Vibration analysis of bearing mechanisms for predictive maintenance

Antony Davi - Co-written with Maria Paula Nascimento

November, 2022

Contents

1 Introduction

Production lines usually do not rely on idle machines. In theory, the shutdown of a machine means a loss of capital for the company concerned. This is why various maintenance methods have been developed since the 20th century to ensure the proper functioning and optimization of industrial production. One of the branches of development, predictive maintenance, focuses on anticipating serious machine problems through constant monitoring. This logic is particularly interesting when it comes to rotating machines. Bearing parts play an essential role. While they are important, their fragility is also a source of concern.

Nevertheless, the extraction of mechanical vibration data and its subsequent analysis, widely used nowadays, can make it possible to detect and prevent possible failures due to bearing wear.

Based on the knowledge and analysis of vibration phenomena measured by specific devices, it is possible to detect excessive vibration changes in equipment, providing diagnosis and failure trend analysis. This signal processing work is part of a logic of preventive maintenance, a major aspect in the industry of our time.

The objective of this work is therefore to create a system for analyzing the mechanical vibration of bearings in an efficient, ergonomic and automatic way to enhance preventive maintenance.

1.1 Objectives

The main objective was to:

- Propose an automatic detection system
- Compare different methods
- Make the system ergonomic
- Have a documented approach

Secondary objectives:

- Link the capture card remotely
- Have a server to store data
- Establish a predictive approach

2 Theory

2.1 Spectral approach to a gear [4] :

The gear signal is modulated in amplitude and frequency by both a periodic signal of period equal to the rotation period of the pinion, and a periodic signal of period equal to the rotation period of the wheel. In general, frequency modulation is much less important than amplitude modulation.

The model established by C. Capdessus gives us the following formula:

$$s_e(t) = \sum_{n=-\infty}^{+\infty} s_c(t - n\tau_e)\cdot\left(1 + \sum_{m=-\infty}^{+\infty} s_{p_1}(t - m\tau_{p_1}) + \sum_{p=-\infty}^{+\infty} s_{r_1}(t - p\tau_{r_1})\right) \qquad (23)$$

avec
- τ_e : période d'engrènement
- $\tau_{p_1} = N_1\tau_e$: période de rotation du pignon et N_1 est le nombre de dents du pignon
- $\tau_{r_1} = N_2\tau_e$: période de rotation de la roue et N_2 est le nombre de dents de la roue
- $s_c(t)$: signal d'engrènement
- $s_{p_1}(t)$: signal périodique de période τ_{p_1} induit par la rotation du pignon
- $s_{r_1}(t)$: signal périodique de période τ_{r_1} induit par la rotation de la roue

The most striking feature of the signal is the amplitude modulation due to the rotation of the wheels.

The spectrum will be composed by a family of frequency lines $kv_e = \frac{k}{t_e}$ due to the fundamental and harmonics of the gear signal (Figure 17). This family of lines is spread over a large part of the spectrum, because the nature of the gear signal is broadband. In addition, amplitude modulation results in the presence of sidebands

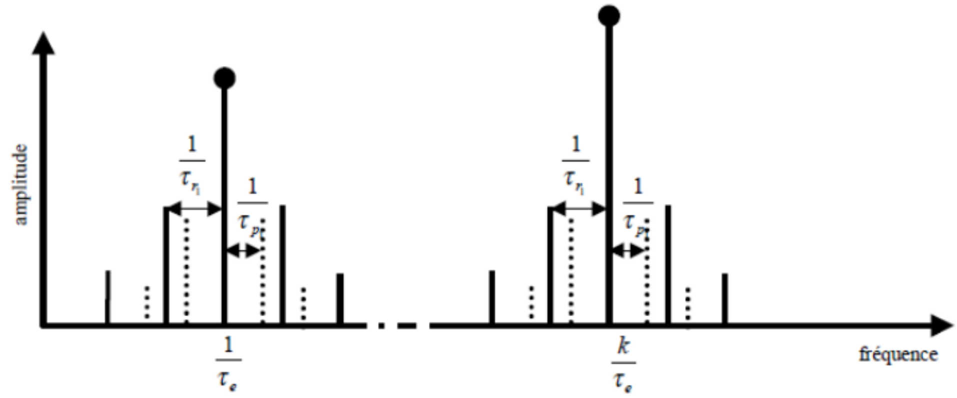


Figure 17: Spectre du signal vibratoire d'un engrenage

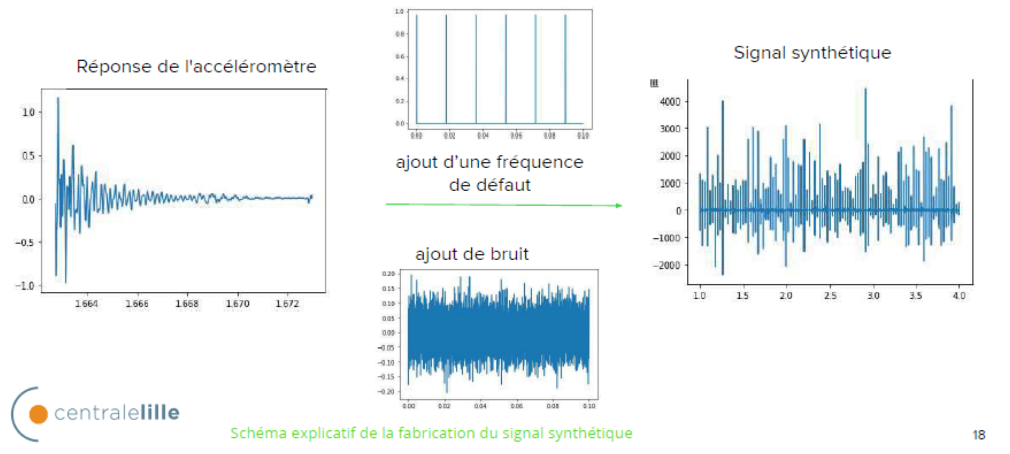
around the gear harmonics, at multiple distances of $\frac{1}{t_{p_1}}$ for pinion modulation, and $\frac{1}{t_{r_1}}$ for wheel modulation (Figure 17).

