POLITECNICO DI TORINO

Collegio di Ingegneria Elettronica, delle Telecomunicazioni e Fisica

Master of Science

in Telecommunications Engineering

Master's Degree Thesis

Intelligent IoT sensing and diagnosis method for rotating machinery based on low-dimensional compressed measurements

Academic Supervisor
Prof. Enrico Magli

Candidate
Martina Cilia

Company Supervisor
Keisuke Yamamoto

A.A. 2017/2018

This thesis has been redacted in partial fulfilment of the requirements for the degree of Master of Science in Telecommunications Engineering, in collaboration with Central Research Laboratory, Hitachi Ltd. 1-280, Higashi-Koigakubo, Kokubunji-shi, Tokyo 185-8601, Japan.

ABSTRACT

Rotating machineries are widely used in modern industry and their continuous functioning is important in several industrial processes. However, their operating conditions have become more sever, involving high speeds, high loads, high temperatures, etc. In consequence, fault conditions are more likely to occur, requiring efficient conditions monitoring techniques and a reliable fault detection method.

Nowadays, fault detection is performed remotely based on industrial Internet of Things (IoT) data processing and Cloud computing. In the traditional fault diagnosis method, data are collected by IoT devices following the Nyquist sampling theorem. Then, they are sent to the Cloud where they are stored and analysed using data analytics techniques. In particular, vibration signals from the rotating devices are commonly used to evaluate the status of the machinery. In fact, many characteristic features can be extracted from the vibration signals that make them the usual choice for condition monitoring and fault detection.

A long-term and continuous monitoring requires the transmission of a continuous stream of data to the Cloud. Due to the complexity of the information embedded in the vibration signals and the ever-growing demand for the information of complex mechanical systems, the features extracted are not sufficient and the transmission and storage of the raw signals must be considered. According, a large amount of data is acquired by multiple sensors with high sampling rates over long operation periods, which imposes heavy burden on the acquisition hardware, data storage and transmission bandwidth.

Taking into account the scenario previously depicted, two problems must be faced for fault detection of industrial machinery:

1. The computational efforts at the Cloud are burdensome;
2. The network traffic is high due to the huge amount of data collected.

In order to overcome the problems related to the traditional fault detection method for industrial machinery, in this thesis, an intelligent IoT sensing system based on compressed sensing theory and Edge computing is proposed.

In such a system compressed sensing is used in order to reduce the amount of measurements acquired by the sensors, while keeping the acquisition process simple. In addition, the compressive sensing mechanism allows maintaining the features of the raw signal, at the expense of a more complex mechanism to recover it from the low-dimensionality measurements. At the same time, according to the Edge computing framework, part of the computation efforts is delegated to devices that lie near the sources of data and that interact with them. Thus, data may be processed in real time, the latency can be reduced and useless data can be discarded.

Specifically, the compressed sensing mechanism works only in presence of a basis, so called

dictionary, able to sparsify the signal of interest. In the proposed approach, the basis is constructed from a set of training signals applying dictionary learning algorithms. The trained dictionary is chosen, instead of a general one, because it is constructed on a specific class of data and it generates a very sparse representation only for signals in the same class as the training ones. Thus, according to the compressed sensing paradigm, the reconstruction quality, for a signal that carries the same features as the training signals, will differ from the reconstruction quality for a signal with different vibration signature.

This characteristic is exploited in the proposed system to generate a fault detection algorithm, based on the low-dimension compressed measurements. This algorithm is supposed to run at the Edge of the network in order to produce a preliminary detection of a change in the machine condition. As a consequence, the latency in the fault detection can be reduced, taking preliminary decision directly at the edge terminal. Moreover, considering this approach, the network traffic to the Cloud can be reduced. In fact, only useful data and useful information on the machinery condition will be sent to the Cloud to be stored and to carry out further analysis.

In this thesis, the feasibility of this system for fault diagnosis is studied in detail. In particular, the study focuses on two aspects: the reconstruction capability of the compressive sensing theory for vibration signals when a dictionary learning approach is followed and the automatic detection of a fault condition at the edge terminal using the compressed measurements.

The first part of the thesis regards the simulation of the reconstruction algorithm for different types of signals and the evaluation of the reconstruction quality and the achievable dimensionality reduction.

To start with, a dictionary is trained for each class of signals in order to catch the different vibration signatures and then later the signals are reconstructed using the dictionary that pertains to their class. As expected, when a correct dictionary is used, the quality of the reconstruction increases as the compression ratio increases. However, since the signal is affected by noise a perfect reconstruction is not possible. In fact, there will always be a certain amount of noise, even if it becomes smaller when the number of measurements is sufficiently high. On the other hand, the lowest bound on the measurements required for a satisfactory reconstruction of the original signal is determined by the sparsity of the signal on the sparsifying basis. For these reasons a trade-off between the desired quality and the desired compression rate must be considered, according to the specific application.

Once the reconstruction performance on a correct dictionary has been verified, in the second part of the thesis the fault detection algorithm based on the lower-dimension compressed measurements is tested.

Compared to a general dictionary, the trained dictionary improves the quality of the sparsification and consequent reconstruction only on its specific class and, for this reason, a specific dictionary is needed for each class under study. When the basis does not encompass the features of the signal to be reconstructed, the latter will not be sparse in the chosen basis; then, the compressive sensing framework fails and the reconstructed waveform will not match the original one. Accordingly, the simulations reveal that, if the compression ratio is fixed, there is a difference in the quality of the

reconstruction on the four dictionaries. This difference may be smaller when there are some shared features between the different classes but it can be exploited to discriminate the fault states.

Since the compressed measurements are supposed to convey the information embedded in the raw signals, the proposed fault detection exploits the reconstruction differences based on the received compressed measurements and on a estimate of the measurements based on the reconstructed signal. It is obvious that, compared to the case in which the discrimination is performed on the original signal, the ability to recognize the state decreases. Nevertheless, the results obtained show that the detection algorithm based on the trained dictionaries allows to achieve a good recognition rate for the fault conditions while decreasing the amount of measurements transmitted.

Even if the presented algorithm considers the availability of some signals to train the dictionary, the presence of an unknown state is also taken into account. This means that actually, it can be supposed that the dictionary is trained only when a change of state is detected. Following a principle similar to the one exploited for fault recognition, then, in the third part of the thesis an online fault detection algorithm at the edge terminal has been considered.

Considering a bi-directional transmission link from the sensor terminal to the edge terminal, a control on the dimension of the data sent from the sensors can be controlled. If the edge terminal does not have a dictionary for the current state, it can request the transmission of the raw data needed for the dictionary learning algorithm. Once the dictionary is trained, a request for the compressed data can be sent and the sensors start the transmission of the low-dimensional measurements. Then, a new training request is sent only when the reconstruction quality goes below a given threshold.

According to the result of the simulations for the states taken into account, the system can follow the change of state when a fault condition appears. Moreover when the new dictionary is evaluated the acquisition of the low-dimensional system can be started again. The dictionaries can be stored at the cloud and they can be used with other machines. At the point then the signals can be completely transmitted at the Cloud for further analysis on the conditions.

The last part of the thesis regards a preliminary study on real signals acquired in a real factory environment. The study was aimed at determining the capability of the reconstruction process and the possibility to use a dictionary for different machines of the same type has been verified.

Concluding, compressive sensing appears to be a promising technique to reduce the burden on the data transmission for condition monitoring. It allow to keep the sensor terminal simple but at the same time not only the features are receive at the Cloud side but also the raw signals can be reconstructed to extract more features. Moreover, the trained dictionary permits a preliminary detection of a fault state at the Edge of the network. In this way a continuous stream of data to the Cloud can be avoided and the processing burden at the Cloud can be reduced. Even if the trained dictionary causes a delay, it can be considered acceptable for systems that do not have frequent fault states. Finally, considering a decrease on the performances or an increase in the compression ratio, the use of the same dictionary for different machines of the same type can be taken into account.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

In today's industry, maintenance is one of the main procedures that has considerably high amount of cost regularly. Moreover, studies reveal that more than one third of the maintenance money is wasted in ineffective methods. Due to the regularity of the expense, condition monitoring and fault diagnosis have always gained a lot of interest from researcher in the industrial field and more efficient and reliable methods are always needed. [1]
The maintenance techniques can be divided into the following three main categories:

- Run-to-failure method: the system is fixed only when it is not able to operate anymore. Nowadays it is not used because it implies a big waste of money, since at that point the failure may be impossible to fix;
- Preventive maintenance: the maintenance is run on a periodic basis according to a fixed time schedule or based on expected life statistics deducted from past data. This is usually applied to critical machinery but the cost is still very high since the production is stopped during the maintenance period even if there are no actual failures;
- Predictive maintenance: the maintenance is run based on the actual monitoring of the machine. Sensors and measurements are used in order to acquire signals used as indicator of the machinery conditions. This is the most effective method since tasks are performed only when needed. It reduces the cost of maintenance but at the same time it is the most difficult approach since it requires reliable techniques.

Even if the three methods presented are usually combined, nowadays, predictive maintenance plays an important role due to the availability of a huge amount of data from low-cost sensors installed in the factories. All this information on the machinery can be used to better plan the maintenance work and several advantages may be envisioned: reduce cost, prevent unexpected equipment failures, increase equipment lifetime, increase plant safety, optimize spare parts handling.
In this thesis work, the focus will be on the condition monitoring of rotating machinery and an intelligent IoT system for fault detection will be analysed.

1.1 Rotating machinery and bearing fault detection

Rotating machinery is widely used in modern industry and their continuous function is important in several industrial processes. However, their operating conditions have become more sever, involving high speeds, high loads, high temperatures, etc. In consequence, fault conditions are more likely to occur, requiring a well-schedule maintenance and a reliable fault detection method.
Bearings are the most critical component in rotating machinery and their operation status

directly affects the overall performance of the equipment. Moreover, bearing faults are often a warning sign of other defects in the machine. For these reasons, the availability of a good predictive maintenance approach is particularly important for rotating machinery.

Vibration signals from the rotating machine are commonly used to evaluate the status of the machinery. In fact, many characteristic features can be extracted from the vibration signals that make them the usual choice for condition monitoring and fault detection.

Features are extracted from the measured signals based on various signal processing techniques in time domain, frequency domain or time-frequency domain. Several techniques have been applied such as time-domain statistical analysis and spectral analysis, time-frequency and other transform domain analysis, adaptive decomposition and entropy.

However, these extracted features may include redundant information. Therefore, some dimensionality reduction method has been also widely applied to identify a lower-dimensional space that efficiently represents the high dimensional data. Some of the techniques commonly used are: Principal Component Analysis (PCA), Kullback-Leibler divergence, feature discriminant analysis.

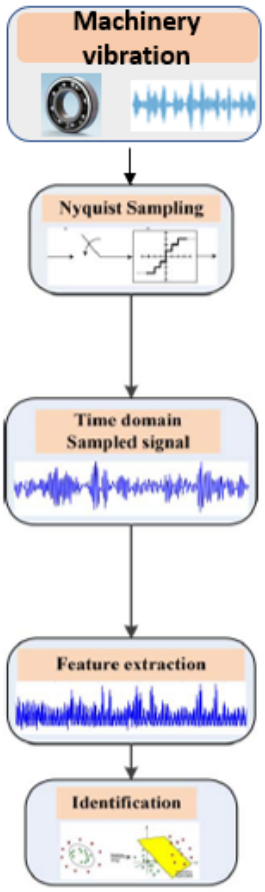


Figure 1. Traditional fault detection method for industrial machinery

All these techniques rely on Nyquist theorem for signal acquisition. According to the theorem the sampling rate must be at least twice the bandwidth of the signal in order to accurately reconstruct it. Then the features can be extracted from the raw data and they are used by some artificial intelligence method, such as Artificial Neural Network (ANN), Support Vector Machine (SVM) or k-nearest neighbour, in order to determine the fault state. In figure 1 this traditional approach is illustrated.

1.2 Cloud and edge computing

Nowadays, fault detection is performed remotely based on industrial Internet of Things (IoT) data processing and Cloud computing.

In the traditional fault diagnosis method, data are collected by IoT devices following the Nyquist sampling theorem. Then, they are sent to the Cloud where they are collected and analysed using analytics techniques.

A long-term and continuous monitoring requires the transmission of a continuous stream of data to the Cloud. Due to the complexity of the information embedded in the vibration signals and the ever-growing demand for the information of complex mechanical systems, the features extracted are not sufficient and the

transmission and storage of the raw signals must be considered. According, a large amount of data is acquired by multiple sensors with high sampling rates over long operation periods, which imposes heavy burden on the acquisition hardware, data storage and transmission bandwidth.
In addition, the measurements acquired from industrial machinery for fault detection hide complex non-linear relationships. Moreover, the bearing fault information will be often interfered or lost in the background noise after the vibration signal has been transferred complicatedly. Thus, the feature extraction requires complex signal processing algorithm and the features extracted are affected by sample size and noise, making the analysis task at the cloud a complex process.
Taking into account the scenario previously depicted, two problems must be faced for fault detection of industrial machinery:

1. The computational efforts at the Cloud are burdensome;
2. The network traffic is high due to the huge amount of data collected.

These two problems can be mitigated considering a decentralized approach. In this sense, Edge Computing may come to the aid.
This methodology can be defined as a set of techniques to optimize cloud computing systems by pushing some portion of an application, its data, or services away from centralized nodes ("the core") to the other logical extreme ("edge") of a network. Figure 2 shows the complementary approach between Cloud Computing and Edge Computing. In the Edge Computing concept, part of the computation efforts is delegated to devices that lie near the sources of data and that interact with them. Thus, data may be processed in real time; the latency can ben reduced and useless data can be discarded.
Even thought it represents an obvious solution to decrease the burden on the Cloud, Edge Computing by itself cannot reduce the burden on the network caused by transmitted data. In fact, as depicted in figure 2, with the presence of billions of sensors and devices, the amount of data that pass through the network is still huge. Thus, in addition, it must be considered a compression technique that can reduce the transmitted data, keeping the embedded information consistent.

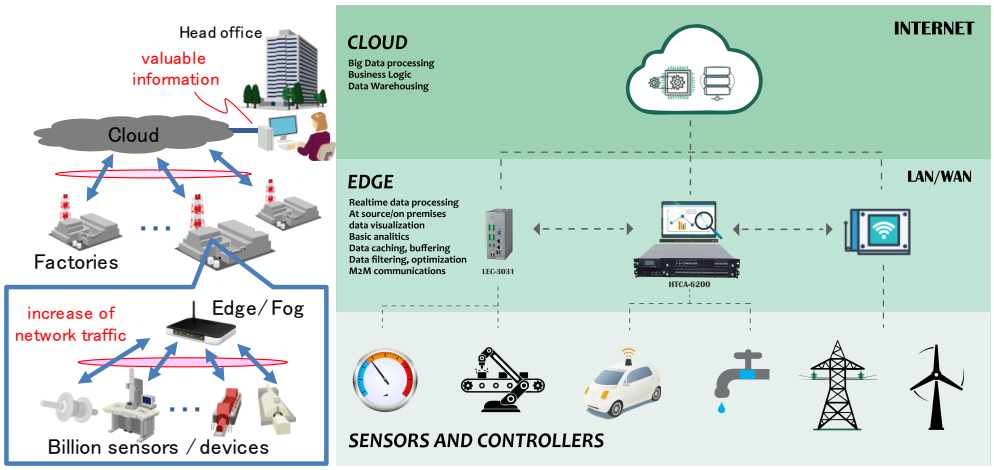


Figure 2. Cloud computing and edge computing

1.3 Intelligent IoT sensing

A system based on the Edge Computing framework is considered in order to try to overcome the problems related to the traditional fault detection method for industrial machinery.

In such a system *compressed sensing* is used in order to reduce the amount of measurement acquired by the sensors keeping the acquisition process simple. In addition, the compressive sensing mechanism allows maintaining the features of the raw signal, at the expense of a more complex mechanics to recover it from the low-dimensionality measurements.

The compressed sensing mechanism works only in presence of a basis, so called dictionary, able to sparsify the signal of interest. In the considered system, the basis is constructed from a set of training signals applying *dictionary learning* algorithms. As the dictionary is constructed on a specific set of signals, then its sparsifying capability will be different on different class of signals. Then, according to the compressed sensing paradigm, the reconstruction quality for a signal in the same class as the training ones will differ from the reconstruction quality for a signal with different vibration signature.

Combining these two techniques, an intelligent IoT sensing system for condition monitoring and fault detection is envisioned. The entire system is illustrated in figure 3, where the three parts of the system are described: sensor terminal, edge terminal and cloud.

At the sensor terminal, the signal is acquired using traditional acquisition methods that follow the Nyquist theorem. Then, the dimensionality of these signals is reduced following compressive sensing theory and the low-dimensional signals can be transmitted at the edge terminal. Only portions of the high-dimensional signals acquired are sent as they are at the edge terminal, as training signals, to construct the dictionary.

Once the dictionary has been trained, it can be used to guarantee a high-quality reconstruction of the sensors data. At the same time, part of the efforts to determine the status of the machinery can be transferred at the edge terminal using the low-dimensional data or the

reconstructed one. As a consequence, the latency in the fault detection can be reduced, taking preliminary decision directly at the edge terminal.

Finally, useful data and useful information on the machinery condition are sent to the cloud to be stored and to carry out further analysis. When a different condition is determined, the previous dictionary is stored on the cloud and new high-dimensional signals are acquired from the sensors to construct a dictionary for the new state.

In the system described, two processes can be considered for fault detection that differs from the traditional one showed in figure 1. On one side, the signal can be reconstructed and a traditional fault detection method may be used on the reconstructed signal. This approach is depicted on the left side of figure 4. On the other hand, the dictionary may be used to determine the state of the machinery at the edge and only valuable information can be sent to the cloud for further analysis. This second approach is depicted on the right side of figure 4.

In this thesis, the feasibility of this system for ball bearing fault diagnosis will be studied in detail. In particular, the study will focus on two aspects: the reconstruction capability of the compressive sensing theory when a dictionary learning approach is followed and the automatic detection of a fault condition at the edge terminal using the compressed measurements.

The remainder of this thesis is organized as follow. Chapter 2 describe the theoretical foundation for compressive sensing acquisition and dictionary learning algorithms. Chapter 3 describes the target signals utilized for machinery monitoring and the reconstruction performance. Chapter 4 analyses the considered algorithm for fault detection based on different trained dictionary for different class of signals. Chapter 5 considers the online detection of an unknown state based on the reconstruction performance on the current state dictionary. Finally, chapter 6 briefly describes the experiment conducted in a real environment as a preliminary analysis of the compressive sensing approach for vibration signals.

