NEURAL MODE ESTIMATION

Peng Sun[1] *Zhenyu Wen*[1] *Yejian Zhou*[1] *Zhen Hong*[1] *Tao Lin*[2]

[1] Zhejiang University of Technology
[2] Westlake University

ABSTRACT

Mode decomposition methods are the current workhorse for the analysis of non-stationary signals. However, current attempts at these methods mainly focus on improving accuracy, leaving computational efficiency untouched. To this end, we leverage the neural mode decomposition technique and propose an open-source Neural Mode Estimation (NME) to deliver a large speedup (at least 50×) while maintaining the accuracy. Specifically, we transform the mode decomposition problem into an extremum problem of a functional in the cosine transform domain, and train a neural network to approximate the solution. We demonstrate in extensive empirical results that NME can provide an improved trade-off between speed and accuracy, enabling fast, high-quality, stable mode decomposition of non-stationary signals.

***Index Terms*—** Signal Processing, Machine Learning, Mode Decomposition

1. INTRODUCTION

The non-stationary signals are omnipresent in society, such as speech signals and multitone sinewave of varied frequency [1]. Characterizing the Time-Frequency (TF) information is the main task in non-stationary signal processing. Various methods [2, 3] were proposed to jointly analyze the time and frequency of a non-stationary signal. However, these methods are inflexible in practice, which requires to pre-set some parameters for processing different signals.

The mode decomposition methods have been proposed to overcome the drawback mentioned above. These methods aim to adaptively extract Amplitude-Modulated-Frequency-Modulated (AM-FM) signals (i.e., narrow-band signals) from a non-stationary signal. Huang *et al.* [4] first propose the Empirical Mode Decomposition (EMD), which decomposes an input signal into a set of narrow-band components via an iterative sifting process.

Given that EMD is still falling short of steadily deployment in practice [5] due to the lacking of theoretical foundation, end effects, sifting stop criterion, extremum interpolation, mode mixing, etc. Several followup methods [6] have been proposed to tackle these challenges. For example, EEMD [7] proposes a noise-assisted data analysis method to improve EMD in terms of mode mixing, which adds white noise to the signal to provide a uniform reference frame in the TF space. BS-EMD [8] is introduced as a mathematically amenable method, in which the knots of the B-splines are taken as the local extremum points of the signal and the envelope mean in the original EMD is replaced with the moving average of B-splines. More recently, Dragomiretskiy and Zosso propose a theoretically sound Variational Mode Decomposition (VMD) [9] that formulates a problem of seeking an ensemble of modes and their respective center frequencies as a convex optimization problem and the goal of the problem is to minimize the collective bandwidth of the signal's components subject to the total reconstruction constraint.

Despite the effectiveness of aforementioned EMD methods, these iterative algorithms require to decompose a signal into modes through a set of iterative processes, resulting in high computational complexity [10]. To this end, we propose a data-driven mode decomposition method, called NME, which performs mode decomposition for a signal in one step by neural networks instead of iterative algorithm (e.g., the Lagrange multiplier used in VMD). Specifically, we theoretically justify the feasibility of transforming mode decomposition to a problem of finding extremum conditions for a functional [11], in which the solution to this problem can be estimated approximately with parametric models. As our solution, we introduce the *Quotient Estimation* and *Recursive Inference* to approximate these models by using neural networks. In inference stage, the trained NME performs mode decomposition 251-740× faster than VMD while maintaining the same decomposition accuracy.

2. METHODOLOGY

Mode decomposition problem aims to extract K narrow-band signals (i.e., modes) from a signal, in which k-th mode is formulated as:

$$u_k(t) = A_k(t) \cdot \cos[\phi_k(t)] \tag{1}$$

Moreover, these modes are concentrated around their respective center frequencies. In order to learn a mode decomposition model, we transfer this problem to a problem of solving a functional (see Section 2.1 and Section 2.2). Then, we train a deep learning model to approximately solve this problem (Section 2.3).