Manifestation of defects:

Consider a gear, if the teeth are correct, the vibrational spectrum will have the same pace as that defined in Figure 17, with side bands of given amplitudes. If one of the two wheels has a deteriorated tooth, then a periodic shock occurs at the frequency of rotation of this wheel. This shock will modulate in amplitude the gear signal, so there will be an increase in the modulation factor of the wheel considered, and therefore an increase in amplitude of its lateral lines.

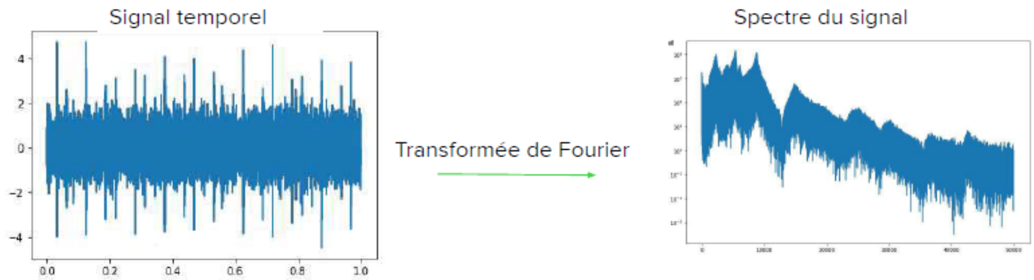
2.2 Envelope frequency analysis:

The envelope detection process has been coded and implemented in python. It is specified in the third part in the face of a concrete case. The signal is parasitized, we want to find the non-stationary frequencies (Gaussian).

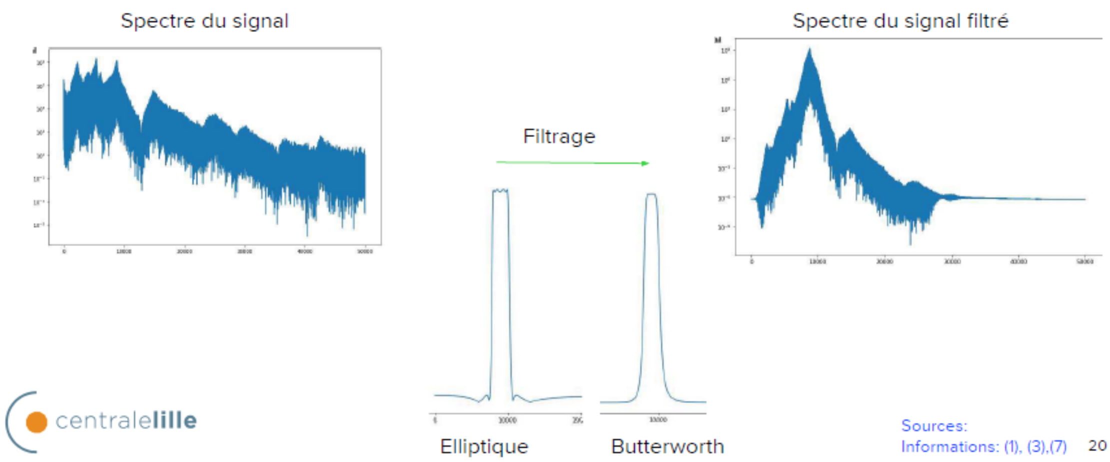


It is organized in several stages:

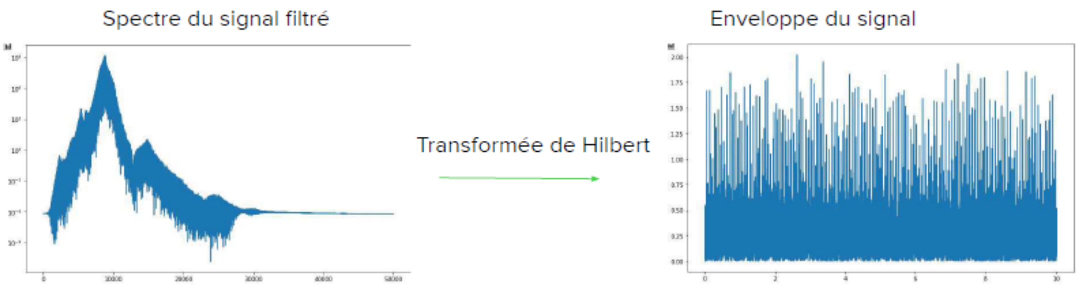
- Fourier transform of the input time signal and identification of the resonance frequencies of the machine.



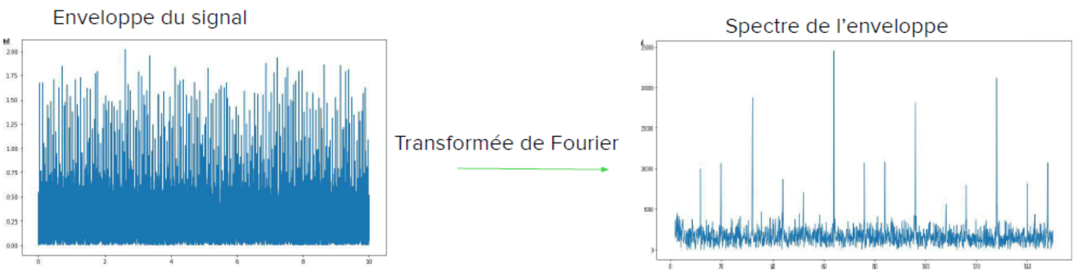
- Signal filtering.