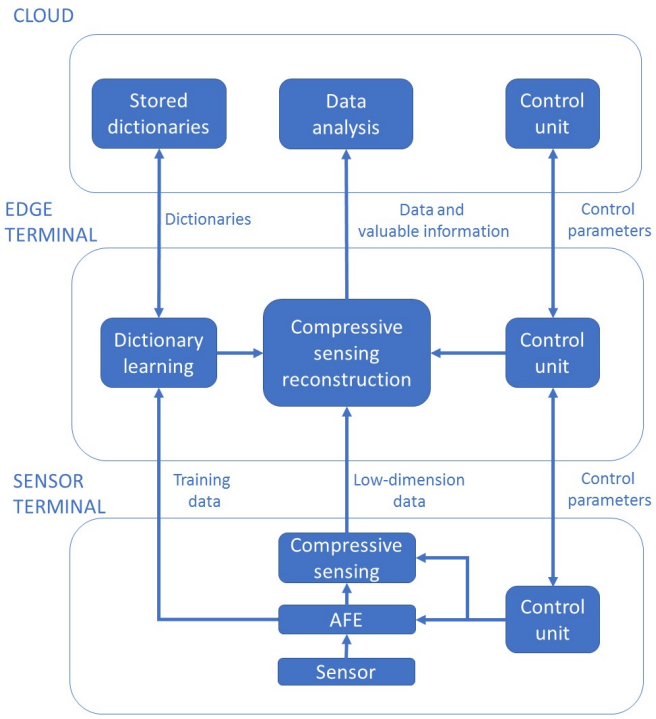


Figure 3. Block diagram of the proposed intelligent IoT sensing system

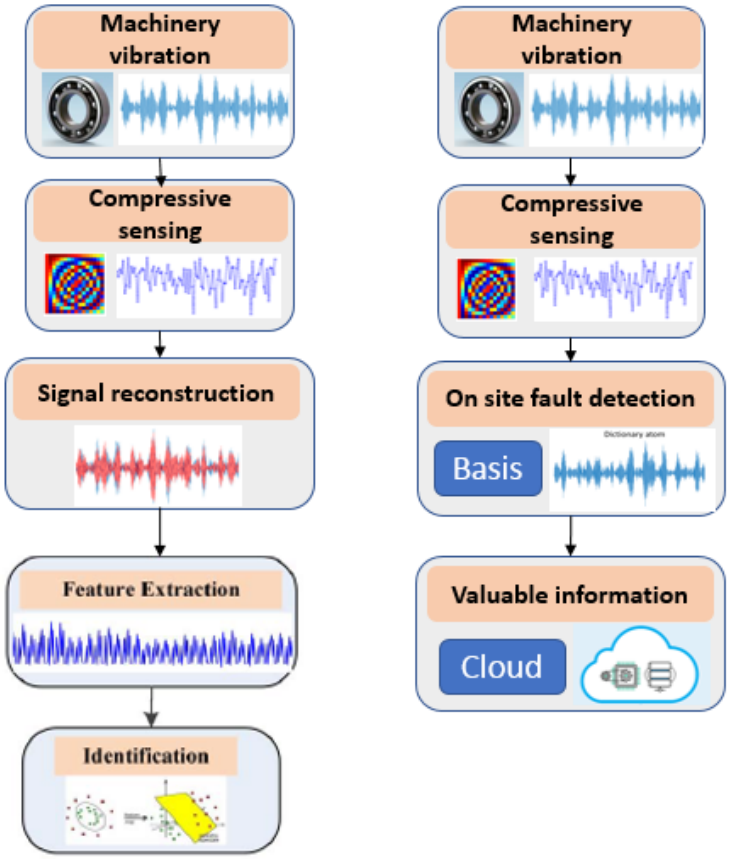


Figure 4. Fault detection process using compressive sensing

2. THEORETICAL BACKGROUND

In this chapter the theoretical foundations of *compressive sensing* and *dictionary learning* will be presented in order to understand how these techniques can be applied for predictive maintenance and condition monitoring.
The following definitions provide a general description of their concepts:

- Compressive sensing: A high dimensional signal can be recovered by means of only few linear measurements, provided that the signal is sparse or nearly sparse;
- Dictionary learning: Representation method which aims at finding a sparse representation of the input data in the form of a linear combination of basic elements as well as those basic elements themselves.

2.1 Compressed sensing

Conventional approaches to acquisition and reconstruction of signals, images, video and many other types of data, follow the basic principle of the Nyquist sampling theorem. This theorem demonstrates that a general continuous-time band limited signal can be exactly recovered from a set of uniformly spaced samples. The necessary condition is that the samples must be acquired at the so-called *Nyquist sampling frequency* f_s, defined as follow

$$f_s \geq 2B \qquad (1)$$

where B is the bandwidth of the signal that is the highest frequency present in the signal.
The Nyquist theorem is so fundamental that nowadays almost every signal acquisition system is based on its principles. Nevertheless, this acquisition system may present two major problems in some applications: the resulting f_s might be so high that a massive number of samples will be acquired, and it might be complex or even impossible to design a device capable to work at the desired f_s. Moreover, the efforts to acquire a huge amount of samples are often useless, since the necessary information that must be conveyed can be compressed. Therefore, it is usual to add a compression stage after the sampling stage. This process leads to a paradox: massive amounts of data are collected at the acquisition stage but they are, in large part, discarded at the compression stage to facilitate storage and transmission.
Compressed Sensing (CS) is a framework that allows overcoming - under certain hypothesis - the limitations that arise when sensing data under a Nyquist theorem-based approach. As Manuel Candès asserts, CS performs as "if it were possible to directly acquire just the important information about the object of interest" [2]. Rather than first sampling at a high rate and then compressing the sampled data, CS allows to find a way to directly sense the data in a compressed form so that - at least in principle – it is possible to obtain super-resolved signals from just a few sensors.

CS relies on two basic principles: the *sparsity* of the signal that must be acquired and the *incoherence* of the sensing process [2]:

- *Sparsity*: the information content of a signal can be expressed using only the values and locations of the largest coefficients of the signal representation on a given basis.
- *Incoherence*: generalization of the uncertainty principle. While the signal of interest has a sparse representation in the given basis, it must be spread out in the acquisition domain. This means that the sensing waveforms have an extremely dense representation in the basis domain.

The CS is a well-established theory that came out from the work of Candès, Romberg and Tao and of Donoho. They showed and demonstrated that a finite-dimensional signal, that has a sparse or compressible representation on a certain domain, can be acquired – in a signal independent way – using a simple and efficient acquisition protocol and it can be recovered from a small set of its linear, incoherent measurements. [3] [4] [5] [6] [7] [8] [9]

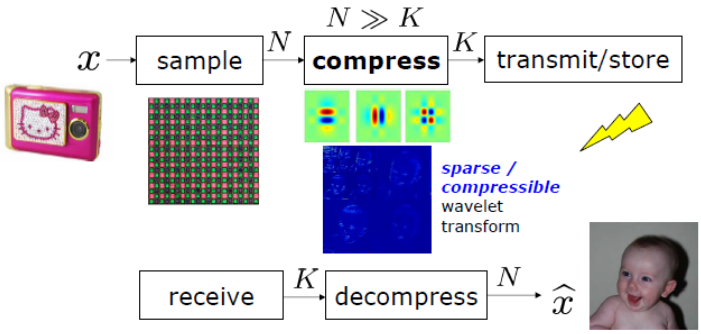


Figure 5. Established paradigm for data acquisition [9]

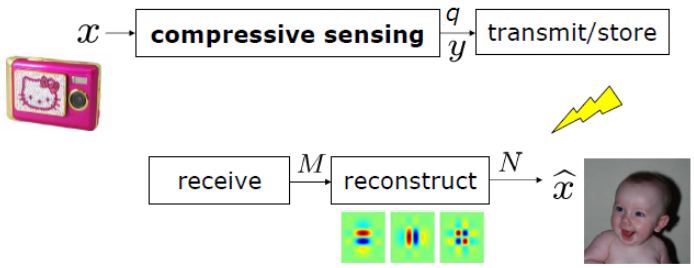


Figure 6. Compressive sensing paradigm for data acquisition [9]

In the following of this paragraph a brief overview of the CS theory will be given.
First of all, normed vector spaced will be reviewed. Secondly, sparsity and incoherence will be widely explained, in order to understand the theoretical guarantees on the reconstruction of the original signal. Finally, the reconstruction algorithms will be explained. [2]

2.1.1 Vector space

Signals generated from common physical systems can be modelled as *vectors* lying in an appropriate *vector space*. Typically, it is useful to consider a *normed vector space,* i.e. a vector space in which a norm is defined.
The general definition of l_p *norm* for $p \in [1, \infty)$ of a signal $\mathbf{x}$ in the signal vector space is the following:

$$\|\mathbf{x}\|_p = \begin{cases} \left(\sum_{i=1}^{n} |x_i|^p\right)^{\frac{1}{p}}, & p \in [1. \infty) \\ \max_{i=1,2,\dots n} |x_i|, & p = \infty \end{cases} \qquad (2)$$

In some cases, it is also convenient to define a norm for $0 < p < 1$. In this way, we obtain a *quasinorm*, since it fails to satisfy the triangle inequality.
A so-called l_0 *norm* is also defined in the following way:

$$\|x\|_0 = |\mathrm{supp}(x)|$$
$$\mathrm{supp}(x) = \{i : x \neq 0\} \qquad (3)$$

Although this quantity does not satisfy the properties of a norm, this abuse of terminology is frequently used. Figure 7 shows a pictographic representation of the most used norm in the compressive sensing theory.

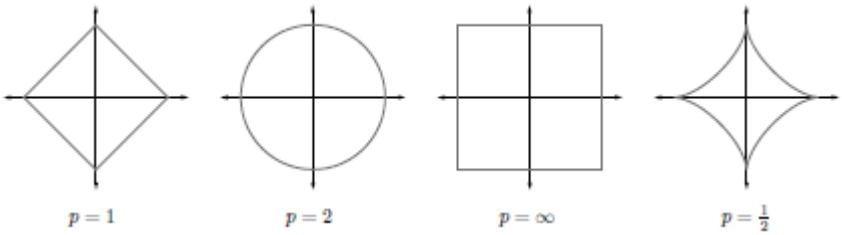


Figure 7. Unit spheres in $\mathbb{R}^2$ for the l_p norms with $p = 1, 2, \infty, 1/2$ [10]

We can also consider the standard *inner product* in $\mathbb{R}^n$, defined as

$$< x, z >= z^T x = \sum_{i=1}^{n} x_i z_i \qquad (4)$$

A fundamental concept in the theory of vector spaces is the one of *basis*. A basis for the vector space $\mathbb{R}^n$ is a set $\{\psi_i\}_{i=1}^n$ of vectors of the space with the following properties:

- the vectors span the entire space $\mathbb{R}^n$;
- the vectors are linearly independent.

This implies that each vector in the space can be uniquely represented as a linear combination

of the vectors in the basis. In particular, for each vector $x \in \mathbb{R}^n$, there exist unique coefficients $\{c_i\}_{i=1}^n$ s.t.

$$x = \Psi c = \sum_{i=1}^{n} c_i \psi_i \tag{5}$$

If the basis is orthonormal, that is

$$< \psi_i, \psi_j >< \psi_i, \psi_j >= \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases} \tag{6}$$

then the coefficients can be easily calculated as follow

$$c_i =< x, \psi_i >$$
$$c = \Psi^T x \tag{7}$$

It is useful to generalize the concept of a basis considering a set of possibly linearly dependent vectors. Such a set of vectors is usually called *frame.* Frames allow for a better representation of a signal due to their redundancy: for a given signal x the coefficients $\{c_i\}_{i=1}^n$ will not be unique anymore.

Note that, in the context of sparse representation theory, basis and frames are usually referred as *dictionary* and *overcomplete dictionary* respectively, while the basis or frame vectors are called *atoms.*

2.1.2 Sparse model

The information contained in a high-dimensional signal is frequently small compared to its actual dimensionality. For this reason, such signals can often be well approximated using far fewer atoms of a dictionary compared to the signal support. If this approximation is exact the signal is said to be *sparse;* if the approximation is not exact the signal is said to be *compressible*.

Mathematically, a signal $x \in \mathbb{R}^n$ is *k-sparse* if

$$\|x\|_0 \leq k \tag{8}$$

Usually, the signals of interest in real applications are not sparse by themselves. However, they can admit a sparse representation in some basis Ψ. In this case, we will express the signal as $x = \Psi c$ and it will be *k-sparse* if

$$\|c\|_0 \leq k \tag{9}$$

In the context of the following discussion, we let

$$\Sigma_k = \{x: \|x\|_0 \leq k\} \tag{10}$$

be the set of all *k*-sparse signals.

Actually, few real-world signals are precisely sparse; rather they are compressible.

It is possible to define a compressible signal considering the rate of decay of its coefficients. Specifically, consider an ordered version of the coefficients of a signal in a given sparsifying basis, that is $|c_1| \geq |c_2| \geq \cdots \geq |c_n|$. Then, a signal is said to be *compressible* if its coefficients obey a power law decay, that is there exist constants $C_1, q > 0$ s.t.

$$|c_i| \leq C_1 i^{-q} \tag{11}$$

The larger q, the more compressible the signal is. Compressible signals can be accurately represented by $k \ll n$ coefficients with a negligible error.

Since the choice of which dictionary atoms are used can change according to the specific signal considered, sparsity is a highly nonlinear model. In fact, a linear combination of two k-sparse signals will generate a signal that may not be k-sparse anymore, since their supports may not coincide.

Figure 8 shows an example of wavelet approximation. The wavelet transform represent a sparsyfing basis Ψ. Even if image are not exactly sparse in the basis, they are compressibile and the approximation on the coefficients results very similar to the original image.

In the theory of compressed sensing, we will consider that it will always exist a sparsifying basis; that is a dictionary of atoms $\{\psi_i\}_{i=1}^n$ that generate a space in which the considered signal is sparse or compressible. Then, in the following of this thesis we will see how to choose this sparsifying basis.

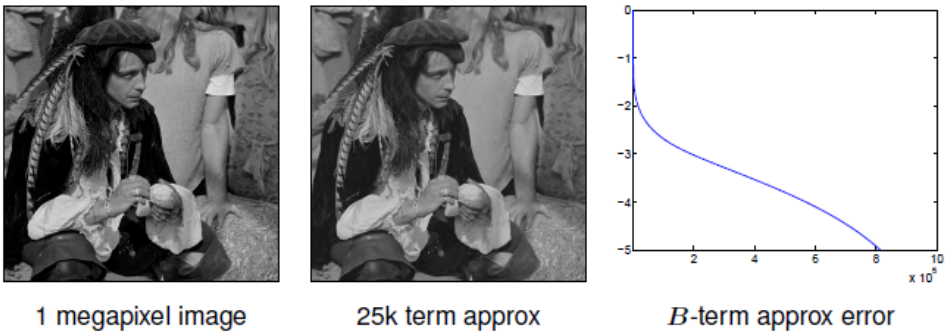


Figure 8. Wavelet approximation [2]

2.1.3 The sensing problem

Let consider $x \in \mathbb{R}^n$ a discrete signal, that is sparse or compressible. For this signal, the sensing mechanism of a compressed sensing framework can be described in the following way:

$$y = \Phi x \qquad i = 1 \dots q \tag{12}$$
$$y_i = < \Phi_i, x >$$

Where y_i for $i = 1 \dots q$ represent the q acquired measurements and $\Phi \in \mathbb{R}^{qxn}$ is the *sensing matrix.*

More precisely, x may also be the sparse representation on a given basis Ψ of a non-sparse

signal $f \in \mathbb{R}^n$; then it will result $f = \Psi x$, in the sense that it is possible to approximate the signal f with is sparse representation x. In this last case the sensing mechanism will be described in a slightly different way:

$$y = \Phi f = \Phi \Psi x \quad (13)$$

Where $\Psi \in \mathbb{R}^{nxn}$ is the *sparsifying basis* or *dictionary*.
Then supposing that the signal of interest is sparse in some basis we can work with the general model:

$$y = Ax \quad (14)$$

considering $A \epsilon \mathbb{R}^{qxn}$
Obviously, in the context of compressive sensing, we are interested in having $q \ll n$; where n is the dimensionality of the signal and it is in general large, while q is typically much smaller than n;
In this scenario, two main questions arise:

- How should we design the sensing matrix in order to be able to acquire all the relevant information contained in x?
- How can we recover the original signal x from the low-dimension measurements?

First of all, we will define a number of properties that the sensing matrix should satisfy to ensure feasibility of the reconstruction. Secondly, we will describe the algorithms used to recover the original high-dimensional signal from the low-dimensional measurements.

2.1.4 Sensing matrices

Null space condition

If we want to be able to recover all sparse signals x from the measurements y, then for any $x, x' \epsilon \Sigma_k$ it must be $Ax \neq Ax'$, otherwise it would be impossible to distinguish x from x' based only on y.
More formally, we can define the *null space* of A as

$$\mathcal{N}(A) = \{z: Az = 0\} \quad (15)$$

Then A uniquely represents all $x \epsilon \Sigma_k$ iif $\mathcal{N}(A)$ contains no vectors in Σ_{2k}.
The most common way to characterize this property is known as *sparks*.

DEFINITION 1.1 The spark of a matrix A is the smallest number of columns of A that are linearly dependent. [11]

This definition lead to the following guarantee:

THEOREM 1.1 For any vector $y \epsilon \mathbb{R}^q$ there exists at most one signal $x \epsilon \Sigma_k$ s.t. $y = Ax$ iif $spark(A) > 2k$ [11]

COROLLARY 1.1 Since $spark(A) \in [2, q+1]$ then Theorem 1.1 asserts that the theoretical number of measurements necessary to obtain a unique solution is $q \geq 2k$.

Theorem 1.1 holds for exactly sparse signals but it does not guarantee the recovery for compressible signals. For the latter, we have to guarantee that $\mathcal{N}(A)$ does not contain any vectors that are too compressible in addition to sparse vectors.

Mathematically, defining $T \subset \{1,2,\dots,n\}$ as a subset of indices and $T^c = \{1,2,\dots,n\} \setminus T$.

The, we can define the *Null Space Property (NSP)*

DEFINITION 1.2 (NSP) A matrix A satisfies the NSP of order k if there exists a constant C>0 s.t.

$$\|h_T\|_2 \leq C \frac{\|h_{T^c}\|_2}{\sqrt{k}} \qquad (16)$$

for all $h \in \mathcal{N}(A)$ and for all T such that $|T| \leq k$.

Generally speaking, the NSP says that the vectors in the null space of A should not be too concentrated on a small subset of indices.

Restricted isometry property

The NSP is a necessary and sufficient condition to guarantee the recovery of a sparse or compressible signal. Nevertheless, it does not take into consideration the noise.

In case of corrupted measurements, Candès and Tao introduced the so called *Restricted Isometry Property (RIP)* [12].

DEFINITION 1.3 (RIP) A matrix A satisfies the RIP of order k if for all $x \epsilon \Sigma_k$ there exists a $\delta_k \epsilon \{0,1\}$ s.t.

$$(1-\delta_k)\|x\|_2^2 \leq \|Ax\|_2^2 \leq (1+\delta_k)\|x\|_2^2 \qquad (17)$$

Generally speaking, if a matrix A satisfies the RIP of order 2k, it means that it approximately preserves the distance between any pair of k-sparse vectors. In fact, it means that the distance between any $x_1, x_2 \epsilon \Sigma_k$ will be preserved by the corresponding measurements since

$$(1-\delta_{2k})\|x_1 - x_2\|_2^2 \leq \|Ax\|_2^2 \leq (1+\delta_{2k})\|x_1 - x_2\|_2^2 \qquad (18)$$

This property assures that, if the measurements matrix satisfies the RIP, then the measurement process will be robust against noise.

