Automatic peak detection algorithm for gearbox monitoring

Diebold Jean-Frédéric [1], Leiber Maxime [1], El Badaoui Mohammed [1,2]

[1] *Safran Tech, Rue des jeunes Bois, 78117 Châteaufort, France*
2 Jean-Monnet Saint-Etienne Univ., LASPI Lab., F-42023, Saint-Etienne, France

Please select category below:
Normal Paper ☒
Student Paper ☐
Young Engineer Paper ☐

Abstract

Health monitoring's purpose is to anticipate failures, minimize unexpected maintenance, and improve overall safety in complex systems. Gearboxes are a critical component where health monitoring can significantly improve reliability. The goal of these techniques is to detect and localize damage in its early stages, enabling proactive maintenance and preventing catastrophic failures.

In gearbox analysis, distinct rotational speeds of gear components are reflected in the spectral signature of accelerometer measurement. They provide a rich source of information for fault detection. However, extracting meaningful insights from these signals can be challenging, particularly in the presence of noise and variability.

Various approaches exist to filter measurements based on theoretical frequency components - such as time synchronous averaging - or isolate characteristic frequency peaks in spectrum. Unfortunately, these methods can be time-consuming, require fine-tuning of parameters, or rely on expert knowledge to target specific areas of interest.

We propose a Topological Peak Identification Algorithm (TPI), leveraging techniques from the Topological Data Analysis domain. This approach offers an efficient and robust method to extract peaks from a signal's spectrum. By creating a filtration and keeping groups based on persistence along the height of local maxima, TPI requires minimal parameterization and can identify subtle changes in spectral features.

We demonstrate the application of TPI to a gearbox with a tooth root crack from HUMS2023 contest dataset, associating retrieved frequencies with specific gearbox components. The results highlight the potential of TPI to enhance condition-based maintenance by providing early fault detection, reducing maintenance costs, and improving overall system reliability.

Keywords: Health Monitoring, spectral analysis, peak detection

Introduction

Health monitoring is a critical aspect of maintaining complex systems, and gearbox analysis is a prime example of its importance. The goal of health-monitoring techniques is to detect and localize damage in its early stages, enabling proactive maintenance and minimizing the risk of catastrophic failures. In gearbox analysis, distinct rotational speeds of gear components are reflected in the spectral signature of accelerometer measurements, providing a rich source of information for fault detection. However, extracting meaningful insights from these signals can be challenging, particularly in the presence of noise and load variability.

Topological Data Analysis (TDA) has emerged as a powerful tool for signal processing [1], particularly in the context of persistence based clustering [2] or shape segmentation [3]. TDA provides a robust and efficient way to extract meaningful features from complex data, with a demonstrated stability toward noise [4]. Several applications in medical domain have proven its usefulness, such as arrhythmia detection from ECG signals [5], or pedestrian movements analysis [6]. Filtrations built using TDA allow identifying emergences in scalar fields. As an example, such approach has been used for identifying 2D peaks in gaz spectrometry [7].

For crafting health monitoring indicators, many approaches aim at separating sources in the signal [8], or at finding modulations in spectral signature mapped to characteristic frequencies of parts of the system [9]. For identifying these frequencies in the spectrum, naive approaches rely on setting absolute amplitude threshold and focus on local areas of the spectrum, which makes them sensitive to noise or to spectrum baseline variations, leading to manual selection of frequencies or iterative processes.

In this article, we propose a Topological Peak Identification algorithm (TPI) based on filtration algorithms of the TDA domain. The algorithm we introduce has already been implemented for identifying peaks in temporal data [10], we use it in this paper for crafting a new vertex representation of the spectrum defined by all the identified peaks on the logarithm of the spectrum. Then, we process this set of peaks separately for building health indicators.

The proposed algorithm is presented in the next section, where we detail its application and usefulness for spectrum segmentation. We then demonstrate the effectiveness of the algorithm in a health-monitoring context.

Filtration algorithm for peak detection in spectrum

The filtration algorithm automatically generates a filtration of the spectrum and extracts its expressive frequencies, characterized by their amplitude values.

Algorithm

The proposed peak detection algorithm, adapted from the 1D version presented in [10], leverages the principles of topological data analysis to efficiently extract significant frequency components from a given spectrum. The algorithm operates on a spectrum defined as a set of points $\{(x_i, y_i), i \in [1, N]\}$, where x is the frequency axis and y is the magnitude axis.

We create groups of points named g defined as a connected set of points; its peak is the point with the highest magnitude and its saddle the one with lowest magnitude (see Fig1a). We define the border of a group as the set of the two external nodes that are directly connected to it.

