

PiPs: a Kernel-based Optimization Scheme for Analyzing Non-Stationary 1D Signals

Jieren Xu¹, Yitong Li², Haizhao Yang³,
David Dunson⁴, Ingrid Daubechies⁴

¹ Google LLC, ² Apple Inc, ³Duke University, ⁴University of Maryland College Park
³jierenxu@google.com, ²yitongli@apple.com, ³ingrid.dunson@duke.com
⁴hzyang@umd.edu

December 12, 2022

Abstract

This paper proposes a novel kernel-based optimization scheme to handle tasks in the analysis, *e.g.*, signal spectral estimation and single-channel source separation of 1D non-stationary oscillatory data. The key insight of our optimization scheme for reconstructing the time-frequency information is that when a nonparametric regression is applied on some input values, the output regressed points would lie near the oscillatory pattern of the oscillatory 1D signal only if these input values are a good approximation of the ground-truth phase function. In this work, *Gaussian Process (GP)* is chosen to conduct this nonparametric regression: the oscillatory pattern is encoded as the *Pattern-inducing Points (PiPs)* which act as the training data points in the GP regression; while the targeted phase function is fed in to compute the correlation kernels, acting as the testing input. Better approximated phase function generates more precise kernels, thus resulting in smaller optimization loss error when comparing the kernel-based regression output with the original signals. To the best of our knowledge, this is the first algorithm that can satisfactorily handle fully non-stationary oscillatory data, close and crossover frequencies, and general oscillatory patterns. Even in the example of a signal produced by slow variation in the parameters of a trigonometric expansion, we show that PiPs admits competitive or better performance in terms of accuracy and robustness than existing state-of-the-art algorithms.

1 Introduction

This paper is concerned with 1D single-channel source separation and estimation for oscillatory signals. Suppose a signal $f(t)$ is defined on a time domain $[0, T]$ with K intrinsic components and non-constant frequencies:

$$f(t) = \sum_{k=1}^K f_k(t) = \sum_{k=1}^K a_k(t) s_k(\phi_k(t)), \quad (1)$$

where $a_k(t)$ and $\phi_k(t)$ are smooth, slowly varying functions representing the latent *amplitude* and *phase functions* of the k th component, $f_k(t)$, for $k = 1, \dots, K$. The derivative of phase function $\phi_k(t)$ is called the *frequency function*, denoted as $\omega_k(t)$ and is also assumed to be smooth. $s_k(t)$ is a periodic *shape* (or *pattern*) function for the k th component, describing a potentially complicated evolution pattern of the signal. We assume $s_k(t)$ to be bounded, continuous, to have periodicity 1 and to satisfy $\int_0^1 s_k(t) dt = 0$, with unit L_2 -norm on $[0, 1]$. The variation of $a_k(t)$ and $\omega_k(t)$ are assumed to be sufficiently small and the magnitude of $\omega_k(t)$ is assumed to be large enough such that the pattern is well defined.

One toy example of Model (1) are trigonometric functions. A more complicated example in (Fig 1(b:bottom)) is, *e.g.*, a real Photoplethysmogram (PPG) signal in Figure 1(a); the PPG signal describes the human cardiac and respiratory cycles with $K = 2$ intrinsic components: the first component (Fig 1(b:middle)) represents the beating of the heart and the second represents the cyclic respiratory behavior of the lungs

