COMILLAS
UNIVERSIDAD PONTIFICIA
ICAI

Degree in Industrial Technologies

Bachelor's final project

Compressive Sensing to predict the health of an aircraft's engine

Author
Alvaro Burgos

Supervised by
Gilles Chardon

Gif-sur-Yvette and Madrid
June 2020

Declaro, bajo mi responsabilidad, que el Proyecto presentado con el título
Compressive Sensing to predict the health of an aircraft'sengine
en la ETS de Ingeniería - ICAI de la Universidad Pontificia Comillas en el
curso académico 2019/2020 es de mi autoría, original e inédito y
no ha sido presentado con anterioridad a otros efectos. El Proyecto no es
plagio de otro, ni total ni parcialmente y la información que ha sido tomada
de otros documentos está debidamente referenciada.

Fdo.: Álvaro Burgos Madrigal Fecha: 30/06/2020

Autorizada la entrega del proyecto

EL DIRECTOR DEL PROYECTO

Fdo.: Gilles Chardon Fecha: 03/07/2020

Thank you

Thank you to my parents, to my professor Gilles Chardon and to an indispensable colleague through the realisation of the project, Nicolas.

COMPRESSIVE SENSING PARA PREDECIR LA SALUD DE UN MOTOR DE AVIÓN

Autor: Burgos Madrigal, Álvaro.

Director: Chardon, Gilles

Entidad Colaboradora: Safran Tech

RESUMEN DEL PROYECTO

Cuando una aerolínea firma un contrato de explotación de un avión con una empresa fabricante de aviones, el riesgo de que el avión sufra una avería debe asumirlo la empresa fabricante y no la aerolínea, dado que son ellos quienes han fabricado la aeronave. Es por ello por lo que las aerolíneas, en lugar de alquilar un avión durante un periodo de tiempo, están alquilando horas de vuelo aseguradas por la empresa fabricante.

Alquilar horas de vuelo supone que el responsable de una avería del avión es la propia empresa fabricante, y por tanto es ella la que debe trabajar en el mantenimiento de los aviones para ahorrar costosas reparaciones, para el fabricante de forma directa, y para la aerolínea en términos de vuelos anulados.

Safran Tech, es una de las empresas fabricantes de aviones que se encarga de optimizar el momento en el que se hace la reparación de un componente de avión para asegurar que el avión no tenga una avería nunca, pero que no hagan recambios de piezas cuando estas no están aún defectuosas.

Existen dos tipos de operaciones de mantenimiento: el mantenimiento programado y el mantenimiento previsto. El primero no es eficiente ya que la pieza que es reemplazada puede no estar lo suficientemente dañada como para eliminarla y aumentará los costes de mantenimiento. Por lo tanto, el mantenimiento predictivo es la clave para reducir los costes de mantenimiento de una aeronave. Se basa en la predicción de cuándo es probable que la pieza se rompa para reemplazarla un poco antes. Este método, aunque es mucho más difícil de implementar que el anterior, es el más eficiente.

Este proyecto entra dentro del mantenimiento previsto. Una de las muchas piezas que pueden estudiarse para predecir su mantenimiento es el motor del avión. Añadiendo sensores que capturen las señales de las vibraciones del motor, es posible intuir cual es la salud del motor.

El problema de esto es que las frecuencias de resonancia del motor son demasiado elevadas como para medir un vuelo de muchas horas, muestreado a una frecuencia superior a la frecuencia de Nyquist. Es un problema de Big Data.

El objetivo de este trabajo es la reconstrucción de esta señal del motor, muestreando únicamente una porción aleatoria de la señal, y sabiendo que la señal es “sparse” (la señal tiene muchos coeficientes despreciables) en una base determinada, que es el dominio frecuencial de la señal.

Para abordar este problema se trabajará con tres modelos: dos modelos deterministas y un modelo probabilístico. Los dos primeros son el FISTA (Fast Iterative Shrinkage-Thresold Algorithm) y el OMP (Orthogonal Maching Pursuit) y el último es el método Bayes Lasso.

El objetivo es conseguir un modelo que reconstruya la señal de la forma más precisa posible, de una forma rápida y que no haga relevante ningún coeficiente que sea despreciable, pues imposibilitaría el estudio de la salud de la señal. En último lugar, los parámetros de la señal no deberían de depender de a salud de la señal, y a ser posible tampoco del ratio de compresión de la señal, para poder fijar un parámetro optimo e inamovible en el método.

Los tres algoritmos existen en la literatura, pero su aplicación en señales de un motor de avión no es común. La reconstrucción de señales bajo unas hipótesis es muy común en la reconstrucción de imágenes.

El algoritmo FISTA es una forma iterativa del operador gradiente de la formulación de Lasso. La formulación de Lasso se define como: $f(x) = ||y - Ax||_2^2 + \lambda||x||_1$, siendo A la matriz de medición, y la señal medida, x la señal real y λ un término que ajusta el peso de la “sparcity” de la señal con el error cuadrático de la señal reconstruida.

El algoritmo FISTA, calcula iterativamente valores de la seña reconstruida x con el gradiente de esta función. Después, toma en cuenta los valores de x_{n-1} y de x_{n-2} para inferir el término x, lo que produce una convergencia más rápida que un método más básico que el FISTA, el ISTA (Iterative Shrinkage-Thresolding Algorithm).

Input: $\Phi, \Psi, y, \lambda, \alpha$
$x_0 \to 0$
$t_0 \to 1$
$z_0 \to x_0$
for $n \leftarrow 0$ *max_iterations* **do**
    $\hat{z}_n \leftarrow \hat{z}_n + \frac{1}{\alpha}\Psi\Phi^T(y - \Phi\Psi^{-1}\hat{x}_n)$
    $_{n+1} \leftarrow soft(\hat{z}_n, \frac{\lambda}{2\alpha})$
    $\mathrm{t}_{n+1} = \frac{1+\sqrt{1+4t_n^2}}{2}$
    $\mathrm{z}_{n+1} = \hat{x}_{n+1} + \frac{t_n - 1}{t_{n+1}}(\hat{x}_{n+1} - \hat{x}_n)$
    $\mathrm{r} \leftarrow y - \Phi\Psi^{-1}\hat{x}_{n+1}$
end
Output: Reconstructed signal $\hat{x}_{n+1}, r$

Algorithm 2: FISTA

El algoritmo OMP parte del supuesto de “sparsity” para reconstruir la señal. Reduce el problema a la búsqueda de los K coeficientes no despreciables de la señal. Para ello, las columnas de la matriz de medida se convierten en vectores (de norma 1) y se selecciona la columna cuyo producto escalar con el resto es mayor. El resto es la diferencia entre la señal medida y su proyección sobre el subespacio vectorial formado por los vectores ya elegidos. Una vez ya se han escogido los K vectores que equivalen a los K coeficientes no despreciables de la señal x, se resuelve el sistema a= Ax, siendo a la proyección mencionada y x de dimensión K. Reordenando los coeficientes, se obtiene la señal x reconstruida.

Input: Λ, y, k
$r \leftarrow y$;
$I \leftarrow []$;
for $i \leftarrow 1\ k$ **do**
  $Rest \leftarrow [\langle \Lambda_j, r \rangle$ for $j = 1...N, \notin I]$;
  $max\_Rest = max(Rest)$;
  $I \leftarrow I + [find\_Ind(max\_Rest)]$;
  $a, u_i = P_{\Lambda_I}(r)$;
  $r = y - a$;
end
Q,R $\leftarrow QR\_Decomposition$($u_{j=1..k}$,A,I) 7;
$x^{(I)} \leftarrow R^{-1} Q^* a$;
for $j \leftarrow 1\ k$ **do**
  $x_{I[j]} \leftarrow x_j^{(I)}$;
end
Output: x

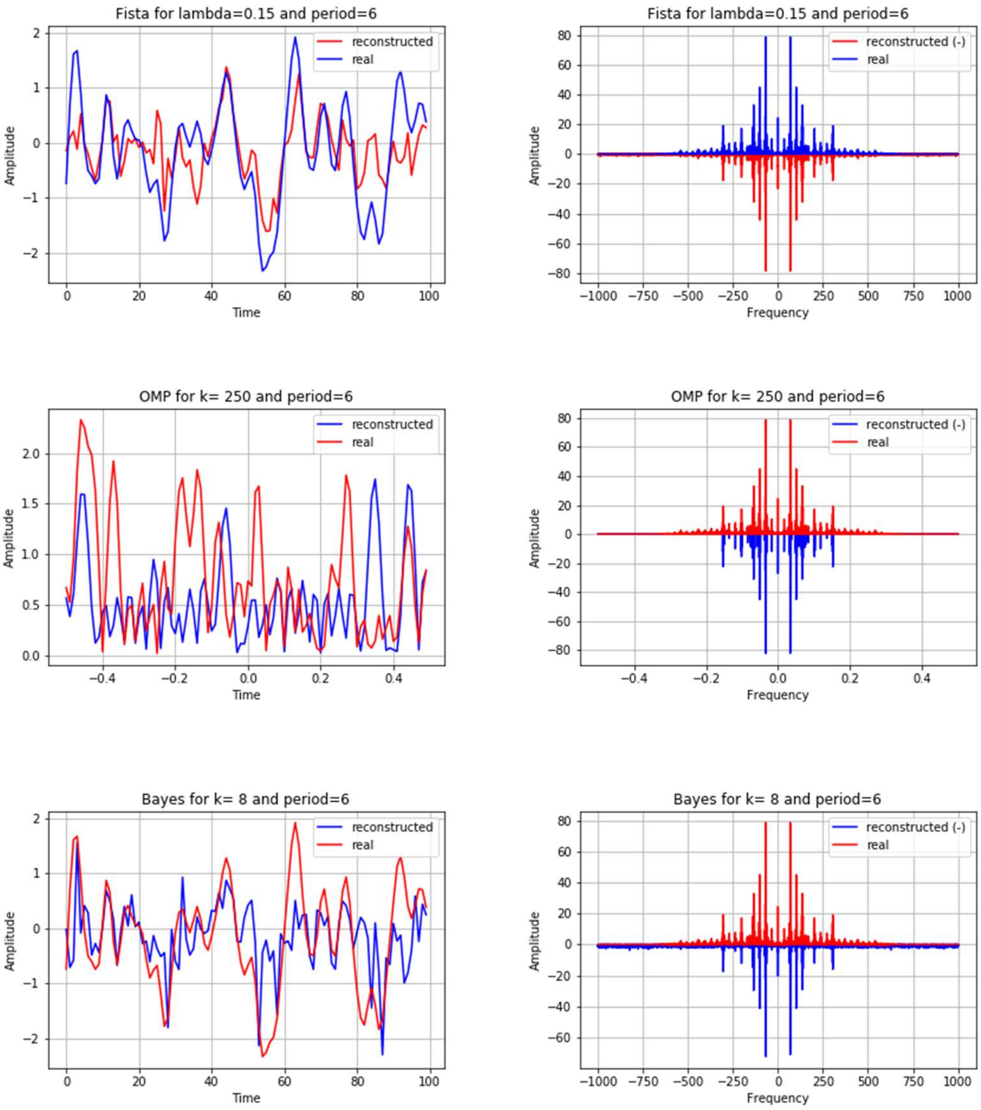
Algorithm 4: Orthogonal Matching Pursuit

El método Bayes Lasso comienza con unas hipótesis sobre la ley probabilística con la que se comportan ciertos parámetros de la señal, para inferir la ley que define el comportamiento de la señal en función dichos parámetros. Dado que esta ley es muy complicada, y no es posible trabajar con ella, el Método de Gibbs es una solución para conseguir un muestreo de la señal reconstruida, que tiende con cada iteración a la ley probabilística de la señal. Finalmente se calcula una esperanza de las n últimas muestras para reducir el error aleatorio. El algoritmo iterativo de Gibbs final es:

