
Non-Oscillatory Pattern Learning for Non-Stationary Signals

Jieren Xu¹, Yitong Li¹, David Dunson¹,
Ingrid Daubechies¹, Haizhao Yang²

¹Duke University, ²National University of Singapore

¹jieren.xu,yitong.li,dunson,ingrid@duke.edu,²matyh@nus.edu.sg

Abstract

This paper proposes a novel non-oscillatory pattern (NOP) learning scheme for several oscillatory data analysis problems including signal decomposition, super-resolution, and signal sub-sampling. To the best of our knowledge, the proposed NOP is the first algorithm for these problems with fully non-stationary oscillatory data with close and crossover frequencies, and general oscillatory patterns. NOP is capable of handling complicated situations while existing algorithms fail; even in simple cases, e.g., stationary cases with trigonometric patterns, numerical examples show that NOP admits competitive or better performance in terms of accuracy and robustness than several state-of-the-art algorithms.

1 Introduction

This paper concerns oscillatory data defined on a time domain $[0, T]$ of the form:

$$f(t) = \sum_{k=1}^K f_k(t) = \sum_{k=1}^K a_k(t) s_k(\phi_k(t)) = \sum_{k=1}^K \sum_n \hat{s}_k(n) a_k(t) e^{in\phi_k(t)}. \quad (1)$$

Here a_k and ϕ_k are the latent instantaneous amplitude and phase functions of the k th component, which are assumed to be smooth over time. The derivative of a phase function ϕ_k is called an instantaneous frequency function and denoted as ω_k . s_k is a periodic shape function with periodicity one satisfying $\int s_k(t) dt = 0$ and it has a unit L2-norm on $[0, 1]$.

Oscillatory data in Model (1) arise in numerous applications [36, 25, 1, 44, 22, 18, 46, 43, 5, 32, 45, 19, 3, 21, 20, 16, 4, 28] and data analysis of this kind has been an active research field for decades. Usually only $f(t)$ (and sometimes K) is available and the goal is to estimate a_k , ϕ_k (or ω_k), and s_k from $f(t)$. Hence, this is a general problem including and generalizing sub-problems like adaptive time-frequency analysis [2, 8, 42], empirical mode decomposition [17, 38, 37], super-resolution [21, 20, 16, 4, 28], pattern recognition [47], etc. In spite of many successful algorithms for solving these sub-problems, to the best of our knowledge, there is no existing algorithm fulfilling the ultimate goal of estimating a_k , ϕ_k (or ω_k), and s_k when $f(t)$ is fully non-stationary with close and crossover frequencies, and general shape functions. Many existing algorithms require high sampling rate for better accuracy and robustness, which is not practical due to the limit of battery capacity of mobile devices that collect oscillatory data, e.g. portable health monitors.

Fig. 1 shows a synthetic example of Model (1) when $K = 2$, $a_k \equiv 1$, $\phi_1(t) = 15t - \cos(t)$, $\phi_2(t) = 20t + \sin(t)$, $s_1(t) = s_1^{eg}(t)$, and $s_2(t) = s_2^{eg}(t)$. $f(t)$ is in fact a superposition of infinitely many deformed planewaves due to the Fourier series expansion of shape functions. Hence, close and crossover frequencies are unavoidable. In terms of frequencies, due to Heisenberg uncertainty principle, time-frequency analysis methods [2, 8, 42] are not able to estimate these instantaneous frequencies; due to the fully non-stationary nature of $f(t)$, existing super-resolution methods [21, 20, 16, 4, 28] are not able to estimate frequencies $\omega_k(t)$. Moreover, existing GP based methods [35, 23,

24, 34, 26] cannot produce reliable source separation in Model (1). Non-compact support of s_k in the Fourier domain (Fig. 2(b)) creates particular challenges. Some other regression methods have also been designed to estimate shape functions [39, 41, 31], but they assume phase functions are known.

In order to solve the challenging inverse problem implicit to Model (1), it is natural to choose priors within a Bayesian model for the latent components. While there are many possible stochastic process choices, Gaussian Processes (GPs) are appealing in enforcing smoothness providing easy inclusion of prior information and leading to computational tractability. This paper proposes a **non-oscillatory pattern (NOP) learning** scheme, the first framework that can estimate a_k , ϕ_k (or ω_k), and s_k simultaneously from $f(t)$. NOP repeatedly applies a two-stage iteration until convergence. In the first stage, fixing rough estimates of shapes, a novel non-stationary GP regression is proposed to estimate amplitude and phase functions. A novel set of auxiliary points, referred to as the pattern inducing points, are introduced for this purpose. As oppose to traditional stationary kernels, non-stationary kernel functions in our GP model approximately transform a non-stationary data analysis problem into a stationary one, greatly reducing the regression difficulty. Furthermore, we propose to embed the information of rough shape estimates into the GP regression model, reducing a deep GP regression problem [47, 7, 6] with a composition of two latent variables (e.g. the composition of a shape function and a phase function) into a simpler problem with only one latent variable, the phase functions. This significantly reduces the computational cost and difficulty in the regression.

In the second stage, fixing rough estimates of amplitudes and phases, a non-stationary GP regression is proposed or other iterative regression methods in [31, 39–41] is applied to estimate shapes. The main difference of the proposed non-stationary GP regression in the second stage to existing methods is that, the variance of the point estimate of shape functions is simple to derive.

The first and second stages of the algorithm are introduced in Sections 2 and 3, respectively. Section 4 summarizes NOP. As we shall see in the numerical examples in Section 5, NOP works for a wide range of signals in Model (1) and the performance is not sensitive to initialization.

2 Estimation of the instantaneous information

In this section, we assume the shape functions $\{s_k\}_{1 \leq k \leq K}$ are known and aim at estimating the phase and amplitude functions $\{\phi_k\}_{1 \leq k \leq K}$ and $\{a_k\}_{1 \leq k \leq K}$, respectively.

2.1 Main ideas

We start with a simple case when the signal $f(t) = s \circ \phi(t)$. Assume a non-stationary GP $f(t) \sim \mathcal{GP}(0, \mathbf{k}(t, t'))$, and a stationary GP $s(t) \sim \mathcal{GP}(0, \mathbf{k}(t, t'))$, where $\mathbf{k}(\cdot, \cdot)$ is the automatic relevance determination (ARD) [27] squared exponential (SE) kernel:

$$\mathbf{k}(x, x') = \beta^{\text{SE}} \exp\left(-\frac{1}{2} \alpha^{\text{SE}} (x - x')^2\right), \quad (2)$$

with kernel parameters β^{SE} and α^{SE} . Now we consider the phase function ϕ as the **latent variable** of the GP f , and claim the resulting GP is periodic and stationary in the domain ϕ . This is because $f(t) = f(\phi^{-1} \circ \phi(t)) = s(\phi(t))$, by plugging in the new input $x = \phi(t)$, the corresponding output is $f(t) = f \circ \phi^{-1}(x) = s(x)$, i.e., any point $(\phi(t), f(t))$ lies on the curve $s(\cdot)$. $t \mapsto \phi \mapsto f$ seems to be a deep/hierarchical GP model [6], but in fact this ‘unwarping’ process directly removes the first GP layer while keeping the second layer stationary. This is a crucial idea for success of our NOP. And we write the one layer GP with input ϕ as $f(\phi_t) \sim \mathcal{GP}(0, \mathbf{k}(\phi_t, \phi_{t'}))$ for $\phi_t = \phi(t)$.

Next, we introduce a point estimate of the phase function $\phi(t)$ as a latent variable. Suppose $\mathbf{y} \in \mathbb{R}^{N1}$ is the observations (contaminated by white-noise with standard deviation σ) of $f(t)$ at time locations $\mathbf{t} \in \mathbb{R}^N$. Denote $\boldsymbol{\phi} = \phi(\mathbf{t}) \in \mathbb{R}^N$ and $\mathbf{f} = s(\boldsymbol{\phi}) \in \mathbb{R}^N$. As can be seen, a direct inference of the latent $\boldsymbol{\phi}$, even when s is given, is not trivial. This motivates us to introduce a new set of auxiliary points for the GP to retrieve the latent input variable $\boldsymbol{\phi}$, and to reveal the underlying patterns for

¹We will use bold font for vectors.

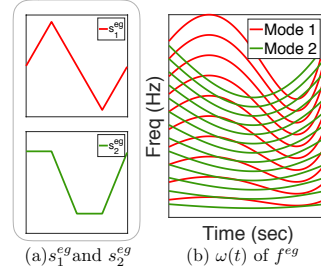


Figure 1: An example of Model (1) with shape functions $s_k = s_k^{eg}$. (a) Shapes s_k^{eg} . (b) instantaneous frequencies $n\phi_1'(t)$ (green) and $n\phi_2'(t)$ (red) for all positive n .

Model (1) when s is not known. We refer to this novel set of auxiliary points as the **pattern inducing points**, denoted as (z, u) , where $z \in \mathbb{R}^M$ is the input variable and $u \in \mathbb{R}^M$ is the respective output evaluated at the inducing locations z . Usually we uniformly discretize and fix z at M phase locations in $[0, L)$, generally $L \geq 1$.

Similar to sparse inducing points [30, 33], the pattern inducing points are treated as variational parameters, and can be updated. They possess the benefits of both sparse inducing points and training points [27], but aim to reveal the true underlying pattern of Model (1) and are endowed with specific meanings. The intuition behind the pattern inducing points (z, u) is that they serve as certain **landmarks**, shown as the green dots (input z as the green triangles) in Fig. 2, of the true underlying pattern s (shaded green line) in Fig. 2, of the true underlying pattern s (shaded green line). The latent input ϕ (red right triangle in Fig. 2) can drive the dataset (ϕ, y) (red dots in Fig. 2) across the phase domain (in Fig. 2(a) from (top) the initial wrong ϕ to (middle) to (bottom) the correct ϕ), to match the underlying pattern $s(t)$ (in green) by these landmarks. The ϕ that matches the dataset (ϕ, y) with the landmark points (z, u) is the correct ϕ since $(\phi(t), f(t)) = (x, s(x))$ for $x = \phi(t)$. So this process is also referred to as a **pattern driven method** for Model (1). Note that setting of $L > 1$ with the SE kernel, is much more stable than the setting of $L = 1$ with the periodic kernel [27], which is highly non-convex and can be trapped at much more local minima validated by our numerical result.