Moreover, it is also possible to consider how many measurements are necessary to achieve the RIP. If we only focus on the dimensions of the problem (n, m and q) then we can determine a lower bound on the number of measurements needed.

THEREOM 1.2 Let $A \epsilon \mathbb{R}^{qxn}$ be a matrix that satisfies the RIP of order 2k with constant $\delta_{2k} \in (0, \frac{1}{2}]$, then

$$q \geq Cklog(\frac{n}{k}) \quad (19)$$

Where $C = \frac{1}{2}\log(\sqrt{24}+1) \approx 0.28$

Coherence

The spark, NSP and RIP provide theoretical guarantee on the reconstruction of the original signal. Nevertheless, it is typically difficult to determine whether a general matrix A satisfies the above conditions. In many cases, it is preferable to define a property of A that is easily computable. For this reason, the *coherence* of a matrix is generally used in the CS framework [11].

DEFINITION 1.4 (COHERENCE) The coherence of a matrix A, $\mu(A)$, is the largest absolute inner product between any two columns a_i and a_j of A:

$$\mu(A) = \max_{1 \leq i < j \leq n} \frac{|< a_i, a_j >|}{\|a_i\|_2 \|a_j\|_2} \quad (20)$$

It can be shown that $\mu(A) \in \left\{\sqrt{\frac{n-q}{q(n-1)}}, 1\right\}$ where the lower bound is called Welch bound. Moreover, if $n \gg q$, the then $\mu(A) \in \left\{\frac{1}{\sqrt{q}}, 1\right\}$.

The coherence is useful because it can be related to the spark and the RIP.

LEMMA 1.1 For any matrix A

$$spark(A) \geq 1 + \frac{1}{\mu(A)} \quad (21)$$

From THEOREM 1.1 it can be deduced that to reduce the number of measurements required to recover the signal, spark(A) must be high so $\mu(A)$ must be as low as possible.
Moreover, merging THEOREM 1.1 with LEMMA 1.1 the following condition on A guarantees the uniqueness of the solution

THEOREM 1.3 If

$$k < \frac{1}{2}\left(1 + \frac{1}{\mu(A)}\right) \quad (22)$$

then for each measurements vector $y \in \mathbb{R}^q$ there exists at most one signal $x \epsilon \Sigma_k$ such that $y = Ax$.

Theorem 1.7 together with the Welch bound, provides an upper bound on the level of sparsity k that guarantees uniqueness using coherence: $k = O(\sqrt{q})$.
Similarly, the coherence can also be related to the RIP

LEMMA 1.2 If A has unit-norm columns and coherence $\mu = \mu(A)$, then A satisfies the RIP of order k with $\delta_k = (k-1)\mu$ for all $k < \frac{1}{\mu}$.

In the case in which $A = \Phi\Psi$, we consider the *mutual coherence* $\mu(\Phi, \Psi)$

DEFINITION 1.5 (MUTUAL COHERENCE) The *mutual coherence* $\mu(\Phi, \Psi)$ is the maximum inner product between any two columns of Φ and Ψ.

$$\mu(\Phi, \Psi) = \max_{1 \le i < j \le n} \frac{|< \phi_i, \psi_j >|}{\|\phi_i\|_2 \|\psi_j\|_2} \quad (23)$$

2.1.5 Sensing matrices construction

Summarizing the results from the previous section, we are searching for a matrix A that has high spark, low coherence and satisfies the RIP. As stated at the beginning of this paragraph, incoherence plays a major role in the robust reconstruction of the original signals.
CS is mainly concerned with low coherence pairs. Examples of such a pair are:

- the canonical spike basis $\phi_k(t) = \delta(t-k)$ as sensing matrix and the Fourier basis $\psi_j = \frac{1}{\sqrt{n}} e^{-i2\pi jt/n}$ as sparsifying matrix that are maximal incoherent in any dimension.
- the noiselets as sensing matrix and the wavelets as sparsifying matrix;

These constructions make use of well know sparsifying basis that achieve very good results in many domains. Then in the construction of the sensing basis we need to take into account the specific sparsifying basis used in order to guarantee the conditions on the reconstruction.
Nevertheless, if sensing with incoherent systems is good, then efficient mechanisms have to acquire correlations with random waveform. This is the reason why early works in CS has promoting the use of *random matrices*.
It can be demonstrated that, a random matrix A whose entries are independent and identically distributed (i.i.d.), with continuous distributions, have $\text{spark}(A) = q + 1$ with probability one. Moreover, it can also be shown that if the entries are chosen to a Gaussian, Bernoulli, or more generally any sub-gaussian distribution, then the resulting matrix will satisfy the RIP with high probability.

This means that random matrices guarantee the reconstruction of the original signal. Moreover, they have many additional benefits. First, it can be shown that, for random constructions, it is possible to recover the signal using any sufficiently large subset of the measurements. Thus, the system can be robust against any lost of corruption of small portions of the measurements. Secondly, as it will be explained in the next paragraph, the sparsifying matrix can also be adapted to the specific signal. In these cases, *universality* is required; that is, the reconstruction must be guaranteed independently from the actual sparsifying basis used. It happens that, random matrices have this useful property. In fact, they result to be maximally incoherent with almost all sparsifying basis. This means that a random matrix will guarantee the reconstruction independently from the domain in which the signal is sparse, if the number of measurements is sufficient.

However, it worth to notice that there are some drawbacks of using random matrices. In fact, a fully random matrix is sometimes impractical to build in hardware. For this reason, they are usually replaced by a pseudorandom approach; though it can be very resource intensive, for large dimension signals.

Despite these disadvantages, in the following it will be assume the usage of a *random gaussian matrix*. For the purpose of this thesis, in fact, the universality property of the random matrices plays an important role. Moreover, it allows concentrating the attention on the reconstruction quality of the signal using a simple mechanics for the acquisition.

Obviously, the sensing matrix is one of the key theoretical aspects of compressive sensing. For this reason, many studies have been done on its construction and it is still an open field of study.

This aspect falls outside the purpose of this thesis, for this reason it will be neglected in this overview. Nevertheless, it can represent an open path for further improvement of the present work.

2.1.6 Signal recovery

The reconstruction problem implies the solution of the under-determined system of equations

$$y = Ax \tag{24}$$

Given the measurements y and the knowledge that our original signal is sparse or compressible, the natural way of attempt to recover x is by solving an optimization problem. The different approaches that can be followed differ from the objective function considered. Common approaches are summarized below:

- *Minimum l_0 norm reconstruction;* the problem to be solved is of the form

$$\hat{x} = \min_{x \epsilon \mathbb{R}^n} \|x\|_0 \ subject\ to\ x \epsilon \mathcal{B}(y) \tag{25}$$

where $\mathcal{B}(y)$ ensure that $\hat{x}$ is consistent with the measurements y. In particular we will consider two cases:

- Absence of noise: $\mathcal{B}(y) = \{x: y = Ax\}$
- Presence of noise: $\mathcal{B}(y) = \{x: \|Ax - y\|_2 \leq \varepsilon\}$

This optimization problem leads to the reconstruction of the original K-sparse signal with 2K random measurements and it is possible to analyze theoretically its performance. Unfortunately, the objective function considered in this case is non-convex and for a general matrix A it results in a NP-hard problem;

- *Minimum l_2 norm reconstruction;* the problem to be solved is of the form

$$\hat{x} = \min_{x \epsilon \mathbb{R}^n} \|x\|_2 \ subject\ to\ x \epsilon \mathcal{B}(y) \qquad (26)$$

The main advantage of this approach is the simplicity of the solution but the solution is almost always an incorrect one. In fact, in this case we are looking for a sparse solution but we will find the least energy one. In the case in which the original signal is not sparse by itself, the sparsest solution does not necessarily have the least energy. For this reason, it may differ from the exact solution given by the l_0 norm reconstruction.

- *Minimum l_1 norm reconstruction;* the problem to be solved is of the form

$$\hat{x} = \min_{x \epsilon \mathbb{R}^n} \|x\|_1 \ subject\ to\ x \epsilon \mathcal{B}(y) \qquad (27)$$

This approach corresponds to a convex relaxation of eq 25. Even if it is not immediately obvious, the solution to eq. 27 will be similar to the solution to eq 25. Even if, a theoretical explanation will be skipped in this context, figure 9 and figure 8 show a geometric interpretation of the usage of l_1 norm and the accurate reconstruction compared to the usage of l_2 norm. It is also possible to show that the exact solution can be recovered by $q \geq Cklog(\frac{n}{k})$ measurements when a random sensing matrix is used. Moreover, provided that $\mathcal{B}(y)$ is convex, this approach is computationally feasible.

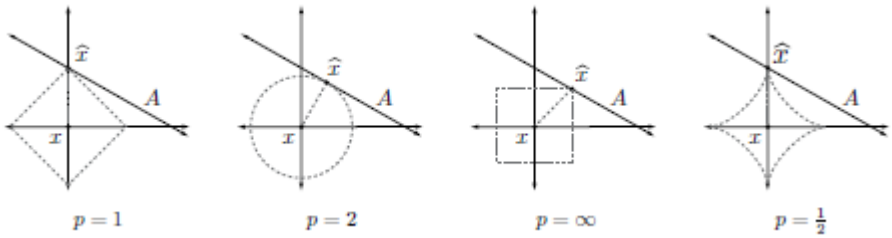


Figure 9. Best approximation of a point in R2 by a one-dimensional subspace using l_p norms for $p = 1,2,\infty,{}^1/_2$ [10]

The l_1 norm represents a powerful tool for recovering sparse signal. The power of l_1 minimization is that not only will it lead to a probably accurate recovery, but the formulation in eq. 27 is a convex optimization problem for which there exist efficient and accurate numerical solvers.

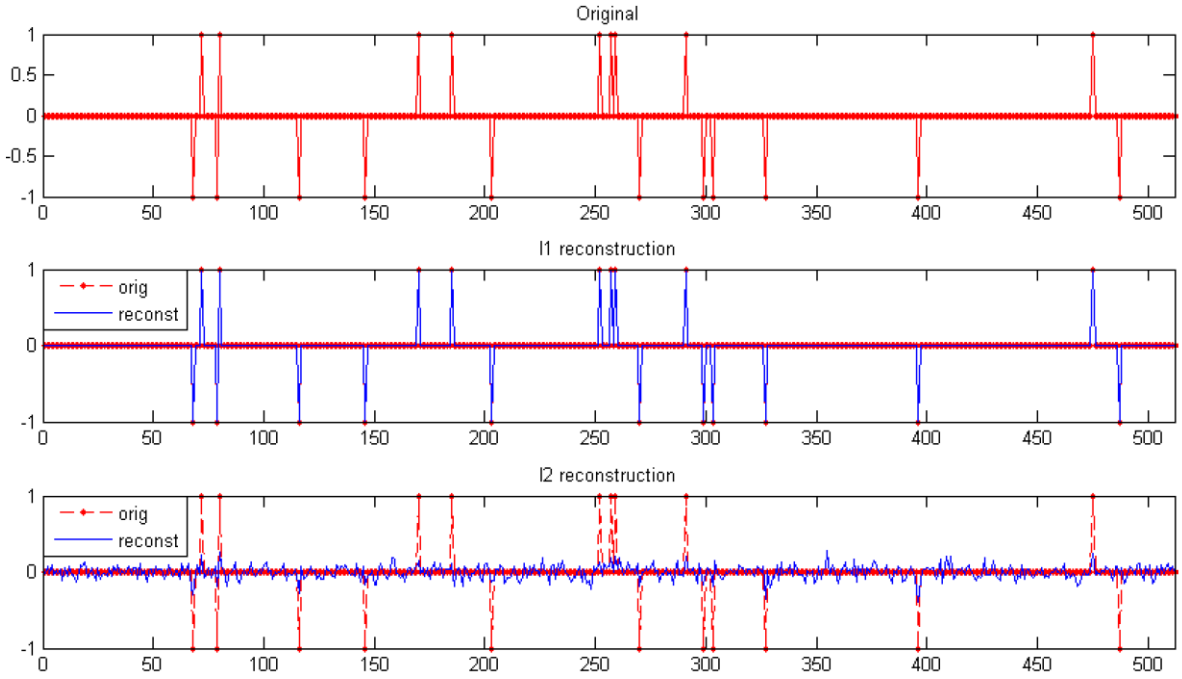


Figure 10. Difference in the reconstruction between l_1-norm and l_2-norm

While the optimization problem could be solved using general-purpose convex optimization software, there now also exist several algorithms designed to explicitly solve this problem in the context of CS. These works have primarily focused on the case where $\mathcal{B}(y) = \{x: \|Ax - y\|_2 \leq \varepsilon\}$. However, many works in literature have also considered the unconstrained version of this problem that is:

$$\hat{x} = \min_{x \epsilon \mathbb{R}^n} \frac{1}{2} \|Ax - y\|_2 + \lambda \|x\|_1 \qquad (28)$$

Even if, convex optimization techniques are powerful methods for computing sparse representation, there is also a variety of greedy algorithms for solving such problems. [13] [14]

Greedy algorithms rely on iterative approximation of the signal coefficients and support, either by iteratively identifying the support of the signal until a convergence criterion is met, or by obtaining an improved estimate of the sparse signal at each iteration that attempts to account for the mismatch to the measured data.

Some greedy method can actually be shown to have performance guarantees that match those obtained for convex optimization problems. However, the techniques required to prove performance guarantees are substantially different.

One example of a greedy algorithm used for sparse signal recovery is the *Orthogonal Matching Pursuit (OMP)* [15]. It begins by finding the column of A most correlated with the measurements. Then, it repeats this step by correlating the columns with the signal residual, which is obtained by subtracting the contribution of a partial estimate of the signal from the original measurements vector.

The algorithm is formally defined in figure 11. In this definition, supp () finds the location of non-zero values in a vector while $H_k(x)$ denotes the hard thresholding operator on x that sets all entries to zero except for the k entries of x with largest magnitude. The stopping criterion can consist of either a limit on the number of iterations, that also limits the number of non-zeros in $\hat{x}$, or a requirement that $y \cong A\hat{x}$ in some sense.

Other examples of greedy algorithms used for CS reconstruction are: *Matching Pursuit, Gradient Pursuit, Iterative Thresholding.* [14]

Algorithm: Orthogonal Matching Pursuit	
Inputs:	Measurement matrix Φ and measurements y
Initialization:	$\hat{x}_0 = 0, r_0 = y, \Gamma_0 = \emptyset, i = 0$
	while stopping criterion not met **do**
	$temp \leftarrow \Phi^T r_i$
	$i \leftarrow i + 1$
	$\Gamma_i \leftarrow \Gamma_{i-1} \cup support(threshold(temp, 1))$
	$\hat{x}_i\|_\Gamma \leftarrow \Phi_\Gamma^\dagger y, \hat{x}_i\|_{\Gamma^C} \leftarrow 0$
	$r_i \leftarrow y - \Phi\hat{x}_i$

Figure 11. OMP algorithm for compression sensing reconstruction

2.2 DICTIONARY LEARNING

The theoretical results described in the previous section are correct under the hypothesis of sparsity of the original signal. Since the compression sensing theory is completely founded upon the sparsity hypothesis, crucial points are the *sparse problem* and the *dictionary selection.*

Considering the general frame in (1.13), the original signal $\mathrm{Y} \in \mathbb{R}^n$ must be represented as

$$\mathrm{Y} = \Psi \mathrm{x} \qquad (29)$$

where $\mathrm{x} \in \mathbb{R}^K$ is the *sparse representation* of Y and $\Psi \in \mathbb{R}^{nxK}$ is the *dictionary* or *sparsifying basis.*

In this context, given the original signal Y, two problems must be faced:

- How to select the dictionary Ψ;
- Given a dictionary Ψ how to calculate the sparse representation x.

In order to better understand the problem, Figure 12 represents a pictorial description of the sparsifying process.

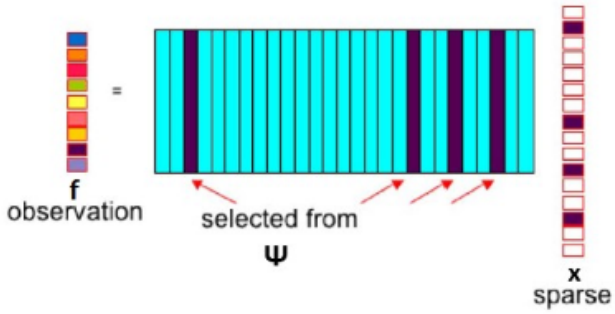


Figure 12. Sparse representation using dictionary Ψ

2.2.1 Sparse representation

If the dictionary Ψ is given, then eq. 29 represents a system of linear equations to be solved in order to find the representation of the signal on the given dictionary. If $K = n$ the representation of signal Y on the dictionary is unique. Although this case is useful to univoquely represent the signal, in the framework of sparse representation, it is much more interesting to consider two other cases:

- *Undercomplete dictionary:* $K < n$;
- *Over-complete dictionary:* $K > n$;

In these conditions, the system of linear equation in eq. 29 is undetermined and it does not exist a unique solution to the representation problem. This allows us to seek the most informative representation of the signal, promoting *sparsity*.
If we define the dictionary atoms as $\psi_i \; i = 1, \dots, K$, then the representation will be sparse if only few atoms of the dictionary Ψ will be actively used in the linear combination.
The problem of finding the atoms that represent the signal can be formulated as an optimization problem in the following way:

$$\min_x \|x\|_0 \text{ subject to } \|f - \Psi x\|_2 \leq \varepsilon \qquad (30)$$

If we compare eq. 30 with eq. 25, it is clear that they represent two faces of the same problem. That is why the techniques used to solve this NP-hard problem are actually the same. In particular, the two classes of algorithms seen in the previous section will be used. The first

includes greedy algorithm such as OMP. In the second group, we find algorithms based on convex relaxation method such as LASSO.
The performance of the algorithms in terms of approximation quality and sparsity level depends not only on the signal but also on the dictionary. In fact, considering a specific class of signals, not all the dictionaries provide the same approximation performance. There exist dictionaries that are more able to describe the intrinsic nature of the class of signals, leading to higher sparse solutions.
Since the compression sensing reconstruction quality is tightly linked to the sparsity level of the signal – according to theorems highlighted in the previous section – then the dictionary itself plays a major role in the compression sensing framework.

2.2.2 Dictionary learning

Designing the dictionary in order to fit the structure in the original data is one of the most challenging tasks in the theory of sparsness.
Earlier works made use of a predetermined set of basis function, such as *Fourier transforms*, *discrete cosine transforms (DCT), wavelets, curvelets*, etc. These dictionaries are simple, always lead to fast algorithms and they can be used for many different kinds of signals. However, their performances largely depend on how adaptive the atoms are to the signal structure. For this reason, when dealing with more complex signal data, it is often difficult to generate a very sparse representation and the representation error would be very large.
In order to overcome these problems, another approach for designing the dictionary is to adapt the dictionary to the data set considered. In this second case, called *dictionary learning*, the dictionary is designed according to a learning process from a set of sample signals. Such a dictionary learning process can capture the inherent characteristic of the signal regardless of any prior knowledge. Moreover, it can reduce the representation sparsity and the representation error, since it is fitted on the class of signal used in the training phase. This means that the dictionary is not general anymore but it can only represent with high performances a specific class of signal. The drawback of this second type of dictionary is that the learning process is usually very slow compared to the fast algorithms available for general purpose dictionaries.
Following this two main branches, in the past decades many techniques have been developed to construct the sparsifying basis. An overview of these methods can be found in [16] [17].
In the following of this thesis, the advantages provided by a learned dictionary will be exploited. For this reason, in this section, the most frequently used learning techniques will be described; that is the *Method of Optimal Directions (MOD)* [18] and the *K-singular value decomposition (K-SVD)* [19]. First of all, a mathematical formulation of the dictionary learning problem is needed.