Input: $\Lambda, y, k, sigma^2, x_r$
$x_0 \rightarrow 0$
$\gamma_0 \rightarrow 1$
for $j \leftarrow 0\ burn{+}iterations$ **do**
  $(\gamma_i | x_i, \sigma^2, k)_{i=1...N} \leftarrow IG(k/2, \frac{|x_i|^2 + k\sigma^2}{2\sigma^2})$
  $S{=}\sigma^2(\Lambda^* \Lambda + D_{\gamma_i}^{-1})^{-1}$
  $(x_i.real, x_i.imag | \gamma_i, \sigma^2)_{i=1...N} = \mathcal{N}(S[i]\Lambda^* y, S[i] I_2) r[j] = ||x_r - x||^2 G$
  if $j > burn$ **then**
    $\hat{x}_s = average(x_i)_{i=burn+1...burn+iterations}$
  else
    $\hat{x}_s = 0$
  end
end
Output: Reconstructed signal $\hat{x}_s$,error matrix r

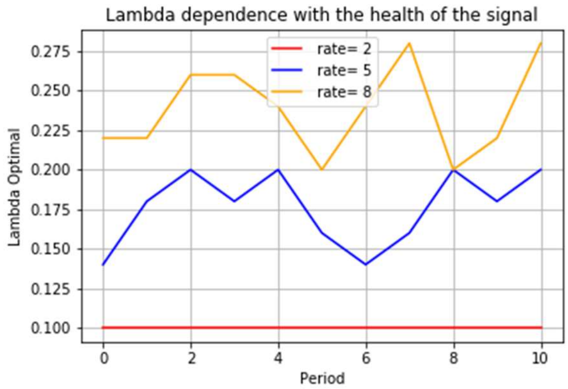
Algorithm 5: Bayesian

En el campo frecuencial puede observarse que la reconstrucción de las frecuencias más fundamentales se mantiene en todos los casos. No obstante, en el plano temporal, es claro que el modelo FISTA es el que mejor reconstruye la señal.

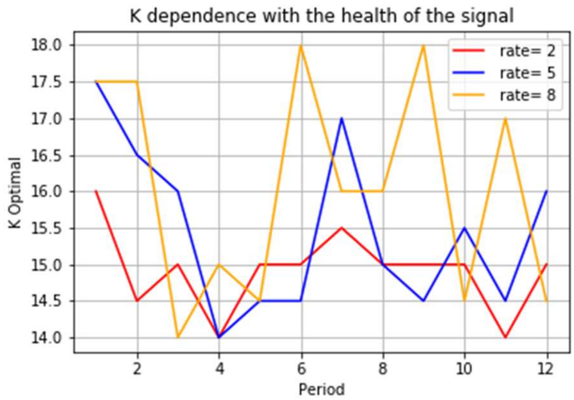


Este hecho también puede reconocerse en la optimización de los parámetros de cada modelo. El error cuadrático de la señal FISTA esta siempre unos órdenes por debajo del error de los demás modelos.

Estudiando la relación de los parámetros con la salud de la señal y con el ratio de compresión, se puede comprobar, para un ratio de compresión de 5, que los modelos OMP y FISTA tienen un parámetro óptimo casi constante.



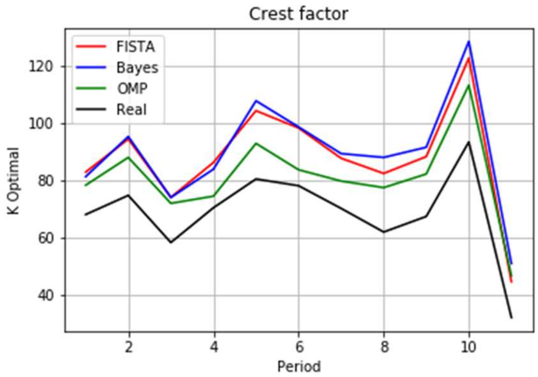
FISTA Model



Bayes Model

OMP Model

Finalmente, se introduce el término Crest Factor, que es el ratio entre el pico más alto de una señal con el valor efectivo de la señal. Este factor ayuda a intuir cuando la señal tiene menos coeficientes no despreciables, pues el error efectivo aumentaría y el Crest Factor disminuiría. Puede comprobarse que la señal se reconstruye bien teniendo en cuenta el Crest Factor:



A partir del periodo 10 la señal está más dañada. Este resultado se ve en todos los casos, aunque debido a la reconstrucción, la función del Crest Factor esta ligeramente desplazada en cada caso.

COMPRESSIVE SENSING TO PREDICT THE HEALTH OF AN AIRCRAFT ENGINE

Autor: Burgos Madrigal, Álvaro.

Director: Chardon, Gilles

Associated institution: Safran Tech

PROJECT SUMMARY

When an airline signs a contract to exploit an aircraft with an aircraft manufacturer, the risk of the aircraft suffering a failure should be assumed by the manufacturer and not by the airline, since they were who built the aircraft. That is why airlines, instead of renting an aircraft for a period of time, are renting flight hours guaranteed by the manufacturer.

Renting flight hours means that the manufacturer itself is responsible for the failure of the aircraft, and therefore it is him that must work on the maintenance of the aircraft to save costly reparations, present on the manufacturer itself, and on the airline in terms of cancelled flights.

Safran Tech, is one of the aircraft manufacturers that searches a model that induces the moment when an aircraft component has to be repaired, not only to ensure that the aircraft does not have a breakdown but also to reduce the amount of reparations done.

There are two types of maintenance operations: scheduled maintenance and predicted maintenance. The first one is not efficient as the part being replaced may not be damaged enough to remove it and will increase maintenance costs. Therefore, predictive maintenance is the key to reducing the maintenance costs of an aircraft. It is based on predicting when the part is likely to break down to replace it a little sooner. This method, despite being much more difficult to implement than the previous one, is the most efficient.

This project is part of predictive maintenance. One of the many parts that may be studied to predict maintenance is the aircraft's engine. The objective is to predict the health of the engine capturing some of the engine's vibrations signals by adding some sensors to the engine frame.

The problem with this is that the resonance frequencies of the engine are too high to measure a flight of many hours, sampled at a frequency higher than the Nyquist frequency. It is a Big Data problem.

The target of this project is the reconstruction of the engine's signal, by only sampling a random portion of the signal, and knowing that the signal is "sparse" (the signal has many negligible coefficients) on a given basis, which is the frequency domain of the signal.

To address this problem, we will work with three models: two deterministic models and a probabilistic model. The first two are the FISTA (Fast Iterative Shrinkage-Thresold Algorithm) and the OMP (Orthogonal Maching Pursuit) and the last one is the Bayes Lasso method.

The aim is to achieve a model that reconstructs the signal as accurately as possible, in a fast way and that does not make any negligible coefficient relevant, as it would make it impossible to study the health of the signal. Finally, the parameters of the signal should not depend on the health of the signal, and if possible, neither on the compression ratio of the signal, to set an optimal and immovable parameter in the method.

All three algorithms exist in the literature, but their application to aircraft engine signals is not common. Signal reconstruction under some hypotheses is very common in image reconstruction.

The FISTA algorithm is an iterative form of the Lasso formulation gradient operator. The Lasso formulation is defined as: $f(x) = ||y - Ax||_2^2 + \lambda||x||_1$, where A is the measurement matrix, y is the measured signal, x is the actual signal and λ is a term that adjusts the weight of the signal sparsity with the quadratic error of the reconstructed signal.

The FISTA algorithm iteratively calculates values of the reconstructed signal x with the gradient of this function. It then takes into account the values of x_{n-1} and x_{n-2} to infer the term x, resulting in faster convergence than a more basic method than FISTA, the ISTA (Iterative Shrinkage-Thresolding Algorithm).

Input: $\Phi, \Psi, y, \lambda, \alpha$
$x_0 \to 0$
$t_0 \to 1$
$z_0 \to x_0$
for $n \leftarrow 0$ *max_iterations* **do**
$\quad \hat{z}_n \leftarrow \hat{z}_n + \frac{1}{\alpha}\Psi\Phi^T(y - \Phi\Psi^{-1}\hat{x}_n)$
$\quad {}_{n+1} \leftarrow soft(\hat{z}_n, \frac{\lambda}{2\alpha})$
$\quad t_{n+1} = \frac{1+\sqrt{1+4t_n^2}}{2}$
$\quad z_{n+1} = \hat{x}_{n+1} + \frac{t_n - 1}{t_{n+1}}(\hat{x}_{n+1} - \hat{x}_n)$
$\quad r \leftarrow y - \Phi\Psi^{-1}\hat{x}_{n+1}$
end
Output: Reconstructed signal $\hat{x}_{n+1}, r$

Algorithm 2: FISTA

The OMP algorithm assumes sparsity to reconstruct the signal. It reduces the problem to searching for the non-negligible K-coefficients of the signal. To do this, the columns of the measurement matrix are converted into vectors (of norm 1) and the column with the highest scaling product with the residual is selected. The residual is the difference between the measured signal and its projection (a) on the subspace formed by the already chosen vectors. Once the K vectors have been chosen, the system a=Ax is resolved, with x being of dimension K. Rearranging the coefficients, the reconstructed signal x is obtained.

Input: Λ, y, k
$r \leftarrow y$;
$I \leftarrow []$;
for $i \leftarrow 1\ k$ **do**
  $Rest \leftarrow [\langle \Lambda_j, r\rangle$ for $j = 1...N, \notin I]$;
  $max\_Rest = max(Rest)$;
  $I \leftarrow I + [find\_Ind(max\_Rest)]$;
  $a, u_i = P_{\Lambda_I}(r)$;
  $r = y - a$;
end
Q,R $\leftarrow QR\_Decomposition(\mathrm{u}_{j=1..k}$,$\Lambda$,I) 7;
$x^{(I)} \leftarrow R^{-1}Q^{*}a$;
for $j \leftarrow 1\ k$ **do**
  $x_{I[j]} \leftarrow x_j^{(I)}$;
end
Output: x

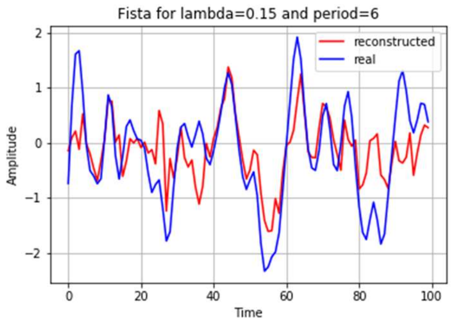
Algorithm 4: Orthogonal Matching Pursuit

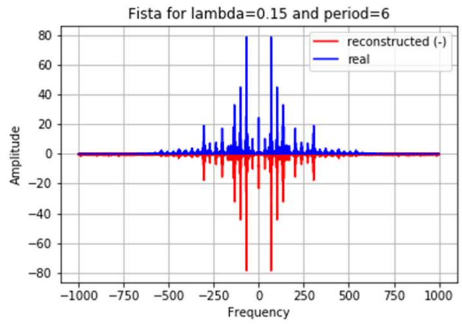
The Bayes Lasso method starts with some hypotheses about the probabilistic law with which certain parameters of the signal behave, to infer the law that defines the behavior of the signal as a function of these parameters. Since this law is very complicated, and it is not possible to work with it, the Gibbs Method is a solution to obtain a sample of the reconstructed signal, which tends with each iteration to the probabilistic law of the signal. Finally, a expectancy of the last n samples is calculated to reduce the random error. The final Gibbs iterative algorithm is:

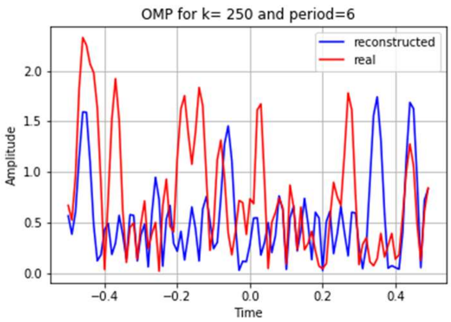
Input: $\Lambda, y, k, sigma^2, x_r$
$x_0 \rightarrow 0$
$\gamma_0 \rightarrow 1$
for $j \leftarrow 0\ burn{+}iterations$ **do**
  $(\gamma_i | x_i, \sigma^2, k)_{i=1...N} \leftarrow IG(k/2, \frac{|x_i|^2 + k\sigma^2}{2\sigma^2})$
  $\mathrm{S}{=}\sigma^2(\Lambda^*\Lambda + D_{\gamma_i}^{-1})^{-1}$
  $(\mathrm{x}_i.real, x_i.imag | \gamma_i, \sigma^2)_{i=1...N} = \mathcal{N}(S[i]\Lambda^* y, S[i] I_2) r[j] = ||x_r - x||^2{}_G$
  if $j > burn$ **then**
    $\hat{x}_s = average(x_i)_{i=burn+1...burn+iterations}$
  else
    $\hat{x}_s = 0$
  end
end
Output: Reconstructed signal $\hat{x}_s$,error matrix r

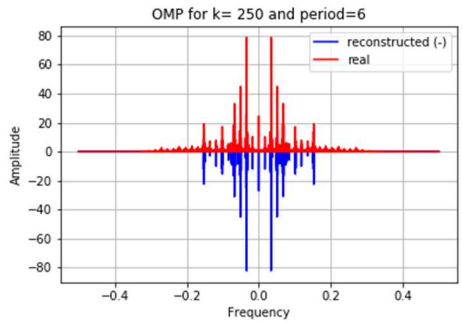
Algorithm 5: Bayesian

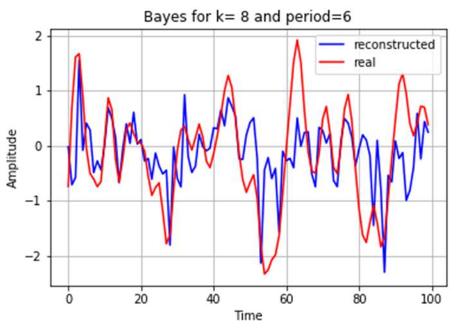
In the frequency domain it can be observed that the reconstruction of the fundamental frequencies is maintained in all cases. However, on the temporal domain, it is undoubtedly the FISTA model the one that best reconstructs the signal.

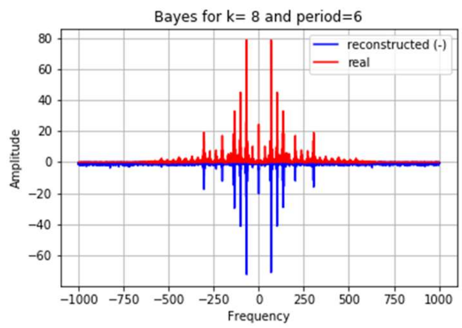






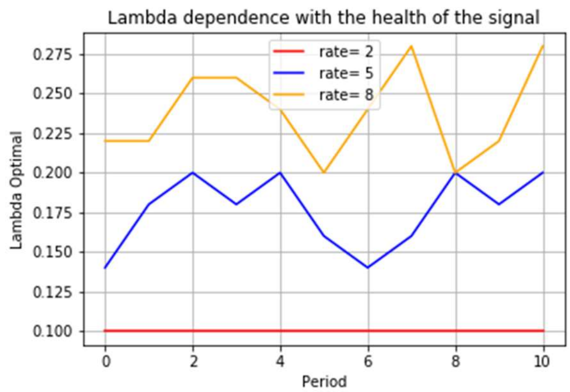




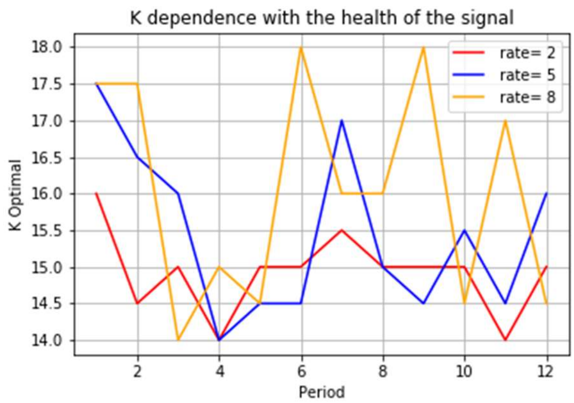


This fact can also be recognized in the optimization of the parameters of each model. The quadratic error of the FISTA signal is always a few orders below the error of the other models.

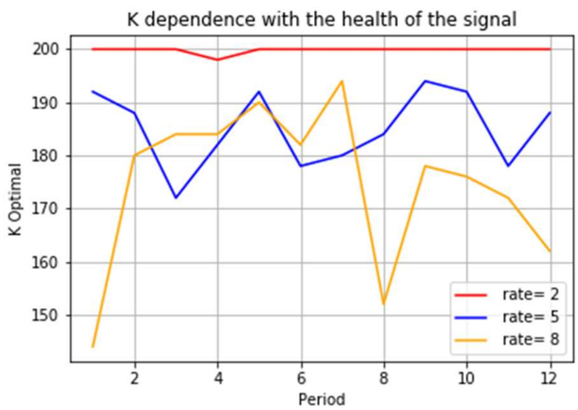
By studying the correlation between the parameters and the health of the signal and the compression ratio, it can be shown, for a compression ratio of 5, that the OMP and FISTA models have an almost constant optimal parameter.



FISTA Model

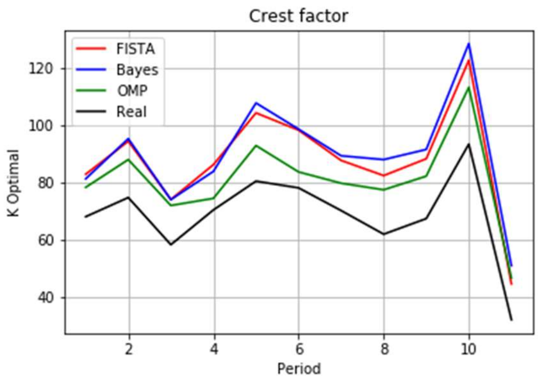


Bayes Model



OMP Model

Finally, the term Crest Factor is introduced, which is the ratio between the highest peak of a signal and the effective value of the signal. This factor helps to sense when the signal has fewer non-negligible coefficients, as the effective error would increase, and the Crest Factor would decrease. The reconstructed signal allows the prediction of the health of the signal by considering the Crest Factor:



From the 10th period the signal starts being damaged. This result is seen in all cases, although due to the reconstruction, the function of the Crest Factor is slightly displaced in each case.

Contents

List of Figures

Chapter 1

Introduction and problem statement

Nowadays the sector of aerial mobility needs an enormous investment to build an aircraft. The aircraft, as it has to give benefits to its shareholders, must be planned to have a long life expectancy so it can complete as much journeys as possible. Nevertheless airlines are not the builders of the aircraft and they do not want to assume the risk of an aircraft's break down. They prefer instead paying for the hours hours the plane is supposed to work in an acceptable operating mode, rather than paying the whole plane.

As a result, enterprises which build the transport may come up with ways of assuring a well maintenance of the airplane, as their role is becoming to maximize the life expectancy. This project was born following this objective.

There are two types of maintenance: scheduled maintenance checks and predicted maintenance. The former is not efficient as the part that may be replaced is not enough damaged to remove it and will rise the maintenance costs. Therefore, predictive maintenance is the key to reducing the costs of maintenance of an aircraft. It is based on predicting when the part is likely to break in order to replace it some time before. This method, even though is far more difficult to implement than the former one, is the most efficient. Engines are studied with this purpose since 1962.

1.1 Safran

Safran [Saf19] is an international high-technology group, operating in the aviation (propulsion, equipment and interiors), defense and space markets. It is a French-registered company but has a global presence, with more than 95,000 employees and sales of 24.6 billion euros in 2019.

Safran is involved in a project that searches to optimize the maintenance of its engines. It has an agreement with an enterprise to undertake a study that predicts when an aircraft's engine should be repaired. The company provides the signal of the vibrations of an engine in their laboratories which models the behaviour of the engines of an airplane in an stationary state. There are twelve different states or periods of the life cycle of the engine, from a healthy engine to a damaged one. This signal follows a certain periodicity as it is related with the frequencies of the angular velocity of the engine's gears.

1.2 Main objectives

The big objective of this project is optimizing the life expectancy of an aircraft's engine and reducing the costs of maintenance. This means, being able to predict the exact moment in which the operator should intervene on the motor's engine to prevent its failure and not to make expensive and inefficient reparations. In this project, identifying when the engine is faulty will be presented. Nevertheless, the objective of this project only concerns the sampling and reconstruction of the signal.

In order to have a faithful representation of the signal, by the Shannon-Nyquist criteria, the signal sampling frequency must be at least twice as big as the frequency of the study. This signal's resonance frequencies are in the order of the frequency of rotation of the gears of the engine. Therefore sampling with the Shannon-Nyquist Criteria over a 4h flight results in a massive amount of data, which is not handy. The signal has to be quickly sampled, transported and processed to retrieve a conclusion about the health of the engine. Therefore respecting the Shannon-Nyquist Criteria is not possible.

1.3 Compressive Sensing

Compressive sensing [Wak08] is a technique that samples a signal under the Shannon-Nyquist frequency. It efficiently reconstructs a signal, by resolving undetermined linear systems (m equations, dimension of the sampled signal, and n variables, the dimension of the real signal, with m¡n). Using the sparcity of the signal as an prerequisite, it is possible to reconstruct it with far fewer samples than it would be needed with the Nyquist-Shannon Theorem.

Compress sensing needs additionally incoherence.It will be shown that a random sampling matrix Φ is enough to meet it.

Compressive Sensing is a problem introduced by:

$$y = \Phi s + \epsilon$$

with:

y: Measured Signal, with n Samples

Phi: Measurement Matrix, of dimension $m \times n$

s: Real Signal in the temporal domain, of dimension n

ϵ: Error, which will be approximated by $\mathbb{N}(0, \sigma^2)$

or:

$$y = \Phi\Psi^{-1}x + \epsilon = Ax + \epsilon$$

with Ψ^{-1} the Inverse Fourier Transform and A the measurement matrix.

Chapter 2

Description of the technologies

2.1 Sparcity

Sparsity expresses the idea that the information rate of a continuous time signal may be much smaller than suggested by its bandwidth, or that a discrete-time signal depends on a number of degrees of freedom which is comparably much smaller than its length. More precisely, CS exploits the fact that many natural signals are sparse or compressible in the sense that they have concise representations when in the proper basis.

Let $\{\psi\}_{i=1}^{N}$ be an orthonormal basis and its basis matrix $N \times N$, $\Psi = [\psi_1|\psi_2|...|\psi_N]$ with the vectors as columns, a signal x can be represented as:

$$x = \sum_{i=1}^{N} \psi_i s_i = \Psi s$$

x and s are both representations of the same signal in different domains. The signal is k-Sparse if and only if, for an orthonormal basis, the signal has n-k zero coefficients. In reality, as the signals have some noise, a coefficient will be asumed to be zero if it is negligible with respect the others.

This propriety is indispensable for a faithful reconstruction of an under-sampled signal. For a real signal of length n and k-Sparse, the dimension m of sampling should verify $n >> m >> k$.

In this project s is the real signal in the temporal domain and x is the real signal in the frequency mode. Therefore $s = \Psi^{-1}x$

2.2 Fourier Transform (FT)

Given that our signal is sparse in the frequency domain it is relevant to know how to switch from the temporal domain to the frequency domain. The Fourier Transform allows us to do so.