2.1. Single-mode Decomposition

This subsection describes a method to solve the single-mode decomposition problem (i.e., $K = 1$) under the cosine transform domain. Assuming that an observed signal f which is consisted of a narrow-band signal $u_1^\dagger$ with a center frequency 0 and a separate Gaussian noise η, i.e., $f = u_1^\dagger + \eta$. To obtain $u_1^\dagger$, we use classical Tikhonov Regularization [12] to formalize a well-posed problem as:

$$u_1^\dagger = \underset{u_1 \in \mathbf{R}}{\arg\min}\{\|f - u_1\|_2^2 + \alpha \,\|\partial_t u_1\|_2^2\} \tag{2}$$

α denotes the regularization coefficient parameter, and $\partial_t u$ represents the varying of u with respect to t.

Transforming Eq 2 to cosine transform domain. [13] has illustrated that both sine and cosine transforms are orthogonal transforms, which can be formalized as Eq 3.

$$\|g(t)\|_2^2 = \|\mathcal{S}[g(t)]\|_2^2 = \|\mathcal{C}[g(t)]\|_2^2 \tag{3}$$

where $g(t)$ represents a certain funtion and $\mathcal{S}$ denotes the sine transform. Therefore, when we replace $g(t)$ with $f(t) - u_1(t)$ in Eq. 3, we have:

$$\|f(t) - u_1(t)\|_2^2 = \left\|\hat{f}(\omega) - \hat{u}_1(\omega)\right\|_2^2 \tag{4}$$

where the hat symbol $(\hat{\cdot})$ indicates the cosine transform (e.g., the $\hat{f}$ denotes the cosine transformation for f, and can be equivalently expressed as $\hat{f} = \mathcal{C}(f)$). Then, according to time domain differential properties of sine transform [13], when we replace $g(t)$ with $\partial_t u_1(t)$ in Eq. 3, we can obtain:

$$\alpha \,\|\partial_t u_1(t)\|_2^2 = \alpha \,\|\mathcal{S}(\partial_t u_1(t))\|_2^2 = \alpha \,\|{-\omega} \cdot \hat{u}_1(\omega)\|_2^2 \tag{5}$$

At last, we add Eq. 4 and Eq. 5, and then we get:

$$\hat{u}_1^\dagger = \underset{\hat{u}_1 \in \mathbf{R}}{\arg\min}\{\left\|\hat{f} - \hat{u}_1\right\|_2^2 + \alpha \,\|{-\omega} \cdot \hat{u}_1\|_2^2\} \tag{6}$$

Next, the $L2$ norm in Eq. 6 can be transformed to integral, then we have:

$$\hat{u}_1^\dagger = \underset{\hat{u}_1 \in \mathbf{R}}{\arg\min}\{\int_{-\infty}^{\infty} (\hat{f} - \hat{u}_1)^2 + \alpha \cdot (\omega \cdot \hat{u}_1)^2 d\omega\} \tag{7}$$

Additionally, when the frequency center of the mode $u_1^\dagger$ in the cosine transform domain shifts to ω_1 from 0, we can perform a change of variable $\omega \leftarrow \omega - \omega_1$ for Eq. 7 to obtain that $\hat{u}_1^\dagger = \underset{\hat{u}_1 \in \mathbf{R}}{\arg\min}\{J(\hat{u}_1)\}$, in which the functional $J(\hat{u}_1)$ is formulated as:

$$J(\hat{u}_1) = \int_{-\infty}^{\infty} (\hat{f} - \hat{u}_1)^2 + \alpha \cdot [(\omega - \omega_1) \cdot \hat{u}_1]^2 d\omega \tag{8}$$

To sum up, the single-mode decomposition comes down to finding the condition for the minimum of Eq. 8. Next subsection, we extend the single-mode decomposition to multi-mode decomposition.