- Hilbert transform .



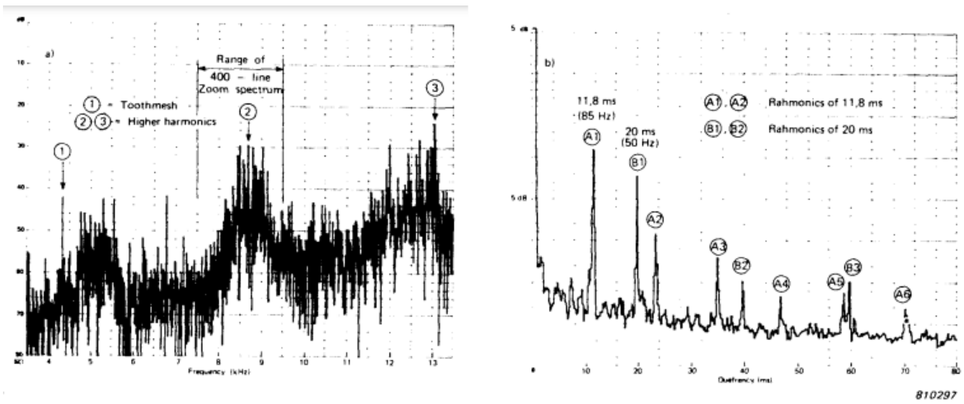
- The Fourier transform of the envelope.



Without preprocessing the time signal, it makes it possible to identify the most glaring defects.
These are at the origin of peaks in the spectrum of relatively high amplitudes. This is a safe approach for periodic fault detection[5].
Implemented in the solution, it is coupled with a cepstral analysis that highlights the periodic frequencies of abnormally high energy.
The Cepstral analysis specified below coupled with a statistical selection is a first step towards the identification of defects.

2.3 Cepstrum approach [4] :

The cesptrum was introduced as a means of detecting echoes. It is known that in rotating mechanical systems, their behavior is periodic and failures generate a system of multiple echoes. The unit of measurement normally used is the quefrency, because it is the spectrum of a spectrum. As a practical application to understand the concept, an example is used from the precursor of the subject, Randall.



The figures above show respectively a power spectrum calculated on an accelerometer signal from a gearbox and the energy cesptrum. Theoretically, the resolution of the energy cesptrum is all the more efficient as the frequencies are low. Secondly, it is much more effective when we have a history of the gear, knowing then the amplitudes of working state and failure.

Note that the amplitude of the peak corresponding to the pinion has increased to the detriment of the amplitude of the peak corresponding to the wheel. This finding is theoretically justified by the fact that there is an increase in signal energy generated by the pinion.

In summary, the energy cesptrum highlights harmonics and fundamentals. The great advantage of this method is that it allows the automatic detection of defects of a periodic nature. Indeed, it is complicated to detect harmonics automatically. Even if we spot frequency peaks, these may be random on the one hand. On the other hand, defects with lower amplitudes are difficult to detect on the side of the spectrum unlike the cesptrum.

These are two reasons for the choice of the application of the cesptrum to the wavelet spectrum.

2.4 Temporal analysis of signals by the wavelet method: Detecting pulses[5]

The main problem with the envelope method is that it does not adapt to the energy of a signal for a particular frequency range. It then makes it possible to detect defects with the highest resonances but notes two main disadvantages:

- It requires human intervention, it is necessary to read the peaks of frequencies with the naked eye.
- It hides lower frequency peaks but still not Gaussian (so potentially defects).

Regarding the first point, the cepstrum is a track to automate peak detection, and in practice it works.

Regarding the second point, it is also possible to detect defects by **distinguishing shock waves and sinusoidal resonances.** The failure having the particularity of being a shock wave, it is not silly to detect shocks within the time signal, this being allowed by the wavelet transform.

It was decided to perform the continuous wavelet transform on the characteristic frequency ranges of the defects and then evaluate their envelope and judge a defect or not by calculating the spectral kurtosis.
The wavelet transform (OT) is interpreted as an adapted multi-scale filtering aimed at finding the moments when the signal most closely resembles a known form "à priori" and this for different dilated versions of this form. Thus, it adapts the size of the analysis window to the local characteristics of the signal: small window when the signal varies rapidly and larger window when its variations are slow.

Here is the greatness that interests us:

$$C_w(a,b) = \int_{-\infty}^{+\infty} s(t)\ \overline{\psi_{a,b}(t)}\ dt$$

It corresponds to the scalar product of the signal and wavelet such that:

$$\Psi_{a,b}(t) = \frac{1}{\sqrt{a}} \Psi\left(\frac{t-b}{a}\right)$$

With the coefficient of proportionality being a coefficient used to have the same energy in each of the analyzing wavelets. The notion of frequency is replaced by the notion of scale and these are inversely proportional.
Such quantities transmit direct information about the temporal and frequency properties of the signal. They therefore make it possible to accurately identify the appearance of a given frequency at a given time in a signal. The choice of the morlet wavelet was directed by [5]. In this paper, they use wavelet transform to identify pulses and calculate the kurtosis of identified pulses to establish a verdict on the severity of the defect.
Thus, the method used is as detailed in the article, the temporal and spectral kurtosis will then be calculated in order to detect defects not listed and / or impulsive.