The Fourier Transform is a method which decomposes a function into the frequencies that compose it. Therefore the "Time Domain" refers to the input and the "Frequency Domain" refers to the output of the FT. The FT of a function results in a function in the complex domain.

The transform function of a function $y(t)$ is defined as

$y : \Re \to \mathbb{C}; F(y(t)(\omega)) = \int_{-\infty}^{\infty} y(t)e^{-2\pi it\omega}$

The inverse Fourier Transform converts a signal that belongs to the Frequency Domain into the Time Domain. It is defined as:

$y : \mathbb{C} \to \mathbb{C}; F(y(\omega)(t)) = \int_{-\infty}^{\infty} y(\omega)e^{2\pi it\omega}$

2.2.1 Discrete Fourier Transform (DFT)

As computers are not able to do continuous calculations,the discretization of the Fourier Transform allows to compute the transform automatically.

The Discrete Fourier Transform is the method that converts a finite sequence of equally-spaced samples of a function in the time domain into a sequence of the same length of equally-spaced samples of the discrete-time Fourier Transform, which is the continuous transform of the sequence of depart.

The Discrete Fourier Transform is:

$Y(k) = DFT[y(n)] = \sum_{n=0}^{N-1} y(n)e^{-j2\pi kn/N}$

The principal drawback of the Discrete Fourier Transform is that its complexity is in $\mathcal{O}(N^2)$. To resolve this issue, the Fast Fourier Transform(FFT) is used in this project.

The FFT is an algorithm that computes de DFT or its inverse in order to reduce the global complexity from $\mathcal{O}(N^2)$ to $\mathcal{O}(N \log N)$

2.3 Measurement Matrix

As the measurement matrix is a selection of m coefficients out of a signal of n coefficients, all the matrix coefficients must be equal to 0 or 1. Additionally:

Let $a_1, ..., a_m$ be the lines of the matrix Φ, then:

$\{\forall a_k \in \Phi \;\; \exists! \; a_{k,j} \in a_k \; s.t. \; a_{k,j} = 1\}$

Moreover, for all the n columns of the matrix Φ there must be at most one element on each column equal to 1.

A representation of a matrix Φ may be:

$$\Phi = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & \dots & 1 \end{pmatrix} \in \{0,1\}^{m \times n}$$

2.4 Incoherence

Mutual coherence of a matrix A is defined as the maximum absolute value of the scalar product of the different columns of a matrix A, assuming the columns are normalized.

Let $a_1, ..., a_m \in \mathbb{C}^d$ be the columns of the matrix A, which are assumed to be normalized such that $a_i^* a_i = 1$ The coherence of a matrix A is defined as:

$$M = \max_{1 \leq i \neq j \leq m} |a_i^* a_j|$$

For small enough compression rates, the matrix $A = \Phi\Psi^{-1}$ is proved to have a small coherence.

2.5 Lasso's equation

Lasso's formulation [Par19] will be implemented to minimize the error in the signal reconstructed.

We have the following equation to minimize:

$$\mathbb{J}(x) = \underbrace{||\Phi\Psi^{-1}x - y||_2^2}_{(f_1)} + \underbrace{\lambda||x||_1}_{(f_2)} \tag{2.1}$$

Where:

λ is the regularisation parameter.

The first term is related to minimizing the quadratic error of the reconstructed signal with respect to the original one. The second term is a consequence of sparsity: It controls the zeroing of the non-important components.

The minimization problem with the l1-norm constraint does not have a closed form solution. It is solved in iterative ways. In order to define an iterative procedure, a non-negative term is added to the cost function. α is chosen bigger than the biggest eigenvalue of the matrix $\Psi\Phi^T\Phi\Psi^{-1}$. This term will not change the minimization solution as it tends to zero as $\hat{x}$ converges. New cost function is:

$$\mathbb{H}(\hat{x}) = \underbrace{||\Phi\Psi^{-1}\hat{x} - y||_2^2}_{(f_1)} + \underbrace{\lambda||\hat{x}||_1}_{(f_2)} + (\hat{x} - \hat{x}_s)(\alpha I - \Psi\Phi^T\Phi\Psi^{-1})(\hat{x} - \hat{x}_s)$$

Where:

$f_1 = ||\Phi\Psi^{-1}x - y||_2^2$ is is the Data Fidelity Term; ensures the accuracy of the signal by the quadratic error

$f_2 = \lambda||x||_1$ is the Regularisation Term, ensures the sparcity of the signal

α is bigger than the maximum eigenvalue of $\Psi\Phi^T\Phi\Psi^{-1}$ $(\alpha > 1)$

Developing f_1:

$f_1 = x^*(\Psi^{-1})^*\Phi^T\Phi\Psi^{-1}x - y^*\Phi\Psi^{-1}x + (y^*\Phi\Psi^{-1}x)^* + ||y||_2^2$

$f_1 = x^*\Psi\Phi^T\Phi\Psi^{-1}x - 2Re(y^*\Phi\Psi^{-1}x) + ||y||_2^2$

Computing the gradient of f_1:

$\nabla f_1 = 2\Psi\Phi^T\Phi\Psi^{-1}x - 2\Psi\Phi^T y$

The gradient of f_2 is representable in: $\{x \in \mathbb{C} - \{0\}\}$:

$$\nabla f_2(x) = \begin{cases} dir(x) & \text{if } x \neq 0 \\ Doesn'texist & \text{if } x = 0 \end{cases} \quad (2.2)$$

Therefore $\nabla H(\hat{x})$ equals:

$\nabla\mathbb{H}(\hat{x}) = 2\Psi\Phi^T\Phi\Psi^{-1}\hat{x} - 2\Psi\Phi^T y + \lambda dir(\hat{x}) + 2(\alpha I + \Psi\Phi^T\Phi\Psi^{-1})(\hat{x} - \hat{x}_s)$

Developing $\nabla\mathbb{H}(\hat{x}) = 0$ we get to:

$\hat{x} + \frac{\lambda}{2\alpha}dir(\hat{x}) = \frac{1}{\alpha}\Psi\Phi^T(y - \Phi\Psi^{-1}\hat{x}_s) + \hat{x}_s$

Getting to the iterative equation:

$\hat{x}_{n+1} + \frac{\lambda}{2\alpha}dir(\hat{x}_{n+1}) = \frac{1}{\alpha}\Psi\Phi^T(y - \Phi\Psi^{-1}\hat{x}_n) + \hat{x}_n$

2.5.1 Soft equation

We solve this iterative equation to get the x_{n+1} term:

$\hat{x}_{n+1} = \frac{1}{\alpha}\Psi\Phi^T(y - \Phi\Psi^{-1}\hat{x}_n) + \hat{x}_n - \frac{\lambda}{2\alpha}dir(\hat{x}_{n+1})$

Therefore we can define: $\hat{x}_{n+1} = soft(k, \mu) = k - \mu dir(\hat{x}_{n+1})$

With:

$\mu = \frac{\lambda}{2\alpha} \in \Re$

$k = \frac{1}{\alpha}\Psi\Phi^T(y - \Phi\Psi^{-1}\hat{x}_n) + \hat{x}_n \in \mathbb{C}$

We will prove $soft(k, \mu) = dir(k)max(|k| - \mu, 0)$ by solving

$$\hat{x}_{n+1} = k - \mu dir(\hat{x}_{n+1})$$

First of all:

$\hat{x}_{n+1} = dir(\hat{x}_{n+1})|\hat{x}_{n+1}|$

Then:

$|\hat{x}_{n+1}|dir(\hat{x}_{n+1}) + \mu dir(\hat{x}_{n+1}) = k$

$\Leftrightarrow dir(\hat{x}_{n+1}) = dir(k)$ (and with the same sense)

and $\hat{x}_{n+1} = dir(k)(|k| - \mu)$

Nevertheless as $dir(\hat{x}_{n+1}) = dir(k)$ and $|\hat{x}_{n+1}| = |k| - \mu$, if $|k| - \mu < 0$ $|\hat{x}_{n+1}|$ has to be null

As a result:

$soft(k, \mu) = dir(k)max(|k| - \mu, 0)$

Graphically, the soft function, as it comes from the l1-norm, ensures sparsity of negligible components of the vector $\hat{x}_n$. Graphically, for the the specific case where $k \in \Re$:

Chapter 3

Description of the developed model

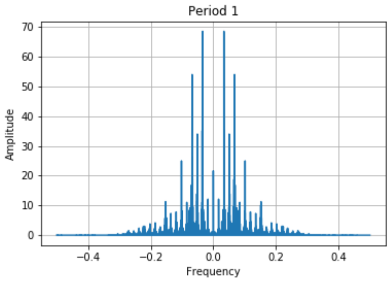
3.1 Objectives and specifications

The objectives that will be pursued are the followings:

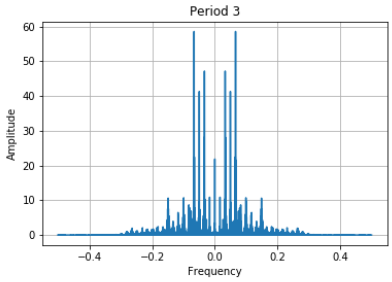
1. Find the model that reconstructs the signal with the best precision
2. Ensure that the reconstructed signal never has a peak where the real signal has not as it may cause a failure in the posterior study of its health.
3. Approach the study of the health of the engine with the reconstructed signals to prove that the reconstruction is acceptable.
4. Study the sensibility of the parameters
5. Study the relation between the parameters and the rate of compression
6. Study the relation between the parameters and the health of the signal

3.2 Data

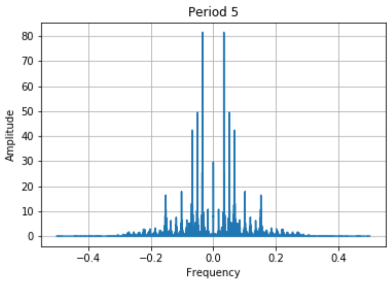
The signal reconstructed appearance, from a healthy state to a damaged one, only odd pictures showed, is the following:



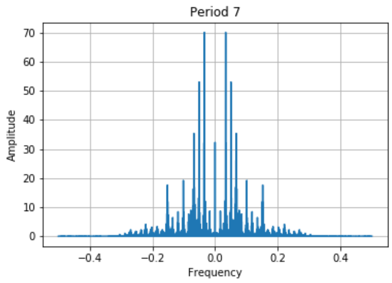
(a) Period 1



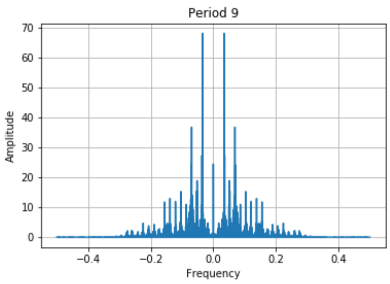
(b) Period 3



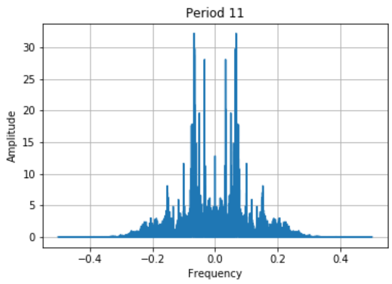
(c) Period 5



(d) Period 7



(e) Period 9



(f) Period 11

Figure 3.1: Real Signal

In all the states of the signal, it is remarkable that most of the coefficients of the signal are negligible. As the health of the signal deteriorates, the amount of non-zero elements increases, which means that the gears of the engine have more and more harmonics in their rotation. At last instance, this vibrations beyond the operating frequencies produce a breakdown on the engine.