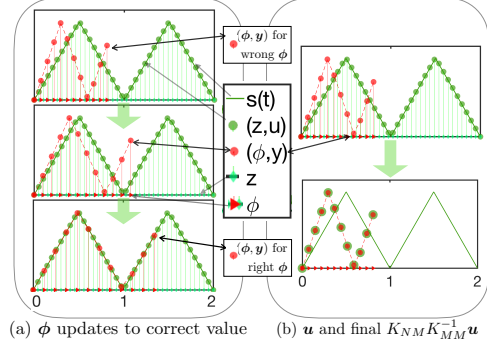


Figure 2: The pattern driven model. (a) the data points are driven and finally match to the underlying pattern s . (b) $K_{NM}K_{MM}^{-1}\mathbf{u}$.

Let $q(\mathbf{u})$ be the variational distribution to approximate the posterior distribution of the inducing variable \mathbf{u} . For the purpose of computational efficiency, we adopt $q(\mathbf{u}) \sim \mathcal{N}(\boldsymbol{\alpha}_u, \Sigma_u)$, where $\boldsymbol{\alpha}_u \in \mathbb{R}^M$ is the mean of \mathbf{u} and the variance matrix $\Sigma_u \in \mathbb{R}^{M \times M}$ of \mathbf{u} is diagonal. In this section since $s(t)$ has been observed, $\boldsymbol{\alpha}_u = \mathbf{u}$ and $\Sigma_u = 0$. When the auxiliary pattern inducing points are ready, we design the following data-pattern matching process in a standard GP setting by encoding the latent variable ϕ in the GP kernel. Conditioning on \mathbf{u} and ϕ , the likelihood of \mathbf{f} becomes

$$p(\mathbf{f}|\mathbf{u}, \phi) = \frac{p(\mathbf{f}, \mathbf{u}|\phi)}{p(\mathbf{u})} = \mathcal{N}(K_{NM}K_{MM}^{-1}\mathbf{u}, K_{NN} - K_{NM}K_{MM}^{-1}K_{MN}), \quad (3)$$

By further marginalizing out \mathbf{f} and \mathbf{u} out, we have

$$p(\mathbf{y}|\phi) = \mathcal{N}(K_{NM}K_{MM}^{-1}\boldsymbol{\alpha}_u, K_{NN} - K_{NM}K_{MM}^{-1}K_{MN} + \Sigma_u + \sigma^2 I_N). \quad (4)$$

Here, I_N is an identity matrix of size $N \times N$, K_{NN} is an $N \times N$ covariance matrix, $(K_{NN})_{ij} = k((\phi)_i, (\phi)_j)^2$ for $\phi \in \mathbb{R}^N$, and similarly $(K_{MM})_{ij} = k((z)_i, (z)_j)$ for $z \in \mathbb{R}^M$, $(K_{MN})_{ij} = k((z)_i, (\phi)_j) = (K_{NM})_{ji}$. Hence, a point estimate of the latent variable ϕ conditioning on $q(\mathbf{u})$ can be computed by Bayesian approach using Eqn. (4).

In the case of multi-components as in Model (1), we denote $\phi_k \in \mathbb{R}^N$, $z_k \in \mathbb{R}^M$, and $\mathbf{u}_k \in \mathbb{R}^M$ for the k th component, and $\mathbf{f} = \sum_{k=1}^K \mathbf{f}_k = \sum_{k=1}^K s_k(\phi_k) \in \mathbb{R}^N$. Assuming the independence among variables of the same type, e.g., among $\{\phi_k\}$, among $\{u_k\}$, etc., Eqn. (4) can be generalized to

$$p(\mathbf{y}|\Phi) = \mathcal{N}\left(\sum_{k=1}^K K_{NM,k}K_{MM,k}^{-1}\boldsymbol{\alpha}_{u_k}, \sum_{k=1}^K (K_{NN,k} - K_{NM,k}K_{MM,k}^{-1}K_{MN,k} + \Sigma_{u_k}) + \sigma^2 I_N\right), \quad (5)$$

where $\Phi \in \mathbb{R}^{N \times K}$ consists of latent variables ϕ_k as the k th column, $K_{NN,k}$, $K_{NM,k}$, $K_{MM,k}$, and $K_{MN,k}$ are defined for covariance matrices of the k th component similarly to the case of one component.

When we have time-varying smooth amplitude functions $a_k(t)$ as in Model (1), let $\mathbf{a}_k = a_k(\mathbf{t})$, then $\{\mathbf{a}_k\}$ are a new set of latent variables of \mathbf{f} . In this case, Eqn. (5) takes the form:

$$p(\mathbf{y}|A, \Phi) = \mathcal{N}\left(\sum_{k=1}^K \mathbf{a}_k \odot K_{NM,k}K_{MM,k}^{-1}\boldsymbol{\alpha}_{u_k}, \sum_{k=1}^K \mathbf{a}_k \mathbf{a}_k^T \odot (K_{NN,k} - K_{NM,k}K_{MM,k}^{-1}K_{MN,k} + \Sigma_{u_k}) + \sigma^2 I_N\right), \quad (6)$$

²Denote $(\cdot)_i$ as the i th entry of a vector, and $(\cdot)_{ij}$ as the ij th entry of a matrix.

where $A \in \mathbb{R}^{N \times K}$ consists of a_k as its k -th column.

To accelerate the computation, instead of applying the Bayesian approach with Eqn. (6), we directly fit the mean value of the Gaussian process with the observations \mathbf{y} via a least square (LS) problem:

$$\min_{A, \Phi} \left\| \mathbf{y} - \sum_{k=1}^K \mathbf{a}_k \odot K_{NM,k} K_{MM,k}^{-1} \boldsymbol{\alpha}_{\mathbf{u}_k} \right\|_2^2, \quad (7)$$

where Φ has been encoded in the kernel matrices in the minimization objective function.

2.2 Adaptive local estimation

The optimization problem in (7) is non-convex and there is no strong prior of ϕ_k and \mathbf{a}_k to provide a good initialization. Note that in most applications, ϕ_k and \mathbf{a}_k usually smoothly vary in time. This motivates us to generate signal patches and locally parametrize the phase and amplitude functions with low-degree polynomials in Eqn. (7). Within each patch, the non-stationary signal becomes much more stationary. As we shall see later, our numerical experiments show that degree $d \leq 2$ is sufficiently good.

For each observation patch (also denoted as \mathbf{y}), let B and $C \in \mathbb{R}^{(d+1) \times K}$ be the matrices consisting of the coefficients of the d -degree polynomials representing the K amplitude and phase functions, respectively. Then we can specify the amplitude and phase functions with B and C by the following LS problem:

$$\min_{B, C} \left\| \mathbf{y} - \sum_{k=1}^K \left(\sum_{j=0}^d (B)_{jk} t^j \right) \odot K_{NM,k} K_{MM,k}^{-1} \boldsymbol{\alpha}_{\mathbf{u}_k} \right\|_2^2, \quad (8)$$

where \odot is the entry-wise dot product, and C has been absorbed in the covariance matrices.

Once the polynomial coefficients B and C of the amplitudes and phases in Eqn. (8) have been specified, they provide a local point estimate of the amplitude and phase functions, which can be used to obtain a global point estimate via a robust curve fitting algorithm [10, 11]. A standard moving average can be applied to estimate the variance of the point estimate.

3 Estimation of shape functions

In this section, we assume the phase and amplitude functions $\{\phi_k\}_{1 \leq k \leq K}$ and $\{a_k\}_{1 \leq k \leq K}$ are known and estimate the shape functions $\{s_k\}_{1 \leq k \leq K}$. We do not aim at closed formulas for $s_k(t)$. As introduced in Section 2, it is sufficient to estimate the pattern inducing variables \mathbf{z}_k and $\mathbf{u}_k \in \mathbb{R}^M$ to represent $s_k(t)$.

When amplitude and phase functions are given, shape function estimation methods have been studied previously in [31, 39–41]. These methods achieves high accuracy when amplitude and phase function estimates are close to the ground truth. There is no quantitative criteria to measure how well the shape function estimate performs when the amplitude and phase function estimate is not very good. This motivates us to apply variational inference, following [33], that explicitly expresses the distribution of the pattern inducing variables in a joint form. Please refer to Supplemental Material (SM) Section A and Section B for details of these three well-established approaches.

4 Overview of NOP

NOP repeatedly applies a two-stage iteration until convergence. We have introduced the first stage in Section 2: given rough estimates of shapes, a non-stationary GP regression is applied to estimate amplitude and phase functions. The second stage has been introduced in Section 3: given rough estimates of amplitudes and phases, several regression methods can be adopted to estimate shapes. Hence, a complete algorithm description can be summarized in Algorithm 1 below. The respective prediction formulation of Algorithm 1 is derived in SM Section C.

In many applications, e.g. ECG and photoplethysmogram data analysis, heuristic properties of the physical system are available and we know the rough range of instantaneous frequencies ω_k . Hence, we can apply a band-pass filter to $f(t) = \sum_{k=1}^K a_k(t) s_k(\phi_k(t))$, and then estimate a_k and ϕ_k of $\hat{s}_k(1) a_k(t) e^{i\phi_k(t)}$ in a certain frequency band using traditional time-frequency analysis methods [?, 8, 42] to initialize $\{a_k\}$ and $\{\phi_k\}$ following the method in [40].

Algorithm 1 NOP

Input: N measurements of (\mathbf{t}, \mathbf{y}) , the number of components K , the maximum iteration number J , and the accuracy parameter ϵ .

Output: estimates of the pattern inducing variables \mathbf{u}_k representing $s_k(t)$, and the latent variables ϕ_k and \mathbf{a}_k representing $\phi_k(t)$ and $a_k(t)$, respectively.

Initialization: $\epsilon_1 = \epsilon_2 = 1$, $\epsilon_0 = 2$, the current iteration number $j = 0$. Initialize the estimates of $\{\mathbf{a}_k\}$ and $\{\phi_k\}$, and let $h = 0$; or initialize the estimate of $\{\mathbf{u}_k\}$ and let $h = 1$.

while $j < J$, $\epsilon_1 > \epsilon$, $\epsilon_2 > \epsilon$, and $|\epsilon_1 - \epsilon_0| > \epsilon$ **do**

if $h = 0$ **then**

 | Given $\{\mathbf{a}_k\}$ and $\{\phi_k\}$ in the previous iteration, compute $\{\mathbf{u}_k\}$ by the method in Section 3.

else

 | Given $\{\mathbf{u}_k\}$ in the previous iteration, compute $\{\mathbf{a}_k\}$ and $\{\phi_k\}$ by the method in Section 2.

if $h = 1$ **then**

 | Given $\{\mathbf{u}_k\}$ in the last step just above, compute $\{\mathbf{a}_k\}$ and $\{\phi_k\}$ by the method in Section 2.

else

 | Given $\{\mathbf{a}_k\}$ and $\{\phi_k\}$ in the last step just above, compute $\{\mathbf{u}_k\}$ by the method in Section 3.