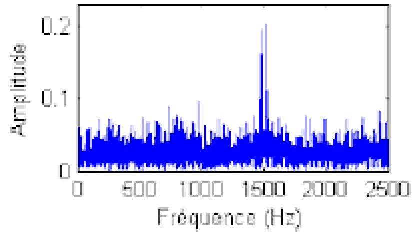


***Figure 3.** Le spectre du signal simulé*

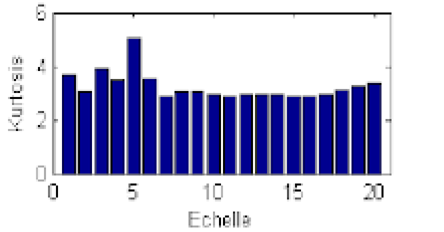


***Figure 4.** Le Kurtosis des coefficients (pas=1)*

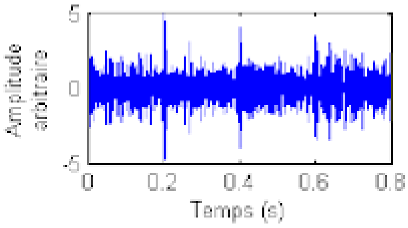


***Figure 5.** Les coefficients d'ondelettes à l'échelle 5*

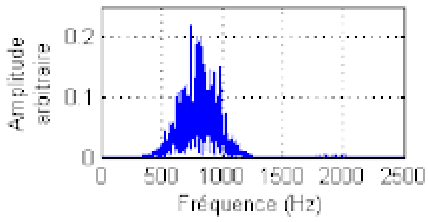
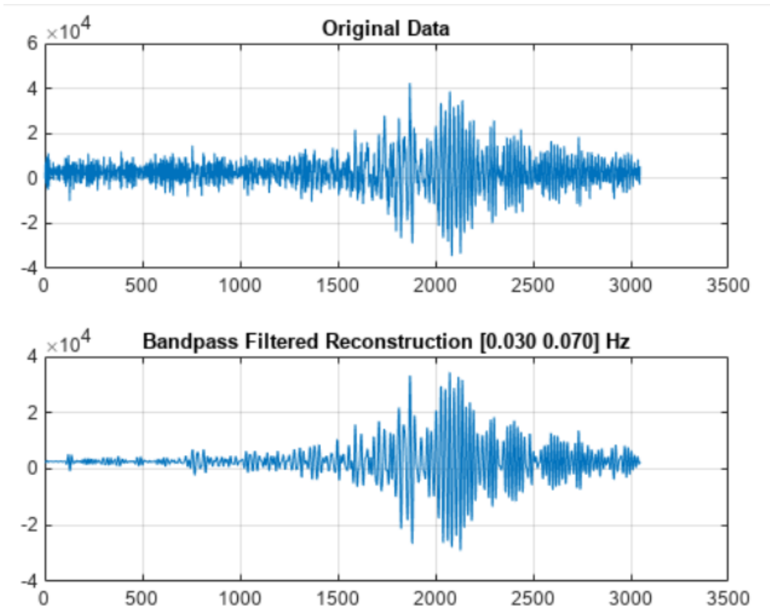


***Figure 6.** Le spectre des coefficients d'ondelettes à l'échelle 5*

Source [5]

The method was developed via matlab [7]. The principle is to perform the wavelet transform of the signal centered on a defined frequency range and then perform the reverse transform in order to reconstruct the signal.



Source: MatLab

2.5 Application of Kurtosis in the identification of impulsive signals:

The advantage of the envelope method is to be able to focus on a very precise range of frequencies and then reduce the surrounding Gaussian noise. The envelope method applied to this new filtered time signal allows for a better visualization of the frequency distribution centered around the defect.

The spectrum obtained can be analyzed using statistical tools to evaluate the curvature (kurtosis) of the frequency peak analyzed to evaluate its stationarity.

The choice to perform a wavelet transform on the Hilbert transform is detailed below.
The reason for the evaluation of kurtosis is based on sources [5] and [6]:

- Articles mention in particular the importance of wavelet transformation in the accuracy of the calculation of kurtosis.
- They also show that defect detection makes sense when it can be assigned a severity scale: the associated kurtosis.

"In several works kurtosis has shown itself to be more sensitive than other scalar indicators, which places it as a preferred indicator when it comes to shock-type defects, especially those of bearings and gears" [6]

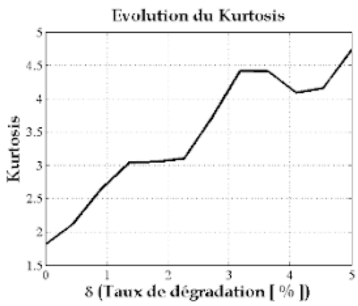
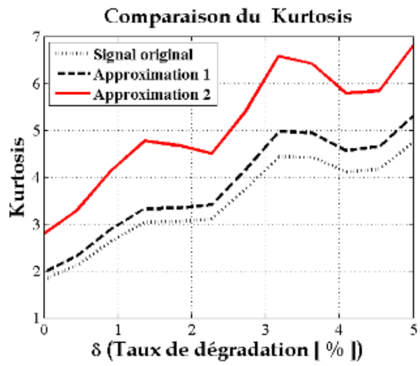