Given a set of training data $\{Y_j\}_{j=1}^{m} \in \mathbb{R}^n$, from the particular class of signals at hand, that are organized into a matrix $Y \in \mathbb{R}^{nxm}$, the dictionary learning algorithm seek to find a dictionary $\Psi \in \mathbb{R}^{nxK}$, in such a way that all training signals have a sufficiently sparse representation in it, and the sparse representation itself $X \in \mathbb{R}^{Kxm}$, where $\{X_j\}_{j=1}^{m} \in \mathbb{R}^K$ represents the sparse representation coefficient vector corresponding to signal Y_j.

The problem can be formulated as an optimization problem of the form:

$$\left[\widehat{\Psi}, \widehat{X}\right] = \underset{\Psi,X}{\operatorname{argmin}} \|Y - \Psi X\|_2^2 \ \text{s.t.} \ \forall j = 1, \dots, M \ \left\|x_j\right\|_0 \leq k \qquad (31)$$

The above problem is not convex with respect to the pair (Ψ, X) and, for this reason, it is difficult to solve. The usual approach that many dictionary learning algorithms follow consists in dividing the problem in two steps and iteratively perform them until a stopping criterion is met.

The two steps are:

1) *Sparse representation:*

$$\widehat{X}^{(i+1)} = \underset{X}{\operatorname{argmin}} \left\|Y - \widehat{\Psi}^{(i)} X\right\|_2^2 \ \text{s.t.} \ \forall j = 1, \dots, M \ \left\|x_j\right\|_0 \leq k \qquad (32)$$

   The dictionary is fixed from the previous iteration and the sparse representations of all the training signals are computed using the current dictionary.

2) *Dictionary update:*

$$\widehat{\Psi}^{(i+1)} = \underset{\Psi}{\operatorname{argmin}} \left\|Y - \Psi \widehat{X}^{(i+1)}\right\|_2^2 \qquad (33)$$

   The dictionary is updated to reduce the representation error of step 1.

Since Step 1 is an ordinary sparse coding problem that can be solved with many different sparse coding algorithms [14], the main difference between many dictionary learning algorithms is Step 2.

K-SVD

K-SVD is a highly efficient dictionary learning method that exploits the Singular Value Decomposition (SVD) to update the dictionary. In the dictionary update stage, only one atom is updated at a time. Moreover, while updating each atom, the non-zero entries in the associated row vector of X are also updated. This leads to a matrix rank-1 approximation problem which is then solved via SVD operation.

As we said previously, the general approach followed consists of two steps. In the K-SVD algorithm, Step 1 makes use of the OMP algorithm described in figure 11, in order to find the sparse representations of the training signals.

Instead, in order to update the dictionary, eq 33 can be rewritten in the following way:

$$\|Y - \Psi X\|_2^2 = \left\| Y - \sum_{j=1}^{K} \psi_j x_j^T \right\|_2^2 = \left\| \left(Y - \sum_{j \neq k} \psi_j x_j^T \right) - \psi_k x_k^T \right\|_2^2 \quad (34)$$

$$= \|E_k - \psi_k x_k^T\|_2^2$$

Where E_k means the error for all signal samples when the atom ψ_k is removed. At this point, we may be tempted to perform SVD on E_k directly and find alternative ψ_k and x_k^T to reduce the error. However, a step must be done before performing SVD to be sure the updated x_i^T is not filled.

A variable ω_k is defined as:

$$\omega_k = \{i | 1 \leq l \leq K, x_k^T(i) \neq 0\} \quad (35)$$

The sample signals $\{Y_i\}$ that use the atom ψ_k can be indexed by ω_k and the positions of non-zero entries in x_k^T can be determined by parameter i in eq 35. Defining Ω_k as a matrix of sizes $\{m, |\omega_k|\}$, with ones on the $(\omega_k(i), i)$ th entries and zeros elsewhere. The multiplication $x_k^R = x_k^T \Omega_k$ changes the length of x_k^T to $|\omega_k|$ by removing the zero entries. In the same way we obtain the matrices $Y_k^R = Y_k \Omega_k$ of sizes $\{n, |\omega_k|\}$ and $E_k^R = E_k \Omega_k$ of sizes $\{n, |\omega_k|\}$ only include the sample signals or error columns that use the atom ψ_i .

Therefore, we define:

$$\|E_k \Omega_k - \psi_k x_k^T \Omega_k\|_2^2 = \|E_k^R - \psi_k x_k^R\|_2^2 \quad (36)$$

Figure 13 visually explains the procedure.

At this point, we can process $E_k^R = U \Delta V^T$ using SVD. In this way, the atom ψ_k can be changed to the first column of U and x_k^R to the first column of V multiplied by $\Delta(1,1)$.

All the atoms in Ψ updated one by one and the iteration of sparse coding and dictionary learning is repeat until converge or the number of iteration reached.

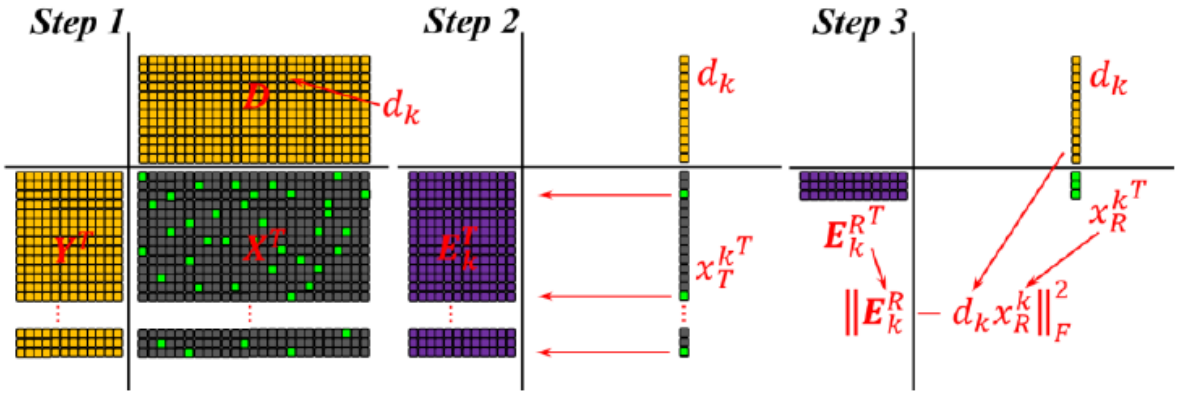


Figure 13. KSVD algorithm procedure [20]

3. COMPRESSED SENSING RECONSTRUCTION OF ROLLING ELEMENT BEARING VIBRATION SIGNALS

3.1 Rolling element bearings and vibration signals

Rolling elements bearing of different sizes are widely used in machinery in variety of different industrial applications: in production line, in electric motors, pumps, and gearboxes.
In general, rolling element bearings are designed to carry axial and/or radial load while minimizing the rotational friction by placing rolling elements such as cylinders or balls between inner and outer races.
There are different types of rolling element bearings. Among all of them, ball bearings are the cheapest, since balls are used instead of cylinders in their construction. Ball bearings have smaller sizes and limited load carrying capacity compared to the other rolling element bearings, but they can support both axial and radial loads. [21]
There are also different types of ball bearings such as thrust, axial, angular, contact and deep groove ball bearing. In the following of this thesis, the latter will be considered and it is depicted in figure 14.

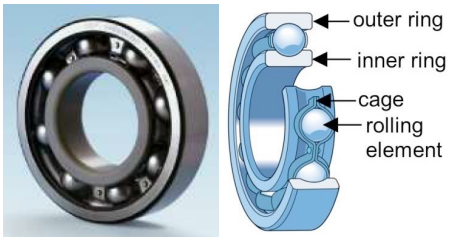


Figure 14. Typical deep-grove ball bearing and its main elements [22]

As it can be observed from figure 14, this type of ball bearing consists of an inner ring, an outer ring, balls and a cage holding the balls apart from each other. In most cases, the outer ring is held stationary while the inner ring and the balls rotate.
There are several types of defects that can occur on a ball bearing including, mechanical damage, crack damage, lubrificant deficiency and corrosion [21]. Moreover, they can affect all the three principal components of the bearing: the outer ring, the inner ring and the rolling elements.
Whichever is the cause that leads to deterioration of the machinery component, these defects will create a specific pattern in the vibration signals generated by the machine. The specific pattern is different according to the different position of the fault. In general, both inner and outer race faults conditions will generate a periodic modulation on the signal. In fact, the

rolling elements will periodically hit the defects, according to a certain frequency that can be related to the rotation period of the bearing. On the other hand, a damage on the rolling element may or may not generate a periodic vibration, depending on some factors such as the load and the degree of the defects.

In figure15, the vibration signals for a fault free bearing (normal working condition), a bearing with an inner race fault, a bearing with an outer race and a bearing with a rolling element fault are represented.

As can be seen, both the inner ring fault and the outer ring fault generate impulses in the vibration signals acquired, compared to the vibration signal in the absence of fault conditions. The outer ring fault signal is different from the inner ring fault signal because, in most case, the former is kept fixed while the latter and the ball rotate. Especially in the case of an outer ring fault, the pattern depends on the position of the load. On the other hand, the ball fault condition modifies the pattern of the vibration signal but it does not generate clear impulses in the signal. For this reason, it might result more difficult to detect this kind of fault compared to a ring fault.

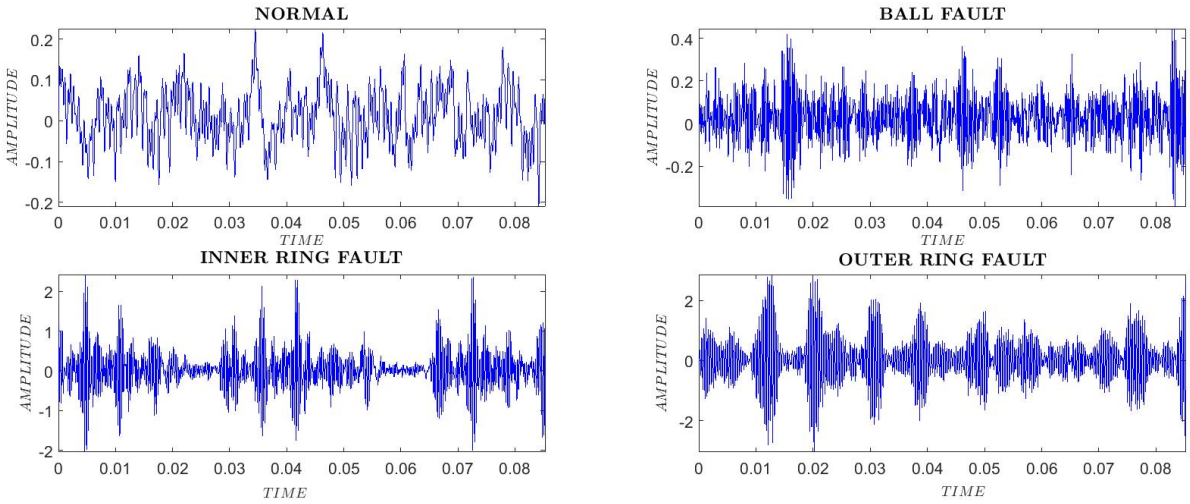


Figure 15. Time signal for different fault conditions

The signal represented in figure 15 are taken from the Bearing Data Center of the Electrical Engineering lab at Case Western Reserve University. This data set will be used for most of the experiments in this thesis since they are commonly used to test bearing faults detection techniques [23]. The signal acquisition set up consists of a 2 hp Reliance Electric motor and the signals were collected using accelerometers, which were attached to the housing with magnetic bases. Motor bearings were seeded with faults using electro-discharge machining (EDM) and the faults were introduced separately at the inner ring, rolling element (i.e. ball) and outer ring. The digital signals were acquired at 12000 samples per second. The experimental set up taken from the Case Western site is shown in figure 16. In table 1, instead,

a summary of the type of signals used is depicted. For further specification on the signals refer to [23].

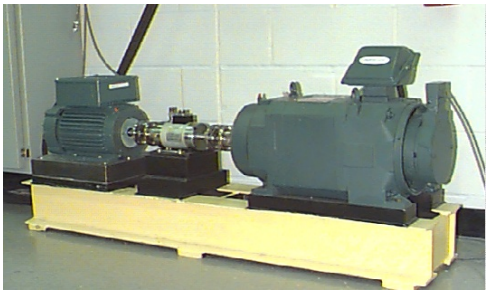


Figure 16. Case western experimental set up [23]

Table 1. Type of signals used

State of the signal	Motor speed (r/min)	State of the signal	Motor speed (r/min)
Normal state	1797	Outer ring fault	1797
	1772		1772
	1750		1750
	1730		1730
Inner ring fault	1797	Ball fault	1797
	1772		1772
	1750		1750
	1730		1730

3.2 Vibration signals reconstruction

In the acquisition of vibration signals of industrial machinery, different factors contribute to the entire signal, such as noise and undesired signal modulations. For this reason, the signal may carry redundant information, for what concern the features necessary to recognize a fault condition. This means that these features components might be sparse in the whole vibration signal or in a certain transform domain.

The important features of the signal can be acquired by means of a learning process that generates a specific basis for the actual signal; this basis captures the underlying structure of the signal and it allows to represent it by its sparse representation only. If the resulting signal is sufficiently sparse in the generated domain then, according to the compressive sensing theory $q \ll n$ samples are sufficient to exactly reconstruct the signal with high probability.

The acquisition and reconstruction procedures, according to the compressive sensing theory explained in Chapter 2, are depicted in figure 17. The vibration signal $Y \in \mathbb{R}^n$ from the machinery is supposed to be sparse on a basis $\Psi \in \mathbb{R}^{n \cdot K}$, being $X \in \mathbb{R}^K$ its sparse representation. This signal is acquired using a compressive sensing approach: by means of a

measurement matrix $\Phi \epsilon \mathbb{R}^{q \cdot n}$ the measurement vector y lies in a low-dimensional space of dimension $q < n$. Then, the acquired measurement can be transmitted and the high-dimensional signal sparse coefficients are reconstructed solving the minimization problem in figure 17. Finally, the high dimensional signal is a linear combination of the atom in the basis Ψ weighted by the corresonding coefficients in the estimated sparse representation vector $\hat{X}$.

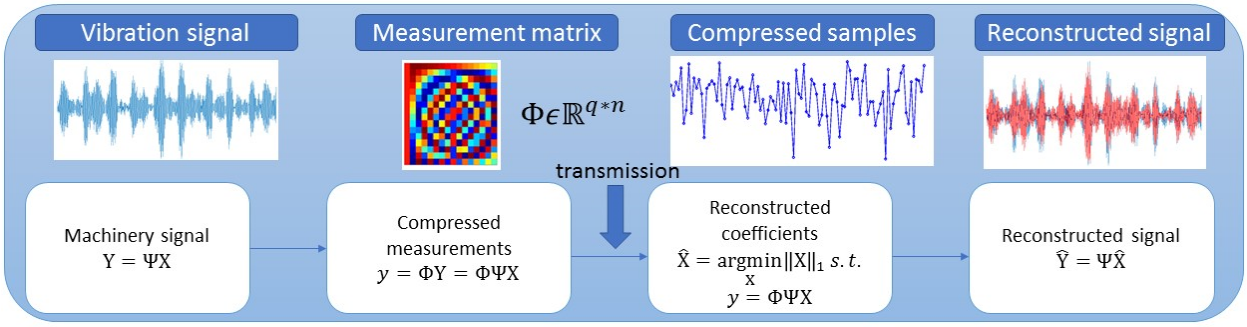


Figure 17. Compressive sensing procedure

A fundamental part of the compressive sensing framework is the availability of a sparsifying basis. For this reason, a dictionary is generated for each state under analysis: normal state, inner ring fault state, outer ring fault state, ball fault state. The dictionary generation procedure is depicted in figure 18.

For each state, the available vibration signal is divided into a set of m overlapping segments $\{Y_j\}_{j=1}^{m} \in \mathbb{R}^n$ with $n = 1024$. As it possible to observe from figure 15, this window size has been chosen because it is sufficient enough to include the signature of each fault state. The resulting matrix $Y \in \mathbb{R}^{n \cdot m}$ with $m = 512$ is used as input for the KSVD algorithm to train the dictionary. The output is a dictionary of size $\Psi \in \mathbb{R}^{n \cdot K}$, where the number of atoms is set to $K = 512$, and the corresponding sparse representation of the training signals: $X \in \mathbb{R}^{K \cdot m}$.

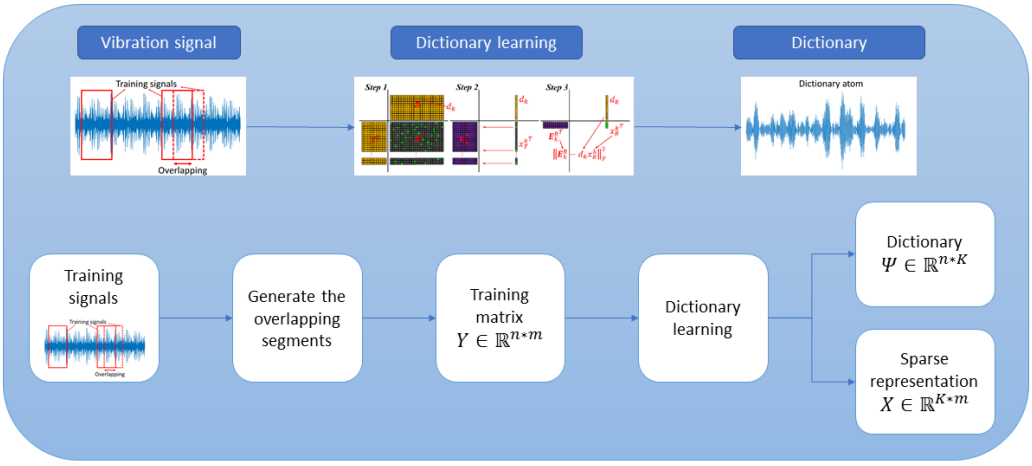


Figure 18. Dictionary learning procedure

In figure 19 one of the generated atoms for each dictionary is compared to a snapshot of the training signals. This figure highlights the characteristic of the atoms for each dictionary. In particular, it is clear how each dictionary has caught the structure of the signal from which it is generated. Moreover, even if the picture illustrates only the dictionary atoms generated using the KSVD algorithm, the previous statement is also true if the MOD algorithm is applied.