 Update $\epsilon_0 = \epsilon_1$, ϵ_1 and ϵ_2 are set to be the ℓ^2 -norm of the difference of the amplitude and phase estimates in the previous and current iteration.

Another simple initialization is to let $\mathbf{u}_k = \sin(\mathbf{z}_k)$. Since we adopt local patch analysis in Section 2, components after Fourier series expansion, $\{\widehat{s}_k(1)a_k(t)e^{i\phi_k(t)}\}_k$, become approximately orthogonal to each other in a short time domain. Hence, the LS in (7) is able to recover the amplitude and phase functions corresponding to $\{\widehat{s}_k(1)\}$ since they usually have the largest K magnitude.

The LS problems are non-convex and hence we cannot guarantee convergence to the global minimizer. However, the iterative scheme seems to provide very good results and the algorithm generally converges after only a few iterations. The good performance of NOP might come from the fact that once the amplitude and phase function estimates are roughly good (might not be the global minimizer of LS problems), the shape estimation step can quickly provide very good estimate of shape functions, which can be guaranteed if we adopt methods in [31, 39, 41]. h in Algorithm 1 specifies whether we update shape functions first or not; if the initialization of shape functions are better than those of amplitude and phase functions, we choose $h = 1$; otherwise, we choose $h = 0$. The global convergence analysis would be an interesting future work.

5 Experiments

In this section, we provide numerical examples to demonstrate the performance of NOP³, especially in the case of super-resolution and adaptive time-frequency analysis. LS problems in all examples are solved by Adam [14] aiming at better local minimizers. We choose degree-1 (or degree-2 when specified) polynomials to approximate local amplitude and phase functions in these LS problems. The hyperparameters of NOP are set as follows: noise level $\sigma = 10^{-0.8}$, $\alpha^{\text{SE}} = 2 \times 10^3$, and $\beta^{\text{SE}} = 1$. In the local patch analysis, we generate signal patches such that each patch contains approximately 3 to 10 periods. In the tests for super-resolution, we repeat the same test with 10 noise realizations for the purpose of using the expectation and variance of estimation error to measure the performance of different algorithms. $\Delta\omega$ and $\Delta\phi$ denote the point-wise estimation error.

5.1 Super-resolution spectral estimation

There has been substantial research for the super-resolution problem that aims at estimating time-invariant amplitudes and frequencies in a signal $f(t) = \sum_{k=1}^K a_k e^{i\omega_k t}$ with $a_k > 0$, $\omega_k > 0$, and $\{\omega_k\}$ are very close. Among many possible choices, the baseline models might be MUSIC [29], ME [4, 12, 13], and ESPRIT [28]. Hence, we will compare NOP with these methods⁴ to show the advantages of NOP. Although the Fourier transform usually fails [29] to identify $\{a_k\}$ and $\{\omega_k\}$, we use its results as the initialization for NOP.

Accuracy and robustness with different spectral gaps In this experiment, we use $f^{\{1\}}(t) = s_1(2\pi\omega_1 t) + s_2(2\pi\omega_2 t) + \mathcal{N}(0, \sigma^2)$, where the two instantiations of $s_1(\cdot)/s_2(\cdot)$ are (i) $s_1 = \cos/s_2 =$

³Code to be appear in <https://github.com/david-dunson>.

⁴Code from <http://people.ece.umn.edu/~georgiou/files/HRTSA/SpecAn.html>.

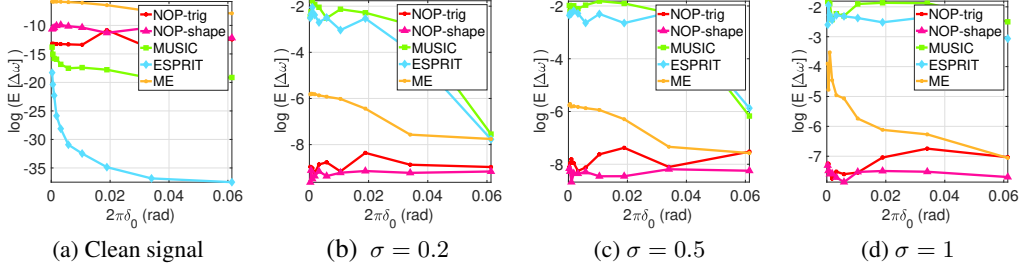


Figure 3: Frequency estimate (absolute) error of $f^{\{1\}}(t) = \cos(2\pi\omega_1 t) + \sin(2\pi\omega_2 t)$, where $\omega_1 = 38.8/1024$ and $\omega_2 = (38.8 + \delta_0)/1024$ with different δ_0 and white noise $\mathcal{N}(0, \sigma^2)$.

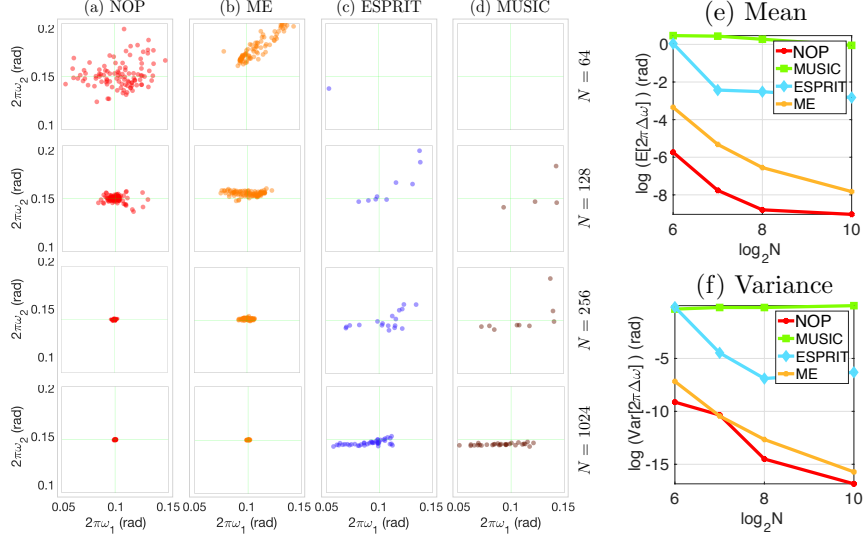


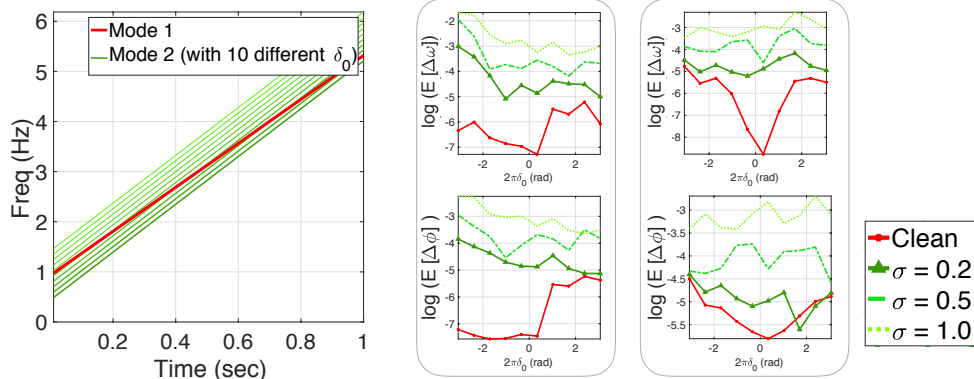
Figure 4: Left: results of $f^{\{2\}}$ with different number of samples $N = 64, 128, 256,$ and 1024 from top to bottom, and by different methods in an order of NOP, ME, ESPRIT, and MUSIC from left to right. The ground truth frequencies are $(2\pi\omega_1, 2\pi\omega_2) = (0.1, 0.15)$. 100 tests with different noise realization were performed and the estimated frequencies are visualized in a 2D domain centered at the ground truth. Right: the expectation and variance of estimation errors for different methods and numbers of samples.

sin (red lines in Fig. 3) and (ii) $s_1 = s_1^{eg}/s_2 = s_2^{eg}$ as in Fig. 1 (pink lines in Fig. 3). Here $\omega_1 = 38.8/1024$ and $\omega_2 = (38.8 + \delta_0)/1024$, δ_0 varies from $0.05/1024$ to $10/1024$, and different noise variance are σ^2 . The sampling rate is 1 Hz and the number of samples is $N = 100$ in this example. Fig. 3 shows the frequency estimation accuracy of NOP, MUSIC, ESPRIT, and ME. As we can see, NOP achieves machine accuracy in the noiseless case and is much more accurate than other methods in all noisy cases. It also reaches almost the same accuracy for the trigonometric (red) and shaped (pink) instantiations in (i) and (ii). Baseline methods are not applicable to instantiation (ii).

Accuracy and robustness with different sampling rates In this experiment, we set $f^{\{2\}}(t) = \sum_{k=1}^2 a_k \sin(\omega_k t)$ with $a_1 = 0.5, a_2 = 1, \omega_1 = 0.1,$ and $\omega_2 = 0.15$. The sampling rate of this signal is still 1Hz and the numbers of samples are $N = 64, 128, 256,$ and 1024 to generate four sets of test data. There are two different kinds of noise to generate noisy test data: 1) white Gaussian noise $\mathcal{N}(0, 0.35)$ is directly added to $f(t)$; 2) a stochastic process in t with i.i.d. uniform distribution in $[0, 2\pi]$ is added to phase functions $\{\omega_k t\}_{1 \leq k \leq 2}$. Fig. 4 summarizes the results of frequency estimates in this experiment. ESPRIT and MUSIC lose accuracy in all tests. NOP and ME achieve high accuracy when the number of samples is large and NOP is slightly better than ME in terms of accuracy and estimation bias.

5.2 Estimation of time-variant frequencies

In this section, we show the capacity of NOP for estimating close and crossover time-varying instantaneous frequencies. An adaptive time-frequency analysis algorithm, ConceFT [9]), is used as a comparison. And local approximation degree is set to $d = 2$ in this section.