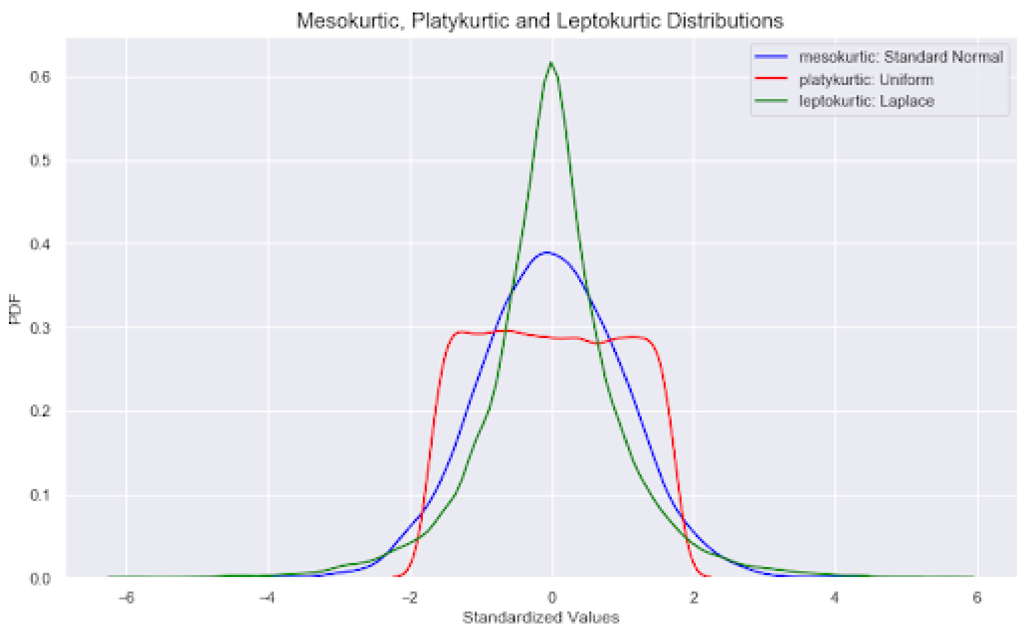
Figure 3: Évolution du Kurtosis - Signal similé bruité



Kurtosis, which is used for the search for resonances, is given by the following formula:

$$K_{urt} = \frac{\frac{1}{N} \sum_{i=1}^{N} \left(s(i) - \bar{s}\right)^4}{\left[\frac{1}{N} \sum_{i=1}^{N} \left(s(i) - \bar{s}\right)^2\right]^2} \quad (\bar{s} \text{ la moyer} \quad (4)$$

Calculating the kurtosis of a distribution is equivalent to evaluating its vertical elongation.
Here is an indicator of the value of kurtosis according to the distribution of points considered.



Two distributions in particular are used in defect detection:

- The temporal distribution of a signal, we speak of temporal kurtosis (detailed in appendix). Its calculation and interest are detailed here [4] p50.
- The spectral kurtosis of the frequency band analyzed, specified here [8].

In the final evaluation method, the spectral curvature of the frequencies is evaluated. This choice was guided by numerous failures in the evaluation of temporal kurtosis which had very few characteristic features when applied to a defect.
Spectral kurtosis has the particularity of being able to detect defects without knowing the associated characteristic frequencies [8]. However, it is not exactly the kurtosis implemented in matlab that is used but another version revisited and more personalized because less ambitious since we do not want to detect unknown defects. The objective being to know if yes or no, the defect identified during the cesptrum is indeed a defect.

Indeed, it aims to identify whether the frequency peak associated with the defect studied has a curvature characteristic of a frequency peak.
The main advantages of kurtosis are:

- To detect defects not listed via spectral kurtosis [8]
- Identify impulsive signals via temporal kurtosis [4]

However, this method, although implemented, has not borne fruit on already known defects of sinusoidal and non-impulsive character.
The rest of the project may focus on the deepening of these two methods but for the achievement of its main objective, it was chosen to opt for another approach to validate the defect character or not of a peak identify on the cepstrum.

2.6 Evaluation of the curvature index of a frequency peak: Is it a defect or not?

The approach is from here personal but has shown its proof:
Within the envelope spectrum of a signal, there are two types of defects:

- Those with Laplacian curvatures.
- Those with Gaussian curvature.

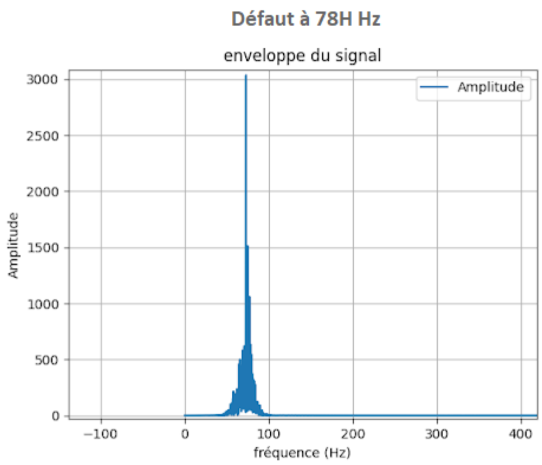
If a frequency peak can be modeled by one of its curves, then we evaluate whether it is leptokurtic tendency by evaluating the curvature by the following formula [11]:

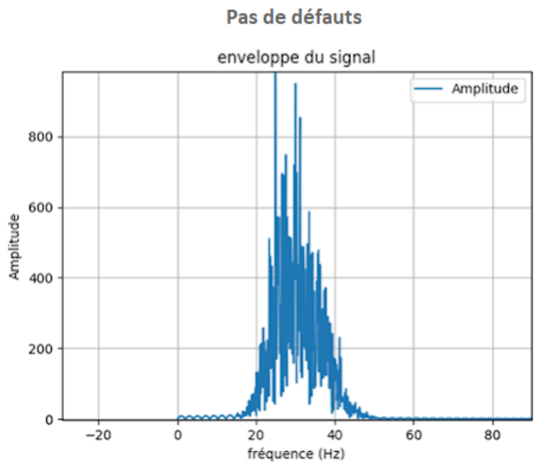
$$\kappa = \frac{|f''(x)|}{\left(1 + [f'(x)]^2\right)^{\frac{3}{2}}}$$

Then for a Gaussian curve of type $X \sim N(\mu, \sigma^2)$, the curvature at the highest point will be proportional to: $\frac{D}{\sqrt{2 \times \Pi \times \sigma}}$
For a Laplacian curve of type $Laplace(\mu, b)$, of scale b, k will be proportional to: $\frac{D}{2 \times b^3 \times \frac{1+D^2}{4 \times b^4}^{\frac{3}{2}}}$

(With D a coefficient related to the height of the peaks.)

