Tracking Multi-Modulated Periodic Sources for Rotating Machines Diagnostic

Nacer Yousfi[1,2], Karim Abed-Meraim[1,*], Maxime Leiber[2], Mohammed Elbadaoui[2,3]
[1]PRISME Lab., Université d'Orléans, [*] IUF member, France
[2] SAFRAN Tech, Dig. Sc. & Tech. Dep., Magny-Les-Hameaux, France
[3] Université de Lyon, UJM-St-Étienne, LASPI, EA3059, F-42023 Saint-Étienne, France

Résumé – Ce travail introduit une nouvelle approche pour récupérer les composantes périodiques des signaux multi-modulés. Nous proposons un algorithme des moindres carrés alternés conçu pour la récupération des signaux périodiques dans des conditions stationnaires. L'estimation initiale de cet algorithme est obtenue en appliquant une régression linéaire sur une forme linéarisée du modèle de signal. De plus, les amplitudes des sources variant dans le temps (généralement associées à l'apparition de défauts) sont mises à jour de manière adaptative à l'aide d'une technique de descente de gradient. Les résultats de simulation montrent que notre approche est efficace, atteignant la borne inférieure de Cramér-Rao. Nos travaux actuels présentent des résultats prometteurs basés sur des données expérimentales issues des signaux de vibration d'un engrenage planétaire, permettant ainsi de décomposer efficacement le signal afin de localiser et suivre la dégradation de l'engrenage défectueux.

Abstract – In this paper, we investigate a model-based approach for recovering periodic components from multi-modulated signals. We propose an alternating least squares (ALS) algorithm designed for periodic signal recovery under stationary conditions. The initial guess for ALS is obtained by running a linear regression algorithm on a linearized form of the signal model. Additionally, the time-varying source amplitudes (typically associated with fault degradation) are adaptively updated using gradient descent technique. Simulation results show that our approach is efficient as it attains the Cramér-Rao lower bound. Our current work demonstrates promising results based on experimental data from planetary gearbox vibration signal, effectively decomposing the signal to localize and track the degradation of the faulty gear.

1 Introduction

Vibration analysis methods have shown good performance in detecting various gearbox failures, as the vibration signals contain critical information regarding the gearbox's health. Rotating machine such as planetary gearbox have vibration signals that exhibit strong modulation characteristics due to the amplitude modulation effect caused by the gear faults [11],[6],[14]. The modulating frequencies are closely associated with the characteristic frequencies of gear faults. Moreover, modulation can be observed in vibration signals under both healthy and faulty conditions. This is due to manufacturing errors, time-varying vibration transmission paths, and unequally distributed loads, among other factors. These issues make it difficult to distinguish fault-related modulations from unrelated ones. Failure to address such challenges can often mislead the diagnosis and localization of faulty components. In rotating machine vibration analysis, the main challenge is isolating a set of signals (known as source separation or source extraction) related to the system's moving parts from the global vibration signal. This challenge has been widely investigated for decades in fields such as audio processing, telecommunications. Intensive research has been conducted to address this issue for linear mixtures, quadratic mixtures, and polynomial mixtures [2, 3, 4]. In health monitoring, accurately distinguishing fault-related modulation components from those unrelated to faults is essential for effective fault diagnosis and localization in rotating gearboxes. Traditional envelope analysis, Hilbert demodulation, and empirical mode decomposition (EMD) methods often struggle to separate modulation contributions from different system components in gearbox vibration signals, which can hinder the monitoring of faulty components [11], [14]. In [12] a source separation method, using a tachometer and the prior knowledge of the kinematic gearbox model, has been introduced to estimate and separate the periodic sources. This method relies on a linearization step, which results in the estimation of a large and redundant set of parameters. As a consequence, the method in [12] is inefficient, with a large loss of performance as compared to the Cramer Rao Bound (CRB).

In this paper, we propose an improved solution which solves the previous issue and meets the CRB. Both batch and adaptive implementations are proposed using alternating least squares optimization technique.