This is the reason why it is essential that the signal does not invent a peak where there is not, because in that case the signal would appear to be more damaged than how it actually is.

3.3 Algorithms

3.3.1 ISTA

The ISTA Algorithm (iterative shrinkage-thresholding algorithms) will be presented to introduce the FISTA. It is used for solving linear inverse problems arising in signal/image processing. This class of methods, which can be viewed as an extension of the classical gradient algorithm, is attractive due to its simplicity and thus is adequate for solving large-scale problems even with dense matrix data.

In order to apply the method, the lasso formulation will be used as the cost function. Its gradient must be computed inducing an iterative function. Applying the ISTA Method:

Input: $\Phi, \Psi, y, \lambda, \alpha$
$x_0 \rightarrow 0$
for $n \leftarrow 0$ *max_iterations* **do**
    $\hat{x}_{n+1} \leftarrow soft(\frac{1}{\alpha}\Psi\Phi^T(y - \Phi\Psi^{-1}\hat{x}_n) + \hat{x}_n, \frac{\lambda}{2\alpha})$;
    r $\leftarrow y - \Phi\Psi^{-1}\hat{x}_{n+1}$;
end
Output: Reconstructed signal $\hat{x}_{n+1}, r$

Algorithm 1: ISTA

3.3.2 FISTA

[Teb09] However the ISTA method is also known by its slowly convergence. That is why the FISTA Algorithm is introduced (fast iterative shrinkage-thresholding algorithms) which will be faster by several orders of magnitude.

The difference between both algorithms is that the FISTA computes a better actualisation of the term x_n by taking into account the term x_n and x_{n-1}

Input: $\Phi, \Psi, y, \lambda, \alpha$
$x_0 \rightarrow 0$
$t_0 \rightarrow 1$
$z_0 \rightarrow x_0$
for $n \leftarrow 0$ *max_iterations* **do**
  $z_n \leftarrow z_n + \frac{1}{\alpha}\Psi\Phi^T(y - \Phi\Psi^{-1}x_n)$
  $\mathrm{x}_{n+1} \leftarrow soft(z_n, \frac{\lambda}{2\alpha})$
  $\mathrm{t}_{n+1} = \frac{1+\sqrt{1+4t_n^2}}{2}$
  $\mathrm{z}_{n+1} = x_{n+1} + \frac{t_n - 1}{t_{n+1}}(x_{n+1} - x_n)$
  $\mathrm{r} \leftarrow y - \Phi\Psi^{-1}x_{n+1}$
end
Output: Reconstructed signal x_{n+1}, r

Algorithm 2: FISTA

3.3.3 Orthognal Matching Pursuit (OMP)

Notations

[YC 93] Let the problem be finding a x of dimension n and k-sparse (k non-zero coefficients) that resolves best $y = Ax + \epsilon$ where y is given as an input and $n >> m >> k$. The OMP Method searches to find the k most relevant coefficients of x by selecting the k columns of A whose scalar products with y are the biggest.

Let the columns of A be normalized and note $A_{i,i=1...N}$ as the i-th column of A. Let I be a vector of selection of the columns of A be of dimension $k < n$ and verify $\{k \in I : \quad 1 < k < N\}$. Let $A^{(I)}$ be the matrix of dimension $m \times k$ created by the k columns of A that I selected, and $x^{(I)}$ its related coefficients. For a subspace V in $\mathbb{C}^N$, let $P_V(y)$ be the orthogonal projection of y in V. At last, let $\mathcal{A}^{(I)}$ be the span of the columns $A^{(I)}$

Since x in K-Sparse, $y - \epsilon$ is in the span of k columns of A . Then the idea is to find the indexes $i_1, i_2...i_k \in I$ of the columns of A that maximizes the scalar product, find the projection a of y into the span $\mathcal{A}^{(I)}$ and resolve the equation $\mathcal{A}^{(I)}x^{(I)} = ax^{(I)}$ to get the non-zero coefficients of $x^{(I)}$. At last, put back in order the coefficients of x.

Finding the I Indexes

In order to find the k Indexes, we will iteratively search k times the biggest column of A, that has not yet been selected, with the scalar product with r as a norm. r is the part of y that is beyond the span of the already selected columns of A.

The algorithm is the following:

Input: A, y, k
$r \leftarrow y$;
$I \leftarrow []$;
for $i \leftarrow 0\ k$ **do**
    $Rest \leftarrow [\langle A_j, r\rangle \text{ for } j = 1...N, \notin I]$;
    $max\_Rest = max(Rest)$;
    $I \leftarrow I + [find\_Ind(max\_Rest)]$;
    $a = P_{\mathcal{A}_I}(y)$;
    $r = y - a$;
end
Output: I, a

Algorithm 3: Finding the first K Indexes

To have the projection of r in the subspace $\mathcal{A}_I$, the Gran Smith method is implemented 6.3

Resolving the equation

Once the k most relevant indexes are selected (and stored in I), and the projection of y in the span of the columns of I is retrieved, one should resolve the equation $A^{(I)}x^{(I)} = a$. This cannot be simply solved by simple numpy methods. Nevertheless, using the QR Decomposition 6.4 this can be easily resolved.

Then:

$$A^{(I)}x^{(I)} = QRx^{(I)} = a$$

and therefore:

$$x^{(I)} = R^{-1}Q^{*}a$$

Global algorithm

As a result from all of the above, this algorithm is retrieved:

Input: A, y, k
$r \leftarrow y$;
$I \leftarrow []$;
for $i \leftarrow 1$ k **do**
    $Rest \leftarrow [\langle A_j, r\rangle$ for $j = 1...N, \notin I]$;
    $max\_Rest = max(Rest)$;
    $I \leftarrow I + [find\_Ind(max\_Rest)]$;
    $a, u_i = P_{\mathcal{A}_I}(y)$;
    $r = y - a$;
end
Q,R $\leftarrow$ $QR\_Decomposition$($u_{j=1..k}$,A,I) 7;
$x^{(I)} \leftarrow R^{-1}Q^{*}a$;
for $j \leftarrow 1$ k **do**
    $x_{I[j]} \leftarrow x_j^{(I)}$;
end
Output: x

Algorithm 4: Orthogonal Matching Pursuit

3.3.4 Bayesian Lasso

Notations

[Cas08] The object of the study is to solve $y = Ax + \epsilon$ where y is the input signal of dimension m, x is the output complex vector of length $N >> m$ and A is assumed to have a small coherence. ϵ is assumed to follow a Gaussian noise of standard deviation σ, so:

$$y|x, \sigma \sim \mathcal{N}(Ax, \sigma^2 I_m)$$

In this section the signal x will be retrieved by inferring some posterior distributions of the parameters of the signal, in order to use the Gibbs' Method to iterate an approximate a solution.

Choice of priors distributions

A common way of expressing the sparsity of the signal x is choosing a prior of parameter k on the variable x_i/σ. However this is not possible as x is complex. It will be then used a Gaussian mixture associated with a Student distribution in the complex case. Using independent $\gamma_{i,i\in 1,N}$ such that $\gamma_i \sim InverseGamma(k/2, k/2)$ and $Re(x_i/\sigma), Im(x_i/\sigma)|\gamma_i, \sigma \sim \mathcal{N}(0, \gamma_i I_2)$ or the equivalent $Re(x_i), Im(x_i)|\gamma_i, \sigma \sim \mathcal{N}(0, \sigma^2\gamma_i I_2)$

There is little to know about σ^2 therefore it will be used a non-informative prior $\sigma^2 \sim \frac{1}{\sigma^2}$

Retrieving the posterior distribution

The Bayes Theorem is defined as:

$p(x|y) = \frac{p(y|x)p(x)}{p(y)}$

Developing $p(x)$:

$p(x) = \prod_{i=1}^{N} p(x_i) = (\prod_{i=1}^{N} p(x_i|\gamma_i, \sigma)p(\gamma_i))p(\sigma)$

Therefore $p(x|y) \sim \frac{p(y|x)(\prod_{i=1}^{N} p(x_i|\gamma_i\sigma)p(\gamma_i))p(\sigma)}{p(y)}$

By the calculations in 6.2 we get to:

$p(x_{i,i\in\{1,N\}}|\sigma, \gamma_{i,i\in\{1,N\}}) \sim \mathcal{N}(M(\gamma), S(\gamma))$

with $S = (\frac{A^H A + D_{\gamma_i}^{-1}}{\sigma^2})^{-1}$

and $M = SA^*y$

$p(\gamma_i|x_i, \sigma^2)_{i=1...N} \sim IG(k/2, \frac{|x_i|^2 + k\sigma^2}{2\sigma^2})$

$p(\sigma^2|x_i, \gamma_i)_{i=1...N} \sim IG(m/2 + n + 1, \frac{||y - Ax||_2^2}{2} + \sum_{i=1}^{N} \frac{|x_i|^2}{2\gamma_i})$

Gibbs Sampling

Gibbs sampling is a Markov Chain Monte Carlo Algorithm for approximating a complicated joint distribution. As it works with random distribution, it is not a deterministic algorithm such as the FISTA or the OMP.

Gibbs sampling returns a Markov Chain of samples, each of them correlated with its neighbours. Nevertheless, the samples from the beginning of the chain should be discarded as they have not converged yet into the desired distribution.

Input: A, y, k, σ^2, x_r
$x_0 \to 0$
$\gamma_0 \to 1$
for $j \leftarrow 0$ *burn+iterations* **do**
  $(\gamma_i | x_i, \sigma^2, k)_{i=1...N} \leftarrow IG(k/2, \frac{|x_i|^2 + k\sigma^2}{2\sigma^2})$
  S=$\sigma^2(A^*A + D_{\gamma_i}^{-1})^{-1}$
  (x$_i.real, x_i.imag | \gamma_i, \sigma^2)_{i=1...N} = \mathcal{N}(S[i]A^*y, S[i]I_2)$
  $(\sigma | \gamma_i, x_i^2)_{i=1...N} = IG(\frac{m}{2} + n + 1, \frac{||y - Ax||_2^2}{} + \sum_{i=1}^{N} \frac{|x_i|^2}{2\lambda_i})$
  r[j] = $||x_r - x||^2$
  if $j > burn$ **then**
    $\hat{x}_s = average(x_i)_{i=burn+1...burn+iterations}$
  else
    $\hat{x}_s = 0$
  end
end
Output: Reconstructed signal $\hat{x}_s$,error matrix r

Algorithm 5: Bayesian

Numeric implantation of the Bayes Theorem

As the real signal is of dimension 60.000, computing the matrix $S = (\frac{A^*A + D_{\gamma_i}^{-1}}{\sigma^2})^{-1}$ it is not possible because of the memory to allocate in the computer.

The incoherence of the matrix A allows to make a simplification. It has been assumed that the matrix A is incoherent and therefore the non-diagonal coefficients of the matrix $A * A$ have been zeroed.

As a result, $S = (\frac{A^*A + D_{\gamma_i}^{-1}}{\sigma^2})^{-1} \approx \sigma^2((m/n)\mathbb{I} + D_{\gamma_i}^{-1})^{-1}$, which is diagonal

3.4 Numeric Implantation

For the same reason, in all the models, working with the measurement matrix A has been avoided. Instead of a Φ matrix as shown in 2.3, a vector containing the m selected coefficients of x improves the efficiency and is less heavy. Instead of the Ψ matrix, it has always been used the DFT exposed in 2.2.1 which is also more efficient.

Chapter 4

Analysis of results

4.1 Introduction

Firstly, the base case results will provide a general idea of the success of the reconstruction of the signal, for both the time and frequency domain. Then, in the sensitivity analysis, it will be inferred the optimal parameter for each model. In the following step, it will be shown the dependence of this optimal parameter with the other variables: The rate of compression and the health of the signal.

Finally, it will be shown an approach of how the reconstruction of the signals can predict an irregularity in the engine.