(a) $\omega_1(t)$ and $\omega_2(t)$ in time-frequency domain (b) Error for $s_k = \sin$ (c) Error for $s_k = s_k^{eg}$

Figure 5: Short signal $f^{\{3\}}$ ($N = 100$) with linear frequencies. (a) visualizes all the instantaneous frequencies of our synthetic components as the spectral gap parameter δ_0 takes the values $(i - 5)/10.24$ for $i = 0, 1, \dots, 9$. (b) is the estimation error for frequency (top) and phase (bottom) estimates when $s_k = \sin$; (c) is for $s_k = s_k^{eg}$.

Close frequencies and phase estimation error We use $f^{\{3\}}(t) = s_1(2\pi(10/10.24t + 230/10.24^2t^2)) + s_2(2\pi((10/10.24 + \delta_0)t + 250/10.24^2t^2))$, where the two instantiations of $s_1(\cdot)/s_2(\cdot)$ are (i) $s_1 = \cos/s_2 = \sin$ (Fig. 5(b)) and (ii) $s_1 = s_1^{eg}/s_2 = s_2^{eg}$ as in Fig. 1 (Fig. 5(c)). δ_0 varies from $-5/10.24$ to $5/10.24$. The white noise σ_0 has standard deviation $\{0, 0.2, 0.5, 1\}$. We apply short-time Fourier transform [15] to identify rough estimates of instantaneous frequencies and use them as the initialization in this test. When instantaneous frequencies are very close, the initialization is very poor; however, NOP still can identify instantaneous frequencies and phases with a reasonably good accuracy. Result is summarized in Fig. 5.

Fig. 5(a) is the ground truth time-frequency representation of ten tested signals with different value of δ_0 on $\omega_2(t)$. The difference between $\omega_2(t)$ (green line) and $\omega_1(t)$ (red line) are pretty difficult to be detected by existing time-frequency methods. The log error of the point-wise averaged frequency estimate is shown in the first row of of Fig. 5 (b) and (c) on different noise levels σ_0 . The log error of point-wise averaged phase estimate (bottom row) is consistently small as δ_0 changes. Under large noise case with $\sigma_0 = 1$, NOP controls the phase error approximate or below the level of 0.05. Existing time-frequency analysis methods usually estimate instantaneous frequencies first and then integrate them to obtain instantaneous phases, which suffers from accumulated error. However, NOP has no accumulated error.

Close and crossover frequencies In this experiment, we generate a signal consisting of two components with close instantaneous frequencies and a signal with two crossover instantaneous frequencies. Fig. 6 visualizes the ground truth instantaneous frequencies, the time-frequency distribution by ConceFT, the initialization and the estimation results of NOP. ConceFT cannot visualize the instantaneous frequencies even if in the noiseless case. We average out the energy distribution of ConceFT to obtain the initialization of NOP. Although the initialization is very poor, NOP is still able to estimate the instantaneous frequencies with a reasonably good accuracy no matter in clean or noisy cases. However, when the number of components K is known, we can average out the energy band to obtain one instantaneous frequency function and initialize all instantaneous frequencies in NOP using this function. We will show how to handle this special initialization in SM Section D.

5.3 Estimation of amplitudes, phases, and shapes simultaneously

Finally, we apply NOP to estimate amplitudes, phases, and shapes simultaneously from a single record. First, we generate a synthetic example $f^{\{6\}}(t) = \sum_{k=1}^2 s_k^{eg}(\omega_k t) + \mathcal{N}(0, 0.2)$, where $\omega_1 = 3.88/1.024$, $\omega_2 = 4.88/1.024$, and the shapes are visualized in Fig. 7. The sampling rate for this signal is 100 Hz and we sample it at 100 locations. The shape estimates are initialized as \cos and \sin for the first and second components, respectively (see Fig. 7 (b) and (c)). The frequency estimates are initialized as *one* constant centered in the peak spectrogram by ConceFT (see Fig. 7 (a)). As we can see in Fig. 7 (b) and (c), NOP is able to estimate shape functions with a reasonably good accuracy and the reconstructed components match the ground truth components very well.

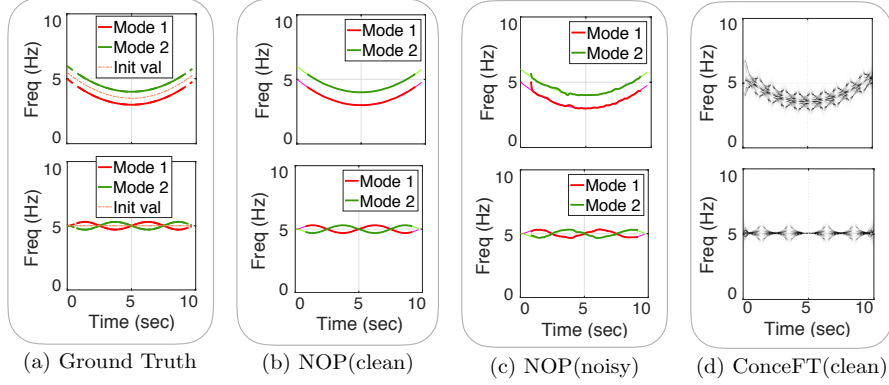


Figure 6: Instantaneous frequency estimates for signals with close and crossover frequencies. (a) ground truth instantaneous frequencies and initialization of NOP. (b) estimated instantaneous frequencies for clean signals. (c) estimated instantaneous frequencies for noisy signals. (d) time-frequency distribution by ConceFT.

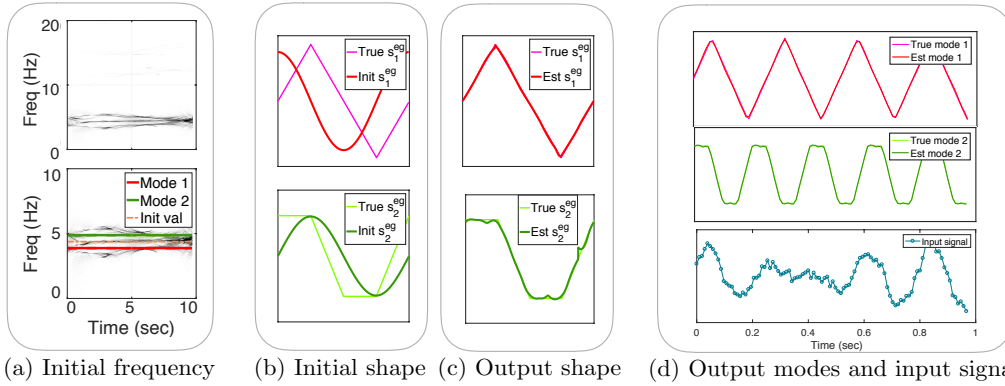


Figure 7: NOP is applied to estimate the amplitude, phase, and shapes of a synthetic signal $f^{\{6\}}(t)$ consisting of two components. (a) the time-frequency distribution of $f^{\{6\}}(t)$ by ConceFT in two different frequency ranges. ConceFT cannot reveal the ground truth instantaneous frequencies (in red and green). But we can initialize NOP by averaging out the distribution (see the dash pink line). (b) and (c) the ground truth shape functions and their estimates. (d) the noisy signal $f^{\{6\}}(t)$ and the reconstructed components by NOP.

In the second example, we apply NOP to a real signal from photoplethysmogram (PPG) (see Fig. 8 (b)). The shape estimates are still initialized as \cos and \sin for the two components, and $N = 100$ samples are involved. The PPG signal contains two components corresponding to the health condition of the heart and lung in a human body.

6 Conclusion

This paper proposes a novel non-oscillatory pattern (NOP) learning scheme for several oscillatory data analysis problems including signal decomposition, super-resolution, and signal sub-sampling. To the best of our knowledge, the proposed NOP is the first algorithm for these problems with fully non-stationary oscillatory data with close and crossover frequencies, and general oscillatory patterns. Numerical examples have shown the advantage of NOP over several state-of-the-art algorithms and NOP is able to handle complicated examples for which existing algorithms fail. NOP could be a very useful tool for pattern analysis for oscillatory data. Although we cannot prove the global convergence of NOP, NOP seems to provide very good results in all of our tests. It is interesting to study the global convergence of NOP in the future.

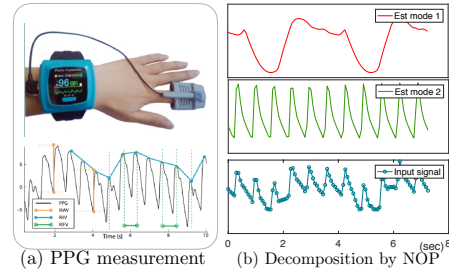


Figure 8: (a) PPG signal. (b) reconstructed components of the PPG signal. These two components were reconstructed from only a small portion of the samples of the original PPG raw data as visualized in the bottom figure.

