elements in $\underline{A}$ are significant. Their indexes form a set called the signal support S, whose cardinality is equal to the number of harmonics in (4.30), i.e., $\|\underline{A}\|_0 = H \ll N$.

The super-resolution CS approach employs this *sparsity* assumption to provide the best estimate of $\underline{A}$ as the sparsest vector that satisfies the condition:

$$\| \underline{Y} - \mathbf{D}\underline{A} \|_2^2 \leq \varepsilon \tag{4.33}$$

where ε depends on the energy related to the underlying EEG trace. In the present case, the GRA-to-EEG ratio is equal to at least 40 dB.

The algorithm crucial point is the recovery of signal support S. This stage is carried out via a singular-value decomposition of the matrix $\underline{Y}\,\underline{Y}^H$, where the superscript denotes transposition and complex conjugation. Decomposition allows to obtain the matrix $\mathbf{V}$ whose columns, that are the eigenvectors of $\underline{Y}\,\underline{Y}^H$, are collapsed into a single vector $\underline{v}$. As shown in [45], the required support S is also the set of indexes of the non-zero elements in vector $\underline{u}$, that is the solution of matrix equation $\underline{v} = \mathbf{D}\underline{u}$. The latter is obtained by a greedy algorithm, such as orthogonal matching pursuit (OMP) [23].

S determines the set of GRA frequencies over the dense grid. Given $s_h \in S$, the corresponding frequency estimate is $f_h = s_h \cdot \Delta_f$, with a resolution of $\pm\Delta_f/2$.

The proposed CS method provides accurate frequency estimation, starting from an observation interval as short as possible. In static conditions the next step would

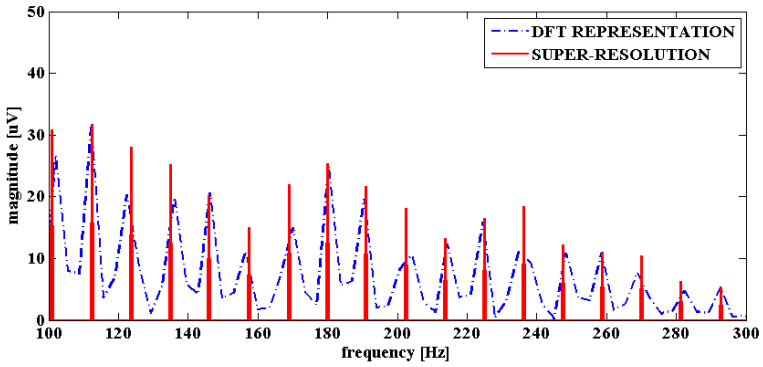


Figure 4.6: Detail of a super-resolved estimate of a GRA occurrence.

be to estimate the Fourier coefficients of (4.30) by solving (4.32) through the matrix pseudo-inverse:

$$\underline{\hat{X}}_S = (\mathbf{D}_S^H \mathbf{D}_S)^{-1} \mathbf{D}_S^H \underline{Y} \tag{4.34}$$

where $\mathbf{D}_s$ is obtained from $\mathbf{D}$ by taking only the columns whose index is in S. A portion of super-resolved spectrum is presented in Fig. 4.6 and compared with standard DFT, showing that leakage and inter-harmonic interference are strongly reduced. However, any DFT-based solution is only the best approximation of the input signal with reference to a static periodic signal model. Conversely, dynamic conditions should not be neglected for complete artefact removal.

TFT-Based Artefact Template

The Taylor-Fourier transform assumes a dynamic signal model, where each harmonic varies according to a specific complex envelope. In this way, the algorithm is able to cope with amplitude and/or phase variations within the observation interval.

In the TFT approach each harmonic is assumed to be modulated by a time function, expressed by a Taylor polynomial expansion of a given order $k > 0$. Thus, TFT coefficients contain both the instantaneous values and the first derivatives of component amplitudes and phases.

Let $\mathscr{L}_2(\mathbb{R})$ be the Hilbert space of complex-valued functions. No constraints are defined with reference to periodicity or amplitude and phase trend. The TFT is de-

fined on a set of linearly independent basis functions:

$$\psi_{(k,h)}(t) = t^k e^{-j2\pi f_h t} \qquad h = 0, \cdots H \quad k = 0, \cdots K \tag{4.35}$$

The proposed algorithm employs the recovered support S to construct a GRA specific basis $\mathbf{B}$, whose vectors are given by:

$$\underline{\psi}_{(k,h)} = \{\psi_{(k,h)}[n]\}_{0 \le n \le N} = (nT_s)^k e^{-j\frac{2\pi hn}{N'}} \tag{4.36}$$

where $h \in S$ and $k = 0, \cdots K$.
In the corresponding subspace, the signal projection is characterized by a set of weighting coefficients $\underline{\theta}_{(k,h)}$, computed by means of weighted least squares approximation:

$$\underline{\theta}_{(k,h)} = (\mathbf{B}^H \mathbf{W} \mathbf{B})^{-1} \mathbf{B}^H \mathbf{W} \underline{Y} \tag{4.37}$$

where $\mathbf{W}$ is a weighting matrix, whose diagonal contains the samples of a Kaiser window ($\alpha = 10$) [90] and $\mathbf{B} = [\underline{\psi}_{(k,h)}]$ with $h \in S$ and $k = 0, \cdots K$.

TFT provides a signal approximation

$$\hat{y}[n] = \sum_k \sum_h \theta_{(k,h)} \psi_{(k,h)}[n] \tag{4.38}$$

where the TFT coefficient $\theta_{(k,h)}$ represents the k-th derivative term of the h-th harmonic component:

$$\theta_{(k,h)} = \frac{T_s^k}{k!} \frac{d^k A_h(t) e^{j\phi_h(t)}}{dt^k}. \tag{4.39}$$

Equation (4.38) represents an accurate GRA template, specific to the signal portion under analysis. It is worth noticing that EEG does not belong to the subspace spanned by vectors of the basis $\mathbf{B}$, so its contribution to the template is negligible. The estimation process accounts for dynamic conditions and does not require further processing, e.g. filtering or averaging steps.

It should be noticed that the vectors in (4.35) span a subspace $\mathscr{T} \in \mathscr{L}_2(\mathbb{R})$ which contains the DFT subspace. Accordingly, TFT approximation error is less than or equal to the DFT approximation error. Even estimates for $k = 0$ improve on the DFT counterpart affected by infiltration of the neglected derivative terms. Taylor expansion terms are not orthogonal, so the TFT basis vectors form only a frame, rather than an orthogonal basis. Vector linear independence is, however, a sufficient condition for the existence of a TFT solution.