4.2 Base case results

The base case results will only be presented for the days 1, 6 and 12, for a time and frequency domain, for each model, with a fixed rate of compression (r=5) and parameter for each model.

4.2.1 FISTA

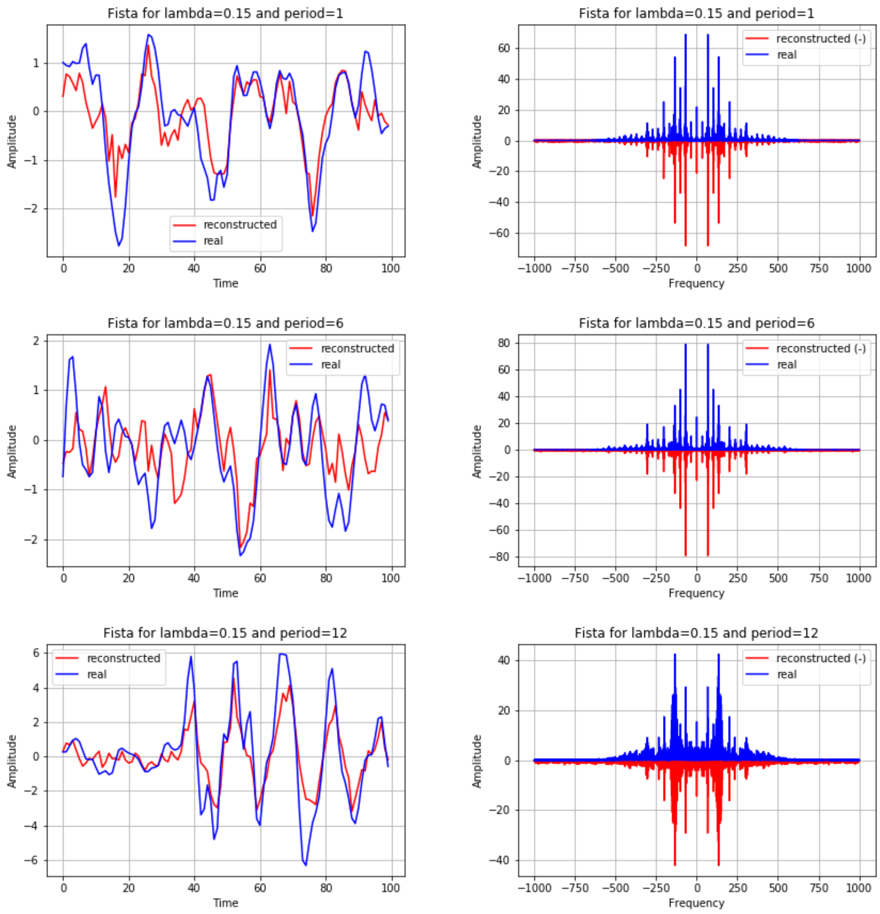


Figure 4.1: FISTA base case results for a rate of compression of 5

4.2.2 OMP

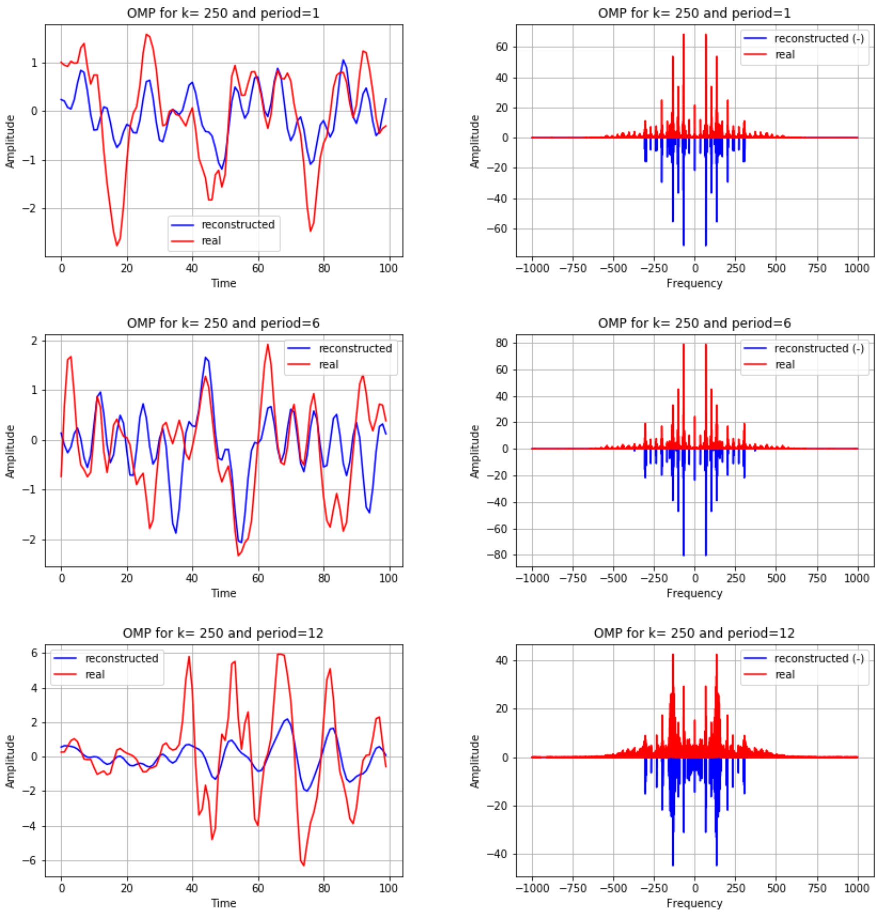


Figure 4.2: OMP base case results for a rate of compression of 5

4.2.3 Bayesian Lasso

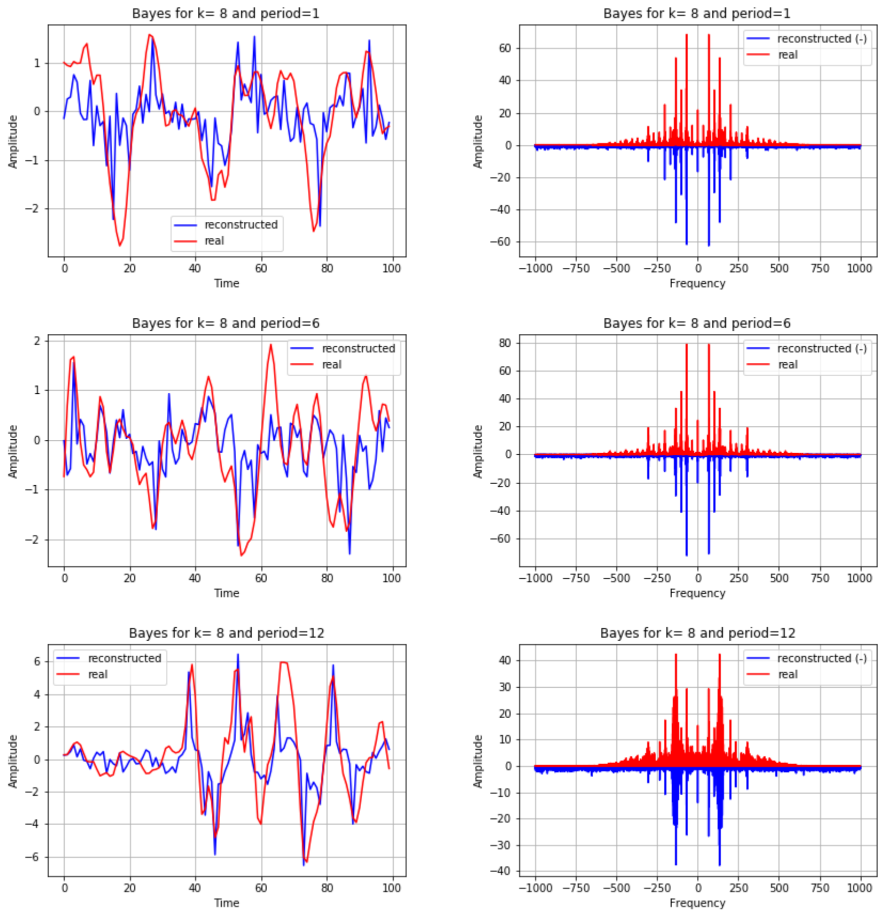


Figure 4.3: Bayesian base case results for a rate of compression of 5

4.2.4 Crest factor

The crest factor is a parameter that represents the extremity of the peaks of a waveform. A crest factor equal to 1 means would be a constant function
It is defined as: $C = \frac{|x_{max}|}{x_{rms}} = \frac{|x_{\infty}|}{x_2}$

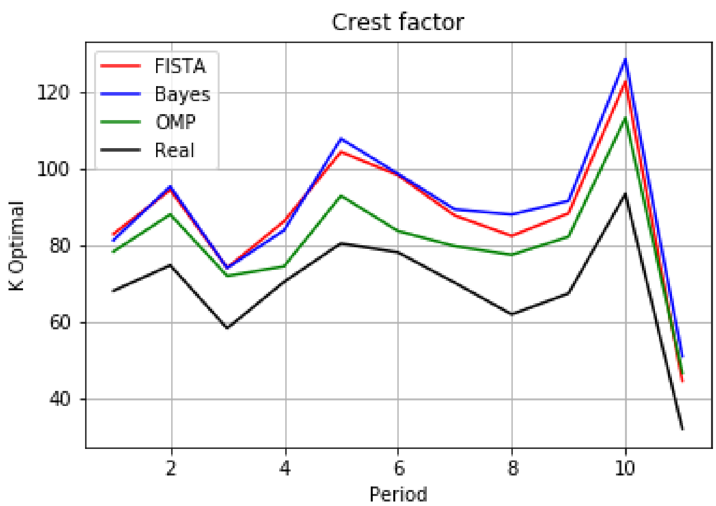


Figure 4.4: Crest Factor

4.3 Sensitivity analysis

4.3.1 FISTA:Optimizing λ

For a rate of compression of 5, it is presented the cuadratic error of the FISTA algorithm as λ increases, in order to reach a minimum quadratic error and fix an optimal λ.