The advantage of using cosine transform. In previous studies [9, 14], it is common to use the Fourier transform [15] to avoid differential function calculation (e.g., $\partial_t(\cdot)$ in Eq. 2). In this paper, however, we use the cosine transform to replace the Fourier transform to speedup the operation. Moreover, we show two advantages of using the cosine transform, comparing to the Fourier transform: 1) The computational load of cosine transform is lower than that of Fourier transform [16]. 2) A function can remain in the real domain $\mathbf{R}$ after cosine transform [17]. This property offers the possibility to utilize the neural network to process a signal after cosine transform, since the neural network is usually used to process data in the real domain. In section 2.3, we will provide the details.

2.2. Multi-mode Decomposition

Assume that the signal f consists of K modes with different central frequencies and a residual r as:

$$f = \sum\nolimits_k^K u_k^\dagger + r \tag{9}$$

In this paper, we aim to decouple the modes $\{\hat{u}_k^\dagger\}$ from a signal f, where $\{\hat{u}_k^\dagger\} := \{\hat{u}_1^\dagger, ..., \hat{u}_K^\dagger\}$. According to the discussion in Section 2.1, these modes can be obtained by seeking the condition for the minimum of a specific functional $J(\{\hat{u}_k\})$ as shown in Eq. 10.

$$\{\hat{u}_k^\dagger\} = \underset{\hat{u}_k \in \mathbf{R}}{\arg\min}\{J(\{\hat{u}_k\})\} \tag{10}$$

where $\{\hat{u}_k\} := \{\hat{u}_1, ..., \hat{u}_K\}$. Moreover, according to the definition of modes in Eq. 1, each mode u_k has a corresponding center frequency ω_k in the cosine transform domain. Therefore, the functional $J(\{\hat{u}_k\})$ can be regarded as a combination of multiple Eq. 8 under different center frequencies, which is formulated as:

$$\begin{aligned} J(\{\hat{u}_k\}) = \int_{-\infty}^{\infty} &(\hat{f} - \sum\nolimits_k^K \hat{u}_k)^2 \\ &+\alpha \sum\nolimits_k^K [(\omega - \omega_k) \cdot \hat{u}_k]^2 d\omega \end{aligned} \tag{11}$$

in which the center frequency ω_k is an unknown variable to be solved. Therefore, we use classical variational method [11] to solve it. Specifically, when Eq. 11 obtains an extreme value, the partial derivative of functional Eq. 11 with respect to all variables is 0. We use this condition to obtain an expression for ω_k with respect to $\hat{u}_k$:

$$\frac{\delta J}{\delta \omega_k} = \int_{-\infty}^{\infty} \alpha \cdot 2(\omega - \omega_k) \cdot (\hat{u}_k)^2 d\omega = 0 \tag{12}$$

then we further transform Eq. 12 to have:

$$\omega_k = \frac{\int_{-\infty}^{\infty} \omega \cdot (\hat{u}_k)^2 d\omega}{\int_{-\infty}^{\infty} (\hat{u}_k)^2 d\omega} \tag{13}$$

Finally, we use Eq. 13 to replace the ω_k in functional Eq. 11 to obtain a functional $J(\{\hat{u}_k\})$ as:

$$J(\{\hat{u}_k\}) = \int_{-\infty}^{\infty} (\hat{f} - \sum\nolimits_k^K \hat{u}_k)^2 + \alpha \sum\nolimits_k^K \left[(\omega - \frac{\int_{-\infty}^{\infty} \omega \cdot (\hat{u}_k)^2 d\omega}{\int_{-\infty}^{\infty} (\hat{u}_k)^2 d\omega}) \cdot \hat{u}_k \right]^2 d\omega \tag{14}$$

To sum up, the mode decomposition comes down to a problem that seeking the condition for the minimum of Eq. 14. Unfortunately, this problem has no analytic solution.