The physical meaning of TFT coefficients allows closed loop control of harmonic frequency estimation. First derivative of phase corresponds to a frequency deviation. Accordingly, harmonic frequencies are iteratively updated

$$f_h = f_h + \frac{d\phi_h(t)}{dt} \tag{4.40}$$

until the stop criterion is satisfied:

$$\frac{\theta_{(1,h)} \cdot NT_s}{\theta_{(0,h)}} \leq 1\% \qquad \forall h \in S \tag{4.41}$$

The computed template is finally subtracted from the input signal, providing a *cleaned* EEG trace $\hat{e}[n]$. In the presence of zero-padded recordings, the template is truncated to the last index of the signal portion affected by GRA.

4.4.4 Algorithm Validation on Simulated and Experimental Data

During concurrent EEG-fMRI acquisition it is hard to define an input signal reference, since GRA entirely covers the underlying EEG. Accordingly, the reconstruction error should be quantified in terms of similarity with other portions of EEG not affected by GRA.

To characterize algorithm performances we employed a data set where EEG had been acquired outside the MRI scanner in a healthy subject, and GRA measured on a phantom was numerically added. Two conditions have been explored, respectively, with patient eyes open and closed. EEG was acquired using a SD-MRI32 Micromed Amplifier (sampling rate: 1024 Hz, digitizer resolution: 22 bits, dynamic range: ±25.6 mV). In order to avoid aliasing phenomena, a hardware band-pass filter from 0.15 to 269.5 Hz has been applied.

Data acquired on a water phantom during fMRI analysis provided a set of GRA profiles. In particular, functional images have been acquired using a Siemens Symphony 1.5 T scanner. The sensing protocol consists of a T2* weighted EPI sequence (number of slices: 36, volume repetition time: 3700 ms). Further details are available in [119].

Spectral analysis of the acquired profiles shows that GRA can be accurately modelled by a sum of harmonic components, whose fundamental frequency is given by (4.29), in this case $f_0 \simeq 11.25$ Hz. Accordingly, a new set of profiles has been simulated, where harmonic amplitudes and phases are drawn, respectively, from a normal and a uniform distribution, i.e., $A_h \sim \mathcal{N}(1 \times 10^4, 5 \times 10^2)$ and $p_h \sim \mathcal{U}(0, 2\pi)$.

Table 4.6: Performance parameters of the algorithm.

Patient state	**Data set**	**Mean residual [uV]**	**St.Dev. [uV]**	**NRMS [%]**	**SNR [dB]**
open	static	-1.806	1.2	22.0	13.1
eyes	dynamic	-0.150	3.3	64.8	3.77
closed	static	1.525	1.3	20.7	11.8
eyes	dynamic	0.189	4.1	54.6	5.23

For algorithm validation, two sets of 10 occurrences have been selected from both the real and simulated data set. Simulated data represent operation in static conditions, since both frequencies and Fourier coefficients do not vary over the observation interval. Conversely, real data include non-ideal phenomena, like unequally spaced frequencies or time-varying Fourier coefficients.

With reference to the super-resolution stage of the algorithm, the observation interval consists of $N = 301$ samples, for which the adoption of an interpolation factor $P = 13$ results in a frequency resolution of 0.26 Hz. The number of harmonics is inferred from the sampling frequency and the expected f_0, i.e., $H = 93$. For the TFT, the Taylor expansion is truncated at the first-order order, i.e., $K = 1$. In static conditions, only the 0-th order coefficients are significant, while in dynamic conditions first derivatives allow the iterative update of harmonic frequencies.

artefact residual after application of the proposed algorithm has been characterized in terms of mean value, standard deviation and normalized root-mean-square (NRMS) ratio, that is defined as:

$$NRMS = \sqrt{\frac{\sum(\hat{e} - e)^2}{\sum(e)^2}} \tag{4.42}$$

where e and $\hat{e}$ represent, respectively, the original and the recovered EEG trace. Its reciprocal is the signal-to-noise ratio (SNR) between the original EEG and the artefact residual, expressed in dB in the following.

For both data sets, Table 4.6 summarizes these performance parameters averaged over 10 GRA occurrences. In static conditions, the residual artefact exhibits low magnitude and variability, and its energy content is about five times lower than the one associated to the EEG trace alone, as shown by NRMS. In dynamic conditions the variability range increases, and the energy content of the residual exceeds

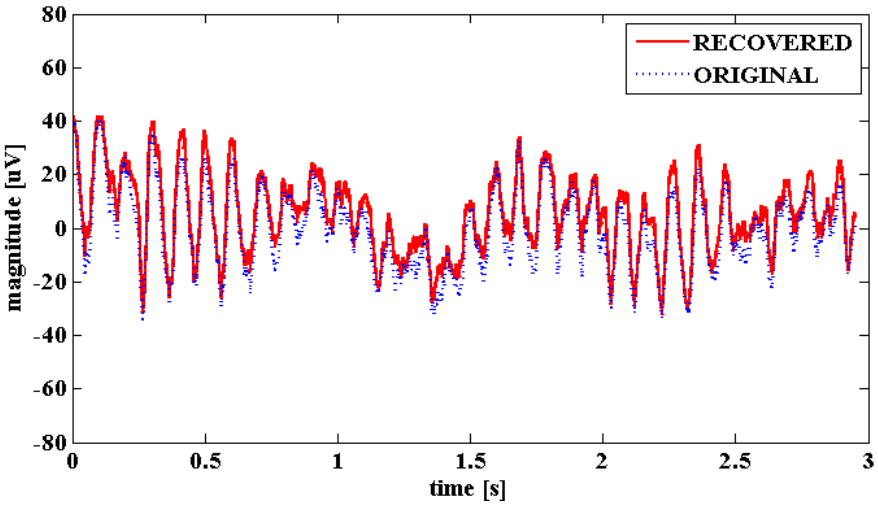


Figure 4.7: Closed-eyes signal recovery (red solid line) vs original EEG trace (blue dotted line) in static conditions [124].

50% of the EEG one. It is worth noticing that SNR is firmly positive in all the considered cases, i.e., the proposed algorithm is always able to estimate GRA components and reduce them.

Examples of closed-eyes signal recovery in static and dynamic conditions are given, respectively, in Fig. 4.7 and Fig. 4.8. Again, better accuracy and smoothness are achieved in static conditions, nevertheless, the residual artefact in dynamic conditions is small enough to avoid significant distortion of the underlying trace.

From a clinical point of view, the main feature of interest is the distribution of EEG signal power over five spectral bandwidth related to specific brain activities: δ ($0.5-4$ Hz), θ ($4-7$ Hz), α ($7-14$ Hz), β ($15-30$ Hz), and γ ($30-100$ Hz or more). Any approach where the template is averaged over more occurrences is prone to distort the actual power distribution, whereas the proposed algorithm constructs a specific GRA template for each short observation interval. Therefore it is interesting to compare the bandwidth powers obtained from the recovered EEG with those of the original.

Fig. 4.9 presents this comparison for open eyes acquisition, while Fig. 4.10 presents the closed eyes case. In static conditions recovery is almost perfect, whereas in dynamic conditions a somewhat larger variability is noticeable. Nonetheless, mutual power ratios are reliably reproduced.