2 Problem Setup

Without loss of generality, we consider the vibration signal of a planetary gearbox system with T samples, described by the following equation [5, 13]:

$$x(t) = s(t) + n(t) \;\; t = 1, \cdots, T \tag{1}$$

$$s(t) = s_1(t)(1 + s_2(t))(1 + s_3(t)) \tag{2}$$

which represents multi-amplitude modulations (AM) between three periodic source signals $\mathbf{s}_{1\leq i\leq 3}$ with estimated (or a priori known) fundamental frequencies $f_i, \;\; i = 1, \cdots, 3$, typically related to the machine's rotation frequencies. $s_1(t)$ represents the meshing component, $s_2(t)$ the modulation at the carrier plate frequency, and $s_3(t)$ the sun gear rotation contribution. $n(t)$ is an additive white Gaussian noise with zero-mean and variance σ_n^2. Consider a stationary regime (i.e. frequencies f_i

are invariant) and let H_i be the number of 'active' harmonics [1] of the i-th source, so that it can be written as:

$$s_i(t) = \sum_{l=1}^{H_i} \sin(2\pi l f_i t) w_{i,l} + \cos(2\pi l f_i t) \tilde{w}_{i,l} \quad (3)$$

where $w_{i,l} = a_{i,l} cos(\phi_{i,l})$ and $\tilde{w}_{i,l} = a_{i,l} sin(\phi_{i,l})$ are the harmonic amplitudes (and phases) to be estimated. Given a set of observations for $t = 1, \cdots, T$, one car rewrite equation (3) in vector form as:

$$\mathbf{s}_i = \mathbf{D}(f_i)\mathbf{w}_i + \tilde{\mathbf{D}}(f_i)\tilde{\mathbf{w}}_i \quad (4)$$

where $\mathbf{s}_i = [s_i(1), \cdots, s_i(T)]^T$, $\mathbf{w}_i = [w_{i,1}, \cdots, w_{i,H_i}]^T$, $\tilde{\mathbf{w}}_i = [\tilde{w}_{i,1}, \cdots, \tilde{w}_{i,H_i}]^T$, and

$$\mathbf{D}(f_i) = \begin{bmatrix} \sin(2\pi f_i) & \cdots & \sin(2\pi H_i f_i) \\ \vdots & & \vdots \\ \sin(2\pi T f_i) & \cdots & \sin(2\pi T H_i f_i) \end{bmatrix}$$

$$\tilde{\mathbf{D}}(f_i) = \begin{bmatrix} \cos(2\pi f_i) & \cdots & \cos(2\pi H_i f_i) \\ \vdots & & \vdots \\ \cos(2\pi T f_i) & \cdots & \cos(2\pi T H_i f_i) \end{bmatrix}$$

which are the source harmonics dictionaries in stationary condition. The main focus then is estimating the Fourier coefficients $\{\mathbf{w}_i, \tilde{\mathbf{w}}_i\}$ of each source and tracking their variations, given the signal $x(t)$ and the sources' fundamental frequencies [2]. Note that, in practice, the number of active harmonics H_i is unknown and is replaced at first by an overestimation, H_{max}, as shown next.

3 Maximum Likelihood (ML) Solution with ALS Optimization

3.1 Cramer Rao Bound

Under the assumption that $n(t)$ is Gaussian, it is straightforward to write the likelihood function of the data and to derive an explicit expression for the Cramer Rao Lower Bound. [3] The estimation error of the source's parameters vectors, denoted $\boldsymbol{\theta}$, is measured via its variance [4] [8]. Let's denote the parameters vector as follows:

$$\begin{aligned} \boldsymbol{\theta} &= \left[\mathbf{w}_1^T, \tilde{\mathbf{w}}_1^T, \mathbf{w}_2^T, \tilde{\mathbf{w}}_2^T, \mathbf{w}_3^T, \tilde{\mathbf{w}}_3^T\right]^T \\ &= [\boldsymbol{\theta}_1^T, \boldsymbol{\theta}_2^T, \boldsymbol{\theta}_3^T]^T \end{aligned} \quad (5)$$

The problem at hand is fully identifiable, thereby the variance of any unbiased estimator $\hat{\theta}$ of θ has a theoretical limit called Cramer Rao Bound (CRB):

$$VAR(\hat{\boldsymbol{\theta}}) \geq CRB(\boldsymbol{\theta}) = F(\boldsymbol{\theta})^{-1} \quad (6)$$

where $F(\boldsymbol{\theta})$ is the Fisher Information Matrix (FIM) which (i, j)-th entry, in this Gaussian noise case, is given by [9]:

$$F_{i,j}(\boldsymbol{\theta}) = \frac{1}{\sigma_n^2} \left[\frac{\partial \mu(\boldsymbol{\theta})}{\partial \theta_i}^T \frac{\partial \mu(\boldsymbol{\theta})}{\partial \theta_j} \right] \quad (7)$$

where $\mu(\theta) = \mathbf{s} = [s(1), \cdots, s(T)]^T$ is the T dimensional mean vector. The derivatives involved in (7) are straightforward and are omitted in this presentation. In the sequel, the CRB is used as a benchmark to illustrate the efficiency of our estimator.