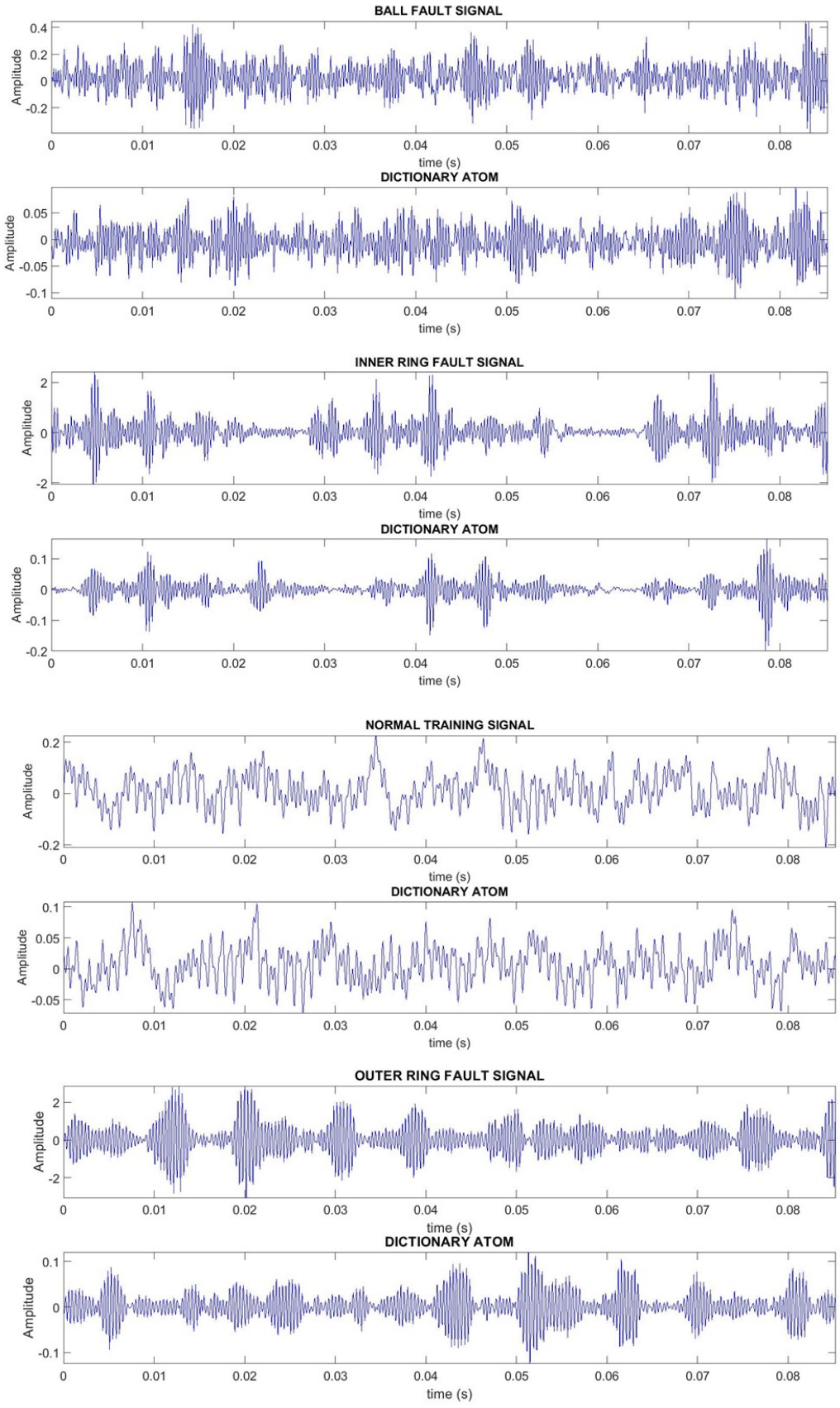


Figure 19. Dictionary atoms for the different type of signals considered.

Once a good dictionary has been acquired, the second step requires testing the reconstruction of the signal from its compressed measurements. These initial results are performed according to the block diagram in figure 17.

The two dictionary learning algorithms described in the previous chapter – MOD and KSVD - are compared according to the quality of the reconstructed signal after applying the compressive sensing framework.

In order to evaluate the accuracy of the reconstruction, in the following of this thesis, the *sample Pearson correlation coefficient* has been used, which is defined as follow:

$$\rho = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \quad (37)$$

where x and y are two vectors of dimensione n and $\bar{x}$ and $\bar{y}$ are their respective sample mean. Moreover, the sensing matrix used in the following is a random matrix of i.i.d. random variables taken from a Gaussian distribution; as explained in chapter 2 this matrix satisfies the RIP and it allows the reconstruction in presence of noise.

It must be highlighted that, in order to evaluate the effectiveness of the reconstruction on the desired class of signal, the compressive sensing algorithm is applied to a signal that does not belong to the training set used for the dictionary learning step.

In figure 20 the effect of the compression ratio is evaluated. In particular, if the number of compressed measurement is q and the original signal consists of n samples, then the *compression ratio* is

$$\eta = \frac{q}{n} \quad (38)$$

The signals are reconstructed using the dictionary that pertains to their class. As expected, when a correct dictionary is used, the quality of the reconstruction increases as the compression ratio increases. However, since the signal is affected by noise a perfect reconstruction is not possible. In fact, there will always be a certain amount of noise, even if it becomes smaller when the number of measurements is sufficiently high.

On the other hand, the lowest number of measurements required for a satisfactory reconstruction of the original signal is determine by the sparsity of the signal on the sparsifying basis.

Figure 20 also highlights that KSVD algorithm results in a better reconstruction compared to the MOD one. That is true for all the signals considered at the expense of a slower dictionary learning procedure. In fact, as showed in table 2 the time needed to train the dictionary with the KSVD algorithm is higher in all the four cases.

Since, in the recognition of the fault states, the reconstruction quality is more important than the difference in computational time present between the two algorithms considered, then in the remaining part of the thesis KSVD will be used.

Table 2. Computational time for dictionary generation

	KSVD	MOD
Normal condition	177.72 s	121.11 s
Ball fault	131.66 s	109.51 s
Inner ring fault	132.82 s	112.23 s
Outer ring fault	167.05 s	152.15 s

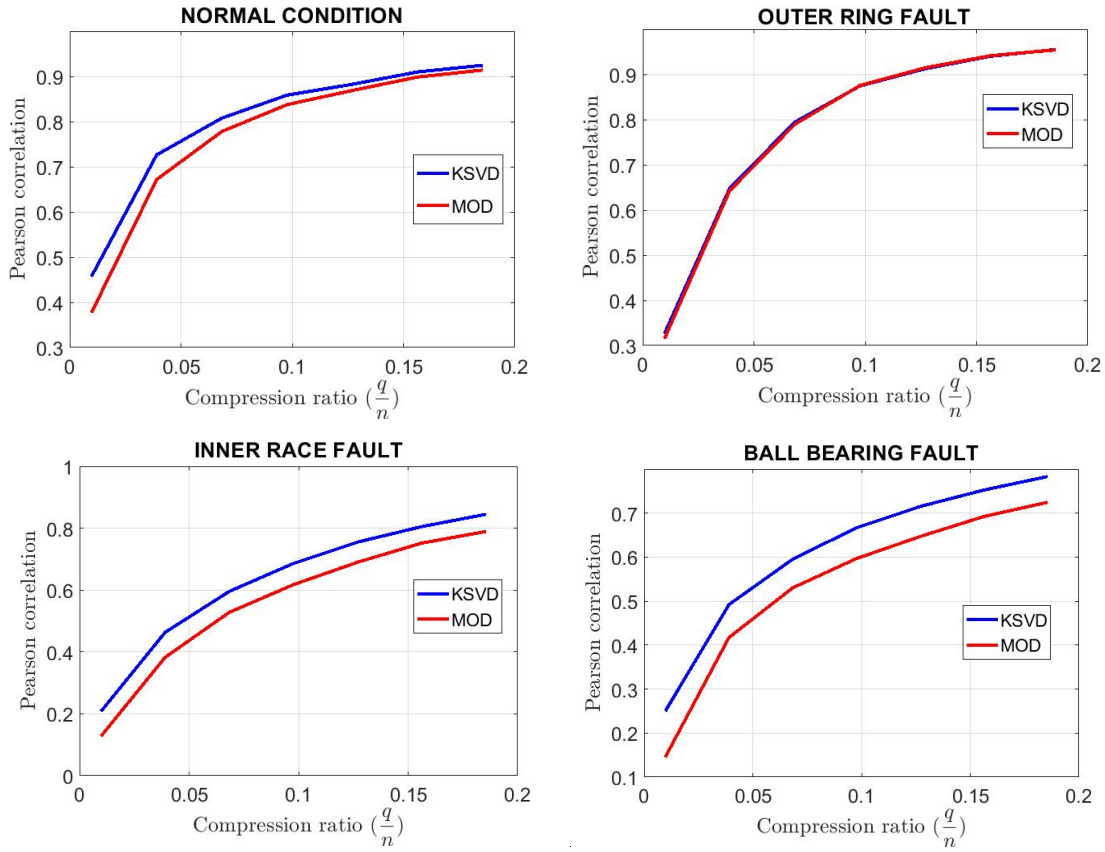


Figure 20. Reconstruction quality at variable compression ratio for the different class considered when two different dictionary learning methods are used

In figure 21, the correlation values for the different states considered are compared when the KSVD algorithm is used to train the dictionary.

According to what has been stated before, the reconstruction is never perfect even if, for what concern the normal condition and the outer ring fault, when $\frac{q}{n} \cong 0.2$ the quality of the reconstruction is very high. On the other, the quality of the reconstruction for the inner race fault signal and the ball fault signal is lower. Then, it is possible to affirm that the outer race fault vibration signal and the normal condition signals have a sparser representation on its

dictionary since, given a certain number of measurements, their performance are better compared to the other signals. This means that, the formers have stronger features that can be easily captured by the dictionary atoms.

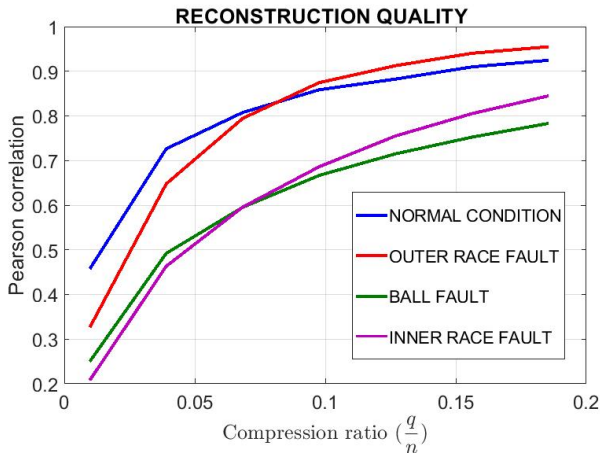


Figure 21. Comparison of the reconstruction quality for the different type of signal considered when the dictionary is trained using KSVD algorithm

Finally, in order to acknowledge the effective reconstruction of the signals considered, in figure 22 a comparison between reconstructed signal and the corresponding original raw vibration signal is showed for Q = 150. Using a compression ratio $\eta \cong 0.15$, the normal condition signal and the outer race fault signal reconstruction has a very high correlation and visually speaking they look almost equal to the original raw signals. On the other hand, the ball fault signal and the inner race fault signal have lower correlation but the reconstruction quality it is sufficient to distinguish the signals and recognize the main characteristics. The correlation values are summarized in table 3.

Table 3. Reconstruction quality for the different class of signal with a compression ratio $\eta \cong 0.15$

SIGNAL	**CORRELATION FOR $\eta \cong 0.15$**
Normal condition	0.9058
Inner ring fault	0.80676
Ball fault	0.75308
Outer ring fault	0.9406

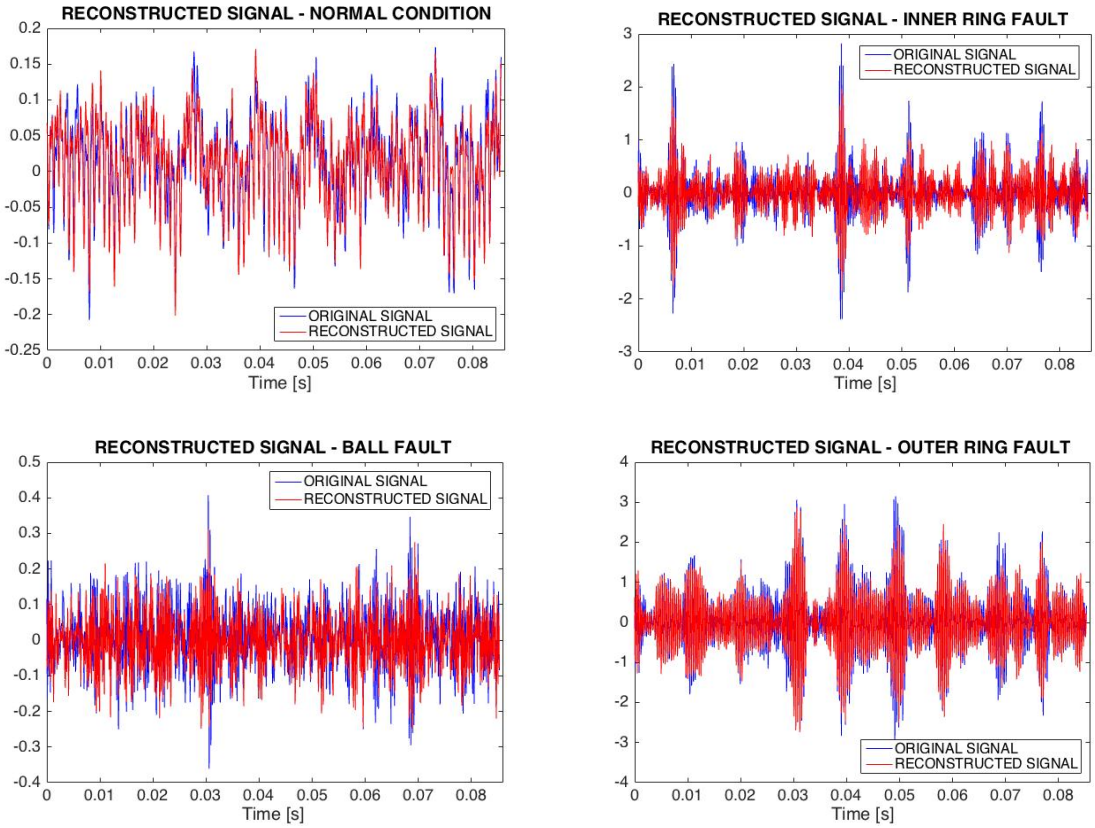


Figure 22. Comparison between the reconstructed signal and the original raw signal for different machine states

4. FAULT DETECTION

The results of the previous chapter show that the signals in a given state can be well reconstructed, from the compressed measurements, when the correct sparsifying basis for that specific state is used. The underlying hypothesis is the sparsity of the analysed class of signal on the trained dictionary. On the other hand, since the dictionary is generated on a specific class of signals, it only encompasses the features of the class considered. This means that, it can only be effective in the sparse decomposition of signals that share the same features as the training one.

Compared to a general dictionary, the trained dictionary improves the quality of the sparsification and consequent reconstruction only on its specific class and, for this reason, a specific dictionary is needed for each class under study. This characteristic can be considered a disadvantage when a fast sparsifying and reconstruction algorithm is needed. In fact, general dictionaries, such as Fourier Transform and Wavelets, are always coupled with fast algorithm and a tuning on the class of signal is not needed. However, in the case of fault condition diagnosis, the uniqueness of the dictionary can be exploited to distinguish the class to which the signal belongs.

In the following of this paragraph, the dictionaries and their sparsification capability will be exploited to recognize the fault states of the bearing based on the vibration signals. [24] [25]

4.1 Reconstruction of a non-sparse signal

In chapter 2, it was highlighted that the signal to be compressed must be sparse in some basis in order to be well-reconstructed from its low-dimension measurements. If the signal is not sparse in the chosen basis, the compressive sensing framework fails and the reconstructed waveform will not match the original one. In this paragraph some simulations are showed in order to better understand what happens when the dictionary is not able to generate a sparse representation of the signal.

First of all, the reconstruction of the signal generated in the normal working condition is analysed. Figure 23 displays the performance when the dictionaries generated from the fault conditions are used compared to the performance when the correct dictionary is used. In particular, the reconstructed waveform when the dictionary generated from the outer race fault signal is used is depicted on the right side of figure 23.

It can be observed that, even if the number of measurements increases, the reconstructed waveform is completely different from the original one. For the signal on the right part of figure 23, considering $\eta \cong 0.15$, the correlation coefficient is $\rho = 0.0311$ when the dictionary is the one for the outer race fault while the correlation is $\rho = 0.9058$ using the

correct dictionary. For this reason, it can be inferred that the signal from a normal working condition presents completely different features compared to the vibration signal acquired when a fault condition arises. As a consequence, the sparsity level of the signal on the given dictionary is not sufficient in order to include its features in the low-dimensional measurements.

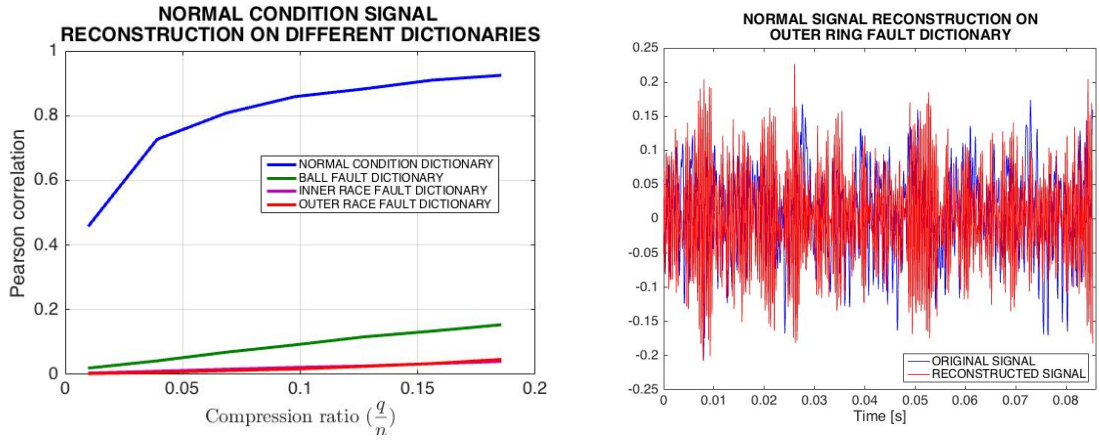


Figure 23. Comparison of the reconstruction quality on different dictionaries for the normal condition signal. On the left: reconstruction performance using the four different dictionaries generated; on the right: Comparison between the reconstructed signal and the original raw signal for a normal condition signal reconstructed using the outer ring fault dictionary with $\eta \cong 0.15$

On the other hand, in figure 24, the performance related to the fault condition signals reconstruction on different dictionaries is related to the one on the correct dictionary for the given state.

Obviously, also in this case, a difference on the quality of the reconstruction and on the reconstructed signal itself can be recognized. In particular, as previously highlighted, this analysis reveals the huge differences between the fault conditions and the normal working conditions and this is true also for the ball bearing, even if in figure 15 the ball bearing appears to be similar to the normal condition signal.

On the contrary, the differences between the three fault conditions considered are not as marked as it was for the normal condition. This behaviour can be ascribed to the presence of some shared feature between the fault conditions that are encompassed on all the dictionaries.