The algorithm traverses all points of the spectrum in decreasing order of amplitude. When an isolated point is selected (i.e. its neighbor points have lower values and hence have not yet been selected), it corresponds to a peak and a new group is created. Therefore each peak will correspond to one group. The groups grow by absorbing all the points connected to it until two groups meet. In this situation, the group with highest peak value absorbs the second one (see Fig 1a: the orange group absorbs the green one then continues growing until meeting the blue one). This second group is recorded with following information: all its constitutive points, its creation peak, and the saddle point where it met the other group. The difference of amplitude between the creation and deactivation points of the group is its relative amplitude.

Algorithm 1 Topological Peak Identification algorithm

Initialization:
$G_{\text{active}} = \{\}$, $G_{\text{deactivated}} = \{\}$
$I = \{i \mid i \text{ is an index of } (x_i, y_i)\}$
Sort I in descending order of y_i
for all $i \in I$ **do**

  if $\nexists g \in G_{\text{active}}$ such that $x_i \in \text{border}(g)$ **then**
    add $g = \{i\}$ to G_{active}, $peak(g) = (x_i, y_i)$

  else if $\exists! g \in G_{\text{active}}$ such that $x_i \in \text{border}(g)$ **then**
    add i to g

  else if $\exists g_1, g_2 \in G_{\text{active}}$ such that $x_i \in \text{border}(g_1)$ and $x_i \in \text{border}(g_2)$ **then**
    identify group with lower peak (say g_1)
    remove g_1 from G_{active} and add it to $G_{\text{deactivated}}$, $\text{saddle}(g_1) = (x_i, y_i)$
    add points of g_1 to g_2
  end if
end for

Algorithm: Topological Peak Detection algorithm

In the end of the algorithm, all recorded groups are retrieved, and we select the groups which relative amplitude exceeds a threshold. By setting a suitable threshold (see Fig 1b), the algorithm effectively segments the spectrum into distinct regions, each representing a significant frequency component.

This approach provides a robust and efficient method for peak detection, particularly in the presence of noise. As shown in Figure 1b, local variations due to noise correspond to groups with small relative amplitudes, while the main components of the spectrum with higher relative amplitudes are isolated in the amplitude histogram. The algorithm is insensitive to baseline variations in the spectrum since the peaks are isolated by their relative amplitude, and it can easily distinguish between noise and significant spectral components.

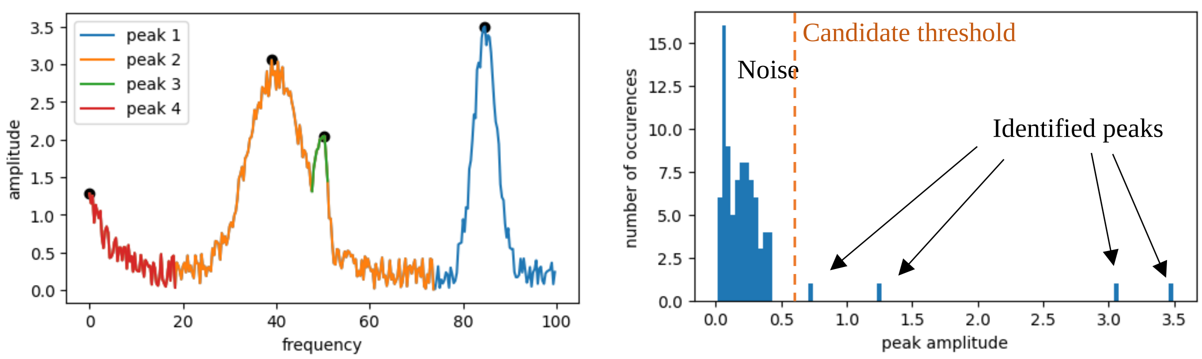


Fig 1a: segmented spectrum and identified peaks, Fig 1b: histogram of relative amplitude

Applications for health monitoring

The extracted points of the spectrum form the basis of a health-monitoring indicator, established through the mapping of these points to specific system components. When applied to the HUMS2023 dataset, this indicator demonstrates its effectiveness in following the condition of a helicopter gearbox with a propagating crack.

HUMS2023 contest dataset for gearbox monitoring

The HUMS2023 data challenge [11] presents a dataset comprising accelerometer measurements aimed at monitoring crack propagation in a planet of a gearbox. The dataset includes 518 recordings of four accelerometers, taken at various time intervals. The objective of the challenge is to detect the presence of the defect in the signal at the earliest possible stage and to accurately track its progression over time.

Peak detection on Gearbox accelerometer signals

In the following, spectrum are presented with frequency order in abscissa (with f = 1. For 1 planet rotation) and ordinate is on logarithmic scale. Upon calibration of a fixed threshold, our algorithm extracts approximately 1400 peaks per file when applied on the logarithm of the spectrum. This number remains stable across the entire dataset.