3.2 Alternating Least Squares

We propose handling the nonlinear model in (2) by leveraging an alternating least squares (ALS) approach to minimize the objective function:

$$\arg\min_{\boldsymbol{\theta}} \sum_{t=1}^{T} ||x(t) - \hat{s}_1(t)(1 + \hat{s}_2(t))(1 + \hat{s}_3(t))||_2^2 \quad (8)$$

where $\hat{s}_i = \mathbf{D}_i \hat{\mathbf{w}}_i + \tilde{\mathbf{D}}_i \hat{\tilde{\mathbf{w}}}_i$. Note that (8) represents the ML estimator in the white Gaussian noise case.

Rather than solving for all $\boldsymbol{\theta}_i$ jointly, the ALS method alternates between minimizing the least squares error for each sources' parameters individually. By doing so, one reduces the non-linear optimization problem into a set of quadratic least squares (LS) problems (with closed form solution) [1]. For an observed data vector $\mathbf{x} = [x(1), \cdots, x(T)]^T$, the ALS algorithm proceeds iteratively, updating the parameters $\boldsymbol{\theta}_{1 \leq i \leq 3}$ at each step using the closed-form analytical solution of the LS problem, as follows:

Algorithm 1: ALS (Alternating Least Squares)

Input: Initial data $\mathbf{x}$, number of iterations $epochs$
Initialization: Randomly initialize parameters $\boldsymbol{\theta}$
$k \leftarrow 0$
while $k \leq epochs$ **do**
  $k \leftarrow k + 1$
  $\boldsymbol{\theta}_1^{(k)} \leftarrow \mathbf{A}_1^{\#}\mathbf{b}_1, \quad \hat{\mathbf{s}}_1 = [\mathbf{D}_1, \tilde{\mathbf{D}}_1]\boldsymbol{\theta}_1^{(k)}$
  $\boldsymbol{\theta}_2^{(k)} \leftarrow \mathbf{A}_2^{\#}\mathbf{b}_2, \quad \hat{\mathbf{s}}_2 = [\mathbf{D}_2, \tilde{\mathbf{D}}_2]\boldsymbol{\theta}_2^{(k)}$
  $\boldsymbol{\theta}_3^{(k)} \leftarrow \mathbf{A}_3^{\#}\mathbf{b}_3, \quad \hat{\mathbf{s}}_3 = [\mathbf{D}_3, \tilde{\mathbf{D}}_3]\boldsymbol{\theta}_3^{(k)}$
end
Return $\theta^{(k)}$

where $()^{\#}$ denotes the Moore-Penrose pseudo inverse and:

$$\mathbf{A}_1 = diag\left((1 + \hat{\mathbf{s}}_2) \odot (1 + \hat{\mathbf{s}}_3)\right)[\mathbf{D}_1, \tilde{\mathbf{D}}_1] \quad (9)$$

$$\mathbf{b}_1 = \mathbf{x} \quad (10)$$

$$\mathbf{A}_2 = diag\left((1 + \hat{\mathbf{s}}_3) \odot \hat{\mathbf{s}}_1\right)[\mathbf{D}_2, \tilde{\mathbf{D}}_2] \quad (11)$$

$$\mathbf{b}_2 = \mathbf{x} - \hat{\mathbf{s}}_1 \odot (1 + \hat{\mathbf{s}}_3) \quad (12)$$

$$\mathbf{A}_3 = diag\left((1 + \hat{\mathbf{s}}_2) \odot \hat{\mathbf{s}}_1\right)[\mathbf{D}_3, \tilde{\mathbf{D}}_3] \quad (13)$$

$$\mathbf{b}_3 = \mathbf{x} - \hat{\mathbf{s}}_1 \odot (1 + \hat{\mathbf{s}}_2) \quad (14)$$

where $diag(\mathbf{x})$ is the diagonal matrix with vector $\mathbf{x}$ as main diagonal and $\odot$ refers to the elementwise multiplication.