References

- [1] E. Alonso, E. Aramendi, D. González-Otero, U. Ayala, M. Daya, and J. K. Russell. Empirical mode decomposition for chest compression and ventilation detection in cardiac arrest. In *Computing in Cardiology 2014* 2014.
- [2] F. Auger and P. Flandrin. Improving the readability of time-frequency and time-scale representations by the reassignment method. *IEEE Transactions on signal processing*, 1995.
- [3] X. Bai, M. Xing, F. Zhou, G. Lu, and Z. Bao. Imaging of micromotion targets with rotating parts based on empirical-mode decomposition. *IEEE Transactions on Geoscience and Remote Sensing* 2008.
- [4] J. P. Burg. The relationship between maximum entropy spectra and maximum likelihood spectra. *Geophysics*, 1972.
- [5] B. Cornelis, H. Yang, A. Goodfriend, N. Ocon, J. Lu, and I. Daubechies. Removal of canvas patterns in digital acquisitions of paintings. *IEEE Transactions on Image Processing* 2017.
- [6] Z. Dai, A. Damianou, J. González, and N. Lawrence. Variational auto-encoded deep Gaussian Processes. *arXiv preprint arXiv:1511.06455*, 2015.
- [7] A. Damianou and N. Lawrence. Deep Gaussian Processes. In *Artificial Intelligence and Statistics*, 2013.
- [8] I. Daubechies and S. Maes. A nonlinear squeezing of the continuous wavelet transform based on auditory nerve models. In *Wavelets in Medicine and Biology*. CRC Press, 1996.
- [9] I. Daubechies, Y. G. Wang, and H.-t. Wu. Concept: concentration of frequency and time via a multitapered synchrosqueezed transform. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 2016.
- [10] D. Garcia. Robust smoothing of gridded data in one and higher dimensions with missing values. *Computational Statistics & Data Analysis*, 2010.
- [11] D. Garcia. A fast all-in-one method for automated post-processing of PIV data. *Experiments in fluids*, 2011.
- [12] T. T. Georgiou. Spectral estimation via selective harmonic amplification. *IEEE Transactions on Automatic Control*, 2001.
- [13] T. T. Georgiou. Spectral analysis based on the state covariance: the maximum entropy spectrum and linear fractional parametrization. *IEEE transactions on Automatic Control*, 2002.
- [14] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio. *Deep learning*, volume 1. MIT press Cambridge, 2016.
- [15] D. Griffin and J. Lim. Signal estimation from modified short-time fourier transform. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 1984.
- [16] M. H. Gruber. Statistical digital signal processing and modeling, 1997.
- [17] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *R. Soc. Lond. Proc. Ser. A Math. Phys. Eng. Sci.*, 1998.
- [18] W. Huang, Z. Shen, N. E. Huang, and Y. C. Fung. Engineering analysis of biological variables: An example of blood pressure over 1 day. *Proc. Natl. Acad. Sci.*, 1998.
- [19] Y. Huanyin, G. Huadong, H. Chunming, L. Xinwu, and W. Changlin. A sar interferogram filter based on the empirical mode decomposition method. In *IGARSS 2001. Scanning the Present and Resolving the Future. Proceedings. IEEE 2001 International Geoscience and Remote Sensing Symposium (Cat. No.01CH37217)*, volume 5, 2001.
- [20] C. E. J. and F.-G. Carlos. Towards a mathematical theory of super-resolution. *Communications on Pure and Applied Mathematics*.

- [21] W. Liao and A. Fannjiang. Music for single-snapshot spectral estimation: Stability and super-resolution. *Applied and Computational Harmonic Analysis*, 2016.
- [22] J. Lu, B. Wirth, and H. Yang. Combining 2D synchrosqueezed wave packet transform with optimization for crystal image analysis. *Journal of the Mechanics and Physics of Solids*, 2016.
- [23] Y. Pan, X. Yan, E. A. Theodorou, and B. Boots. Prediction under uncertainty in sparse spectrum Gaussian Processes with applications to filtering and control. In *ICML*, 2017.
- [24] G. Parra and F. Tobar. Spectral mixture kernels for multi-output Gaussian Processes, 2017.
- [25] E. Pinheiro, O. Postolache, and P. Girão. Empirical mode decomposition and principal component analysis implementation in processing non-invasive cardiovascular signals. *Measurement*, 2012. Special issue on Electrical Instruments.
- [26] J. Quiñero-Candela, C. E. Rasmussen, A. R. Figueiras-Vidal, et al. Sparse spectrum Gaussian Process regression. *Journal of Machine Learning Research*, 2010.
- [27] C. E. Rasmussen and C. K. Williams. *Gaussian Processes for machine learning*, volume 1. MIT press Cambridge, 2006.
- [28] R. Roy and T. Kailath. Esprit-estimation of signal parameters via rotational invariance techniques. *IEEE Transactions on Acoustics, Speech, and Signal Processing* 1989.
- [29] R. Schmidt. Multiple emitter location and signal parameter estimation. *IEEE transactions on antennas and propagation*, 1986.
- [30] E. Snelson and Z. Ghahramani. Sparse Gaussian Processes using pseudo-inputs. In *Advances in Neural Information Processing Systems*, 2006.
- [31] G. Tang and H. Yang. A fast algorithm for multiresolution mode decomposition. *arXiv:1709.06880*, 2017.
- [32] J. B. Tary, R. H. Herrera, J. Han, and M. van der Baan. Spectral estimation-What is new? What is next? *Rev. Geophys.* 2014.
- [33] M. K. Titsias. Variational learning of inducing variables in sparse Gaussian Processes. In *International Conference on Artificial Intelligence and Statistics*, 2009.
- [34] K. R. Ulrich, D. E. Carlson, K. Dzirasa, and L. Carin. GP kernels for cross-spectrum analysis. In *NIPS*, 2015.
- [35] A. Wilson and R. Adams. Gaussian Process kernels for pattern discovery and extrapolation. In *ICML*, 2013.
- [36] H. Wu, Y.-H. Chan, Y.-T. Lin, and Y.-H. Yeh. Using synchrosqueezing transform to discover breathing dynamics from ECG signals. *Applied and Computational Harmonic Analysis*, 2014.
- [37] Z. Wu and N. E. Huang. Ensemble empirical mode decomposition: a noise-assisted data analysis method. *Advances in Adaptive Data Analysis*, 2009.
- [38] Z. Wu, N. E. Huang, and X. Chen. The multi-dimensional ensemble empirical mode decomposition method. *Adv. Adapt. Data Anal.*, 2009.
- [39] J. Xu, H. Yang, and I. Daubechies. Recursive diffeomorphism-based regression for shape functions. *SIAM Journal on Mathematical Analysis*, 2018.
- [40] H. Yang. Synchrosqueezed wave packet transforms and diffeomorphism based spectral analysis for 1D general mode decompositions. *Applied and Computational Harmonic Analysis*, 2015.
- [41] H. Yang. Multiresolution mode decomposition for adaptive time series analysis. *arXiv:1709.06880*, 2017.
- [42] H. Yang. Statistical analysis of synchrosqueezed transforms. *Applied and Computational Harmonic Analysis*, 2017.

- [43] H. Yang, J. Lu, W. Brown, I. Daubechies, and L. Ying. Quantitative canvas weave analysis using 2-D synchrosqueezed transforms: Application of time-frequency analysis to art investigation. *Signal Processing Magazine, IEEE* 2015.
- [44] H. Yang, J. Lu, and L. Ying. Crystal image analysis using 2D synchrosqueezed transforms. *Multiscale Modeling & Simulation*, 2015.
- [45] H. Yang and L. Ying. Synchrosqueezed curvelet transform for two-dimensional mode decomposition. *SIAM Journal on Mathematical Analysis*, 2014.
- [46] C. Zhang, Z. Li, C. Hu, S. Chen, J. Wang, and X. Zhang. An optimized ensemble local mean decomposition method for fault detection of mechanical components. *Measurement Science and Technology*, 2017.
- [47] B. Zhu and D. B. Dunson. Locally adaptive Bayes nonparametric regression via nested Gaussian Processes. *Journal of the American Statistical Association*, 2013.

A Joint pattern describing points

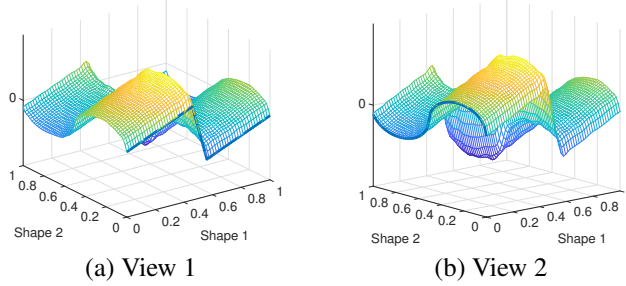


Figure 9: Pattern inducing points of a signal containing two shape functions $s_1(t)$ and $s_2(t)$ with \mathbf{z} supported and arranged in the one \mathbb{R}^2 block. **(ab)**: are two views of these rearranged auxiliary points.

For method (iii), when the influence of the amplitude function is negligible, another natural approach is to apply the variational inference, following [33]. We can explicitly express the distribution of the pattern inducing points in a joint form (Z, \mathbf{u}) , where $\mathbf{u} \in \mathbb{R}^M$, $Z \in \mathbb{R}^{M \times K}$ and $M = \prod_{k=1}^K M_k$ with $\mathbf{u}, \mathbf{z} \in \mathbb{R}^{M_k}$. The m th row of Z is denoted as $Z_m = ((z_1)_{\tau(m,1)}, \dots, (z_K)_{\tau(m,K)})$, and the respective m th entry of the joint $\mathbf{u} = \sum_{k=1}^K \mathbf{u}_k$ is expected to approximate $\sum_{k=1}^K s_k((z_k)_{\tau(m,k)})$. One exemplified (Z, \mathbf{u}) is shown in Fig. 9. The goal is to maximize a lower bound of $p(\mathbf{y}|\Phi)$ (Φ is fixed and thus dropped for now) given as

$$\log p(\mathbf{y}) = \log \int p(\mathbf{y}, \mathbf{f}, \mathbf{u}) d\mathbf{f} d\mathbf{u} = \log \int q(\mathbf{f}, \mathbf{u}) \frac{p(\mathbf{y}|\mathbf{f})p(\mathbf{u})}{q(\mathbf{u})} d\mathbf{f} d\mathbf{u} \quad (9)$$

$$\geq \int q(\mathbf{u}) \log \frac{p(\mathbf{u}) \log \mathcal{N}(\mathbf{y}; K_{NM} K_{MM}^{-1} \mathbf{u}, \sigma^2 \mathbf{I}_N)}{q(\mathbf{u})} d\mathbf{u} - \frac{1}{2\sigma^2} \text{Tr}(K_{NN} - K_{NM} K_{MM}^{-1} K_{MN}). \quad (10)$$

To reduce the unnecessary computational complexity, the periodic kernel takes the form of

$$\mathbf{k}(\mathbf{x}, \mathbf{x}') = \beta^{\text{Period}} \exp \left(-\frac{1}{2} \sum_{k=1}^K \alpha_k^{\text{Period}} \sin^2(\pi |x_k - x'_k|) \right). \quad (11)$$

By applying an inverse Jensen equality on Eqn. (10), we obtain the variational maximizer of the lower bound as

$$q(\mathbf{u}) \propto \mathcal{N}(\sigma^{-2} K_{MM} \Sigma K_{MNY}, K_{MM} \Sigma K_{MM}), \quad (12)$$

where $\Sigma = (K_{MM} + \sigma^{-2} K_{MN} K_{NM})^{-1}$. Then we can further get $\boldsymbol{\alpha}_u = \sigma^{-2} K_{MM} \Sigma K_{MNY}$ and $\Sigma_u = K_{MM} \Sigma K_{MM}$. The mean field approximation and a K -dimensional ANOVA is applied to retrieve $q(\mathbf{u}_k) = \mathcal{N}(\boldsymbol{\alpha}_{u_k}, \Sigma_{u_k})$, for $k = 1, \dots, K$, in closed forms from the joint estimation of $q(\mathbf{u})$. The separated estimates of $q(\mathbf{u}_k)$ can be approximated for the respective shape functions s_k . Computational details are listed in the following section.