Two empirical cut-off values of curvature are defined which then consider the frequency as a defect if its curvature (depending on its type) exceeds one of these values.
To better understand, let's compare a frequency associated with a bearing defect at 78 Hz and a frequency that is not (after wavelet filtering) at the same scale:



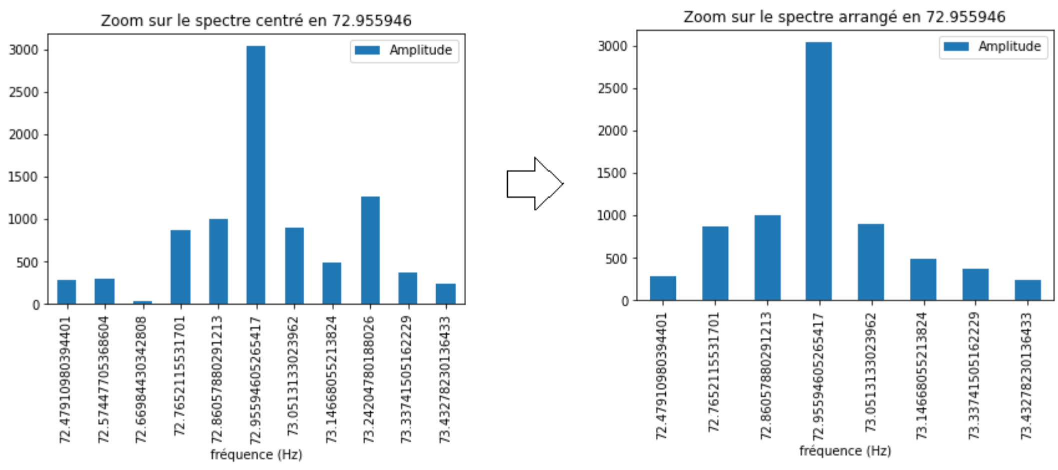


Two major differences:

- The amplitude of a defect tends to be particularly high.
- The distribution of frequencies over a small range around the characteristic frequency is very tight: it is kurtosis.

It is then possible to approximate by a Laplace law (in this case), the distribution of points over a range of one Hz around the evaluated frequency.

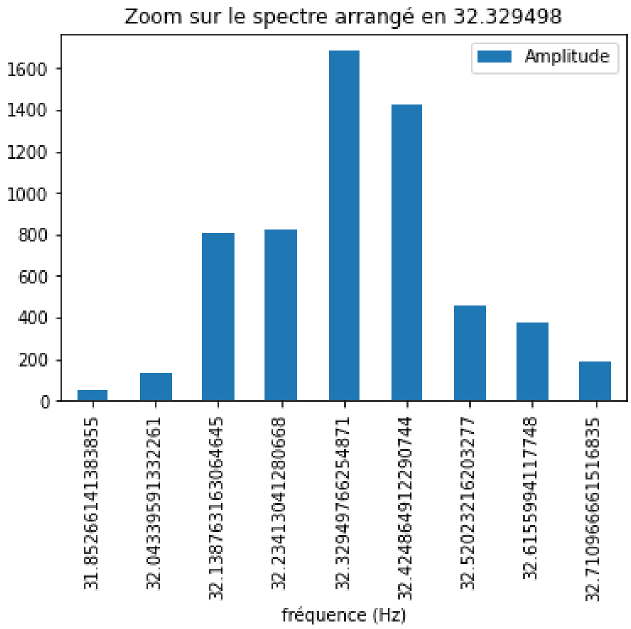
- Zoom on the fault frequency over a range of one Hz and elimination of modulation frequencies to best approach the signal.

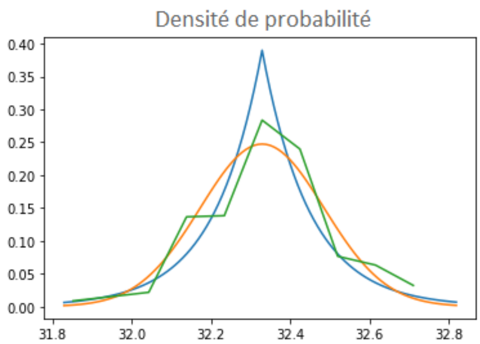


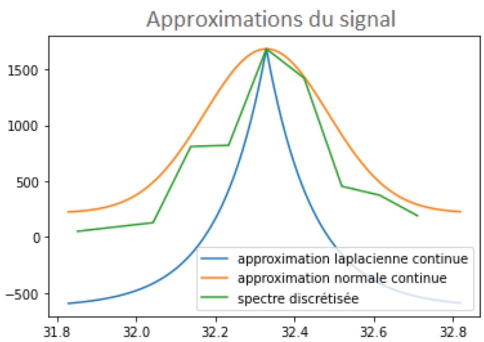
*Certaines fréquences ont été retirées.

Then approximation by a Laplace law:
If the two curves are too far apart, the approximation is rejected and the algorithm tries again by taking fewer frequencies around, if this is still the case the default hypothesis is rejected. When approximation is possible, we evaluate the curvature of the distribution , if the threshold value is exceeded: This is a defect. Otherwise it is not.

- For the non-defect, the approximation will be Gaussian but its curvature will be below the proposed threshold. For a Gaussian peak (default at 32 Hz):







The kappa returned is from 18296. The approximations then identify the fundamentals and harmonics of the defects correctly. The approximations are based on the model proposed by scipy [9].
Thus, by identifying the curvature of each supposed defect, we are able to eliminate the supposed defects that are not and automatically find all the defects.
As indicated, the curvature of each defect is also kept to detect their aggravation.