3.3 Algorithm's initialization

Random initialization leads from time to time to ill convergence (i.e. convergence to local minima). Hence, to overcome this issue and to enhance the convergence rate too, we replaced the random initialization by the weights given by LASSO applied to a linearized form of (2). The LASSO-like objective function is defined as follows:

$$\arg\min_{\mathbf{w}} \frac{1}{2} \sum_t ||\mathbf{x}(t) - \mathbf{D}\mathbf{w}(t)||_2^2 + \lambda ||\mathbf{w}||_1 \quad (15)$$

1. In practice, the vibration signal periodic sources have a finite number of harmonics, effectively contributing to the considered signal.
2. The frequencies are determined based on a measure of shaft rotation speed and machine's kinematic model.
3. When the Gaussian hypothesis is not valid, the algorithm of this paper is still applicable, but it will no longer provide the Maximum Likelihood estimate.
4. Here we consider only unbiased estimators.

where $\mathbf{D}$ represents the dictionary containing all possible combinations of the sources' harmonics according to the model (2) (see [12] for details), $\mathbf{w}$ denotes the modulated harmonic weights of the sources, and λ is the ℓ_1-norm regularization penalty coefficient, which ensures the selection of the most significant source harmonics. Indeed, in practice, the number of harmonics is unknown, and so we build the dictionary $\mathbf{D}$ with an over-estimated number of harmonics for each source. However, the authors of [12] demonstrate that this approach does not achieve the theoretical Cramér-Rao bound, exhibiting a significant performance gap with the CRB, which will also be shown later. This result is not surprising, as this method estimates a quite large and redundant number of parameters, i.e., a total of $2H_1(1 + H_2)(1 + H_3)$ weight coefficients instead of $2(H_1 + H_2 + H_3)$ weight coefficients for our method.

3.4 Adaptive algorithm

When material degradations start occurring, the number of harmonics and their amplitudes and phases vary over time, in a continuous pattern[5]. Health monitoring and abnormalities detection require considering an adaptive process to estimate these time-varying parameters, initialized with the closed-form solution provided by LASSO or ALS at the initial time step.

Many possible solutions can be chosen for the adaptive algorithm. In this work, we have considered using a gradient-like method, namely ADAM optimizer[5] (see [10] for the algorithm's details), applied on a mini-batch of observations of size N, i.e. $\mathbf{x}_t = [x((tN+1), \cdots, x((t+1)N)]^T$, and initialized with the parameter value at the previous iteration.

Our adaptive algorithm is summarized in the following table:

Algorithm 2: Time-varying Parameters Estimation

Input: Observations $\mathbf{x}_0, \mathbf{x}_1, \ldots,$
Initialize: $\boldsymbol{\theta}_0 = \text{ALS}(\mathbf{D}, \mathbf{x}_0),$
for $t \geq 1$ **do**
| $\boldsymbol{\theta}_t = ADAM(\boldsymbol{\theta}_{t-1}, \mathbf{D}, \mathbf{x}_t)$
end
Return: θ_t

Algorithm 3: Time-varying Parameters Estimation

Input: Observations $\mathbf{x}_0, \mathbf{x}_1, \ldots,$
search $f_1(t=0)$
for $t \geq 1$ **do**
| $f_1(t) = ADAM(f_1(t-1), \mathbf{D}, \mathbf{x}_t)$
end
Initialize: $\boldsymbol{\theta}_0 = \text{ALS}(\mathbf{D}, \mathbf{x}_0),$
for $t \geq 1$ **do**
| $\boldsymbol{\theta}_t = ADAM(\boldsymbol{\theta}_{t-1}, \mathbf{D}, \mathbf{x}_t)$
end
Return: $\boldsymbol{\theta}_t$

4 Simulation results

4.1 Synthesized data

To assess the performance of the proposed approach and compare it against the Cramer Rao theoretical limit, we consider a synthetic signal $\mathbf{x}$ sampled at $f_s = 1000$ Hz with $T = 200,000$ samples. The signal consists of $M_1 = 3$ carrier harmonics with a fundamental frequency $f_1 = 99$ Hz. The first modulated source, $\mathbf{s}_2$, contains $M_2 = 3$ harmonics with a fundamental frequency $f_2 = 1$ Hz, and the second modulated source, $\mathbf{s}_3$, has a fundamental frequency $f_3 = 2.83$ Hz and contains also $M_3 = 3$ harmonics. To evaluate the robustness of the approach against the noise, white Gaussian noise has been added to the simulated signal $\mathbf{s}$ such that the input signal-to-noise ratio varies from -5 dB to 10 dB.