B Update pattern inducing points

We start from (iii) under the situation that the influence of the amplitude function $a_k(t)$ is very weak and a point estimate of the latent input Φ is given.

One natural approach is to apply the variational inference following [33], which explicitly expresses the distribution of inducing points (Z, \mathbf{u}) by maximizing a lower bound of $p(\mathbf{y}|\Phi)$ (Φ is dropped in

the rest part of this section) :

$$\log p(\mathbf{y}) = \log \int p(\mathbf{y}, \mathbf{f}, \mathbf{u}) d\mathbf{f} d\mathbf{u} = \log \int p(\mathbf{f}, \mathbf{u}) p(\mathbf{y}|\mathbf{f}) d\mathbf{f} d\mathbf{u}. \quad (13)$$

$$= \log \int q(\mathbf{f}, \mathbf{u}) \frac{p(\mathbf{y}|\mathbf{f}) p(\mathbf{u})}{q(\mathbf{u})} d\mathbf{f} d\mathbf{u} \quad (14)$$

$$\geq \int q(\mathbf{f}, \mathbf{u}) \log \frac{p(\mathbf{y}|\mathbf{f}) p(\mathbf{u})}{q(\mathbf{u})} d\mathbf{f} d\mathbf{u} \quad (15)$$

$$= \int q(\mathbf{u}) \left(\int p(\mathbf{f}|\mathbf{u}, \phi) \log p(\mathbf{y}|\mathbf{f}) d\mathbf{f} + \frac{p(\mathbf{u})}{q(\mathbf{u})} \right) d\mathbf{u}. \quad (16)$$

which is

$$\log p(\mathbf{y}) = \int q(\mathbf{u}) \log \frac{p(\mathbf{u}) \log \mathcal{N}(\mathbf{y}; K_{NM} K_{MM}^{-1} \mathbf{u}, \sigma^2 \mathbf{I})}{q(\mathbf{u})} d\mathbf{u} - \frac{1}{2\sigma^2} \text{Tr}(K_{NN} - K_{NM} K_{MM}^{-1} K_{MN}). \quad (17)$$

By applying the inverse Jensen Inequality on the first term, the variational maximizer of the lower bound is derived as

$$q(\mathbf{u}) \propto p(\mathbf{u}) \exp \log \mathcal{N}(\mathbf{y}; K_{NM} K_{MM}^{-1} \mathbf{u}, \sigma^2 \mathbf{I}_N) \quad (18)$$

$$\propto \mathcal{N}(\sigma^{-2} K_{MM} \Sigma K_{MN} \mathbf{y}, K_{MM} \Sigma K_{MM}) \quad (19)$$

In the above equations, $\Sigma = (K_{MM} + \sigma^{-2} K_{MN} K_{NM})^{-1}$, $\alpha_{\mathbf{u}} = \sigma^{-2} K_{MM} \Sigma K_{MN} \mathbf{y}$ and $\Sigma_{\mathbf{u}} = K_{MM} \Sigma K_{MM}$. The mean field approximation is applied to . Then, we optimize $\alpha_{\mathbf{u}_k}$ and the diagonal version of $\Sigma_{\mathbf{u}_k}$, denoted as $\Sigma_{\mathbf{u}_k}^{\text{diag}}$ by

$$\tilde{\alpha}_{\mathbf{u}_1}, \dots, \tilde{\alpha}_{\mathbf{u}_K} = \underset{\alpha_{\mathbf{u}_1}, \dots, \alpha_{\mathbf{u}_1}}{\text{argmin}} \sum_{m=1}^M \left\| (\alpha_{\mathbf{u}})_m - \sum_{k=1}^K (\alpha_{\mathbf{u}_k})_{\tau(m,k)} \right\|_2, \quad (20)$$

$$\text{and } \widetilde{\Sigma}_{\mathbf{u}_1}^{\text{diag}}, \dots, \widetilde{\Sigma}_{\mathbf{u}_K}^{\text{diag}} = \underset{\Sigma_{\mathbf{u}_1}^{\text{diag}}, \dots, \Sigma_{\mathbf{u}_1}^{\text{diag}}}{\text{argmin}} \sum_{m=1}^M \left\| (\Sigma_{\mathbf{u}}^{\text{diag}})_m - \sum_{k=1}^K (\Sigma_{\mathbf{u}_k}^{\text{diag}})_{\tau(m,k)} \right\|_2, \quad (21)$$

The above optimization problem is essentially a K dimensional ANOVA and can be solved in closed form.

The separated estimates of $q(\mathbf{u}_k)$ serves as an approximation for the respective shape functions s_k .

Method (i) can be implemented by a nonlinear band truncation on the respective time-frequency domain obtained by the ConceFT, STFT, etc. Fig. 10 shows an example of such nonlinear truncation. As can be seen, after this adaptive frequency filtering, only K respective low frequency part of each mode f_k is kept. f_k is transformed from the original $f_k = \sum_{k=1}^K s_k(\phi_k(t))$ to $\tilde{f}_k \approx \sum_{k=1}^K \cos(\phi_k(t) + \phi_k^0)$, or $\tilde{f}_k \approx \sum_{k=1}^K \cos(\phi_k(t) + \phi_k^0 + \pi/2)$. ϕ_k^0 denotes for unknown initial value of the phase function.

The general idea of the approaches in Method (ii) is to apply a nonlinear diffeomorphism on the signal according to the updated latent phase function ϕ_k . Then apply adaptive regression to further guarantee certain properties of the underlying pattern. We refer the readers to [39, 41, 31].

C Prediction

In this part, we introduce how to make predictions on the signals or to generate new samples by NOP with updated output \tilde{A} , $\tilde{\Phi}$, and (Z, \mathbf{u}) or $q(\mathbf{u}|Z)$. Since the amplitude function $a_k(t)$ and the frequency function $\omega_k(t)$ are assumed to be smooth, we can put a GP prior on both of them and treat the updated \tilde{A} , $\tilde{\Phi}$ as respective learning points. The updated patterning inducing points (Z, \mathbf{u}) are directly plugged in for inference.

If we want to predict on d new time points $\mathbf{t}_0 \in \mathbb{R}^d$, the respective predicted value of $\mathbf{f}_0 \in \mathbb{R}^d$ can be obtained as following. The first sample of $\Phi_0 \in \mathbb{R}^{N \times K}$ with respect to \mathbf{t}_0 is drawn from $p(\Phi|\mathbf{t}_0, \mathbf{y})$. $A_0 \in \mathbb{R}^{N \times K}$ of \mathbf{t}_0 from $p(A|\mathbf{t}_0, \mathbf{y})$ is drawn by the following equation

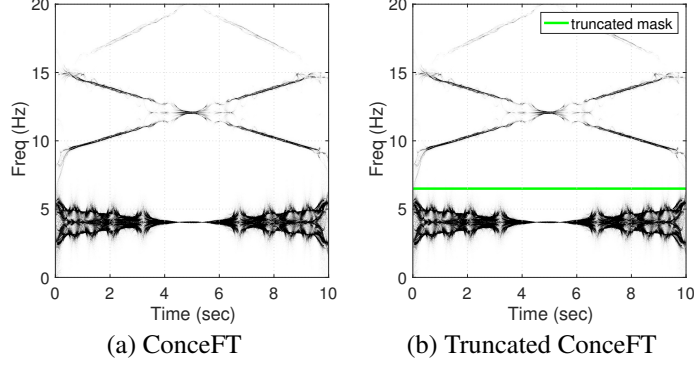


Figure 10: Applying an adaptive filter to change $\sum_{k=1}^K s_k(\phi_k(t))$ into $\sum_{k=1}^K \cos_k(\phi_k(t) + \phi_k^0)$. (a): are the original and truncated time-frequency domain.

$$\Phi_0 \sim \mathcal{N}(K_{dd}K_{dN}\tilde{\Phi}, K_{dd} - K_{dN}K_{NN}^{-1}K_{Nd}), \quad (22)$$

and

$$A_0 \sim \mathcal{N}(K_{dd}K_{dN}\tilde{A}, K_{dd} - K_{dN}K_{NN}^{-1}K_{Nd}), \quad (23)$$

where $(K_{dd})_{ij} = k((\Phi_0)_i, (\Phi_0)_j)$, $(K_{NN})_{ij} = k(\tilde{\Phi}, \tilde{\Phi})$, and $(K_{dN})_{ij} = k((\Phi_0)_i, \tilde{\Phi}) = (K_{Nd})_{ji}$.

Since

$$s_k(\phi_{0,k}) \sim \mathcal{N}(K_{dM,k}K_{MM,k}^{-1}\mathbf{u}_k, K_{dd} - K_{dM,k}K_{MM,k}^{-1}K_{Md,k}), \quad \text{for } k = 1, \dots, K, \quad (24)$$

where $(K_{MM,k})_{ij} = k(\mathbf{z}_k, \mathbf{z}_k)$, and $(K_{dM,k})_{ij} = k(\phi_{0,k}, \mathbf{z}_k) = (K_{Md,k})_{ji}$, thus

$$\tilde{\mathbf{a}}_k \odot s_k(\phi_{0,k}) \sim \mathcal{N}\left(\tilde{\mathbf{a}}_k \odot K_{dM,k}K_{MM,k}^{-1}\mathbf{u}_k, \tilde{\mathbf{a}}_k \tilde{\mathbf{a}}_k^T \odot \left(K_{dd} - K_{dM,k}K_{MM,k}^{-1}K_{Md,k}\right)\right). \quad (25)$$

Since different components are independent, we have

$$\begin{aligned} \mathbf{f} &= \sum_{k=1}^K \tilde{\mathbf{a}}_k \odot s_k(\phi_{0,k}) \\ &\sim \mathcal{N}\left(\sum_{k=1}^K \tilde{\mathbf{a}}_k \odot K_{dM,k}K_{MM,k}^{-1}\mathbf{u}_k, \sum_{k=1}^K \left(\tilde{\mathbf{a}}_k \tilde{\mathbf{a}}_k^T \odot \left(K_{dd} - K_{dM,k}K_{MM,k}^{-1}K_{Md,k}\right)\right)\right) \end{aligned} \quad (26)$$

Therefore, \mathbf{f}_0 can be drawn from distribution (27).