2.7 Reflection journey

Intellectual paths related to signal processing respond to:

- Why go through the cesptrum to read the peaks? [10]
- Why perform the wavelet and not rely solely on the vine? [5] and [6]
- When to perform the wavelet? The cesptrum?
- Why not calculate directly on the original envelope the curvature index rather than going through the wavelet?
- Discrete or continuous wavelet? [5] and [6]
- Why choose one such kurtosis over another? (spectral / temporal)
- Why calculate kappa instead of kurtosis?

3 Code in action

The code follows this schematic to detect the listed defects.

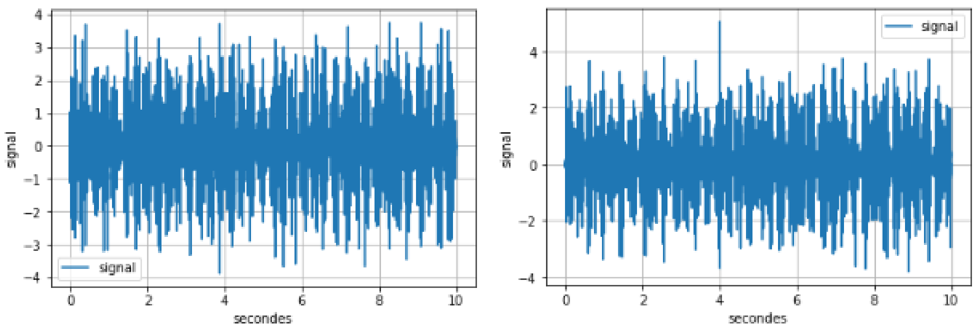
- We obtain the spectrum of the envelope of the time signal.
- We identify via its cesptrum the frequencies likely to be defects by comparing them to the frequency of known defects.
- It is then necessary to check that these are indeed defects.
- The curvature is evaluated one by one of each assumed defect.
- We turn over the real defects as well as their curvature to analyze their evolution.
- The wavelet method has been implemented and allows to analyze spectral and temporal kurtosis which are tools with different vocations.

3.1 Signal processing

3.1.1 Envelope method

The envelope method consists of filtering the time signal into bandpass and performing the Fourier transform, highlighting only the defects of the bearing (high frequency).

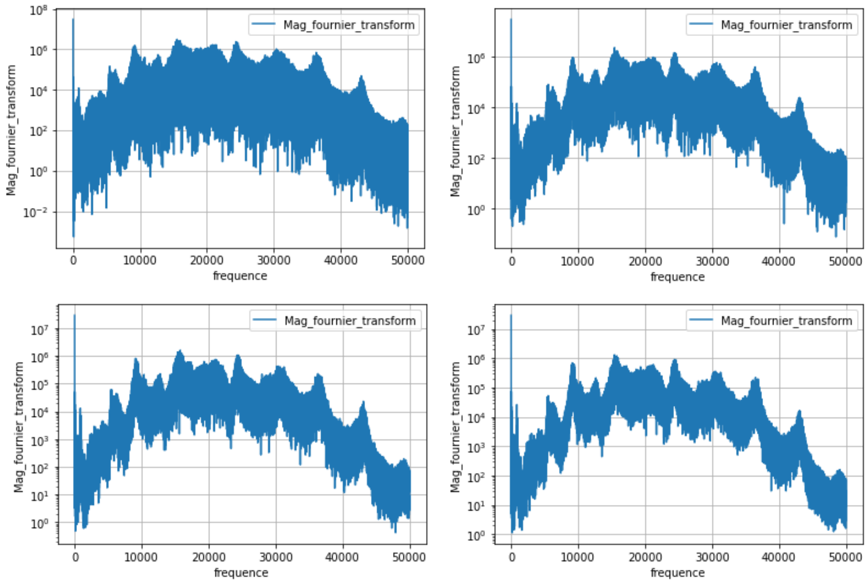
Application: After the accelerometer response and as said before, the signal must pass through a bandpass filter in order to eliminate unwanted (low frequency) signals, which can act as noise and impact the robustness of the method analysis.



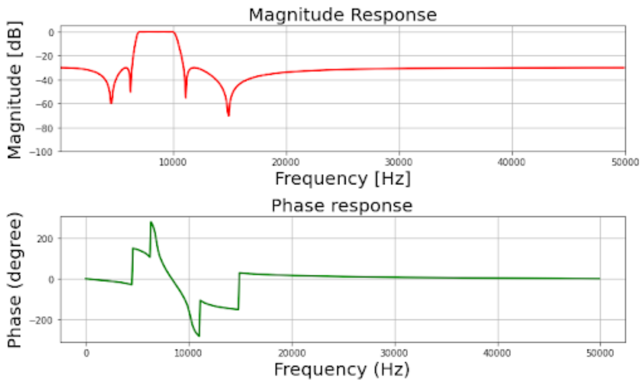
Also in terms of filtering, the Fourier transform is used to calculate the signal spectrum to obtain its frequency profile. The graph of the amplitude of the vibration as a function of frequency will allow conclusions to be drawn. With filtering at the relevant frequency level, defects, which have periodic behavior, become more visible. The Fourier transform is governed by the following function:

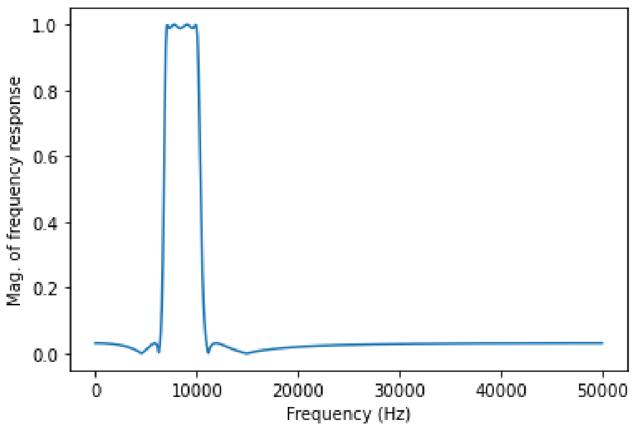
$$\hat{f}(\omega) \equiv F(\omega) \equiv \mathcal{F}\{f(t)\} \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} f(t)\, e^{-i\omega t}\, dt$$

In code, the **prefiltering** function is responsible for performing this step.