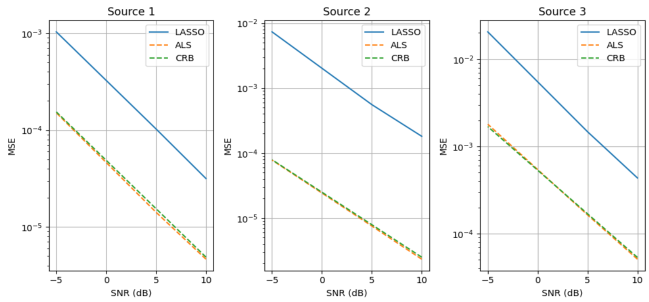

Figure 1 – Our approach-based ALS performance against the CRB with low level of SNR over 100 run.

Figure 1 shows the results of running LASSO on our model with a sparsity penalty. The mean square error (MSE) of the source parameters exhibits a large gap from the CRB. However, using LASSO as an initialization for ALS later leads to a good match with the CRB. In addition to the statistical efficiency of our solution, it also demonstrates robustness to harmonic number overestimation errors. Indeed, the plots in Figure 1 have been obtained with an overestimated number of harmonics $M = 5$ (for all sources) while the exact number was $M = 3$.

In our second experiment, we generate data with variable harmonic amplitudes and phases for both the carrier and the modulations, then run our tracking (adaptive) algorithm. ALS is first applied to an initial batch of data before attempting to track the parameters. The performance evaluation for SNR = 10 is shown in Figure 2, where we present the results over time. We conclude that starting with an optimal initialization using ALS allows for an easy gradient descent. In the figure, we observe that some parameters are initially set to zero during the first time steps (source 3), illustrating a situation of harmonic number overestimation. However, over time, they gradually increase due to degradation, and our tracking algorithm captures this behavior. This is crucial to system health monitoring.

4.2 Planetary gearbox vibration analysis

In this section, we examine the ALS approach to analyze vibration signal recorded during a fatigue test on a helicopter planetary gearbox. The test was designed to investigate the phenomenon of fatigue cracking in a planet gear. The sampling frequency is $f_s = 11583$ Hz. The characteristics of the planetary gearbox are as follows: the ring gear has 99 teeth, the planet gear has 35 teeth, the sun gear has 27 teeth, and there are 4 planet gears. Only the last 7 days have been loaded, during which we have 526 files of 100-second data acquisitions each. For more details on the dataset, refer to [7]. Our focus is to individually estimate the meshing signal $\mathbf{s}_1$ and the planet carrier related signal $\mathbf{s}_2$ and the planet gear related signal $\mathbf{s}_3$.

5. ADAM stands for Adaptive Moment Estimation.

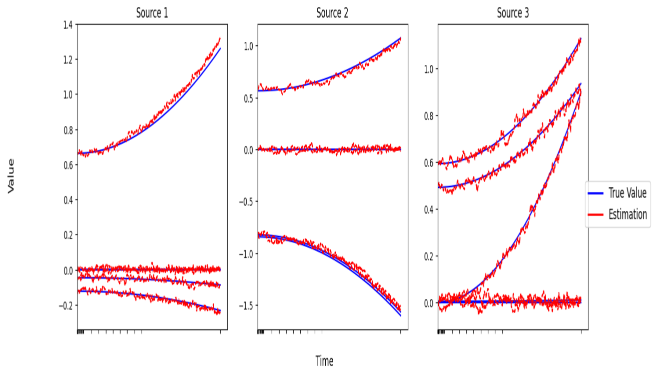

Figure 2 – Tracking of source parameters over time with $SNR = 10\ dB$. Each plot shows the tracking of five harmonics of the considered source.