Further, if $q(\mathbf{u})$ is approximated by the mean field method, $q(\mathbf{u}_k)$ can be simply integrated out from distribution (27).

D Non-distinguishable Problem

In this section, we introduce an *ad hoc* approach to tackle the non-distinguishable problem arises when the initialized instantaneous frequency of different components are set to a same function. For the estimated frequencies (t, ω_1) and (t, ω_2) , where $t, \omega_1, \omega_2 \in \mathbb{R}^N$, we want to distinguish components with instantaneous frequency $(\omega_1)_i$ or $(\omega_2)_i$ for each time point t_i .

The naive method is to sort $(\omega_1)_i$ and $(\omega_2)_i$ firstly. Then we cluster components by their values. This actually works pretty well when no crossover happened for $\omega_1(t)$ and $\omega_2(t)$, which is shown in Fig. 11 (a). However, when the crossover happens, as with the second row in Fig. 7, this simple clustering method will fail, as shown in Fig. 11 (b).

One simple way to tackle this situation is stated as following, based on the smoothness assumption placed on the frequency functions $\omega_k(t)$. First we identify all the possible crossover points of $\omega_k(t)$, which is the robust local minimum of $|\omega_1 - \omega_2|$. Then the initial clustering obtained from the simple sorting, as shown in Fig. 11 (a)(b), is switched whenever t_i passes a potential crossover point, in the order of increasing time. This simple reorder trick will output the new clustering as shown in Fig. 11 (c), which is satisfactory for a further refinement to get a point estimate of $\omega_k(t)$.

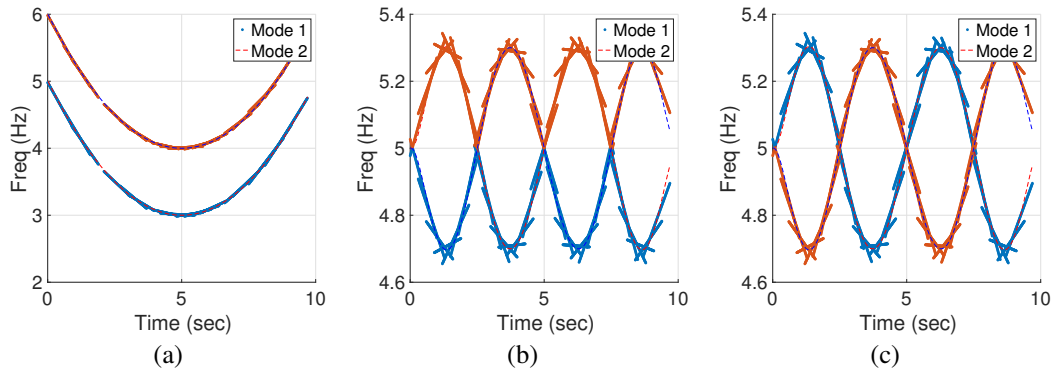


Figure 11: Solution for the non-distinguishable problem. (a)(b) are obtained by the simple ordering clustering. (c) is obtained by the corrected ordering clustering based on the cross-over points identification.

E Signals Plots

This section gives the plots of signals used in the main text.

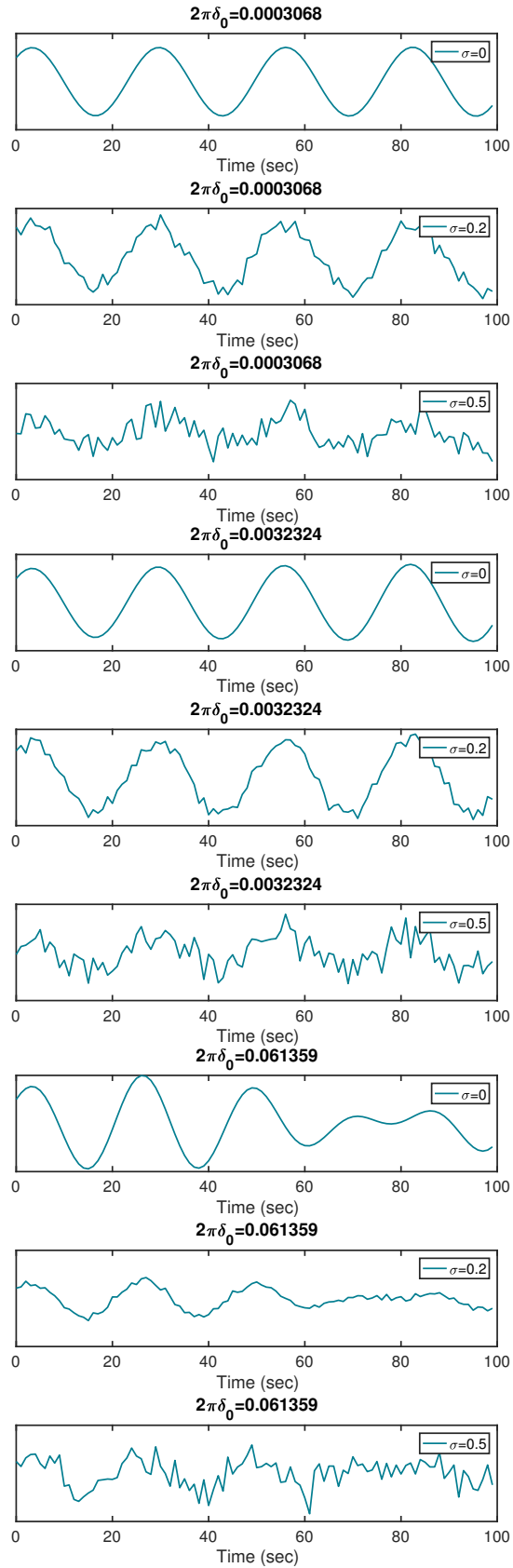


Figure 12: cos version of $f^{\{1\}}$.

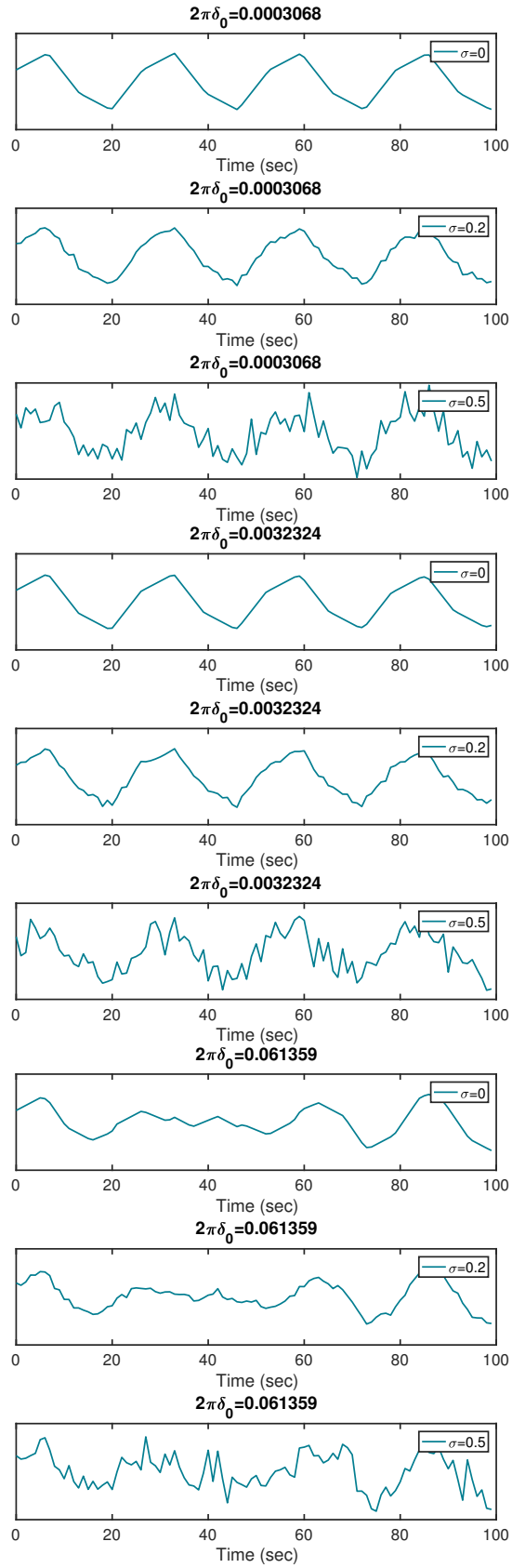


Figure 13: Shape version of $f^{\{1\}}$.

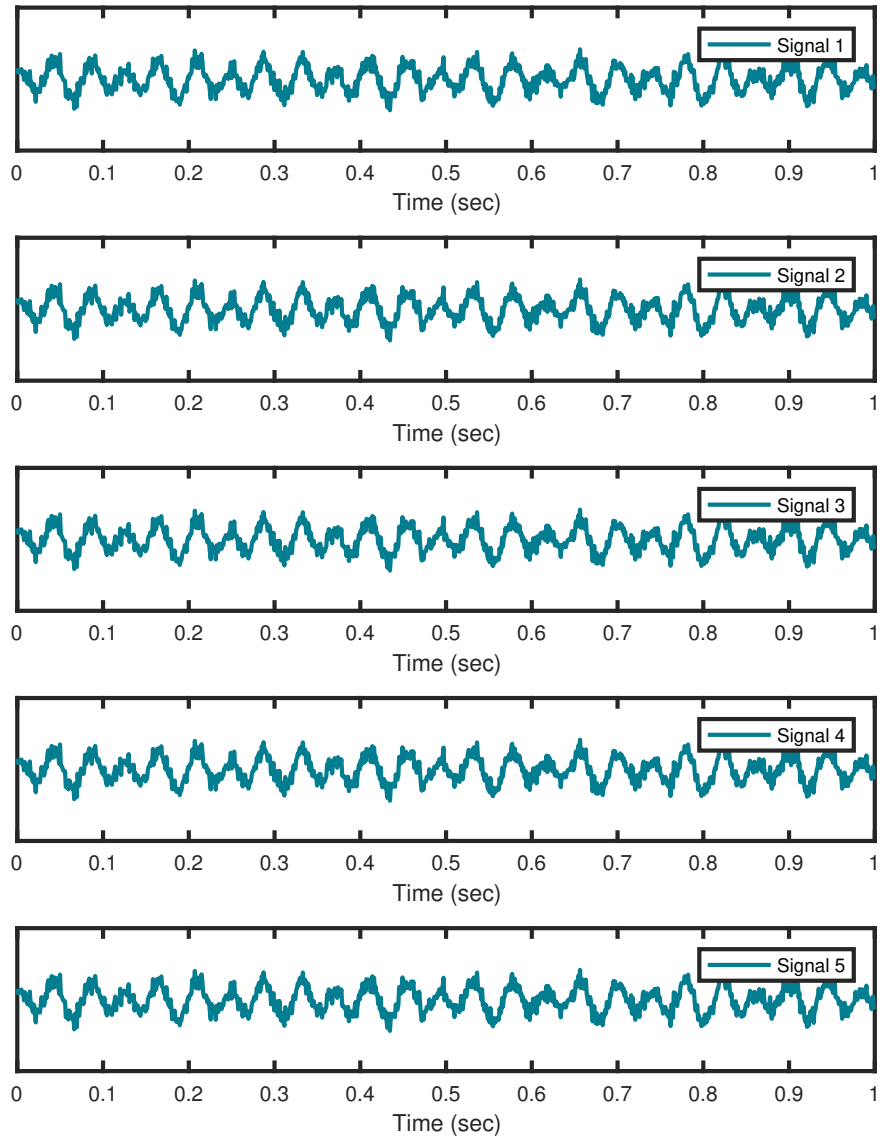


Figure 14: 5 of the 100 random signals in form of $f^{\{2\}}$.