As an example, in figure 25, the outer ring fault signal is considered. On the left side of figure 25, the outer ring fault signal is reconstructed using the normal dictionary. The correlation coefficient is $\rho = 0.0311$ and it is very low compared to its value in table 2, when the correct dictionary is used. On the right side of figure 25, the same signal is reconstructed using the dictionary for an inner fault condition. In this case, the correlation is $\rho = 0.6339$. As explained before, when the dictionary pertains to a different fault condition the correlation difference is not as prominent as before.

Nevertheless, the simulations show a difference in the reconstruction and this means that the sparse representation on the different dictionaries is different and this characteristic may be exploited for the fault detection.

In particular, it is evident that the fault condition signals carry a signature in contrast with the normal condition. For this reason, the recognition of a deviation from the normal behaviour may be simple. On the other hand, the possibility to recognize the different types of fault from the compressed measurements must be better explored.

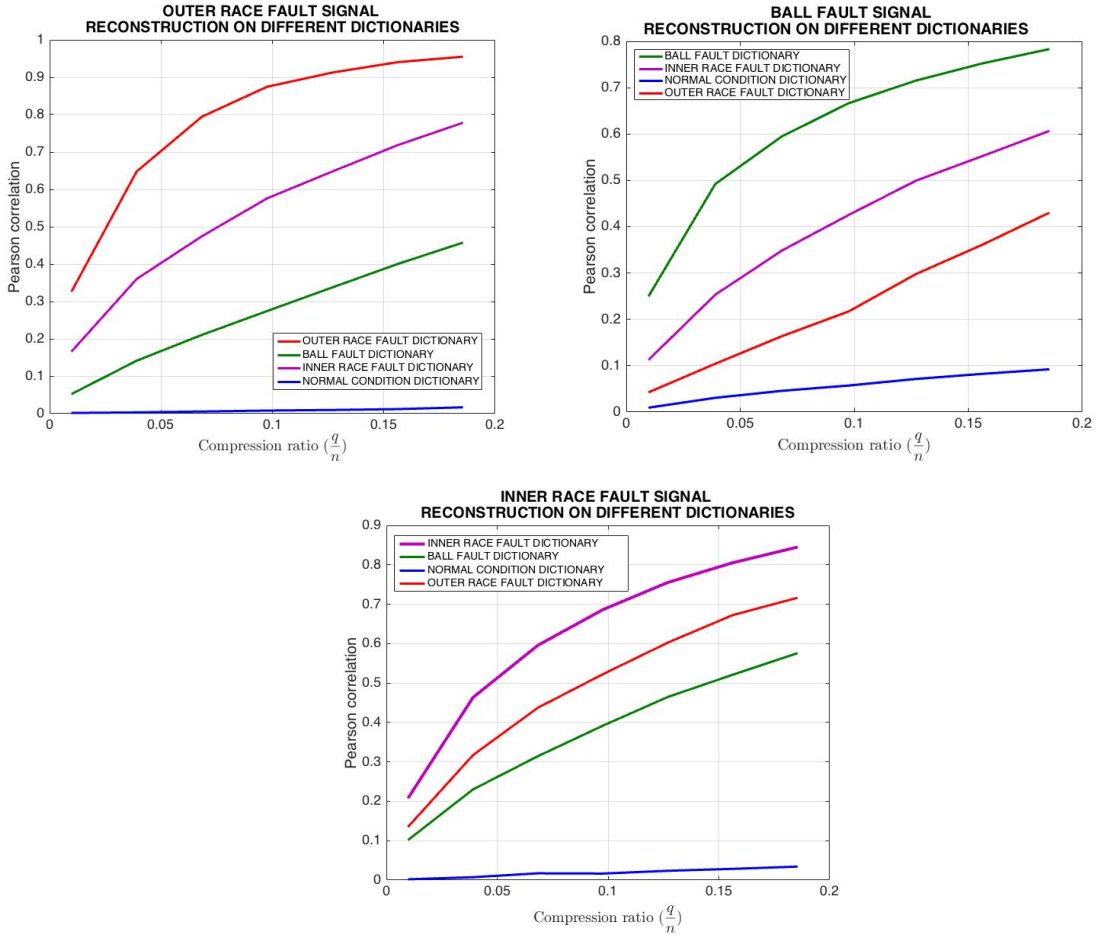


Figure 24. Comparison of the reconstruction quality on different dictionaries for the outer race fault signal, the ball fault signal and the inner race fault signal

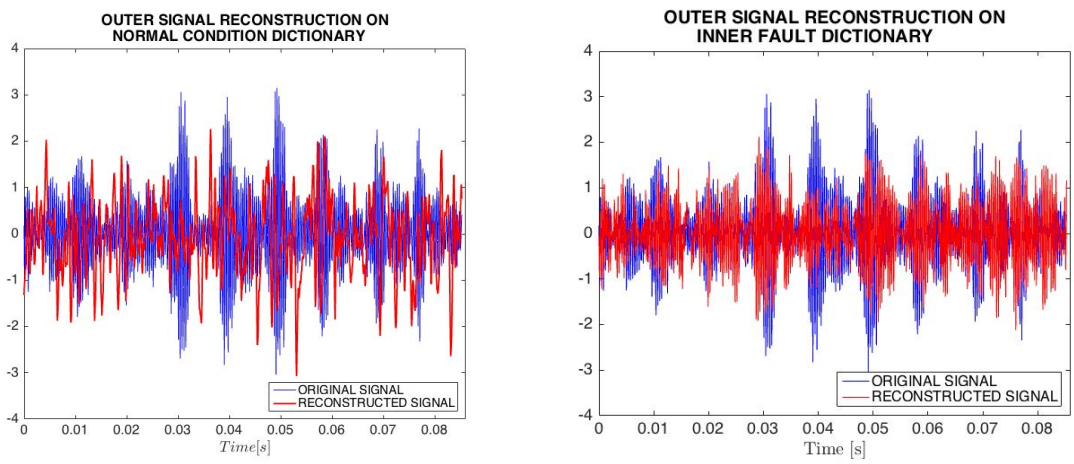


Figure 25. Comparison between the reconstructed signal and the original raw signal for an outer race fault signal with $\eta \cong 0.15$. *On the left: the normal condition dictionary is used; on the right: the inner fault dictionary is used.*

4.2 Fault detection on the compressed measurements

In the previous paragraphs, we have analysed how the compression sensing framework and the dictionaries trained from different class of signals interact. A low-dimensional sensed signal is sufficient to reconstruct, even in presence of noise, the original signal when the correct dictionary is used. On the other hand, the reconstruction differs when the dictionary from a different training states is used. This means that, the trained dictionary might be exploited not only to improve the quality of the reconstruction compared to a general dictionary, but also to recognize the class it pertains to. In this paragraph the detection procedure is first described and then it is tested on the class of faults previously analysed.

4.2.1 Fault detection algorithm

In order to perform a fault detection using the compressive sensing framework and the dictionary learning techniques, first of all a dictionary is trained for each fault condition, using the historical operating vibration data of the bearing. For each class of signals, the training procedure follows the block diagram in figure 18.

The vibration signals from the bearing data centre are divided into two sets: a training set and a test set. In order to generate these sets, the time series of the vibration is divided into overlapping segments of length n=1024. The dictionary is trained with signals from the training set and they are considered to be the historical operating signals. On the other hand, the detection algorithm is analysed on the test sets in order to validate the performance. Moreover, for each class of signal different working condition for different load are taken into account, since this is a typical condition in a factory environment.

In the training procedures a set of m=512 signals are extracted from the training set. These

signals are organized into a matrix $\{Y_j\}_{j=1}^{m} \in \mathbb{R}^n$. Then, the given matrix is used to train the dictionary according to the KSVD algorithm described in chapter 2.

At the end of the training procedure, d dictionaries $\{\Psi_j\}_{j=1}^{d} \in \mathbb{R}^{n*K}$ are available, one for each state considered. In particular, in our experiments 4 dictionaries will be available: normal working condition state (Ψ_1), ball bearing fault state (Ψ_2), inner race fault state (Ψ_3), outer race fault state (Ψ_4).

Once the dictionaries are trained, the compressive sensing framework is applied. The low-dimensional measurements $y \in \mathbb{R}^q$ are acquired from the high-dimensional signals in the test set using the sensing matrix $\Phi \in \mathbb{R}^{qxn}$. The measurements are used to reconstruct the sparse coefficients of the signal on the trained dictionaries Ψ_j for $j = 1 \dots d$. Denoting the reconstructed coefficients on the j-th dictionary as $\hat{x_j}$, than the original signal estimation from the j-th dictionary will be $\hat{Y_j} = \Psi_j \hat{x_j}$.

The fault procedure relies on the different representation values when the selected dictionary is the correct one and when the selected dictionary is the wrong one. In particular, the representation value δ_j for the j-th dictionary is related to the compressed measurements and the reconstructed sparse representation by the following formula:

$$\delta_j = corr\{y, \Phi\Psi_j \hat{x_j}\} \qquad (39)$$

where the correlation used is the one in eq. 37

In this way we can obtain *d* representation errors δ_j for $j = 1 \dots d$ and we can consider obtaining a good representation on the j-th dictionary when the δ_j value is high.

According to the way in which the dictionaries are trained, then, theoretically, the correlation will be maximum on the dictionary correspondent to the signal class, while it will be smaller for the other dictionary. Consequently, it can be inferred that the dictionary with the higher correlation corresponds to the signal class.

Actually, a particular case must be considered: the signal may not pertain to any of the considered class. In order to cope with an unknown state, a threshold on the maximum correlation is set for each class, that is τ_j for $j = 1 \dots d$.

The algorithm depicted in figure 26 implements the fault detection method based on the difference in the representation error.

The algorithm steps are the following:

- The dictionaries Ψ_j for $j = 1 \dots d$ are trained using the high-dimensional signals from the training set;
- A threshold τ_j for $j = 1 \dots d$ for each class of signal is set according to the representation value;
- The measurements $y \in \mathbb{R}^q$ are acquired from a high dimensional signal in the test set using the sensing matrix $\Phi \in \mathbb{R}^{qxn}$;
- The estimate of the signal sparse representation on all the available dictionaries is reconstructed from the compressed measurements based on the compressive sensing

theory; for the j-th dictionary the reconstructed sparse representation will be $\hat{x_j}$;

- The representation value on the different dictionaries is evaluated as $\delta_j = corr\{y, \Phi\Psi_j\hat{x_j}\}$ for $j = 1 \dots d$;
- The maximum correlation $\delta = \max\{\delta_1 \dots \delta_d\}$ is evalauted;
- If $\delta = \delta_j$ and $\delta \geq \tau_j$ then the machine is in state j;
- If $\delta = \delta_j$ and $\delta < \tau_j$ then the machine is in an unknown state.

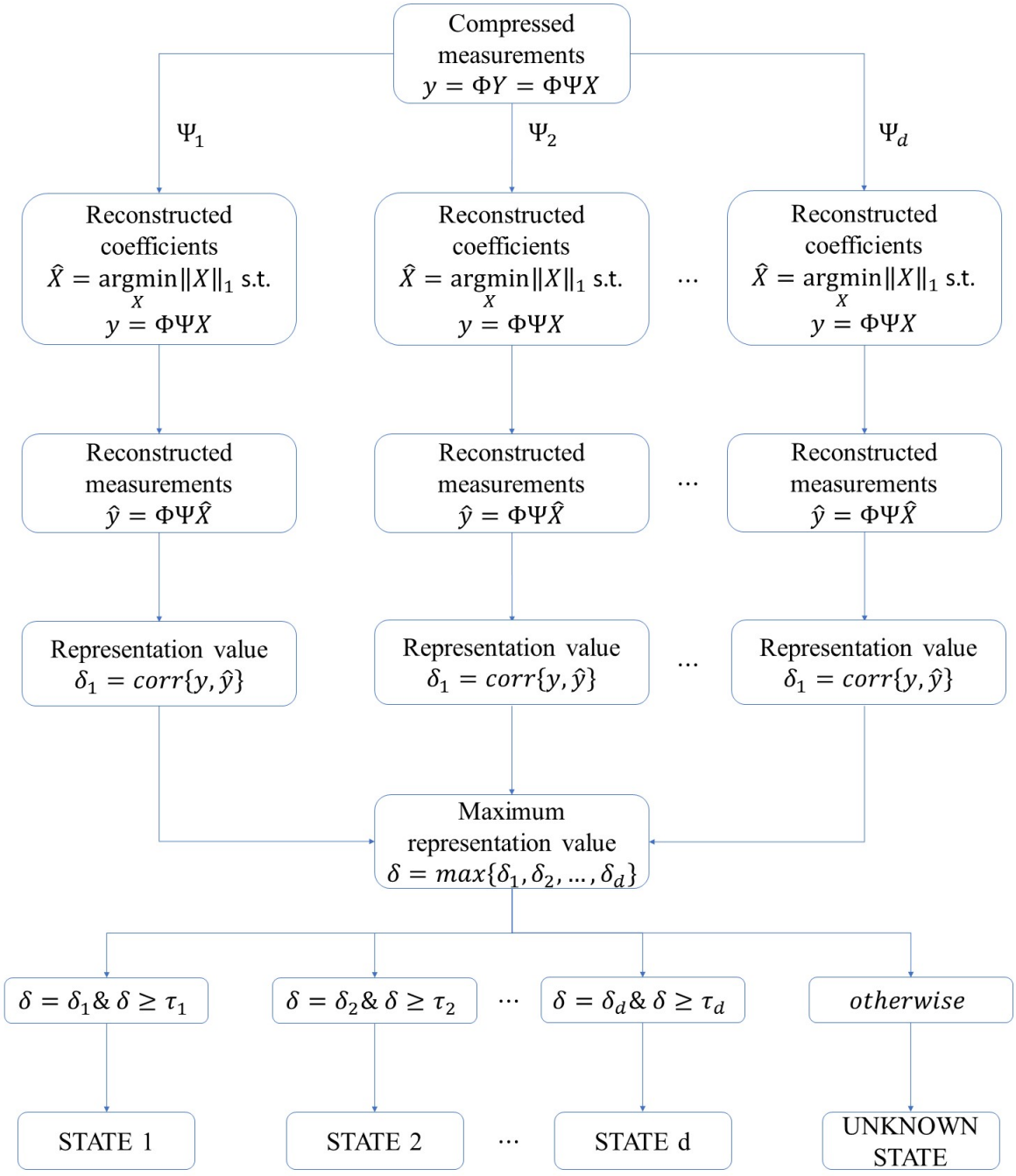


Figure 26. Fault detection algorithm block diagram

4.2.2 Fault detection experiments

In order to evaluate the algorithm depicted in figure 26 the initial study will be performed for a fixed value of number of compressed measurements set to q=90, that means a compression ratio $\eta \cong 0.09$.

The low-dimensional signal $y \in \mathbb{R}^q$ are measured and they are used to reconstruct the signal coefficients $\hat{x}$ on all the dictionaries available. Then, the representation values δ_j on the dictionaries is evaluated.

The results for the different signals considered are shown in figure 27. The top left figure represents the evaluation for the vibration signal related to a normal working condition; the top right represents the evaluation for the vibration signal related to an outer race fault condition; the bottom left represents the evaluation for the vibration signal related to a ball fault condition; the bottom right represents the evaluation for the vibration signal related to an inner race fault condition.

The data can be divided in two groups; the blue group represents the reconstruction quality on the correct dictionary, while the red group represents the reconstruction quality on all the other dictionaries. Ideally, the two groups must be completely separated and the threshold must be set as the division limit between the reconstruction quality on the correct dictionary and the reconstruction quality on the wrong dictionaries. However, in the practical situation the reconstructions values are not perfectly separated, as can be seen in figure 27. For this reason, using the historical data, the following principle will be used in order to set the thresholds: for each state, the error threshold must be less than most of the values of reconstruction correlation on the correct dictionary. In the following tests a value of 95% will be use. This means that the threshold will be set in order to be less then 95% of the blue circles in the figures. As a results, the thresholds values are the following: $\mathrm{tau}_{\mathrm{ball}} = 0.4$, $\mathrm{tau}_{\mathrm{ir}} = 0{,}45$, $\mathrm{tau}_{\mathrm{norm}} = 0{,}7$ and $\mathrm{tau}_{\mathrm{or}} = 0{,}65$.

As considered in the previous paragraph, it is evident in figure 27 that the normal state can be well distinguished when its reconstruction is done with a wrong dictionary. While for the fault conditions, the reconstruction quality using a wrong dictionary may be sometimes confused with the one using the correct dictionary.

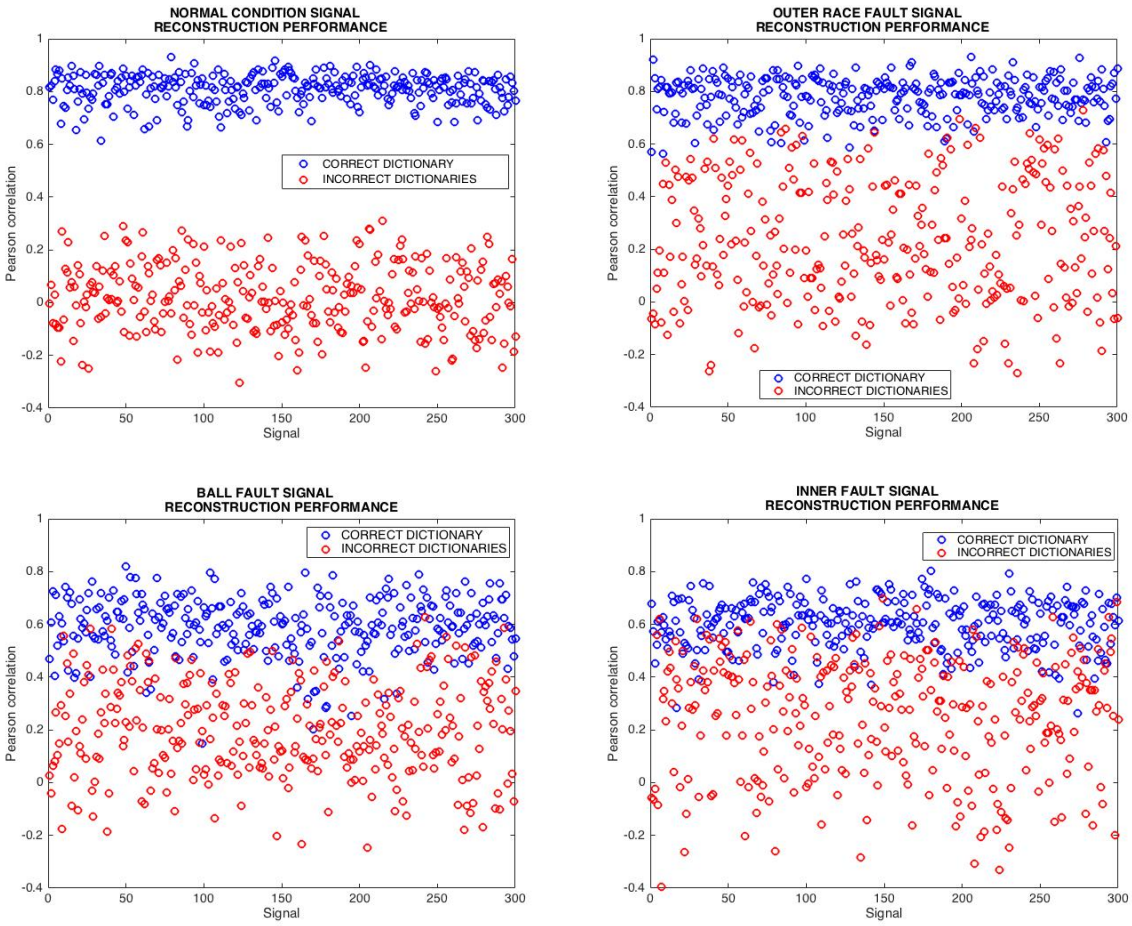


Figure 27. Reconstruction performance for signals from the different classes on all the dictionaries and a compression rate $\eta \cong 0.09$. *Blue label refers to the correct dictionary; red label refers to the incorrect dictionaries. On the top left: normal condition signals; on the top right: outer race fault signals; on the bottom left: ball fault signals; on the bottom right: inner fault signals.*

Once the threshold has been set, then the algorithm is applied to a set of random signals in the test bench. Table 4 shows the recognition rate and the false alarm rate for each class considered.