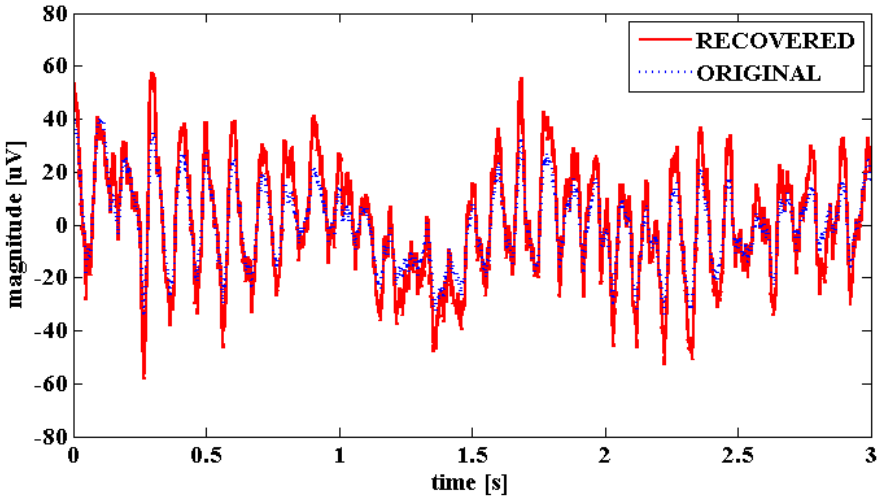


Figure 4.8: Closed-eyes signal recovery (red solid line) vs original EEG trace (blue dotted line) in dynamic conditions [124].

4.4.5 Results Discussion and Computational Considerations

The proposed algorithm is able to cope successfully with time-varying GRA. Each artefact occurrence is processed singly, providing more accurate and specific subtraction templates. A closed loop control based on harmonic phase derivatives ensures correct tracking of the harmonic frequencies, as the observation interval scans the artefact occurrence.

Preliminary experimental results collected both in static and dynamic conditions show a significant reduction of spurious components, with residual artefact

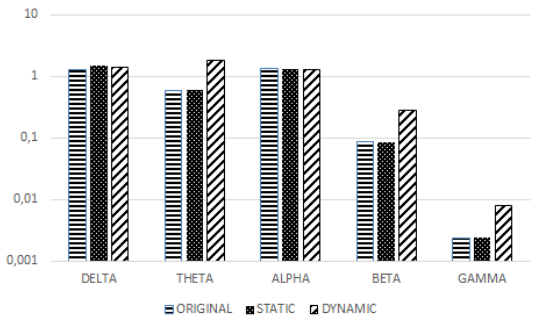


Figure 4.9: Power spectral distribution during open eyes acquisition (logarithmic power scale) [124].

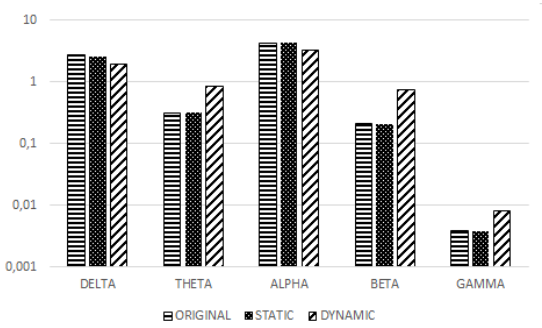


Figure 4.10: Power spectral distribution during closed eyes acquisition (logarithmic power scale) [124].

magnitude almost one order of magnitude lower than the EEG trace under investigation. The resulting SNR is positive in all the considered operating conditions. Moreover, no significant distortions are introduced in spectral power distribution, potentially allowing a reliable clinical interpretation of the recovered EEG trace.

The novelty of this approach is the combination of a super-resolution method based on compressive sensing theory with the Taylor-Fourier transform. TFT in combination with the Kalman filter has been previously employed, in a biomedical context, for the estimation of blood pressure waveforms [125]. Conversely, in this work TFT is directly employed to project the acquired data over the subspace which best represents the GRA occurrence under investigation.

From a computational load point of view, the crucial points are represented by the computation of the cross-correlation matrix $\underline{YY}^H$ and the frequency update iterative loop (4.40). Nevertheless, the capability of operating on a reduced sample set lightens the computational burden. At the moment, the algorithm is conceived for post-processing implementation but, since no averaging procedure or prior information are necessary, once a signal portion affected by GRA is detected, the algorithm can immediately provide the specific template. A suitable compromise between segment size and required estimate accuracy might allow on-line realization.

Chapter 5

Conclusions

This Ph.D. thesis builds on the field of Compressive Sensing and illustrates how sparsity and incoherence properties can be exploited to design efficient sensing strategies, or to intimately understand the sources of uncertainty that affect measurements. Specifically, the research activity has been concerned with definition, analysis and implementation of sparse signal acquisition paradigms.

In this thesis, the main specific contributions include: a common theoretical framework of Compressive Sensing, which can easily fit any sparse acquisition scenario; a CS-based algorithm for providing accurate preliminary information and suitably preprocessed data for a vector signal analyser or a cognitive radio application; a CS-based frequency super-resolution technique, relying on an overcomplete dictionary which explicitly accounts for spectral leakage effect; an off-the-grid frequency estimation approach, which properly combines CS-based super-resolution and DFT coefficients polar interpolation; an enhanced spectral content model for spectral analysis applications in dynamic conditions by means of Taylor-Fourier transform (TFT) approaches.

To conclude, it is worth reflecting on achieved results, remaining open problems and possible directions for future research.

5.1 General Framework of Compressive Sensing Theory

In chapter 1, a mathematical model for sparse signal acquisition has been presented and thoroughly characterized. An annotated dictionary of theorems and cri-

teria has been provided and is expected to be extremely helpful in designing efficient and accurate sensing systems. In this context, two fundamental approaches have been briefly summarized and compared: on one side, the canonical approach based on the Shannon's theorem; on the other side, the CS approach which acquires the signal directly in a compressed form. Particular attention is devoted to the computational efforts required both at sensor and receiver side.

By definition, the CS paradigm has to deal with an under-determined system of equations, whose solution requires ad hoc algorithms. Two main algorithmic classes have been identified, relying respectively on proper Lebesgue norm approximation and greedy iterative search. Their behaviour has been characterized both in noiseless and noisy conditions, devoting particular attention to solution accuracy and computational complexity. All things considered, a general guideline has been proposed: if a blind approach is required, norm approximation represents the most suitable option; whereas greedy search exhibits lower computational complexity at the expense of lower solution accuracy.

On the other hand, it is worth noticing that CS theory is still a dynamic and expanding research field. Just the last year, it celebrated its first tenth anniversary. In a such short time, countless are the algorithms and the applications which have been designed, validated and, in some cases, even implemented and put on the market. In this regarding, CS theory candidates itself not only as an original and fascinating research topic, but also as an occasion for significant innovation in the field of technology development. Many exciting solutions have been yet presented. Others might just be around the corner.