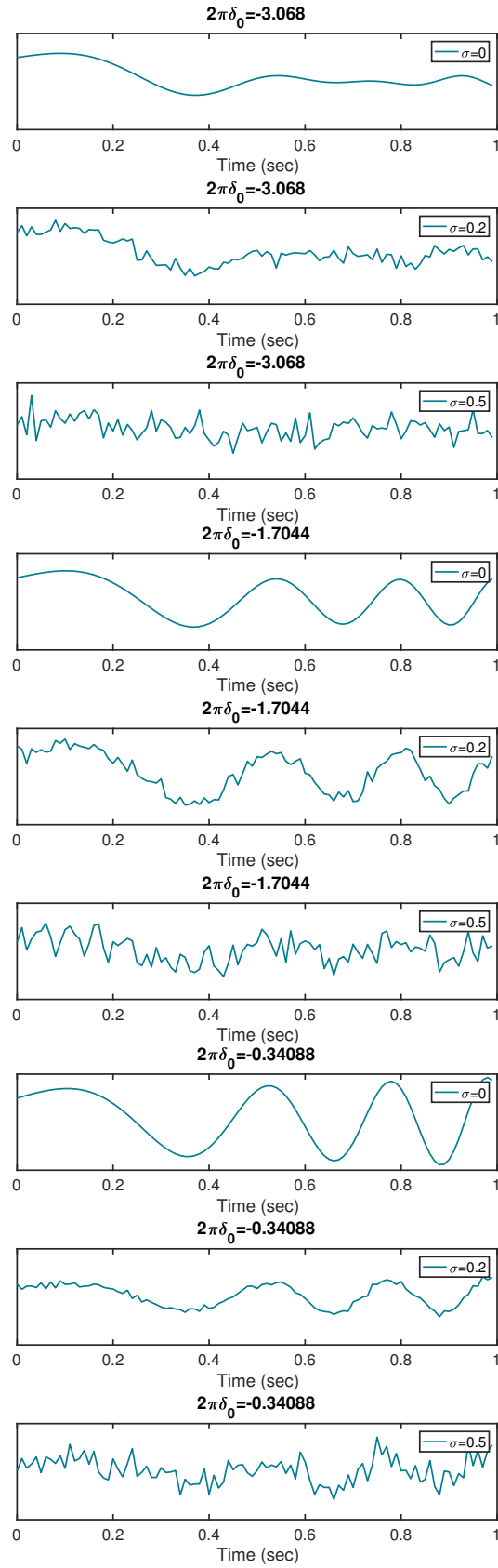


Figure 15: cos version $f^{\{3\}}$.

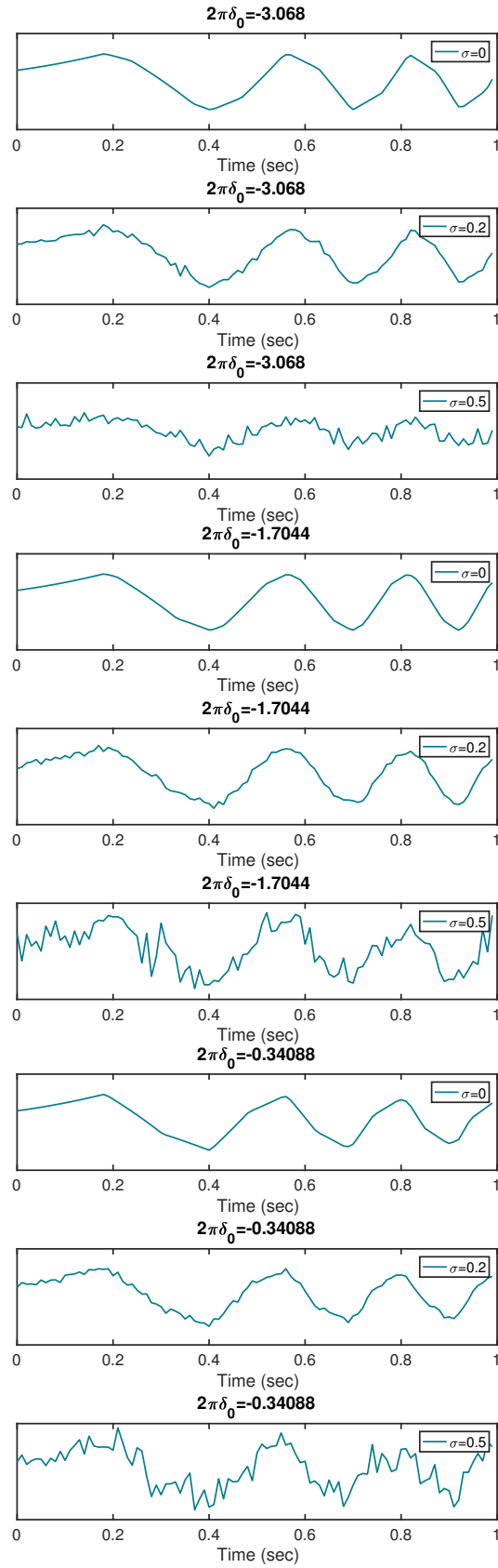


Figure 16: Shape version of $f^{\{3\}}$.

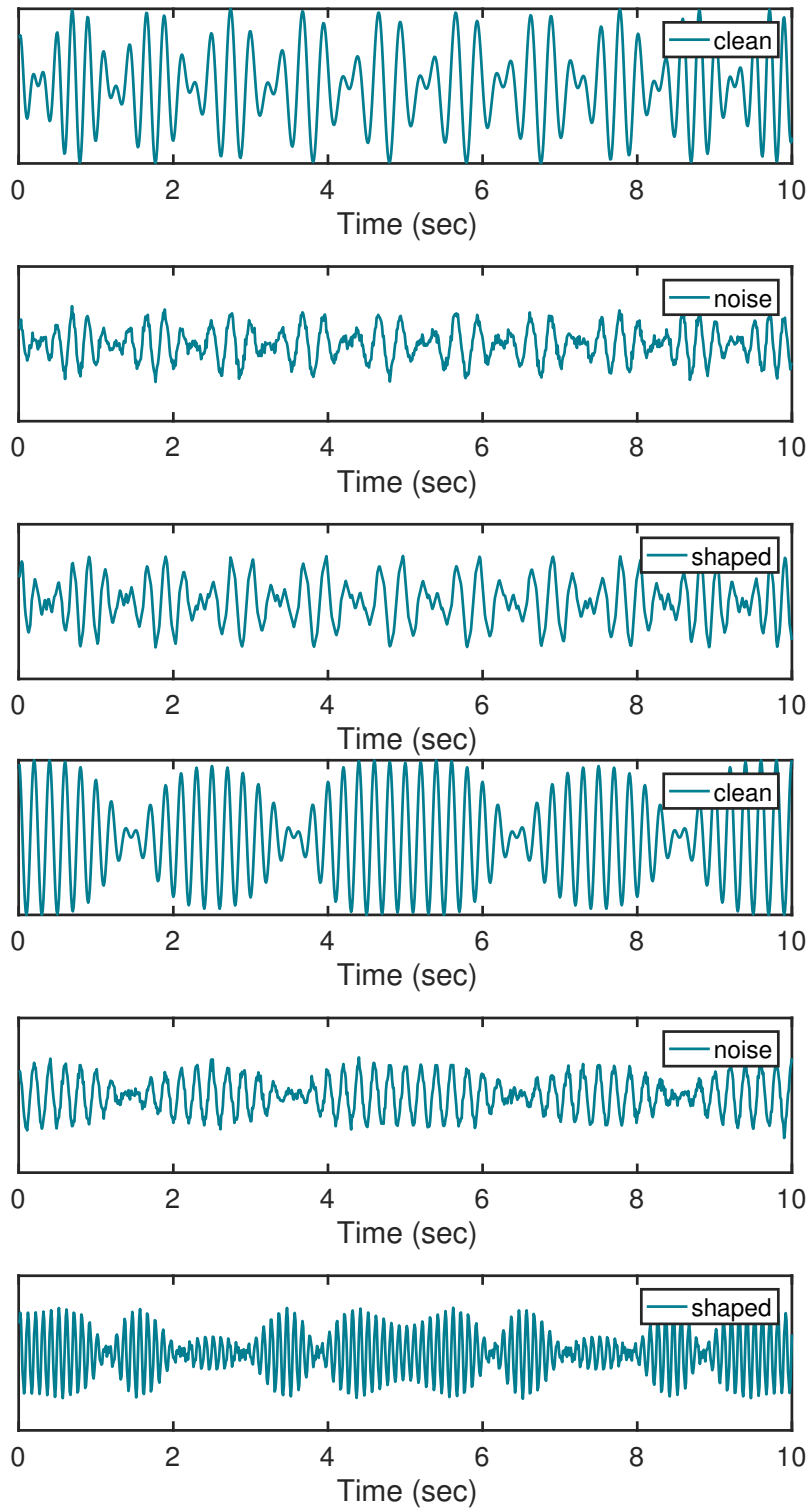


Figure 17: Top three for $f^{(4)}$ and bottom three for $f^{(5)}$.