2.3. Neural Estimation of Functional

NME aims to estimate the solution to the problem mentioned above and thus decompose the a input signal f into K modes. Specifically, NME maps from a input Banach space $\mathcal{F}$ to K ouput Banach spaces $\mathcal{U}_1, ..., \mathcal{U}_K$ [11] in real domain $\mathbf{R}$. Then, we construct K corresponding parametric models $G_{\theta_1} : \mathcal{F} \to \mathcal{U}_1, ..., G_{\theta_K} : \mathcal{F} \to \mathcal{U}_K$. We aim at finding the optimal parameters $\theta_1^\dagger, ..., \theta_K^\dagger$ from some finite-dimensional parameter space Θ, so that $\{G_{\theta_k^\dagger}(f)\} = \{\hat{u}_k^\dagger\}$ for any input signal f, where $\{G_{\theta_k^\dagger}(f)\} := \{G_{\theta_1^\dagger}(f), ..., G_{\theta_K^\dagger}(f)\}$. Consequently, according to Eq. 10, we obtain:

$$\min_{\{\hat{u}_k\}} \{J(\{\hat{u}_k\})\} = J(\{\hat{u}_k^\dagger\}) = J(\{G_{\theta_k^\dagger}(f)\}) \tag{15}$$

Moreover, for any $\theta_1, ..., \theta_K \in \Theta$, the following inequality should always hold:

$$J(\{G_{\theta_k^\dagger}(f)\}) = \min_{\{\hat{u}_k\}} \{J(\{\hat{u}_k\})\} \leq J(\{G_{\theta_k}(f)\}) \tag{16}$$

where $\{G_{\theta_k}(f)\} := \{G_{\theta_1}(f), ..., G_{\theta_K}(f)\}$. Suppose we have observations $\{f_i\}_{i=1}^N$ (i.e., training set), where $f \sim \mu$ is an i.i.d. sequence from the probability measure μ supported on $\mathcal{F}$. Consequently, according to Eq. 16, we obtain:

$$\mathbf{E}_{f\sim\mu}\{J(\{G_{\theta_k^\dagger}(f)\})\} \leq \mathbf{E}_{f\sim\mu}\{J(\{G_{\theta_k}(f)\})\} \tag{17}$$

Therefore, we design a loss function $\mathcal{L}$ according to Eq. 17, and make the parameters $\theta_1, ..., \theta_K$ approach the optimal parameters $\theta_1^\dagger, ..., \theta_K^\dagger$ by minimizing this loss function:

$$\min_{\theta_k \in \Theta} \{\mathcal{L}\} = \min_{\theta_k \in \Theta} \mathbf{E}_{f\sim\mu}\{J(\{G_{\theta_k}(f)\})\} \tag{18}$$

In the following, we describes how to construct K parametric models $\{G_{\theta_k}(f)\}$ by neural networks $T_{\theta_1}, ..., T_{\theta_K}$. A plain idea is that use K neural networks as these models directly (i.e., let $G_{\theta_k}(f) = T_{\theta_k}(f)$ for each k). However, we find a problem that such models can not accurately perform mode decomposition. This problem comes from two aspects: 1) These models estimate the modes independently without supporting each other. 2) The solution is too complex to estimate directly. To tackle these issues, we exploit *recursive inference* to let a certain model refer to the modes decomposed by other models when performing mode estimation, and then develop a *quotient estimation* to simplify the estimation.

1) Recursive Inference. According to Eq. 3, we have the following transformation.

$$f = \sum\nolimits_k^K u_k^\dagger + r \Rightarrow \hat{f} = \sum\nolimits_k^K \hat{u}_k^\dagger + \hat{r} \tag{19}$$

Therefore, it inspires us to utilize recursive inference to perform mode estimation. According to Eq. 19, when we get $\hat{u}_1^\dagger$, we can use $\hat{f} - \hat{u}_1^\dagger = \sum_{k=2}^K \hat{u}_k^\dagger + \hat{r}$ as input to estimate $\hat{u}_2^\dagger$ and continue to use $\hat{f} - \sum_{k=1}^2 \hat{u}_k^\dagger = \sum_{k=3}^K \hat{u}_k^\dagger + \hat{r}$ as input to estimate $\hat{u}_3^\dagger$. This estimation can be repeated until the last mode.

2) Quotient Estimation. We use a neural network T_{θ_k} to estimate the quotient between input and output (i.e., let $T_{\theta_k}(Input) = Output/Input$), which is equivalent to amplitude-frequency gain in the signal community. This technique is widely used in some classical signal processing methods (e.g., the Butterworth filter [18]).