5.2 CS-Based Pre-Processor for VSA or CR Applications

In chapter 2, a CS algorithm has been proposed as a vector signal analysis preprocessor and has been thoroughly characterized. Current VSA architectures routinely allow access to sampled data acquired at the output of the instrument IF filter, therefore it should be possible to integrate the algorithm within the set of signal processing functions normally accessible to a user. The added capabilities considered in this work can enhance VSA performances in wireless network monitoring and support measurement for cognitive radio applications.

A number of open research issues will have to be addressed in future research on the characterisation of the CS algorithm. These include the optimal selection of the decimation factor L and an investigation into the relationship between reconstruction accuracy and the number m of parallel channels. These aspects are important, as they help understand what could be the actual compression level that can be achieved in a practical VSA measurement application. A thorough analysis of performances with regards to different modulation formats is also important to determine whether the algorithm can be considered an effective general purpose tool.

Although some of the measurement algorithms proposed in this work are rather sophisticated, their implementation within the firmware of modern measuring equipment such as a VSA is well within the instrument capabilities in term of computing power and processing time. The support recovery stage in the CS algorithm is the most demanding in terms of computational effort, involving the singular value decomposition of the measurement correlation matrix. This effectively places an upper bound on the number of MWC channels, although this does not seem to be too critical as far as multiple source detection is concerned.

At the time, the CS-based sensing algorithm appears to be beyond the reach of a CR receiver computing power. Nevertheless, research into more computationally efficient CS algorithms could overcome this limitation. The CS algorithm provides a VSA with blind spectrum sensing capabilities at no additional hardware cost and can determine some of the basic parameters of a modulated signal. Capitalizing on the de-noising effects of the CS algorithm, the histogram analysis part of the proposed measurement algorithms provides the basis for detection of modulation formats taken from a dictionary. Thus, VSA capabilities are considerably enhanced and allow instruments of this class to tackle some of the demanding requirements of CR measurement.

5.3 CS-Based Super-Resolution Technique in the DFT Domain

In chapter 3, the spectral estimation of multi-tone signals has been addressed by a CS-based approach (CS-DFT) which provides one order of magnitude super-

resolution, independently from the considered data record length. Referring to a finite grid may appear as a limitation compared to interpolation and parametric methods where, in principle, frequency can be considered a continuous variable. However, finite signal-to-noise ratio (SNR) places a lower-bound as well, which is equivalent for all practical purposes to just considering a discrete fine grid. While application of the CS algorithm does not eliminate spectral leakage, by reducing granularity on the frequency axis it increases, in normalised terms, the frequency separation between adjacent components. This virtually eliminates long-range frequency interference effects, leaving only scalloping loss to be dealt with.

The main features of the proposed algorithm make it of interest for application in smart microgrid projects. In fact, its basic structure can be adapted to a range of PQA and PMU measurements by a comparatively simple set of management and control rules, while the underlying hardware may remain unchanged. This goes towards the aim of realizing low-cost, high-performance measurement nodes in distribution-level smart microgrids, which are one of the targets of current research in the field.

Finally, the problem of continuous estimation has been taken into account. In fact, the granularity issue, although less significant, is not entirely overcome by the aforementioned super-resolution technique.

This has led to consider a two-stage approach, which combines CS-based super-resolution and DFT coefficients polar interpolation. Results obtained so far support the idea that the proposed approach can be considered an interesting alternative for high-resolution multi-tone waveform analysis and can attain super-resolution performance, provided the computing power necessary for this class of algorithms is available.

5.4 Quasi-Stationary Sparse Signal Acquisition Models

In chapter 4, the notion of quasi-stationary signal model has been introduced and its implications to spectral analysis have been briefly discussed, with particular focus on TFT formulas and properties. Some common measurement approaches related to TFT have been summarized. In this context, the proposal of a joint CS-TFT approach has been introduced as a novel signal-independent approach for spectral analysis in quasi-stationary operative conditions.

Two challenging operative scenarios have been selected to validate the proposed algorithm.

Firstly, this combined approach has been applied to power measurements in smart grid scenario. More precisely, CS-TFT has been proposed as a Phasor Measurement Unit (PMU) algorithm. Its performance has been characterized under different static and dynamic conditions, using composite test signals that include various disturbances, such as interharmonics, harmonics or additive noise, in an attempt to represent a severe operating environment. Experimental results have shown that CS-TFT allows, while keeping the observation interval reasonably low, to detect with a good resolution the relevant spectral components of the signal, even when they are changing with time. The algorithm is promising in its ability to cope with severe operating conditions, like the concurrent effect of dynamic conditions and interharmonic interference, with remarkable performance in terms of estimation accuracy.

Secondly, the same approach has been applied to the problem of gradient artefact removal from concurrent EEG-fMRI recordings. The proposed algorithm has been proven to be a promising alternative to averaging techniques widely employed in the literature. In fact, it is able to cope successfully with time-varying GRA. Each artefact occurrence is processed singly, providing more accurate and specific subtraction templates. A closed loop control based on harmonic phase derivatives ensures correct tracking of the harmonic frequencies, as the observation interval scans the artefact occurrence. Preliminary experimental results collected both in static and dynamic conditions show a significant reduction of spurious components, with residual artefact magnitude almost one order of magnitude lower than the EEG trace under investigation. The resulting SNR is positive in all the considered operating conditions. Moreover, no significant distortions are introduced in spectral power distribution, potentially allowing a reliable clinical interpretation of the recovered EEG trace.

Both the practical implementation of CS-TFT have confirmed the goodness of the proposed approach. Thus, plausible directions for future research are represented by its application to similar acquisition problems and its implementation on programmable boards with sufficient computing resources.

Bibliography

[1] C. E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, Jul. 1948.

[2] C. Shannon, "Communication in the presence of noise," *Proceedings of the IRE*, vol. 37, no. 1, pp. 10–21, Jan 1949.

[3] E. Candes and M. Wakin, "An introduction to compressive sampling," *Signal Processing Magazine, IEEE*, vol. 25, no. 2, pp. 21–30, March 2008.

[4] M. Vetterli, P. Marziliano, and T. Blu, "Sampling signals with finite rate of innovation," *Signal Processing, IEEE Transactions on*, vol. 50, no. 6, pp. 1417–1428, Jun 2002.

[5] Z. Ben-Haim, T. Michaeli, and Y. Eldar, "Performance bounds and design criteria for estimating finite rate of innovation signals," *Information Theory, IEEE Transactions on*, vol. 58, no. 8, pp. 4993–5015, Aug 2012.

[6] E. Grossi, "An introduction to compressive sensing - sparsity property and acquisition model," Seminar Proceedings, 2010.

[7] J. M. Whittaker, "The fourier theory of the cardinal function," *Proceedings of the Edinburgh Mathematical Society Series 2*, vol. 1, pp. 169–176, 7 1928.