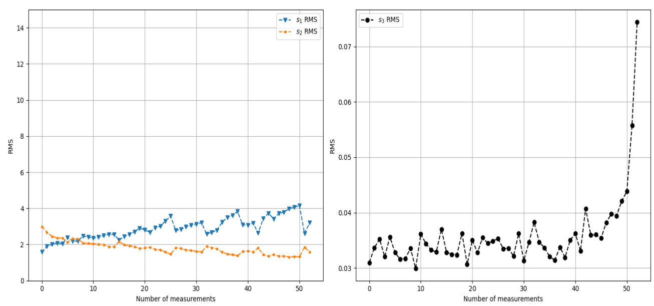

Figure 3 – The root-mean-square (RMS) of the estimated sources s_1, s_2, s_3. The modulation signal s_1 exhibits a monotonic trend correlated with the increasing planetary gear degradation, especially in the final measurements, where an artificial crack is introduced along the planetary gear. In contrast, the meshing and modulation signal s_2 fluctuate over time without a clear monotonic pattern.

We set the number of overestimated harmonics for each source s_i to 15. The vibration signal model of such planetary gearbox is the same given by (2).We performed ALS on our dataset, analyzing the measurements from 1 to 526 with an increment of 10. We then calculated the root mean squares (RMS) of the sources $s_{i \le i \le 3}$, as shown in Figure 3. The RMS of the meshing source s_1 and the carrier plate signal s_2 fluctuates. However, the RMS of the planet signal exhibits a monotonic behavior, which informs about an occurrence of fault in the planet gear that matches with the planet introduced crack. The crack increases drastically in the last days, as the HUMS23 Team [7] made the crack more severe during that period. With our simple proposed method, one can easily track the defective part while monitoring its degradation over time.

5 Conclusion

In this study, we present a simple approach of a multmodulation decomposition using an alternating least squares approach, providing a vibration signal and a set of system rotation frequencies. The extracted sources have crucial information about the related components, which could aid in early fault detection and localization. A synthetic signal was used to illustrate the statistical efficiency of the proposed approach, i.e., it attains the theoretical CRB limit. Additionally, we ran a use-case application with a planetary gearbox, where we successfully identified the defective part as the planet gear and tracked its degradation over time. However, further development is needed to extend this method to non-stationary vibration signals (e.g., variable speeds and loads). This will require incorporating a frequency estimation and tracking step alongside amplitude estimation in the nonlinear model to dynamically adapt to changing conditions.

References

[1] KJ. Astrom and B. Wittenmark. Adaptive control, (2nd edition) addison wesley. *Reading*, 1995.

[2] A. Belouchrani, K. Abed-Meraim, J-F Cardoso, and E. Moulines. A blind source separation technique using second-order statistics. *IEEE Transactions on signal processing*, 45(2):434–444, 1997.

[3] Y. Deville and LT Duarte. An overview of blind source separation methods for linear-quadratic and post-nonlinear mixtures. In *International Conference on Latent Variable Analysis and Signal Separation*, pages 155–167. Springer, 2015.

[4] Y. Deville, LT. Duarte, and S. Hosseini. *Nonlinear blind source separation and blind mixture identification: methods for bilinear, linear-quadratic and polynomial mixtures*. Springer Nature, 2021.

[5] E. Hubert. *Vibration monitoring of an aeronautic power transmission system*. PhD thesis, Lyon Univ., 2019.

[6] E. Hubert, A. Barrau, and M. El Badaoui. New multi-carrier demodulation method applied to gearbox vibration analysis. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2141–2145. IEEE, 2018.

[7] HUMS. Hums2023_data_challenge_dataset, 2023.

[8] W. Qiu K. Abed-Meraim and Y. Hua. Blind system identification. *Proc. of the IEEE*, 85(18), August, 1997.

[9] S. Kay. *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory*. 1st Ed., Pearson, 1993.

[10] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *CORR, abs/1412.6980 and also at arXiv preprint arXiv:1412.6980*, 2014.

[11] PD McFadden. Detecting fatigue cracks in gears by amplitude and phase demodulation of the meshing vibration. *Journal of Vibration and Acoustics*, 108(2):165–170, 1986.

[12] N. Yousfi, K. Abed-Meraim, Y. Marnissi, and M. EL Badaoui. A sparse coding approach for planetary gearbox vibration signal decomposition. In *The Int. Society for Air Breathing Engines*. ISABE, 2024.

[13] N. Yousfi, K. Abed-Meraim, Y. Marnissi, M. Leiber, and M. EL Badaoui. Neural network-based symbolic regression for empirical modeling of the behavior of a planetary gearbox. In *ICASSP*, 2024.

[14] D. Yu, J. Cheng, and Y. Yang. Application of emd method and hilbert spectrum to the fault diagnosis of roller bearings. *Mechanical systems and signal processing*, 19(2):259–270, 2005.