As a result, the k-th mode $\tilde{u}_k = G_{\theta_k}(f)$ can be formulated as:

$$G_{\theta_k}(f) = \begin{cases} T_{\theta_k}(\hat{f}) \cdot (\hat{f}), k = 1 \\ T_{\theta_k}(\hat{f} - \sum_{i=1}^{k-1} \tilde{u}_i) \cdot (\hat{f} - \sum_{i=1}^{k-1} \tilde{u}_i), k \geq 2 \end{cases} \tag{20}$$

At last, we use Eq. 20 to replace G_{θ_k} in Eq. 18 to obtain the final loss function $\mathcal{L}$. We utilize the widely used MLP [19] as the neural network in NME. Then, we train NME by minimizing this loss function through a certain stochastic optimization method such as Adam [20].

3. EVALUATION

To evaluate the effectiveness of NME, we compare the performance of NME with that of two baseline algorithms (i.e., EMD and VMD), and the experiments are conducted on three public real-world datasets.

Metrics. Our evaluations are based on two metrics: latency of mode decomposition(section 3.1) and the accuracy of mode decomposition(section 3.2). The accuracy is measured by orthogonality, which demonstrates the non-repetitiveness of the decomposed modes.

Datasets. Our experiments are conducted on the following 3 datasets, which are widely used in 3 fields (i.e., medicine, industry and engineering), respectively. **EEG-4097** [21] is divided into five subsets: Z, O, N, F, S. Each subset contains 100 temporal series with a sampling frequency of 173.6 Hz and a duration of 23.6 seconds. **Vibration-4000** [22] is a bearing data set under time-varying rotational speed, which consists of signals collected in three health conditions under four manners of speed varying. **Radio-1024** [23] is an over-the-air radio signal which provides 24 types of digital and analog modulations.

Additionally, the test set and training set are randomly divided in a ratio of $3/7$. All experiments are conducted in a desktop PC with Windows OS, 10 cores and 16GB RAM.

		Bearing-5000	
Method	Ratio	Std	Avg(seconds)
VMD	1×	4.0193E-01	6.3785E-01
EMD	1×	7.6288E-01	6.3652E-01
NME	251×	0	2.5880E-03
		Radio-1024	
VMD	1×	7.8845E-02	2.4970E-01
EMD	15×	3.5192E-02	1.6264E-02
NME	740×	0	3.3763E-04
		EEG-4097	
VMD	1×	2.9639E-01	7.8693E-01
EMD	3×	5.1681E-01	2.6165E-01
NME	373×	0	2.1106E-03

Table 1: We calculate the average latency (Avg in table) and the standard deviation (Std in table) of the latency of performing mode decomposition by VMD, EMD and NME over three datasets. Moreover, we further calculate the ratio of these methods against VMD (i.e., the relative speed to VMD).

3.1. Latency Comparison

Table 1 demonstrates that NME is much faster than two baselines. In Radio-1024 dataset, NME is 740 times faster than VMD and 50 times faster than EMD. Also, NME's latency is very stable and the stand deviation of latency over multiple experiments equals 0. This stability comes from the fact that the time cost of NME is only related to the length of the input signal. In contrast, the time cost of VMD and EMD is affected by many factors, e.g., the length, content, and noise of the input signal.

3.2. Orthogonality

Orthogonality is an important criterion used to evaluate the effectiveness of various decomposition algorithms [4]. According to the definition of a mode in Eq. 1, each decomposed mode from a signal should be unique and non-repetitive.

In the following, we first define the metric of orthogonality. Then, we analyze the experimental results and verify that the orthogonality of NME is sufficient.