[8] V. A. Kotel'nikov, "On the transmission capacity of 'ether' and wire in electric communications," *Physics-Uspekhi*, vol. 49, no. 7, pp. 736–744, 2006. [Online]. Available: http://ufn.ru/en/articles/2006/7/g/

[9] H. Nyquist, "Certain topics in telegraph transmission theory," *American Institute of Electrical Engineers, Transactions of the*, vol. 47, no. 2, pp. 617–644, April 1928.

[10] P. L. Butzer, "A survey of the whittaker-shannon sampling theorem and some of its extensions," *Journal of Mathematical Research & Exposition*, vol. 1983, no. 1, pp. 185–212, 1983.

[11] J. J. Benedetto, "Frames, sampling and seizure prediction," in *Advances in Wavelets*. Springer, 1999.

[12] J. C. F. Diaz, W. E. Carter, R. L. Shrestha, C. L. Glennie, M. P. Sartori, and A. Singhania, "Early results from a high-resolution hybrid terrestrial and bathymetry mapping lidar," in *2012 IEEE International Geoscience and Remote Sensing Symposium, Munich, Germany, July 22-27, 2012*, 2012, pp. 4994–4997.

[13] M. Mishali and Y. Eldar, "From theory to practice: Sub-nyquist sampling of sparse wideband analog signals," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 4, no. 2, pp. 375–391, April 2010.

[14] W. Rudin, *Functional analysis*, 2nd ed., ser. International Series in Pure and Applied Mathematics. New York: McGraw-Hill Inc., 1991.

[15] L. Jacques, L. Duval, C. Chaux, and G. Peyré, "A panorama on multiscale geometric representations, intertwining spatial, directional and frequency selectivity," *Signal Processing*, vol. 91, no. 12, pp. 2699 – 2730, 2011, advances in Multirate Filter Bank Structures and Multiscale Representations.

[16] E. Candès and J. Romberg, "Sparsity and incoherence in compressive sampling," *Inverse Problems*, vol. 23, no. 3, p. 969, 2007. [Online]. Available: http://stacks.iop.org/0266-5611/23/i=3/a=008

[17] M. A. Davenport, "Random observations on random observations: Sparse signal acquisition and processing," Master's thesis, Rice University, Aug. 2010.

[18] D. Donoho, M. Elad, and V. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *Information Theory, IEEE Transactions on*, vol. 52, no. 1, pp. 6–18, Jan 2006.

[19] D. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *Information Theory, IEEE Transactions on*, vol. 47, no. 7, pp. 2845–2862, Nov 2001.

[20] M. H. Wright, "The interior-point revolution in optimization: history, recent developments, and lasting consequences," *Bull. Amer. Math. Soc. (N.S*, vol. 42, pp. 39–56, 2005.

[21] S. S. Chen, D. L. Donoho, Michael, and A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing*, vol. 20, pp. 33–61, 1998.

[22] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY, USA: Cambridge University Press, 2004.

[23] S. Mallat and Z. Zhang, "Adaptive time-frequency decomposition with matching pursuits," in *Time-Frequency and Time-Scale Analysis, 1992., Proceedings of the IEEE-SP International Symposium*, Oct 1992, pp. 7–10.

[24] T. Blumensath and M. E. Davies, "Iterative hard thresholding for compressed sensing," *Applied and Computational Harmonic Analysis*, vol. 27, no. 3, pp. 265 – 274, 2009.

[25] Y. Pati, R. Rezaiifar, and P. Krishnaprasad, "Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition," in *Signals, Systems and Computers, 1993. 1993 Conference Record of The Twenty-Seventh Asilomar Conference on*, Nov 1993, pp. 40–44 vol.1.

[26] T. Cai, L. Wang, and G. Xu, "New bounds for restricted isometry constants," *Information Theory, IEEE Transactions on*, vol. 56, no. 9, pp. 4388–4394, Sept 2010.

[27] R. Baraniuk, M. Davenport, R. Devore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constr. Approx*, vol. 2008, 2007.

[28] D. Donoho and J. Tanner, "Observed universality of phase transitions in high-dimensional geometry, with implications for modern data analysis and signal processing," *Philosophical Transactions A: Mathematical, Physical and Engineering Sciences*, vol. 367, no. 1906, pp. 4273–4293, 2009.

[29] I. Carron, "Another surprise in the donoho-tanner phase transition?" available online: http://nuit-blanche.blogspot.it/2010/10/another-surprise-in-donoho-tanner-phase.html, 2010.

[30] J. D. Blanchard, C. Cartis, J. Tanner, and A. Thompson, "Phase transitions for greedy sparse approximation algorithms," *Applied and Computational Harmonic Analysis*, vol. 30, no. 2, pp. 188 – 203, 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1063520310000862

[31] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *Information Theory, IEEE Transactions on*, vol. 55, no. 5, pp. 2230–2249, May 2009.

[32] D. Needell and J. A. Tropp, *CoSaMP: Iterative signal recovery from incomplete and inaccurate samples*, ser. Technical report (California Institute of Technology, Pasadena). Applied and Computational Mathematics Department, CalTech, 2008.

[33] N. Devroye, M. Vu, and V. Tarokh, "Cognitive radio networks," *Signal Processing Magazine, IEEE*, vol. 25, no. 6, pp. 12–23, November 2008.

[34] M. Bertocco, G. Frigo, G. Giorgi, and C. Narduzzi, "A modulation detector based on compressive sensing for vector measurement in cognitive radio," in *Instrumentation and Measurement Technology Conference (I2MTC), 2013 IEEE International*, May 2013, pp. 795–800.

[35] Agilent, *Agilent Spectrum Analysis Basics*, 2006.

[36] R. Baraniuk, E. Candes, R. Nowak, and M. Vetterli, "Compressive sampling [from the guest editors]," *Signal Processing Magazine, IEEE*, vol. 25, no. 2, pp. 12–13, March 2008.

[37] R. Baraniuk, E. Candes, M. Elad, and Y. Ma, "Applications of sparse representation and compressive sensing [scanning the issue]," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 906–909, June 2010.

[38] Y. Eldar and G. Kutyniok, *Compressed Sensing: Theory and Applications*, ser. Compressed Sensing: Theory and Applications. Cambridge University Press, 2012. [Online]. Available: http://books.google.it/books?id=Gm3ihcJwN0YC

[39] S. Engelberg, "Compressive sensing [instrumentation notes]," *Instrumentation Measurement Magazine, IEEE*, vol. 15, no. 1, pp. 42–46, February 2012.

[40] M. Bertocco, G. Frigo, C. Narduzzi, and F. Tramarin, "Resolution enhancement by compressive sensing in power quality and phasor measurement," *Instrumentation and Measurement, IEEE Transactions on*, 2014, accepted for publication.

[41] J. Haupt and R. Nowak, "Compressive sampling for signal detection," in *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on*, vol. 3, April 2007, pp. III–1509–III–1512.