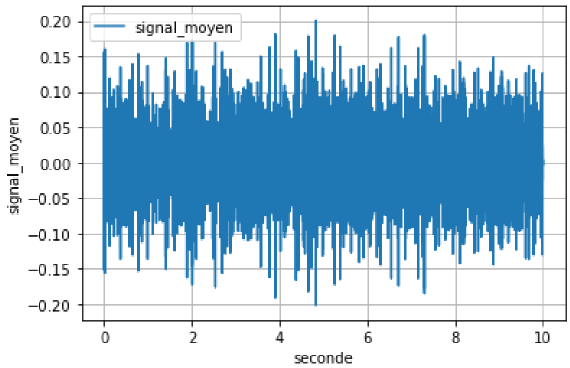


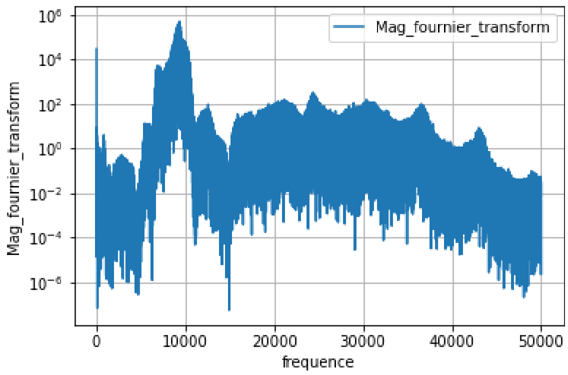
Due to the resonance of the system between the frequencies, 7,000 and 10,000 Hz, this range becomes the reference for analysis.





By filtering the signal:

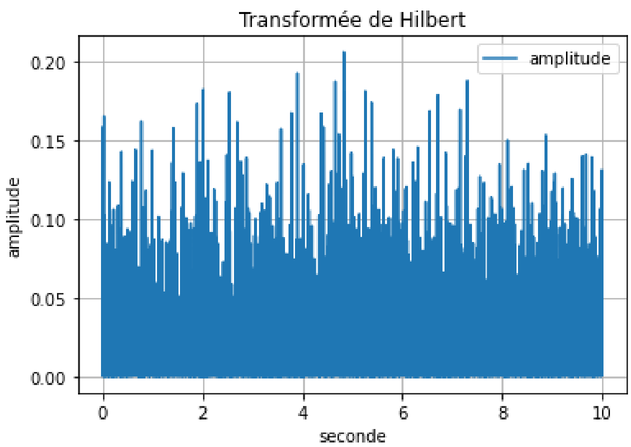




With the result, it will be possible to start the analysis of the envelope with the Hilbert transform, in order to demodulate the signal.
The Hilbert transform of a function is defined by:

$$\mathcal{H}\{f(x)\} = \hat{u}(x) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{f(u)}{u - x} \, du$$

Then a spectral analysis of the complex envelope resulting from the previous transformation is performed, so that the Fourier transform is used again. In code, the **envelope** function is responsible for performing this step.



Faults have their own frequency characteristics depending on the type of bearing used. The resulting values are in the *detect.default* function. The value of the working uncertainty is 0.5 Hz.

3.2 More on the code written ...

The code is available on the github , there is also a powerpoint specifying all the steps.

References

[1] Chapelot M. (EMS) et Richard A., consultants au CETIM surveillance des machines tournantes, guide d'achat Mesures N° 757, septembre 2003.

[2] Baptiste Trajin. Analyse et traitement de grandeurs électriques pour la détection et le diagnostic de défauts mécaniques dans les entraînements asynchrones. Application à la surveillance des roulements à billes. Automatique / Robotique. Institut National Polytechnique de Toulouse - INPT, 2009. Français.

[3] Marie-Line Zani. Tendance Les roulements, des composants à surveiller de près, guide d'achat Mesures N$^{\underline{o}}$ 754 - avril 2003.

[4] : Mohamed El Badaoui, Contribution au Diagnostic Vibratoire des Réducteurs Complexes à Engrenages par l'Analyse Cepstrale. 2008.

[5] : K Belaid, A Miloudi, M Silmani. Utilisation du Kurtosis dans le diagnostic des défauts combinés d'engrenages par la transformée continue en ondelettes

[6] : M. Merzoug , A. Miloudi. Amélioration de la sensibilité du Kurtosis en utilisant le débruitage par ondelettes

[7]: Frequency- and Time-Localized Reconstruction from the Continuous Wavelet Transform. https://fr.mathworks.com/

[8]:Rolling Element Bearing Fault Diagnosis. https://fr.mathworks.com/

[9]: scipy.stats.laplace. https://docs.scipy.org/

[10] : A Short Tutorial on Cepstral Analysis for Pitch-tracking. http://flothesof.github.io/

[11] : Curvature. tutorial.math.lamar.ed.

[12] : Jason Champagne, chaîne yt formation vidéos

[13]: Approximations probabilistes https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.norm.html