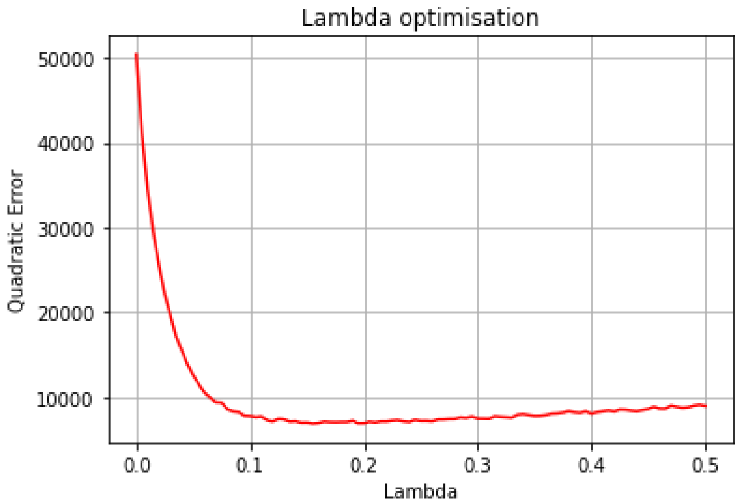
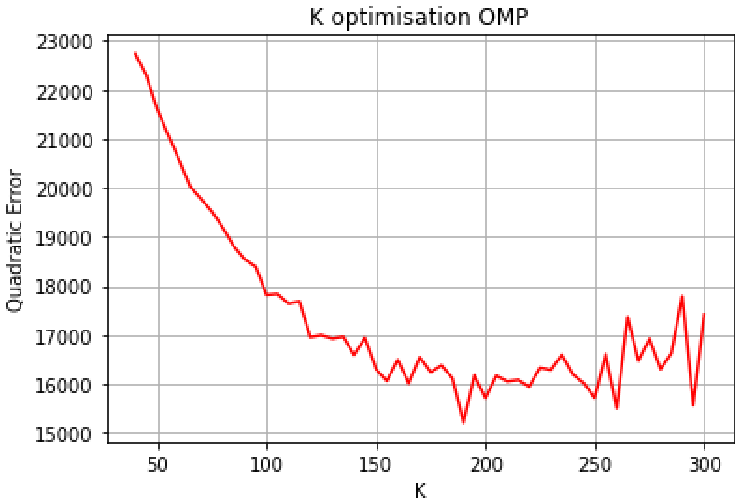


Figure 4.5: Lambda Optimisation for a rate of compression of 5 and period=1

4.3.2 OMP: Optimizing k

For a rate of compression of 5, it is presented the cuadratic error of the OMP algorithm as k increases, in order to reach a minimum quadratic error and fix an

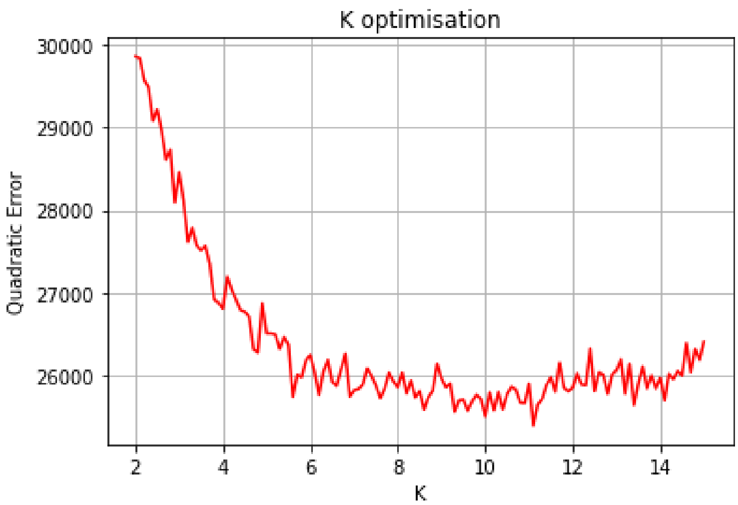


optimal k.

Figure 4.6: OMP optimisation for a rate of compression of 5 and period=1

4.3.3 Bayesien: Optimizing k

For a rate of compression of 5, it is presented the cuadratic error of the Bayesian algorithm as k increases, in order to reach a minimum quadratic error and fix an



optimal k.

Figure 4.7: Bayes Lasso optimisation for a rate of compression of 5 and period=1

4.3.4 Dependance of the parameters

4.3.5 FISTA

In the figure below, for three different ratios of compression, and the twelve periods of the signal, it is retrieved the optimal λ parameter for the FISTA algorithm in order to study its dependence with the ratio of compression and the health

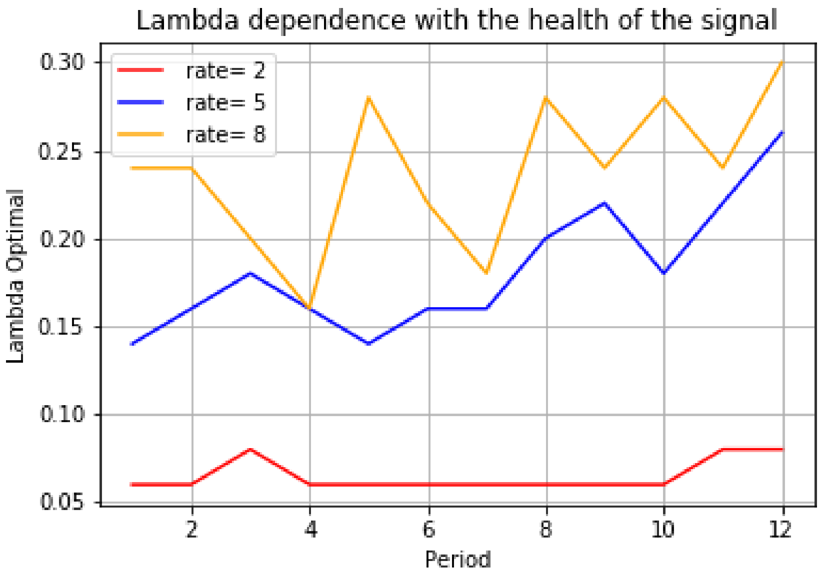


Figure 4.8: Optimal Lambda for three compression rates in the twelve periods

As there is some difference between the optimal λ parameters, there are two figures below evaluating the quadratic error in two cases: Choosing a parameter average of the 12 optimal cases, and choosing the parameter of the 12^{th} period

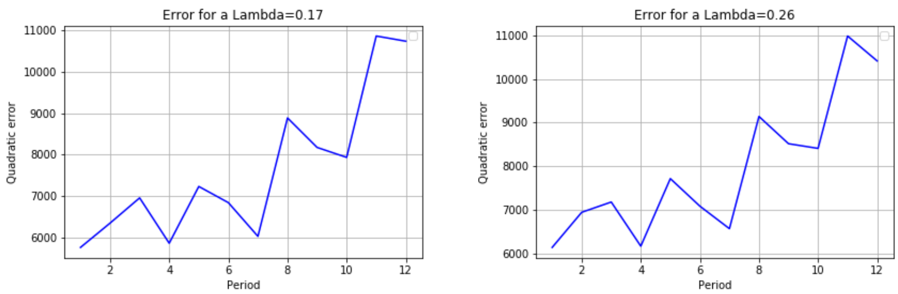
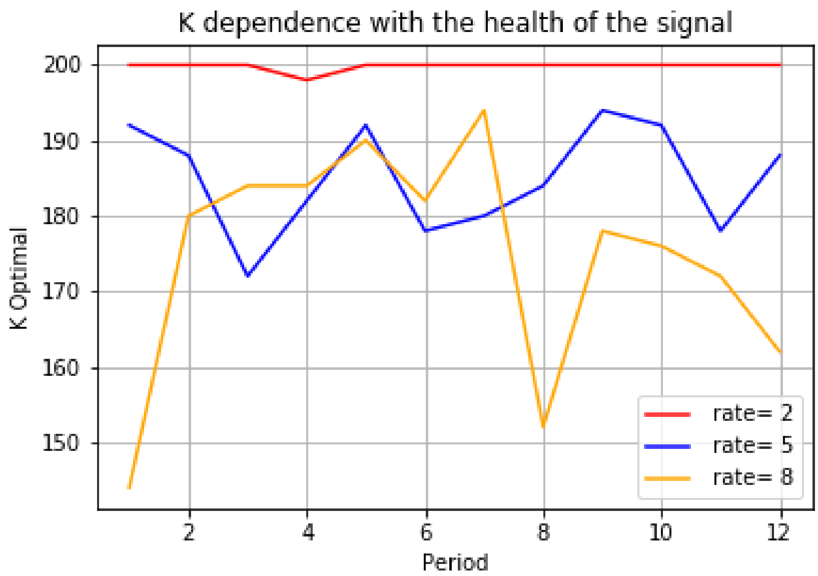


Figure 4.9: Quadratic error for a fixed lambda equal to the average of all the optimal lambdas, and for the optimal lambda in period 12

4.3.6 OMP

In the figure below, for three different ratios of compression, and the twelve periods of the signal, it is retrieved the optimal k parameter for the OMP algorithm in order to study its dependence with the ratio of compression and the



health

Figure 4.10: Optimal K for three compression rates in the twelve periods

As there is some difference between the optimal k parameters, there are two figures below evaluating the quadratic error in two cases: Choosing a parameter average of the 12 optimal cases, and choosing the parameter of the 12^{th} period

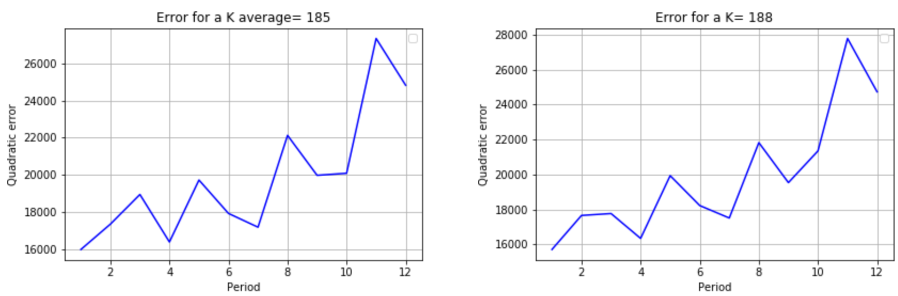


Figure 4.11: Quadratic error for a fixed K equal to the average of all the optimal K, and for the optimal k in period 12

4.3.7 Bayesian

In the figure below, for three different ratios of compression, and the twelve periods of the signal, it is retrieved the optimal k parameter for the Bayesian algorithm in order to study its dependence with the ratio of compression and the health

Figure 4.12: Optimal K for three compression rates in the twelve periods

As there is some difference between the optimal k parameters, there are two figures below evaluating the quadratic error in two cases: Choosing a parameter average of the 12 optimal cases, and choosing the parameter of the 12^{th} period

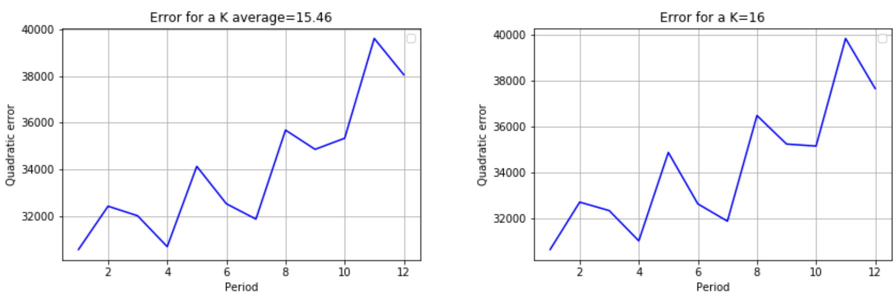


Figure 4.13: Quadratic error for a fixed k equal to the average of all the optimal k, and for the optimal k in period 12

Chapter 5

Conclusions

5.1 Conclusions about methodology

Without any doubt this report is interesting, as it shows how to efficiently reconstruct a signal that is up to 8 times is real value. Even though the study has focused in the engines of the aircrafts, lots of improvements could be done in other ways of mobility, or even in domestic machines or machines of a factory. Any machine that follows a periodicity.

Regarding how the project has been approached, it is clear that the simplification of the computation has been crucial for the project. Avoiding matrices of $60k \times 60k$ is necessary to undertake the project successfully.

Even in this case, final computations to tackle the sensitivity of all the parameters as a function of all the constraints take an enormous amount of time.

5.2 Conclusions about the results

5.2.1 Conclusions about the base case results

Even though observing the temporal domain one would say that the reconstruction of the signal is not accurate enough, the frequency domain is the relevant one, as the matter of the project is identifying the signal with lots of non-zero components, thus the damaged engines.

In the frequency mode it is noticeable that the quality of reconstruction seems acceptable as the differences between a healthy engine and a damaged engine are clear. (That is the signals on period 1 or 6 and the period 12).

5.2.2 Crest Factor

The issue mentioned before is present in the crest factor, which represents the extremity of the peaks of a signal. This is related with the amount of non-negligible components in a signal.

After the period 10 the crest factor decreases drastically, which means that the maximum value of the signal is getting lower in relation with the norm 2 of the signal. This is explained as the signal has more non-zero components and therefore the norm 2 is bigger even though the peak value is similar.

5.2.3 Sensibility

As the FISTA method follows a gradient descend, it is understandable that the cost function is the smoothest. For the case represented **??** the λ optimal is clearly is around 0.15.