[42] M. Mishali, Y. Eldar, O. Dounaevsky, and E. Shoshan, "Xampling: Analog to digital at sub-nyquist rates," *Circuits, Devices Systems, IET*, vol. 5, no. 1, pp. 8–20, January 2011.

[43] G. Fudge, R. Bland, M. Chivers, S. Ravindran, J. Haupt, and P. Pace, "A nyquist folding analog-to-information receiver," in *Signals, Systems and Computers, 2008 42nd Asilomar Conference on*, Oct 2008, pp. 541–545.

[44] M. Mishali and Y. Eldar, "Wideband spectrum sensing at sub-nyquist rates [applications corner]," *Signal Processing Magazine, IEEE*, vol. 28, no. 4, pp. 102–135, July 2011.

[45] ——, "Reduce and boost: Recovering arbitrary sets of jointly sparse vectors," *Signal Processing, IEEE Transactions on*, vol. 56, no. 10, pp. 4692–4702, Oct 2008.

[46] J. Chen and X. Huo, "Theoretical results on sparse representations of multiple-measurement vectors," *Signal Processing, IEEE Transactions on*, vol. 54, no. 12, pp. 4634–4643, Dec 2006.

[47] D. L. Donoho and M. Elad, "Optimally sparse representation in general (non-orthogonal) dictionaries via ℓ_1 minimization," in *Proc. Natl Acad. Sci. USA 100 197–202*.

[48] Y. Zeng and Y.-C. Liang, "Eigenvalue-based spectrum sensing algorithms for cognitive radio," *Communications, IEEE Transactions on*, vol. 57, no. 6, pp. 1784–1793, June 2009.

[49] I. M. Johnstone, "On the distribution of the largest eigenvalue in principal components analysis," *Ann. Statist.*, vol. 29, no. 2, pp. 295–327, 04 2001. [Online]. Available: http://dx.doi.org/10.1214/aos/1009210544

[50] N. Karoui, *On the Largest Eigenvalue of Wishart Matrices with Identity Covariance when N, P and P/n - [infinity]*, ser. Technical report (Stanford University. Dept. of Statistics). Department of Statistics, Stanford University, 2003. [Online]. Available: http://books.google.it/books?id=DjTQHAAACAAJ

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

[52] G. Pfander, H. Rauhut, and J. Tropp, "The restricted isometry property for time–frequency structured random matrices," *Probability Theory and Related Fields*, vol. 156, no. 3-4, pp. 707–737, 2013. [Online]. Available: http://dx.doi.org/10.1007/s00440-012-0441-4

[53] M. Davenport, J. Laska, J. Treichler, and R. Baraniuk, "The pros and cons of compressive sensing for wideband signal acquisition: Noise folding versus dynamic range," *Signal Processing, IEEE Transactions on*, vol. 60, no. 9, pp. 4628–4642, Sept 2012.

[54] J. Tropp and A. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *Information Theory, IEEE Transactions on*, vol. 53, no. 12, pp. 4655–4666, Dec 2007.

[55] Y. Eldar, P. Kuppinger, and H. Bolcskei, "Block-sparse signals: Uncertainty relations and efficient recovery," *IEEE Trans. Signal Process.*, vol. 58, no. 6, pp. 3042–3054, Jun. 2010.

[56] T. Keysight. (2014) Setting up a digital demod measurement.

[57] D. Belega and D. Dallet, "Accurate normalized frequency estimation by the three-point interpolated DFT method with rectangular window," in *17th European Signal Processing Conference - EUSIPCO 2009*, United Kingdom, Aug. 2009, pp. 1329–1333. [Online]. Available: https://hal.archives-ouvertes.fr/hal-00413660

[58] S. Krishnan and C. Seelamantula, "On the selection of optimum savitzky-golay filters," *Signal Processing, IEEE Transactions on*, vol. 61, no. 2, pp. 380–391, Jan 2013.

[59] D. C. Rife and R. R. Boorstyn, "Multiple tone parameter estimation from discrete-time observations," *Bell System Technical Journal*, vol. 55, no. 9, pp. 1389–1410, 1976. [Online]. Available: http://dx.doi.org/10.1002/j.1538-7305.1976.tb02941.x

[60] V. Jain, W. L. Collins, and D. C. Davis, "High-accuracy analog measurements via interpolated fft," *Instrumentation and Measurement, IEEE Transactions on*, vol. 28, no. 2, pp. 113–122, June 1979.

[61] C. Offelli and D. Petri, "The influence of windowing on the accuracy of multifrequency signal parameter estimation," *Instrumentation and Measurement, IEEE Transactions on*, vol. 41, no. 2, pp. 256–261, Apr 1992.

[62] D. Agrez, "Weighted multipoint interpolated dft to improve amplitude estimation of multifrequency signal," *Instrumentation and Measurement, IEEE Transactions on*, vol. 51, no. 2, pp. 287–292, Apr 2002.

[63] S. M. Kay, *Modern spectral estimation : theory and application*, ser. Prentice Hall signal processing series. Englewood Cliffs (N.J.): Prentice Hall, 1988. [Online]. Available: http://opac.inria.fr/record=b1081251

[64] S. L. Marple, *Digital Spectral Analysis: With Applications*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1986.

[65] P. Stoica and R. Moses, *Spectral Analysis of Signals*. Pearson Prentice Hall, 2005. [Online]. Available: http://books.google.it/books?id=h78ZAQAAIAAJ

[66] D. Donoho, "Superresolution via sparsity constraints," *SIAM Journal on Mathematical Analysis*, vol. 23, no. 5, pp. 1309–1331, 1992. [Online]. Available: http://dx.doi.org/10.1137/0523074

[67] E. J. Candeè and C. Fernandez-Granda, "Towards a mathematical theory of super-resolution," *CoRR*, vol. abs/1203.5871, 2012.

[68] M. F. Duarte and R. G. Baraniuk, "Spectral compressive sensing," *Applied and Computational Harmonic Analysis*, vol. 35, no. 1, pp. 111–129, Jul. 2013.

[69] A. Fannjiang and W. Liao, "Coherence pattern-guided compressive sensing with unresolved grids," *SIAM J. Img. Sci.*, vol. 5, no. 1, pp. 179–202, Feb. 2012. [Online]. Available: http://dx.doi.org/10.1137/110838509

[70] ——, "Super-resolution by compressive sensing algorithms," in *Signals, Systems and Computers (ASILOMAR), 2012 Conference Record of the Forty Sixth Asilomar Conference on*, Nov 2012, pp. 411–415.

[71] E. J. Candes, Y. C. Eldar, D. Needell, and P. Randall, "Compressed sensing with coherent and redundant dictionaries," *Applied and Computational Harmonic Analysis*, vol. 31, no. 1, pp. 59–73, 2011.

[72] D. Rife and R. Boorstyn, "Single tone parameter estimation from discrete-time observations," *Information Theory, IEEE Transactions on*, vol. 20, no. 5, pp. 591–598, Sep 1974.