The signal state is correctly recognized for more of the 85% of the cases for each state, with a compression rate of $\eta \cong 0.09$. This means that even if the transmitted signal is highly compressed, it is still possible to recognize a fault condition and a normal working condition.

The outer race fault presents the strongest features and for this reason is recognition rate is higher compared to the other classes. On the other hand, the inner race fault state has some shared features with the other dictionaries and for this reason sometimes it is difficult to distinguish it from the other classes.

Specifically, most of the incorrect detections lie in the unknown class. According to the algorithm, these values strongly depend on the threshold that has been set based on the historical data. It is possible to affirm that the insertion of the threshold in the algorithm

decreases the detection rate but, on the other hand, it is necessary to take into consideration working condition that differs from the evaluated ones.

Table 4. Fault detection percentage for a compression rate $\eta \cong 0.09$

		PREDICTED CLASS				
ACTUAL CLASS		**Ball fault**	**Inner fault**	**Outer fault**	**Normal**	**Unknown**
	Ball fault	92%	4%	0	0	4%
	Inner fault	0	86%	2%	0	12%
	Outer fault	0	0	96%	0	4%
	Normal	0	0	0	95%	5%

Considering what previously stated, it is important to evaluate the role of the threshold. Using the same parameter as before, in figure 28, the evaluation of the recognition rate to each state is depicted, for various threshold values $\mathrm{tau}_{\mathrm{norm}}$, $\mathrm{tau}_{\mathrm{or}}$, $\mathrm{tau}_{\mathrm{ball}}$ and $\mathrm{tau}_{\mathrm{ir}}$.
It is evident that, a lower value of the threshold will allow a higher detection rate on the considered state. However, a drawback must be considered. In fact, at the same time, the misjudgement rate will increase. This means that also the amount of wrong fault detection will increase when the threshold is lowered.

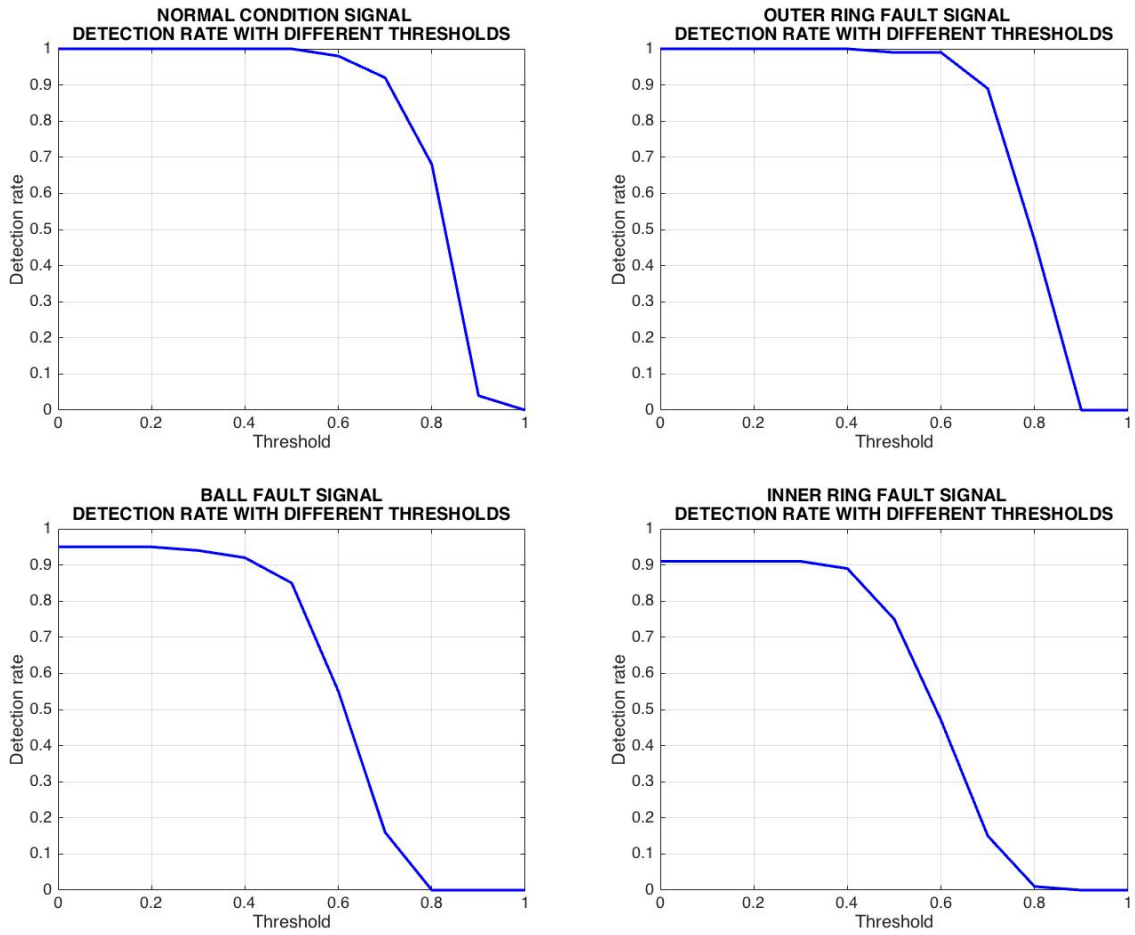


Figure 28. Detection rate of the different class of signals for variable threshold. Top left: normal condition signal; top right: outer ring fault signal; bottom left: ball fault signal; bottom right: inner ring fault signal.

In order to evaluate the effect of the threshold on the detection of an unknown state, one of the given states will not be considered known and the corresponding dictionary will not be used. Then, the algorithm will be applied on signals from the unknown class considered. Obviously, to have a correct detection, the algorithm must recognize the unknown state for all the signals studied.

As an example, the ball fault signals are supposed to be unknown and the detection algorithm uses only the following dictionaries: the normal condition dictionary, the outer ring fault dictionary, and the inner ring fault condition. Then the misjudgement rate is evaluated varying the threshold for the inner ring fault $\mathrm{tau}_{\mathrm{ir}}$ and keeping the threshold for the normal condition $\mathrm{tau}_{\mathrm{norm}}$ and the threshold for the ball fault $\mathrm{tau}_{\mathrm{ball}}$ fixed.

Following the procedure described, in figure 29 the misjudgement rate for the ball fault signals is showed. It is evident that a very small threshold will not be able to recognize an unknown state. In fact, the algorithm will consider the unknown state to be one of the known ones, leading to a wrong decision. On the other hand, as depicted in figure 28 the detection rate will be very high for a lower value of the threshold. Thus, it is obvious that a trade-off

must be taken into account in setting the threshold in order to have a good recognition rate while keeping the misjudgement rate low. For what concern the results previously shown, the threshold was set to $\mathrm{tau}_{\mathrm{ball}} = 0.4$ reaching a detection rate of 92% with a misjudgement rate of 30%.

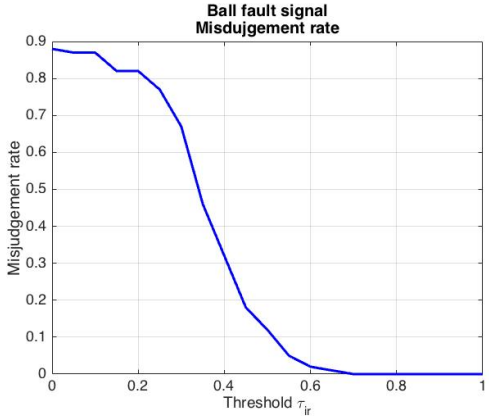


Figure 29. Misjudgement rate of the ball fault signal for variable threshold values τ_{ir}

An important point is the strong dependence of the fault detection rate on the number of compressed measurements acquired. Obviously, the aim of this method is to reduce the burden on the data transmission; for this reason, it is desirable that the number of measurements will be as lower as possible. However, the information conveyed will be decreased when the compression ratio increases and the fault detection may fail.

First of all, the thresholds are set according to the same principle used before: in the following test the 95% rule will be followed. Then, considering the same values as before for the high-dimensional signal and for the dictionary atoms, the recognition rate is evaluated when the compression ratio increases. In figure 30 the detection rate of the different class of signals for variable compression ratio is depicted.

As expected, the detection rate exhibits a general increasing trend when the number of compressed measurements acquired increases. In fact, more measurements will correspond to an increase in the information acquired on the machinery state. This is particularly true for the states that are more difficult to recognize; that is the inner race fault and the ball fault state. On the other hand, the normal working state exhibits a good detection rate even with few acquired measurements. In fact, as we have previously highlighted, the normal condition signal displays sharper difference with respect to the fault conditions. It is also true that the outer ring fault signals have strong features compared to the other fault states in fact, in some cases, its detection rate is even higher than the normal working state.

Considering the experiments conducted, it is possible to affirm that the detection algorithm based on the trained dictionaries allows to achieve a good recognition rate for the fault conditions while decreases the amount of measurements transmitted.

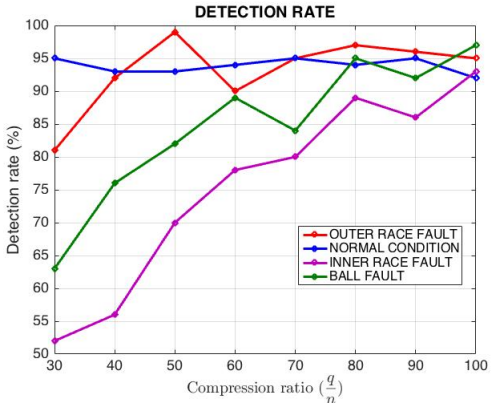


Figure 30. Detection rate of the different class of signals for variable compression ratio

5. ONLINE FAULT DETECTION

The recognition of the type of fault that is affecting the bearing is based on a-priory training signals that allow generating the dictionaries for each fault condition. This can be useful for known fault conditions for which signals are available in a database and they can be used offline. On the other hand, it can happen that the fault condition is not known and it is necessary to recognize any deviation from the normal condition that may occur.
The previous analysis shows that there is a difference in the reconstruction quality for the fault signals when a wrong dictionary is used. Then, the dictionary itself may be used at the edge to evaluate any change in the state of the machine.

5.1 Online fault detection system

The system considered is the one depicted in figure 3. The edge terminal and the cloud are supposed to work together in order to monitor the machinery; at the same time, the transmission burden and the computation burden must be reduced.
In the previous chapters, the compressive sensing technique has been applied and the experimental results show that the signal can be reconstructed, either at the edge or at the cloud, with a high reconstruction quality. Then, the signal can be used to extract the statistical features necessary to detect the fault condition.
The compressive sensing framework allows reducing the amount of measurements acquired at the sensors terminal but the computational burden for the features extraction cannot be reduced. Thus, a fault detection algorithm based on the trained dictionary has been examined.
The results presented in the previous chapters reveal that different dictionaries for different fault conditions can discriminate between the machinery states and reveal the presence of unknown state. However, historical data are needed for the fault conditions in order to train the dictionaries.
Following a principle similar to the one exploited for fault recognition, an online fault detection algorithm at the edge terminal can be envisioned. In this way, the data are previously analysed at the edge in order to detect a change in the machine state; then cloud detection is performed only if the preliminary analysis at the edge reveals an anomaly. This procedure permits to reduce the vibration data transmitted at the cloud, since they are not continuously sent, and it permits to transfer part of the complexity for the fault detection at the edge.
A duplex transmission link is needed: the data flows from the sensor terminal to edge, while some request an control information are sent from the edge terminal to the sensor one. In figure 31 b, a time diagram of the transmitted data between edge terminal and sensor terminal

is depicted.

At the beginning the edge terminal sends a training request to the sensor terminals. The sensors terminal will then send back the raw signals in order to train the dictionary for the current state. After the dictionary has been training, the edge terminal will request to the sensor terminal the transmission of the low-dimensional signals according to the compressive sensing paradigm. Then, the sensors will send only the compressed measurements until when a change of the state is detected at the edge terminal. At that point, the dictionary must be trained again for the new state; thus, the edge sends to the sensors a training request and the sensors will reply with new training signals for dictionary learning. Once the new dictionary has been acquired, the edge will be ready to receive the new compressed measurements.

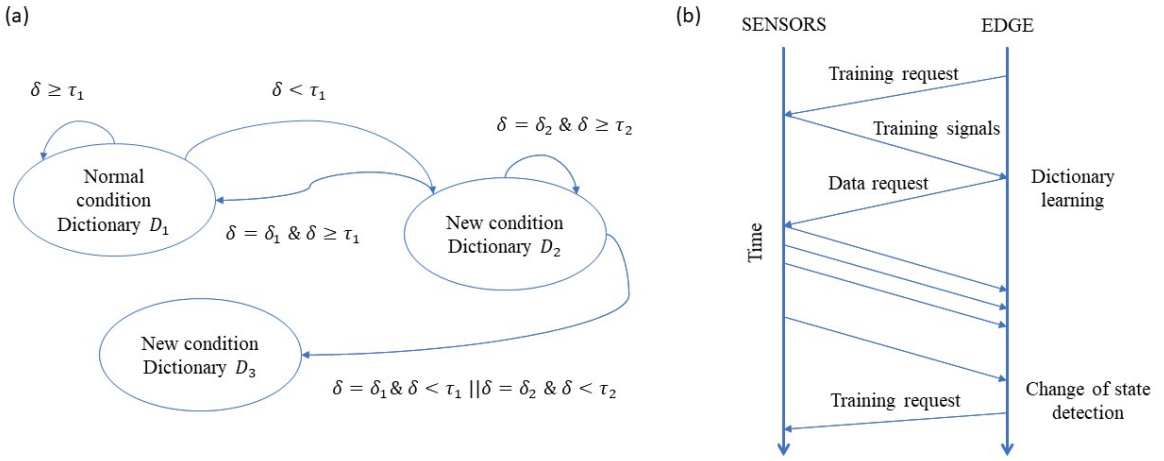


Figure 31. Intelligent online fault detection at the edge terminal. (a) fault detection algorithm; (b) time evolution of the data exchanged between edge terminal and sensors terminal

In figure 32, the data frame is shown according to the change of the state. During the training period, the raw signals are sent to the edge terminal. When the training period ends, the low-dimension measurements are sent according to the compressive sensing mechanism. When a change of the state is detected, the raw signals will be sent for state B and the compressed measurements will be transmitted right after the training phase.

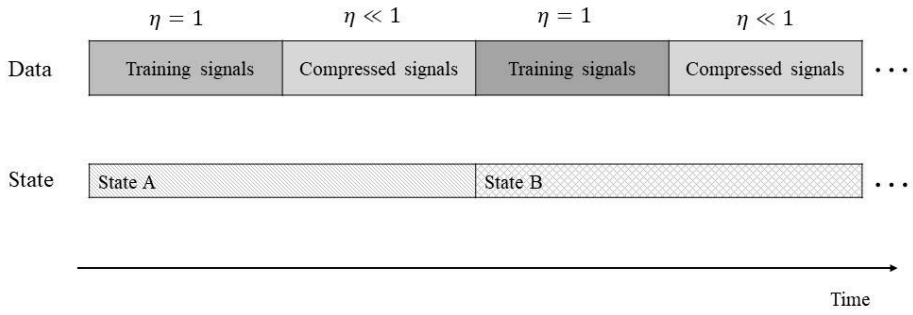


Figure 32. Transmitted data format

5.2 Online fault detection algorithm and experiments

The most important part of the system previously described is the definition of the change of state detection method. The online detection algorithm is based on the algorithm in figure 26. This means that, the detection will be based on the following value: $\delta = \text{corr}\{y, \boldsymbol{\Phi}\Psi\hat{x}\}$; where Ψ represents the current dictionary and $\hat{x}$ represents the sparse estimation on the current dictionary. However at the beginning only the current state dictionary will be available and the normal state will be considered as starting state of the system.

The quality of the reconstruction based on the correlation will be then compared to a threshold τ_1 that is evaluated based on the previous signals from the current state. If the correlation goes below the given threshold for few consecutive times then the machine is considered in a new state: a new dictionary will be trained and a new threshold τ_2 will be evaluated. At this point, two dictionaries are available so, the detection procedure will be performed considering $\delta = \max\{\delta_1, \delta_2\}$. The procedure will then be similar to the one in figure 26: the system will be consider in state 1 or state 2 according to the maximum correlation value δ and to the thresholds calculated, if a unknown state is determine than a new dictionary is evaluated and the procedure continues.

The algorithm described is depicted in the state diagram in figure 31 a, while figure 33 shows some experiments with different fault conditions.

The experiments are conducted using raw signals of dimension $n = 1024$ and a compression ratio $\eta \cong 0.09$; this means that during the training phase the entire signals is transmitted, while, during the compressed phase, the transmitted signals have length $q = 90$.

First, all the different states are considered. Starting from the normal condition the machine is supposed to change from one state to the other. Then, the algorithm is applied in order to follow the signals change. The results are shown on the top left part of figure 33.

Once the dictionary is trained and the threshold τ_{norm} is set up, the machine state is evaluated using the correlation. The machine is supposed to change state only when the

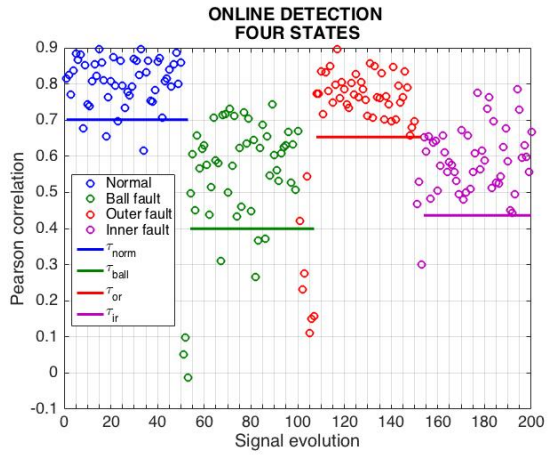
correlation is lower than τ_{norm} for three consecutive signals. Then the new dictionary is trained, the new threshold τ_{ball} is evaluated and the machine is considered to be in the new state. The same procedure is repeated for the other states.