We define the set of frequencies associated to each part: $\{n.f_i, n \in \mathbb{N}\}$, where f_i is the frequency associated to the part (1. for the planets and 0.35353 for the planet carrier), and we call these sets “modulations”.

A significant proportion of the extracted points correspond to modulations of the planets and their carrier, as evident in Figure 2a. Notably, when all points from the entire dataset are superimposed, robust frequency patterns emerge, manifesting as distinct vertical lines (see fig 2b). This point representation of the spectra enables the monitoring of system evolution using a condensed representation, facilitating the development of health-monitoring indicators.

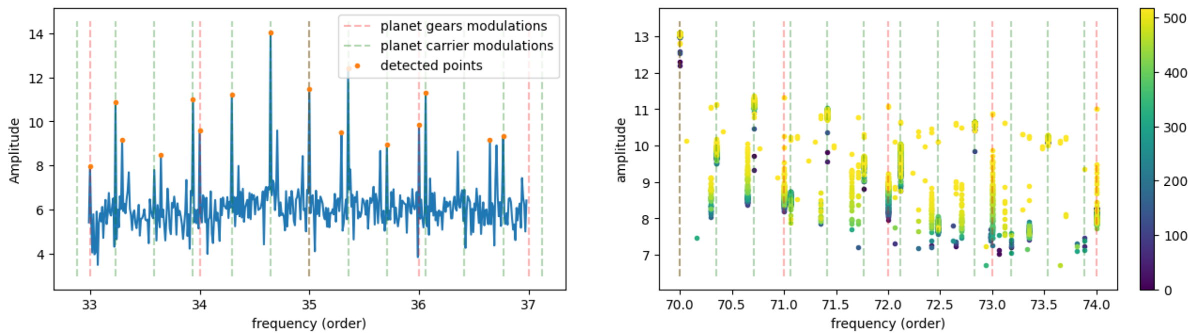
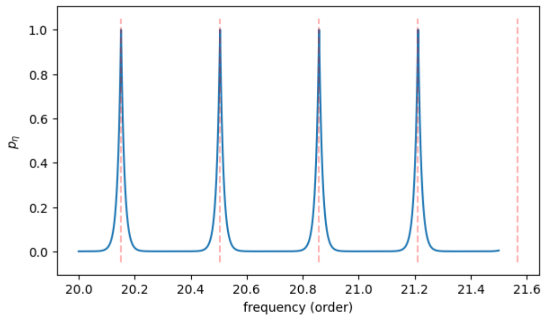


Fig 2a: peaks detected on spectrum, Fig 2b: all detected peaks with $f \in [70,74]$

Health monitoring indicator

We develop a health-monitoring indicator based on these points, grounded in the hypothesis that the amplitude distribution of these points within each file will evolve over time. Specifically, we anticipate that points corresponding to defective components will exhibit an increase in amplitude, whereas those corresponding to non-defective components will remain stable or exhibit a decrease. To operationalize this hypothesis, we divide the dataset into two groups based on the point’s amplitude values, and then identify points that can be reliably associated with the planet carrier in the high-amplitude group using a probability law p_η (see

Fig3a) and the criterion $p_\eta(x_i) > 0.999$. Notably, this criterion reveals that for all four sensors in the dataset, more than 95% of the points in the high-magnitude group can be associated to the planet carrier (see Fig3b), and this proportion decreases over time, thereby constituting a meaningful indicator of the system's health evolution.



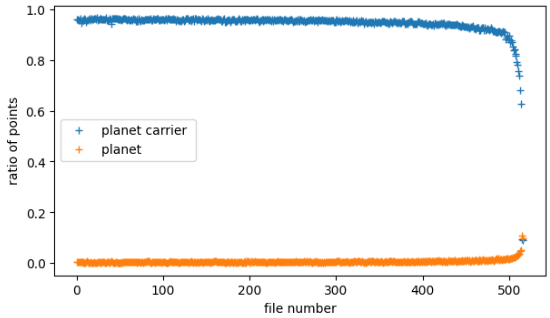
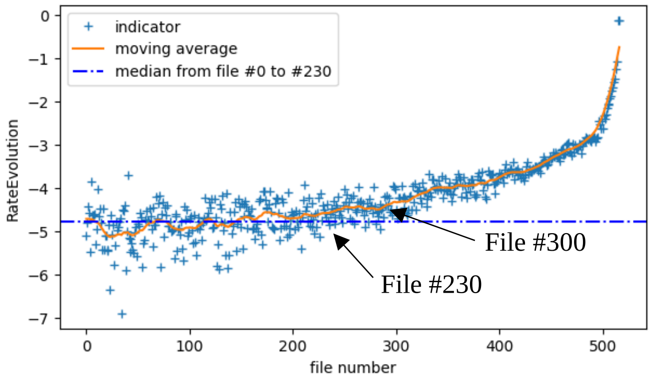


Fig 3a: probability law p_η Fig 3b: rate of points in the higher group mapped to each part

To monitor this decrease with higher accuracy, we average the rates of the four accelerometers, and then follow its evolution in logarithmic scale; this is our indicator (see Fig 4a). The indicator formula is: $RateEvolution = \log(\max_i rate_i - rate_i + \epsilon)$, ϵ is set to avoid $\log(0)$ value.