[73] F. Harris, "On the use of windows for harmonic analysis with the discrete fourier transform," *Proceedings of the IEEE*, vol. 66, no. 1, pp. 51–83, Jan 1978.

[74] S. Cotter, B. Rao, K. Engan, and K. Kreutz-Delgado, "Sparse solutions to linear inverse problems with multiple measurement vectors," *Signal Processing, IEEE Transactions on*, vol. 53, no. 7, pp. 2477–2488, July 2005.

[75] B. James, B. Anderson, and R. Williamson, "Characterization of threshold for single tone maximum likelihood frequency estimation," *Signal Processing, IEEE Transactions on*, vol. 43, no. 4, pp. 817–821, Apr 1995.

[76] M. Bertocco, G. Frigo, and C. Narduzzi, "On compressed sensing and super-resolution in dft-based spectral analysis," in *Proceedings 19th IMEKO TC-4 Symposium and 17th IWADC Workshop Advances in Instrumentation and Sensors Interoperability*, Jul 2013, pp. 615–620.

[77] C. Muscas, "Power quality monitoring in modern electric distribution systems," *Instrumentation Measurement Magazine, IEEE*, vol. 13, no. 5, pp. 19–27, October 2010.

[78] G. D'Antona, C. Muscas, P. Pegoraro, and S. Sulis, "Harmonic source estimation in distribution systems," *Instrumentation and Measurement, IEEE Transactions on*, vol. 60, no. 10, pp. 3351–3359, Oct 2011.

[79] D. Fabri, C. Martins, L. Silva, C. Duque, P. Ribeiro, and A. Cerqueira, "Time-varying harmonic analyzer prototype," in *Harmonics and Quality of Power (ICHQP), 2010 14th International Conference on*, Sept 2010, pp. 1–7.

[80] T. Tarasiuk, "Comparative study of various methods of dft calculation in the wake of iec standard 61000-4-7," *Instrumentation and Measurement, IEEE Transactions on*, vol. 58, no. 10, pp. 3666–3677, Oct 2009.

[81] A. Phadke and J. Thorp, *Synchronized Phasor Measurements and Their Applications*, ser. Power Electronics and Power Systems. Springer US, 2008. [Online]. Available: http://books.google.it/books?id=sCk3mK4kCWwC

[82] J. Burnett, R.O., M. Butts, and P. Sterlina, "Power system applications for phasor measurement units," *Computer Applications in Power, IEEE*, vol. 7, no. 1, pp. 8–13, Jan 1994.

[83] "Ieee standard for synchrophasor measurements for power systems," *IEEE Std C37.118.1-2011 (Revision of IEEE Std C37.118-2005)*, pp. 1–61, Dec 2011.

[84] D. Belega, D. Dallet, and D. Petri, "Accuracy of sine wave frequency estimation by multipoint interpolated dft approach," *Instrumentation and Measurement, IEEE Transactions on*, vol. 59, no. 11, pp. 2808–2815, Nov 2010.

[85] M. Bertocco, C. Narduzzi, and F. Tramarin, "Characterization of microgrid smart metering: Phasor estimation under impaired conditions," in *Instrumentation and Measurement Technology Conference (I2MTC), 2013 IEEE International*, May 2013, pp. 1170–1175.

[86] D. Macii, D. Petri, and A. Zorat, "Accuracy analysis and enhancement of dft-based synchrophasor estimators in off-nominal conditions," *Instrumentation and Measurement, IEEE Transactions on*, vol. 61, no. 10, pp. 2653–2664, Oct 2012.

[87] A. Testa, M. Akram, R. Burch, G. Carpinelli, G. Chang, V. Dinavahi, C. Hatziadoniu, W. Grady, E. Gunther, M. Halpin, P. Lehn, Y. Liu, R. Langella, M. Lowenstein, A. Medina, T. Ortmeyer, S. Ranade, P. Ribeiro, N. Watson, J. Wikston, and W. Xu, "Interharmonics: Theory and modeling," *Power Delivery, IEEE Transactions on*, vol. 22, no. 4, pp. 2335–2348, Oct 2007.

[88] J. de la O Serna, "Dynamic phasor estimates for power system oscillations and transient detection," in *Power Engineering Society General Meeting, 2006. IEEE*, 2006, p. 7.

[89] W. Premerlani, B. Kasztenny, and M. Adamiak, "Development and implementation of a synchrophasor estimator capable of measurements under dynamic conditions," *Power Delivery, IEEE Transactions on*, vol. 23, no. 1, pp. 109–123, Jan 2008.

[90] M. Platas-Garza and J. de la O Serna, "Dynamic harmonic analysis through taylorfourier transform," *Instrumentation and Measurement, IEEE Transactions on*, vol. 60, no. 3, pp. 804–813, March 2011.

[91] G. Barchi, D. Macii, and D. Petri, "Synchrophasor estimators accuracy: A comparative analysis," *Instrumentation and Measurement, IEEE Transactions on*, vol. 62, no. 5, pp. 963–973, May 2013.

[92] D. Belega and D. Petri, "Performance of synchrophasor measurements provided by the weighted least squares approach," in *Instrumentation and Measurement Technology Conference (I2MTC), 2013 IEEE International*, May 2013, pp. 946–951.

[93] J. de la O Serna, "Synchrophasor estimation using prony's method," *Instrumentation and Measurement, IEEE Transactions on*, vol. 62, no. 8, pp. 2119–2128, Aug 2013.

[94] H. Farhangi, "The path of the smart grid," *Power and Energy Magazine, IEEE*, vol. 8, no. 1, pp. 18–28, January 2010.

[95] F. Katiraei, R. Iravani, N. Hatziargyriou, and A. Dimeas, "Microgrids management," *Power and Energy Magazine, IEEE*, vol. 6, no. 3, pp. 54–65, May 2008.

[96] M. Bertocco, G. Frigo, C. Narduzzi, and F. Tramarin, "Resolution enhancement in harmonic analysis by compressive sensing," in *Applied Measurements for Power Systems (AMPS), 2013 IEEE International Workshop on*, Sept 2013, pp. 40–45.

[97] J. Schoukens and J. Renneboog, "Modeling the noise influence on the fourier coefficients after a discrete fourier transform," *Instrumentation and Measurement, IEEE Transactions on*, vol. IM-35, no. 3, pp. 278–286, Sept 1986.

[98] G. Frigo and C. Narduzzi, "Compressive sensing with an overcomplete dictionary for high-resolution dft analysis," in *Signal Processing Conference (EUSIPCO), 2014 Proceedings of the 22nd European*, Sept 2014, pp. 1766–1770.

[99] C. Ekanadham, D. Tranchina, and E. Simoncelli, "Recovery of sparse translation-invariant signals with continuous basis pursuit," *Signal Processing, IEEE Transactions on*, vol. 59, no. 10, pp. 4735–4744, Oct 2011.

[100] K. Fyhn, H. Dadkhahi, and M. Duarte, "Spectral compressive sensing with polar interpolation," in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, May 2013, pp. 6225–6229.