We utilize the overall index of orthogonality proposed in [4] to evaluate the orthogonality of the modes:

$$IO = \left| \int \frac{\sum_j^K \sum_k^K u(t)_k u(t)_j}{\sum_k^K u(t)_k^2 + 2\sum_j^K \sum_k^K u(t)_k u(t)_j} dt \right| \quad (21)$$

where $u(t)_k$/$u(t)_j$ denotes the k-th/j-th decomposed mode from a signal. It should be noted that orthogonality is sufficient when the IO value is low (e.g., the modes are absolutely orthogonal when $IO = 0$). Note that the orthogonality is to compare the non-repetitiveness of the decomposed modes. For a given signal, EMD can not decouple it into a fixed number of modes. The number of decomposed modes by using EMD is influenced by some characteristics of the input signal (e.g., the noise in the signal). Therefore, it is not fair to compare orthogonality with EMD, as the number of modes can affect the IO value.

	Bearing-5000	Radio-1024	EEG-4097
VMD	1.44E-1	7.46E-2	3.31E-1
NME	1.28E-1	7.36E-2	3.18E-1

Table 2: We calculate the average IO value of the modes decomposed by VMD and NME on three datasets.

		Signal-1	
Method	Mode-1	Mode-2	Rec-sig
VMD	2.30E-3	9.83E-3	1.60E-3
NME	2.00E-3	8.20E-3	5.27E-5
		Signal-2	
VMD	2.52E-2	5.42E-2	4.14E-2
NME	2.18E-2	4.03E-2	6.27E-5

Table 3: We perform mode decomposition for two instance signals (i.e., the f_{sig1} and f_{sig2} in Eq. 22) by NME and VMD, respectively. Then, we calculate the REs between the decomposed modes and the groundtruth (Mode-1, Mode-2 in table). Then, we calculate the RE between the reconstructed signal and the original signal (Rec-sig in table).

Table 2 compares the orthogonality values obtained by both NME and VMD. The average IO value of the modes decomposed by NME is slightly lower than that of VMD on three datasets. Therefore, decomposition of NME is more orthogonal than that of VMD

3.3. Case Study

To further evaluate the effectiveness of NME, we use two synthetic signals to demonstrate the correctness of mode decomposition via NME. The benefit of using synthetic signals is that we have the groundtruth of the modes of a given signal. To this end, we use Eq 22 to generate two synthetic signals as [9] illustrated.

$$\begin{aligned} f_{Sig1} &= \cos(4\pi t) + \frac{1}{4}\cos(48\pi t) \\ f_{Sig2} &= \frac{1}{1.2 + \cos(2\pi t)} + \frac{\cos(32\pi t + 0.2\cos(64\pi t))}{1.5 + \sin(2\pi t)} \end{aligned} \quad (22)$$

The accuracy is measured estimated relative error: As a performance measure, the estimated relative error $RE = \frac{\|\hat{s}-s\|_2}{\|s\|_2}$, where $\hat{s}$ and s denote the estimated and theoretical value (i.e., groundtruth), respectively. Table 3 clearly illustrates that the decomposed modes by NME is closer to the groundtruth, compared to that decomposed by VMD.

4. CONCLUSION

In this paper, we proposed a mode decomposition model based on neural networks, which can perform fast and stable mode decomposition. To our best knowledge, NME is the first mode decomposition model without iterative algorithm. We look forward to more mode decomposition models based on neural estimation.

5. REFERENCES

[1] Rajib Sharma, Leandro Vignolo, Gastón Schlotthauer, Marcelo A Colominas, H Leonardo Rufiner, and SRM Prasanna, "Empirical mode decomposition for adaptive am-fm analysis of speech: A review," *Speech Communication*, vol. 88, pp. 39–64, 2017.

[2] Dennis Gabor, "Theory of communication. part 1: The analysis of information," *Journal of the Institution of Electrical Engineers-part III: radio and communication engineering*, vol. 93, no. 26, pp. 429–441, 1946.

[3] Vasile V Moca, Harald Bârzan, Adriana Nagy-Dăbâcan, and Raul C Mureșan, "Time-frequency super-resolution with superlets," *Nature communications*, vol. 12, no. 1, pp. 1–18, 2021.

[4] Norden E Huang, Zheng Shen, Steven R Long, Manli C Wu, Hsing H Shih, Quanan Zheng, Nai-Chyuan Yen, Chi Chao Tung, and Henry H Liu, "The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis," *Proceedings of the Royal Society of London. Series A: mathematical, physical and engineering sciences*, vol. 454, no. 1971, pp. 903–995, 1998.