The Bayesian Method 4.7 has also a minimum in the boundaries even though the curve is not as smooth as the previous one. In fact, the model is probabilistic and therefore the results have a component of random error, even though the reconstructed signal is an average of the last iterations.

The factor k in the OMP Method reflects the amount of coefficients that the program should consider non-zero. It reaches a minimum around 230, even though the curve is less precise as k increases, as the resolution of the last equation start giving errors.

In conclusion, observing the error given for the optimal parameters in each case, it is clear that the model in which the quadratic error is minimal is the FISTA. The OMP has almost twice the error of FISTA and the Bayesian almost three times.

5.2.4 Dependance of the parameters

Ideally, the optimal model would be a model in which its parameter would not change with the rate of compression nor the health of the engine. If this was the case, the same parameter would be used no matter what situation and the quadratic error of the signal would be always minimal.

Observing the results 4.8 4.12 4.10, it is acceptable to find an average parameter that reconstructs the signal within an acceptable error in each case.

For the model FISTA, and a ratio of 5, the relative distance between the biggest and lowest parameter is 0.3, which may seem huge. Nevertheless, computing the error with an average λ and the maximum λ in figure 4.9, it is observable that the difference between the errors is not huge. The error increases always as the period increases, so it will be advisable to choose the λ optimal in period 12, $\lambda = 0.2$ For the OMP model and the Bayesian model it occurs the same. It would be advisable to choose $k = 17$ in the Bayesian and $k = 176$ in the OMP.

5.2.5 Choice of model

Undoubtedly, the model which suits the best is the FISTA. Firstly, the quadratic error of the reconstructed signal is lower than in the other models. This is remarkable comparing the error in figure 4.5 with the corrresponding figures of the other models: 4.7 and 4.6, but also looking at the base case results, 4.1, where even the temporal domain gives good approximations.

The convergence of its parameter λ is clear, as the curve is smooth, which helps optimising the quadratic error.In the rest of the models, there is a convergence, but both figures have some noise which difficulties searching for the optimal parameters.

In terms of fixing a parameter for a given rate of compression, it has been shown in figure 4.8 that for a fixed rate of compression, the optimal parameter does not change as the engine ages. Additionally, for a rate of compression of 5, it is shown in 4.9, that the quadratic error as a function of the aging of the engine for some fixed lambdas, is lower at any case that the best cases in the rest of the models.

Regarding the crest factor, the three models seem to follow approximately the same tendency, which means that the posterior study of the health of the engine is possible.

Nevertheless, the Fista Method also has som drawbacks. It is by far the model which takes the most time to compute one iteration. Its input parameter, lambda, balances the importance between sparcity and the quadratic error, but it does not represent any physical feature of the signal.

To conclude, it may be deceptive that what it seems to be the most obvious proposition, is in fact, the most optimal.

5.3 Recommendation for further studies

As it has been just mentioned, it is surprising that the model which seem more obvious is in fact the most optimal. More models could be implemented in order to find an original idea that induces the convergence faster than the FISTA method and whose final solution is more precise.

However, the clear path to follow after this project is the study of the diagnosis of the health of the engine, which is a subject briefly covered here.

To do so, some additional studies could be implemented in addition to the crest factor, even some much more naive. For instance, it could be measured the amount of samples that are bigger than the biggest peak divided by a fixed rate.

Chapter 6

Appendix

6.1 Sustainable development Goals of the United Nations

In addition of reducing the costs of the maintenance of the aircraft, predictive maintenance meets some of the Sustainable Development Goals presented by the United Nations in September 2015. [Nat15]

In the domain 12, responsible consumption and production, complains about the present consumption of resources, which will be impossible to meet if the humanity follows the actual rates of growth.

Specifically, the project meets the goal 12.5, which is: "By 2030, substantially reduce waste generation through prevention, reduction, recycling and reuse".

6.2 Calculus to retrieve the posterior distribution of x, σ^2 and γ_i

$p(x|y) = \mathcal{N}(Ax, \sigma^2 I_m) \prod_{i=1}^{N} (p(Re(x_i), Im(x_i)|\gamma_i, \sigma) IG(k/2, k/2)) p(\sigma)$

$= \frac{1}{(\sqrt{2\pi\sigma^2})^m} exp(-\frac{1}{2}(y - Ax)^T (\sigma^2 I_m)^{-1} (y - Ax)) \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2\gamma_i}} exp(-\frac{1}{2}\frac{|x_i|^2}{\sigma^2\gamma_i})$

$\frac{(k/2)^{k/2}}{\Gamma(k/2)}(1/\gamma_i)^{k/2+1+1/2}exp(-\frac{k}{2\gamma_i})$

Reorganising the terms and deleting terms not related with σ, x or γ:

$p(x|y) = \frac{1}{(\sigma^2)^{m/2+N/2}} \exp(-\frac{1}{2\sigma^2}[(y-Ax)^T(I_m)^{-1}(y-Ax)+\sum_{i=1}^N x_i^* \frac{1}{\gamma_i} x_i])$
$\prod_{i=1}^N (1/\gamma_i)^{k/2+3/2} \exp(-\frac{k}{2\gamma_i})$

Working with the exponential term (parameters affected by x) :
$-\frac{1}{2\sigma^2}[(y-Ax)^T(I_m)^{-1}(y-Ax)+\sum_{i=1}^N x_i^* \frac{1}{\gamma_i} x_i]$
$= -\frac{1}{2\sigma^2}[y^Ty - 2y^TAx + x^*A^*Ax + x^*D_{\gamma_i}^{-1}x] \sim -\frac{1}{2}[(x-M)^*S^{-1}(x-M)]$

with $S = (\frac{A^*A+D_{\gamma_i}^{-1}}{\sigma^2})^{-1}$ and $M = SA^*y$

Therefore:

$$p(x_{i,i\in\{1,N\}}|\sigma, \gamma_{i,i\in\{1,N\}}) \sim \mathcal{N}(M(\gamma), S(\gamma))$$

Working with the terms related with γ_i:

$\sim \frac{1}{\gamma_i}^{\frac{k}{2}+2} e^{-(x_i^T x_i/\sigma^2+k/2)/\gamma_i} \sim IG(\frac{k}{2}+1, \frac{|x_i|^2+k\sigma^2}{2\sigma^2})$

$$p(\gamma_i|x_i,\sigma^2)_{i=1...N} \sim IG(k/2, \frac{|x_i|^2+k\sigma^2}{2\sigma^2})$$

Working with the terms related with σ^2:

$\sim \frac{1}{\sigma^2}^{m/2+n+2} e^{-\frac{1}{2\sigma^2}[||y-Ax||_2^2+\sum_{i=1}^N \frac{|x_i|^2}{2\gamma_i}]}$

$p(\sigma^2|x_i,\gamma_i)_{i=1...N} \sim IG(m/2+n+1, \frac{||y-Ax||_2^2}{2} + \sum_{i=1}^N \frac{|x_i|^2}{2\gamma_i})$

6.3 Gran Smith

In other words:

Input: k linearly independent vectors, $v_{i=1..k}$
$u_1 \leftarrow \frac{v}{\langle v,v\rangle}$
for $i \leftarrow 2\ k$ **do**
$e_n = v_n - \sum_{j=1}^{n-1} \langle v_n, u_j\rangle u_j$
$\mathrm{u}_j \leftarrow \frac{e_j}{\langle e_j,e_j\rangle}$
end
Output: k orthonormal vectors, $u_{i=1..k}$

Algorithm 6: Gran Smith

$$\begin{aligned} \mathrm{e}_1 &= v_1 & u_1 &= \tfrac{e_1}{|e_1|} \\ e_2 &= v_2 - \langle v_2, u_1\rangle u_1 & u_2 &= \tfrac{e_2}{|e_2|} \\ &\vdots & &\vdots \\ e_n &= v_n - \textstyle\sum_{i=1}^{n-1} \langle v_n, u_i\rangle u_i & u_n &= \tfrac{e_n}{|e_n|} \end{aligned}$$

6.4 QR Decomposition

The QR Decomposition is the decomposition of a matrix A into a product of matrices Q and R, where Q is an orthogonal matrix (columns and rows are unitary vectors orthogonal between each other) and R and upper triangular matrix. This method is supported by the Gran Smith Method. 6.3

Let $v_{i=1...n}$ be the n columns of the matrix A. Then, using the Gran Smith Method 6.3, n orthonormal vectors are retrieved $u_{i=1...n}$

Then, Q and R are the following:

$$\mathrm{Q} = \begin{pmatrix} \mathrm{u}_1 & \mathrm{u}_2 & \cdots & \mathrm{u}_n \end{pmatrix}$$

$$\mathrm{R}= \begin{pmatrix} \langle u_1, v_1\rangle & \langle u_1, v_2\rangle & \cdots & \langle u_1, v_n\rangle \\ 0 & \langle u_2, v_1\rangle & \cdots & \langle u_2, v_n\rangle \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \langle u_n, v_n\rangle \end{pmatrix}$$

6.4.1 Clarifications about the matrix used $A^{(I)}$

The matrix decomposed is $A^{(I)}$. Therefore, there is some issues to be clarified:

1. As $A^{(I)}$ is a matrix of complex numbers, $Q^*Q = I$
2. As the matrix $A^{(I)}$ is a $m \times k$ matrix, with $m > k$, we have:
   Q= $\begin{bmatrix} \mathrm{Q}_1, Q_2 \end{bmatrix}$ with Q_1 of dimensions $m \times k$ and $Q_2 = 0$
   and R= $\begin{bmatrix} \mathrm{R}_1 \\ R_2 \end{bmatrix}$ with R_1 of dimensions $k \times k$ and $R_2 = 0$
   So that $A^{(I)} = Q_1 R_1$

As a result, the following algorithm is used:

Input: $u_{i=1..k}$ orthonormal vectors of length m, $v_{i=1..k}$ linearly independent vectors of length m
$Q \leftarrow [0] * (m \times k)$;
$R \leftarrow [0] * (k \times k)$;
for $i \leftarrow 1$ k **do**
  $Q_{:,i} \leftarrow u_i$ **for** $j \leftarrow 1$ k **do**
    $R_{i,j} \leftarrow \langle u_i, v_j \rangle$;
  end
end
Output: Q,R

Algorithm 7: QR Decomposition

Bibliography

[Par19] Thomas Denimal Joël Gehim Tetiana Bovkum Colin Parellier. *MS Big Data, Compressive Sensing.* Télécom ParisTech, 2018-2019. Chap. 1-3.

[YC 93] P.S. Krishnaprasad Y.C. Pati R. Rezaiifar. *Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition.* 1993. URL: https://ieeexplore.ieee.org/document/342465.

[Cas08] Trevor Park George Casella. "The Bayesian Lasso". In: *Journal of the American Statistical Association* 103.482 (2008). URL: http://archived.stat.ufl.edu/casella/Papers/Lasso.pdf.

[Wak08] Emmanuel J. Candès Michael B. Wakin. *An Introduction to Compressive Sampling.* 2008. URL: https://authors.library.caltech.edu/10092/.

[Teb09] Amir Beck Marc Teboulle. *A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems.* 2009. URL: https://people.rennes.inria.fr/Cedric.Herzet/Cedric.Herzet/Sparse_Seminar/Entrees/2012/11/12_A_Fast_Iterative_Shrinkage-Thresholding_Algorithmfor_Linear_Inverse_Problems_(A._Beck,_M._Teboulle)_files/Breck_2009.pdf.

[Nat15] United Nations. *Sustainable Development Goals.* 2015. URL: https://www.un.org/sustainabledevelopment/. (accessed: 25.06.2020).

[Saf19] Safran. *Group — Safran at a glance.* 2019. URL: https://www.safran-group.com/group-0. (accessed: 25.06.2020).