[101] A. Gupta and J. Preisig, "Adaptive sparse optimization for coherent and quasi-stationary problems using context-based constraints," in *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*, March 2012, pp. 3413–3416.

[102] J. Allen, "Short term spectral analysis, synthesis, and modification by discrete fourier transform," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 25, no. 3, pp. 235–238, Jun 1977.

[103] G. Frigo, "Compressive sensing plus taylor-fourier transform for synchrophasor estimation," available online: http://desl-pwrs.epfl.ch/workshop, 2014.

[104] *IEEE Standard for Synchrophasor Measurements for Power Systems – Amendment 1: Modification of Selected Performance Requirements*, IEEE Std C37.118.1a-2014 (Amendment to IEEE Std C37.118.1-2011), Apr. 2014.

[105] R. Mai, Z. He, L. Fu, B. Kirby, and Z. Q. Bo, "A dynamic synchrophasor estimation algorithm for online application," *Power Delivery, IEEE Transactions on*, vol. 25, no. 2, pp. 570–578, April 2010.

[106] J. A. de la O Serna, "Dynamic phasor estimates for power system oscillations," *IEEE Trans. Instrum. Meas.*, vol. 56, no. 5, pp. 1648–1657, Oct. 2007.

[107] D. Belega and D. Petri, "Accuracy analysis of the multicycle synchrophasor estimator provided by the interpolated DFT algorithm," *IEEE Trans. Instrum. Meas.*, vol. 62, no. 5, pp. 942–953, May 2013.

[108] A. J. Roscoe, I. Abdulhadi, and G. Burt, "P and M class phasor measurement unit algorithms using adaptive cascaded filters," *IEEE Trans. Power Del.*, vol. 28, no. 3, pp. 1447–1459, Jul. 2013.

[109] D. Petri, D. Fontanelli, and D. Macii, "A frequency-domain algorithm for dynamic synchrophasor and frequency estimation," *IEEE Trans. Instrum. Meas.*, preprint.

[110] P. Castello, J. Liu, C. Muscas, P. A. Pegoraro, F. Ponci, and A. Monti, "A fast and accurate PMU algorithm for P+M class measurement of synchrophasor and frequency," *IEEE Trans. Instrum. Meas.*, vol. 63, no. 12, pp. 2837–2845, Dec. 2014.

[111] P. Romano and M. Paolone, "Enhanced interpolated-dft for synchrophasor estimation in fpgas: Theory, implementation, and validation of a pmu prototype," *IEEE Trans. Instrum. Meas.*, vol. 63, no. 12, pp. 2824–2836, Dec. 2014.

[112] P. Castello, M. Lixia, C. Muscas, and P. A. Pegoraro, "Adaptive Taylor-Fourier synchrophasor estimation for fast response to changing conditions," in *Instrumentation and Measurement Technology Conference (I2MTC), 2012 IEEE International*, May 2012, pp. 294–299.

[113] ——, "Impact of the model on the accuracy of synchrophasor measurement," *IEEE Trans. Instrum. Meas.*, vol. 61, no. 8, pp. 2179–2088, Aug. 2012.

[114] P. Castello, P. Ferrari, A. Flammini, C. Muscas, and S. Rinaldi, "A new ied with PMU functionalities for electrical substations," *IEEE Trans. Instrum. Meas.*, vol. 62, no. 12, pp. 3209–3217, Dec. 2013.

[115] A. Roscoe, "Exploring the relative performance of frequency-tracking and fixed-filter phasor measurement unit algorithms under c37.118 test procedures, the effects of interharmonics, and initial attempts at merging p-class response with m-class filtering," *Instrumentation and Measurement, IEEE Transactions on*, vol. 62, no. 8, pp. 2140–2153, Aug. 2013.

[116] M. Bertocco, G. Frigo, C. Narduzzi, C. Muscas, and P. A. Pegoraro, "Compressive sensing plus Taylor-Fourier Transform for synchrophasor estimation," in *Applied Measurements for Power Systems Proceedings (AMPS), 2014 IEEE International Workshop on*, Sep. 2014, pp. 1–5.

[117] P. Castello, J. Liu, A. Monti, C. Muscas, P. Pegoraro, and F. Ponci, "Toward a class "P + M" phasor measurement unit," in *Applied Measurements for Power Systems (AMPS), 2013 IEEE International Workshop on*, Sep. 2013, pp. 91–96.

[118] J. Liu, F. Ni, P. Pegoraro, F. Ponci, A. Monti, and C. Muscas, "Fundamental and harmonic synchrophasors estimation using modified taylor-kaiman filter," in *Applied Measurements for Power Systems (AMPS), 2012 IEEE International Workshop on*, Sep. 2012, pp. 1–6.

[119] E. Sartori, E. Formaggio, S. Storti, A. Bertoldo, P. Manganotti, A. Fiaschi, and G. Toffolo, "Gradient artifact removal in co-registration eeg/fmri," in *World Congress on Medical Physics and Biomedical Engineering, September 7 - 12, 2009, Munich, Germany*, ser. IFMBE Proceedings, O. Dössel and W. Schlegel, Eds. Springer Berlin Heidelberg, 2010, vol. 25/4, pp. 1143–1146.

[120] K. J. Mullinger, W. X. Yan, and R. Bowtell, "Reducing the gradient artefact in simultaneous eeg-fmri by adjusting the subject's axial position," *NeuroImage*, vol. 54, no. 3, pp. 1942 – 1950, 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1053811910012814

[121] P. J. Allen, O. Josephs, and R. Turner, "A method for removing imaging artifact from continuous {EEG} recorded during functional {MRI}," *NeuroImage*, vol. 12, no. 2, pp. 230 – 239, 2000. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1053811900905998

[122] M. Moosmann, V. H. Schönfelder, K. Specht, R. Scheeringa, H. Nordby, and K. Hugdahl, "Realignment parameter-informed artefact correction for simultaneous eeg–fmri recordings," *NeuroImage*, vol. 45, no. 4, pp. 1144 – 1150, 2009. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1053811909000640

[123] F. Freyer, R. Becker, K. Anami, G. Curio, A. Villringer, and P. Ritter, "Ultrahigh-frequency {EEG} during fmri: Pushing the limits of imaging-artifact

correction," *NeuroImage*, vol. 48, no. 1, pp. 94 – 108, 2009. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1053811909006375

[124] G. Frigo and C. Narduzzi, "Eeg gradient artifact removal by compressive sensing and taylor-fourier transform," in *Medical Measurements and Applications (MeMeA), 2014 IEEE International Symposium on*, June 2014, pp. 1–6.

[125] J. de la O Serna, W. Van Moer, and K. Barbe, "Using alternating kalman filtering to analyze oscillometric blood pressure waveforms," *Instrumentation and Measurement, IEEE Transactions on*, vol. 62, no. 10, pp. 2621–2628, Oct 2013.