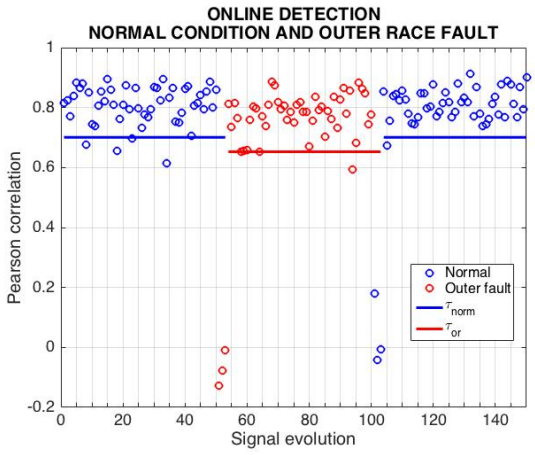
According to figure 33, the change of state is correctly detected when the signals changes from the first state to the second one and from the third state to the fourth one. On the other hand, the change from the second state to the third one is correctly detected only after few signals are considered to be in the wrong state. However, this delay can be accepted since few signals correspond to few seconds in the detection. It is important to underline that the correlation values may pass the threshold for some signals but they do not cause a change in the dictionary since a chain of bad reconstruction is needed.

In the second part of the experiment, the online detection algorithm is used to determine a deviation from the normal working condition. As explained in the previous chapters, the normal state has very peculiar characteristics compared to the other signals; thus, the evaluation of a different working behaviour from the normal one may be simpler. Figure 33, on the top right, bottom left and bottom right corners shows what happens when the fault condition appears while the machine is in the normal state.

The three fault conditions are correctly evaluated and then the machine is supposed to return in a normal working condition after the maintenance is performed. The transition seems to be correctly detected in all the cases. The only difficulty is encounter when the ball fault is recovered. However, as in the previous case, the effect is a small delay in the detection that can be accepted.

Finally, the usage of pilot signal, inserted during the compressed data transmission, can be contemplated to better estimate the condition of the machine on the entire raw signal.





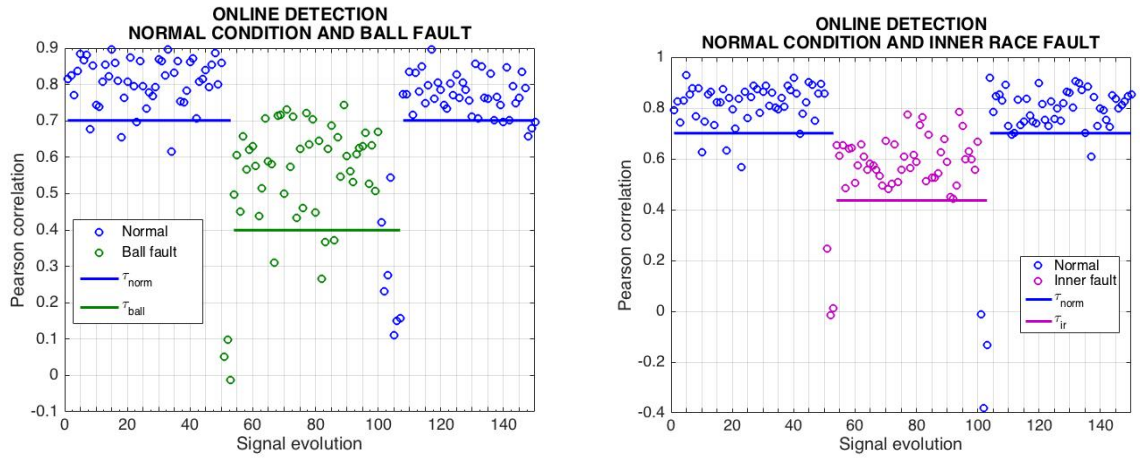


Figure 33. Online fault detection experiments. Top left: four states machine; top right: detection of an outer race fault from a normal condition; bottom left: detection of an ball fault from a normal condition; bottom right: detection of an inner race fault from a normal condition

6. IWAMIZAWA FACTORY

The previous chapters presented the simulation results for a bearing fault diagnosis method based on compression sensing theory and dictionary learning technique. In order to actually implement this method in a real environment, a preliminary study has been performed on real signals acquired from compressors in Hitachi Power's factory in Iwamizawa.
The aims of this preliminary study are the following:

- Confirm the field of the environment;
- Confirm the time required for dictionary learning;
- Confirm the reconstruction quality and the compression ratio required.

6.1 Experimental setup

The vibration signals considered are acquired from two compressors systems of the model Hitachi HiScrew11 used inside Iwamizawa factory.
In order to acquire the vibration signals two accelerometers have been positioned inside the compressor: the first one on the motor of the compressor and the second one on the pump of the compressor. Figure 34 shows the position of the sensors inside the compressor on the left and the experiment environment on the right.

Figure 34. Compressor and sensors setting

The sensors used are 1-axis sensors made in Hitachi, which specifications are described in table 5

Table 5. Sensor specifications

Size	20 x 20 x 5 mm
Weight	2.8 g
Substrate	glass epoxy (t=0.9)
Shield	Bras(C2801P, t=0.3) w/ tin plating
Output	Analog, BNC-connector
Resonance frequency	90 KHz

The signal is acquired using the Picoscope 4224 oscilloscope, that offers a high resolution of 12 bits and 20 MHz bandwidth, and then filtered by an EF122 low-pass filter with a -3dB cutting frequency of 20 KHz.

In figure 35, a general diagram of the experimental set-up is displayed. While, figure 36 represents the real-time acquired signals: the red line is the signal from the motor and the blue line is the signal form the pump.

Figure 35. Experimental setup

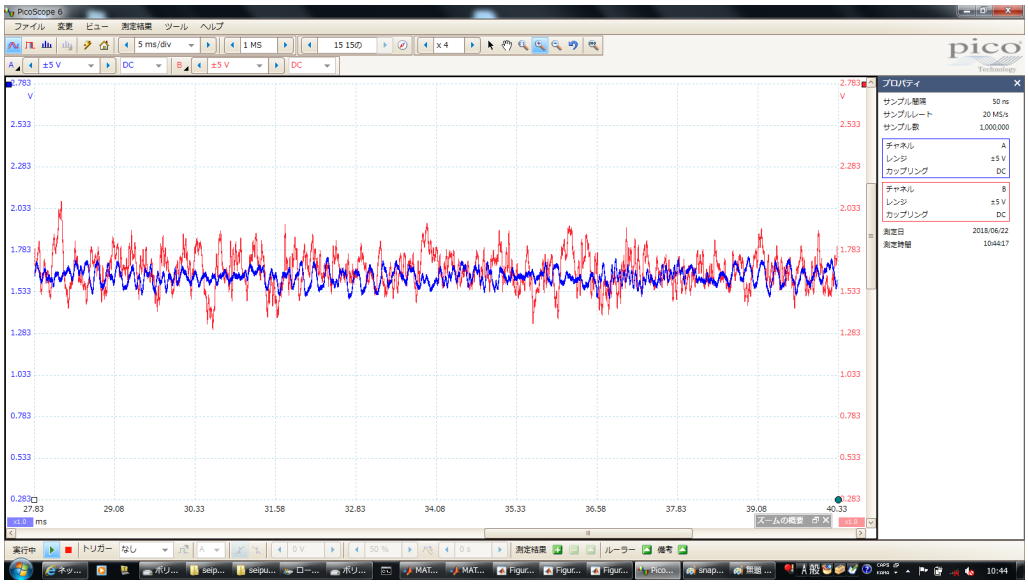


Figure 36. Real-time acquired signals. Blue line: motor sensor; red line: pump sensor

In this preliminary experiment the fault conditions are not generated so the acquired signals are related to the normal working condition of the air compressor with variable load.
In total 4 different time series have been acquired:

- Motor sensor signal from Compressor 1;
- Pump sensor signal from Compressor 1;
- Motor sensor signal from Compressor 2;
- Pump sensor signal from Compressor 2;

6.2 Preliminary study

During the acquisition process a dictionary is generated in order to verify the reconstruction performance using low-dimension measurements. The high-dimensional signal is first acquired in order to allow further analyses; then, from the high dimensional sampled signal, low-dimensional measurements are generated from which the signal is reconstructed using the trained dictionary.
The spectra of the acquired signals are represented in figure 37. The vibration signals of the different compressors have the same spectrum components, even if their power spectrum densities differ. This means that if the dictionary is acquired for a given compressor then the reconstruction quality will differ for a signal acquired from the second compressor, but the features of the signals may be similar so the dictionary should be able to sparsify signals from different compressors.

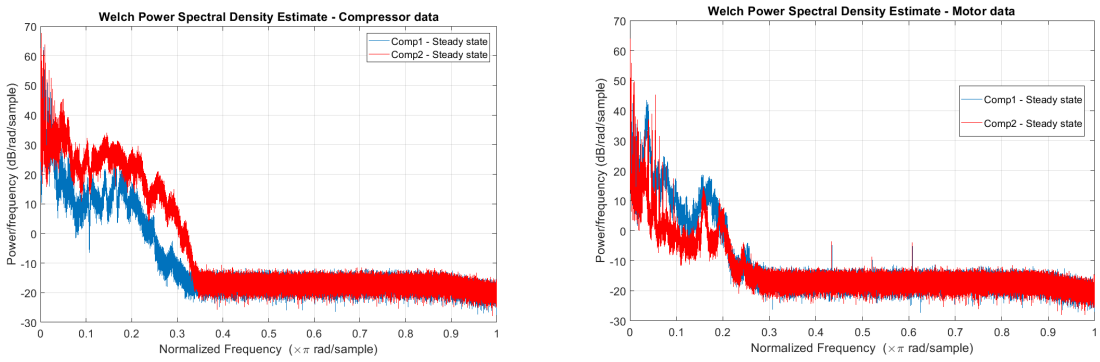


Figure 37. Acquired signals spectrum

The signals acquired are used as input for the KSVD algorithm in order to train the dictionaries. In this way, the time generation of the dictionaries is tested online and it appears to be similar to the one showed in table 2. Since the change of state of the machine is not fast, it is possible to confirm the feasibility of the application of this technology in a real environment.

Then, the compressive sensing framework is applied to evaluate the reconstruction quality for the acquired signals. The results are shown in figure 38.

Comparing this graph with the one in figure 21, the performance for the acquired signals seems to be better. In fact, at least for what concern the signals from the motor, with a compression ratio $\eta = 0.06$, the correlation coefficient already reaches a value greater than 0.90. On the other hand, the performance from the pump signals are not as good as the one related to the motor signals. This means that the position of the sensors makes a difference in the quality of the signal considered and for this reason only the signals from the motor are consider in this work.

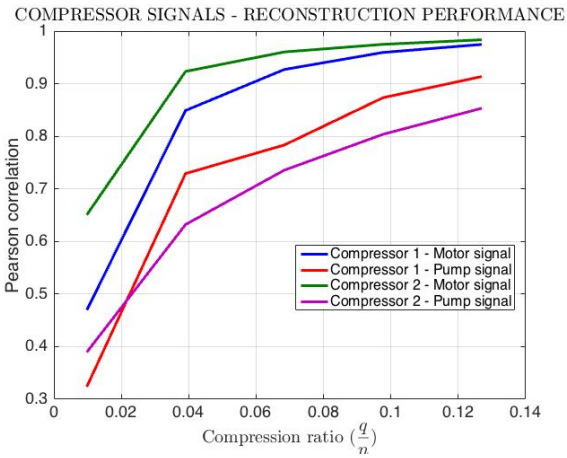


Figure 38. Reconstruction performance of the compressive sensing framework applied to the signals from Iwamizawa factory

The availability of two different machines permits to better investigate the capacity of the dictionary to reconstruct signal that pertain to a similar class; that is, the signal taken from the

motor from another compressor. For this purpose, some of the signal from compressor 1 has been reconstructed using a dictionary trained with signals from dictionary 2 and vice versa; the results are shown in figure 39.

As can be seen, the reconstruction performances are degraded when the dictionary is trained with signals acquired from another machine. However, the class of signal consider is still the same and, for this reason, the signal still can be reconstructed correctly but with a higher compression ratio.

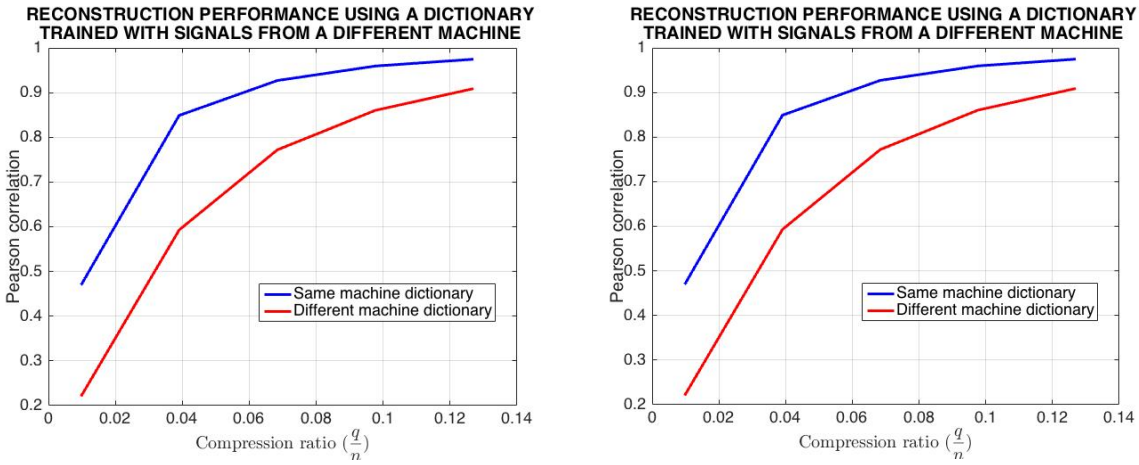


Figure 39. Reconstruction performance when the dictionary trained with signals from another machine is used. Left: Signal from compressor 1 and dictionary from compressor 2; right: Signal from compressor 1 and dictionary from compressor 2.

This final analysis reveals that, not only the burden on the transmitted data can be actually decrease but also the training complexity can be reduced acquiring only one dictionary for the same type of machine.

7. CONCLUSIONS

Traditional methods for condition monitoring and fault detection present two main problems that must be faced in many factories. First of all, the computation is complex and it is mostly implemented at the cloud causing latency; secondly, the amount of data necessary is high and it generates a big traffic on the network.

In this thesis, an intelligent IoT system based on edge computing has been considered for condition monitoring and fault detection of rotating machinery. The system implements a fault detection method supported by the compressive sensing theory and making use of trained dictionaries, in order to solve the problems of traditional methods.

As the analysis shows, the compressive sensing framework represents a useful solution to compress the amount of data transmitted on the network while keeping the features of the signal intact. On the other hand, the trained dictionary can be used to discriminate between the different working conditions of the machine.

It was shown that, the availability of a dictionary for each fault condition, trained with historical data, allows detecting the fault condition with high probability while using only the low-dimensional measurements received from the sensors. Moreover, setting a threshold based on the data can also allow detecting the presence of unknown fault conditions.

This type of detection is greatly influenced by the amount of measurements acquired since a trade-off is needed between the reconstruction quality and the transmission burden. However, it seems to be effective in the case considered and the performances improve when the compression ratio is higher.

The same principle can be also used to consider a first online detection at the edge terminal, to recognize a change in the machinery state. The simulations performed show that the data can be used to set a threshold on the quality of the reconstruction. Since the quality decreases when the signals features change, this can be considered as a sign that something is happening. Thus, this consideration permits to avoid a continuous stream of data to the cloud while sending only valuable information when needed.

It is evident that one of the most important parts of the considered algorithm is the dictionary itself. The capability of the dictionary to sparsify the signals must be high in order to improve the detection performance. Moreover, the compressive sensing framework includes many parameters that can be modified and a further analysis must be performed on the types of sensing matrices that can be used.

Finally, it is possible to say that the techniques considered seem to be promising for fault detection and future study may consider the possibility of using traditional artificial intelligence methods directly on the compressed measurements.

TABLE OF ABBREVIATIONS

ANN	Artificial Neural Networks
CS	Compressed Sensing
DCT	Discrete Cosine Transform
i.i.d.	Independent and Identically Distributed
IoT	Internet of Things
K-SVD	K-Singular Value Decomposition
LASSO	Least Absolute Shrinkage and selection operator
MOD	Method of Optimal Directions
MP	Matching Pursuit
NSP	Null Space Property
OMP	Orthogonal Matching Pursuit
PCA	Principal Component Analysis
RIP	Restricted isometry property
SVD	Singular Value Decomposition
SVM	Support Vector Machine

BIBLIOGRAPHY

[1] Fluke Corporation, "www.euro-index.nl," [Online].
Available:
https://www.euro-index.nl/media/wysiwyg/Documenten/NL/Application-notes/Fluke_The_basics_of_predictive_or_preventive_maintenance_AN_EN.pdf.

[2] E. Candès and M. Wakin, "An introduction to compressive sampling".

[3] E. Candès, "Compressive sampling".

[4] E. Candès, J. Romberg and T. Tao, "Robust uncertainty principles: exact signal reconstruction from higly incomplete frequency information".

[5] E. Candès and T. Tao, "Near optimal signal recovery from random projections: universal encoding strategy?".

[6] J. Romberg, E. Candès and T. Tao, "Stable signal recovery from incomplete and inaccurate
measurements".

[7] D. Donoho, "Compressed sensing".

[8] J. Romberg and M. Wakin, "Compressed sensing: a tutorial".

[9] R. Baraniuk, J. Romberg and M. Wakin, "Tutorial on compressive sensing".

[10] Y. Eldar and G. Kutyniok, Compressed sensing: theory and application.

[11] D. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via l1 minimization".

[12] E. Candès and T. Tao, "Decoding by linear programming".

[13] J. Tropp, "Greed is good: algorithmic results for sparse approximation".

[14] J. Tropp and S. Wright, "Computational method for saprse solution of linear inverse problems".

[15] Y. Pati, R. Rezaiifar and P. Krishnaprasad, "Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition".

[16] R. Rubinstein, A. Bruckstein and M. Elad, "Dictionaries for sparse representation modeling".

[17] I. Tošić and P. Frossard, "Dictionary learning".

[18] K. Engan, S. Aase and J. Hakon-Husoy, "Method of optimal direction for frame design".

[19] M. Aharon, M. Elad and A. Bruckstein, "K-SVD: an algorithm for designing of overcomplete dictionaries for sparse representation".

[20] T. Han, D. Jiang, X. Zhang and Y. Sun, "Intelligent diagnosis method for rotating machinery
using dictionary learning and singular value decomposition".

[21] T. Harris, Rolling bearing analysis, 2001.

[22] SFK, "www.sfk.com," SKF, [Online]. Available: www.sfk.com.

[23] Case Western University, "Western Reserve University Bearing Data Center Website," [Online]. Available: https://csegroups.case.edu/bearingdatacenter/pages/welcome-case-western-reserve-university-bearing-data-center-website.

[24] X. Zhang, N. Hu, H. Lei and L. Chen, "A bearing fault diagnosis method based on sparse decomposition theory".

[25] X. Zhang, N.-q. Hu, H. Lei and L. Chen, "A bearing fault diagnosis method based on the low-dimesional compressed vibration signal".

[26] J. Sun, C. Yan and J. Wen, "Intelligent bearing fault diagnosis method combining compressed data acquisition and deep learning".