[5] Genda Chen and Zuocai Wang, "A signal decomposition theorem with hilbert transform and its application to narrowband time series with closely spaced frequency components," *Mechanical systems and signal processing*, vol. 28, pp. 258–279, 2012.

[6] Naveed ur Rehman, Cheolsoo Park, Norden E Huang, and Danilo P Mandic, "Emd via memd: multivariate noise-aided computation of standard emd," *Advances in adaptive data analysis*, vol. 5, no. 02, pp. 1350007, 2013.

[7] Zhaohua Wu and Norden E Huang, "Ensemble empirical mode decomposition: a noise-assisted data analysis method," *Advances in adaptive data analysis*, vol. 1, no. 01, pp. 1–41, 2009.

[8] Qiuhui Chen, Norden Huang, Sherman Riemenschneider, and Yuesheng Xu, "A b-spline approach for empirical mode decompositions," *Advances in computational mathematics*, vol. 24, no. 1, pp. 171–195, 2006.

[9] Konstantin Dragomiretskiy and Dominique Zosso, "Variational mode decomposition," *IEEE transactions on signal processing*, vol. 62, no. 3, pp. 531–544, 2013.

[10] Vinícius R Carvalho, Márcio FD Moraes, Antônio P Braga, and Eduardo MAM Mendes, "Evaluating five different adaptive decomposition methods for eeg signal seizure detection and classification," *Biomedical Signal Processing and Control*, vol. 62, pp. 102073, 2020.

[11] Zhongjing Ma and Suli Zou, "Extrema of a functional via the variational method," in *Optimal Control Theory*, pp. 39–97. Springer, 2021.

[12] Gene H Golub, Per Christian Hansen, and Dianne P O'Leary, "Tikhonov regularization and total least squares," *SIAM journal on matrix analysis and applications*, vol. 21, no. 1, pp. 185–194, 1999.

[13] Vladimir Britanak, Patrick C Yip, and Kamisetty Ramamohan Rao, *Discrete cosine and sine transforms: general properties, fast algorithms and integer approximations*, Elsevier, 2010.

[14] Victor Namias, "The fractional order fourier transform and its application to quantum mechanics," *IMA Journal of Applied Mathematics*, vol. 25, no. 3, pp. 241–265, 1980.

[15] Mary L Boas, *Mathematical methods in the physical sciences*, John Wiley & Sons, 2006.

[16] John Makhoul, "A fast cosine transform in one and two dimensions," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 28, no. 1, pp. 27–34, 1980.

[17] Nasir Ahmed, T_{-} Natarajan, and Kamisetty R Rao, "Discrete cosine transform," *IEEE transactions on Computers*, vol. 100, no. 1, pp. 90–93, 1974.

[18] Ivan W Selesnick and C Sidney Burrus, "Generalized digital butterworth filter design," *IEEE Transactions on signal processing*, vol. 46, no. 6, pp. 1688–1694, 1998.

[19] Rudolf Kruse, Sanaz Mostaghim, Christian Borgelt, Christian Braune, and Matthias Steinbrecher, "Multilayer perceptrons," in *Computational Intelligence*, pp. 53–124. Springer, 2022.

[20] Diederik P Kingma and Jimmy Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[21] Ralph G Andrzejak, Klaus Lehnertz, Florian Mormann, Christoph Rieke, Peter David, and Christian E Elger, "Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state," *Physical Review E*, vol. 64, no. 6, pp. 061907, 2001.

[22] Huan Huang and Natalie Baddour, "Bearing vibration data collected under time-varying rotational speed conditions," *Data in brief*, vol. 21, pp. 1745–1749, 2018.

[23] Timothy James O'Shea, Tamoghna Roy, and T Charles Clancy, "Over-the-air deep learning based radio signal classification," *IEEE Journal of Selected Topics in Signal Processing*, vol. 12, no. 1, pp. 168–179, 2018.