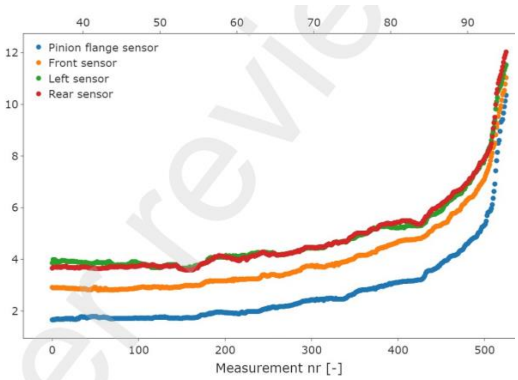


Fig 4a: RateEvolution indicator along time, Fig 4b: Blind deconvolution approach employed by the winning team of Vrije Universiteit [11]

Interpretation

The indicator increases along the test. A first notable threshold is reached at file #230, after which most of the points exceed the median of the previous ones, showing a statistical rupture in the point distribution. The moving average then increase significantly after file #300, which constitutes a second threshold.

We interpret this result as the fact that the crack generates or increases amplitude of peaks related to the planets or other frequencies, resulting in a decrease of the proportion of peaks linked to the planet carrier among the peaks with highest amplitude in the spectrum.

In contrast to energetic indicators, this indicator is more robust to variations in experimental conditions during testing (such as signal amplitude variations), as it is unaffected by scaling transform on accelerometric values since the points are detected in amplitude on a logarithmic scale and then compared to each other. Along [11], with traditional signal processing methods, earliest convincing detection from the data challenge was at file number #263 while more numerical approaches identify the crack emergence at file #175 (see Fig4b).

Conclusion

This paper introduces an approach to automatic peak detection in spectra using topological data analysis, which allows exhaustive peak extraction with a linear complexity. By identifying a large number of peaks in the spectrum, we propose a new representation of the spectra, which remains robust over time. This representation enables the creation of indicators based on the cardinal of characteristic frequencies that detect subtle changes in spectral features. Health monitoring indicators developed using this approach exhibit high robustness to energy variations during testing. For future developments, the concise representation of the spectrum enables the development of new spectrum indicators using powerful mathematical tools, such as statistical model inference or optimal transport.

References

1. Chazal, F. and Bertrand, M., "An introduction to topological data analysis: fundamental and practical aspects for data scientists." Frontiers in artificial intelligence 4, 2021.

2. Chazal, F., et al. "Persistence-based clustering in Riemannian manifolds." Journal of the ACM (JACM) 60.6, 2013.

3. Ovsjanikov, M., et al. "Functional maps: a flexible representation of maps between shapes." ACM Transactions on Graphics (ToG) 31.4, 2012.

4. Chazal, Frédéric, Vin De Silva, and Steve Oudot. "Persistence stability for geometric complexes." Geometriae Dedicata 173.1, 2014.

5. Dindin, M., Umeda, Y., & Chazal, F. "Topological data analysis for arrhythmia detection through modular neural networks." Advances in Artificial Intelligence: 33rd Canadian Conference, Proceedings 33. Springer International Publishing, 2020.

6. Beaufils, B., et al. "Robust stride detector from ankle-mounted inertial sensors for pedestrian navigation and activity recognition with machine learning approaches." Sensors 19.20, 2019.

7. Hadi, P., Christmann, J., and Weller, P., "Automated 2D peak detection in gas chromatography-ion mobility spectrometry through persistent homology." Analytica Chimica Acta 1289, 2024.

8. Soualhi, A., et al. "PHM survey: Implementation of signal processing methods for monitoring bearings and gearboxes." International Journal of Prognostics and Health Management 9.2, 2018.

9. Aherwar, A. "An investigation on gearbox fault detection using vibration analysis techniques: A review." Australian Journal of Mechanical Engineering 10.2, 2012.

10. Huber, S. "Persistent homology in data science." Data Science–Analytics and Applications: Proceedings of the 3rd International Data Science Conference–iDSC2020. Springer Fachmedien Wiesbaden, 2021.

11. Peeters, C., et al. "Fatigue crack detection in planetary gears: Insights from the HUMS2023 data challenge." Mechanical Systems and Signal Processing 212, 2024.