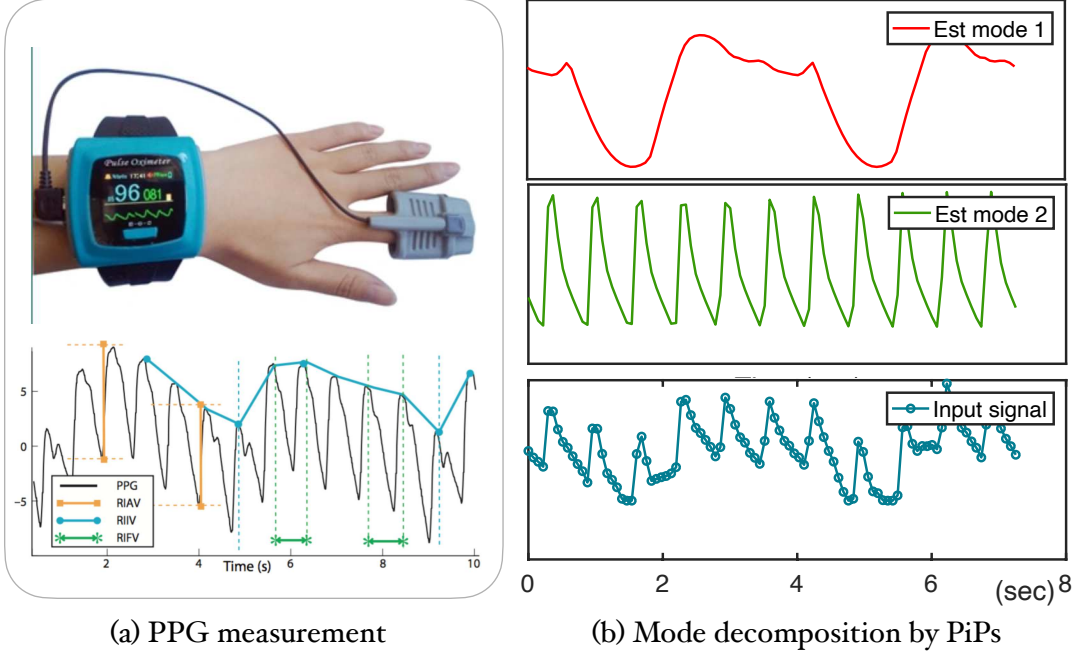


Figure 1: (a) Measurement and several explicitly handcrafted properties of Photoplethysmogram (PPG) signal as a motivating example of non-stationary 1D signals. (b) Reconstructed components (top two figures for Fig 1(b)) of the PPG signal (bottom figure of Fig 1(b)). These two components were reconstructed from only a small portion (100 points) of the samples of the original PPG raw data as visualized in the bottom figure.

(Fig 1(b:top)). Model (1) includes a large family of approximately periodic signals in real applications [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15].

Solving Equation (1) (i.e., identifying amplitude, phase, and shape functions from $f(t)$ in Equation (1)) is a general task that involves several sub-problems: (i) spectral estimation when $\omega_k(t)$ is linear; (ii) adaptive time-frequency analysis [16, 17, 18] that aims to retrieve time-variant information $a_k(t)$, $\phi_k(t)$, $\omega_k(t)$; (iii) mode decomposition [19, 20, 21] that targets the extraction of $f_k(t)$; (iv) pattern recognition [22] to reconstruct $s_k(t)$, etc. Generally, $f(t)$ and K are fed as input information for the above-mentioned approaches.

Despite many successful algorithms for solving these sub-problems, to the best of our knowledge, no algorithm in the literature satisfactorily fulfill the ultimate goal of estimating $a_k(t)$, $\phi_k(t)$ (or $\omega_k(t)$), and $s_k(t)$ when $f(t)$ is fully non-stationary with close and crossover frequencies, and general patterns. Moreover, many existing algorithms require a high sampling rate, which is not always practical (e.g.) for oscillatory data collected by mobile devices, such as portable health monitors (see Figure 1(a)), due to the limit of battery capacity.

This paper proposes a framework that can estimate $a_k(t)$, $\phi_k(t)$ (or $\omega_k(t)$), and $s_k(t)$ simultaneously from relatively few samples of $f(t)$. The algorithm requires a prior input of (1) the number of intrinsic components K ; (2) a rough estimate of the frequency and pattern functions. The estimate in (2) can be quite rough: for instance, for the PPG signal in Figure 1 we initialize the patterns for both components as a sine function, with frequencies of 15 periods/min and 95 beats/min, which is out of common-knowledge rule-of-thumb approximations; the output gives the respective shapes and frequencies for the heart and lung components with the desired accuracy (Figure 1(b)).

Our framework applies a two-stage iteration scheme until convergence: one stage to update phase functions (and amplitudes) and the other stage to update oscillatory patterns. The phase updating stage is the core part of the algorithm. There are four components for nonparametric regression: the input and output of the training points and the input and output of the testing points. Our key intuition is that when a nonparametric regression is applied on some input values, the output regressed points would lie near the oscillatory pattern only if these input values are a good approximation of the ground-truth phase function. The nonparametric regression here is implemented by the *Gaussian Process (GP)*, since the GP-regression-based implementation

shows more robustness compared to several other standard nonparametric regression approaches [23].

In this stage, first, we encode the prior knowledge of the patterns using the *Pattern-inducing Points (PiPs)*. Then we formulate a GP-regression-based optimization problem to retrieve the phase functions by treating the PiPs as training (input-output) points, while the phase functions as the latent testing input and the original signal samples as the testing output respectively. As the input of a GP, the targeted phase function is fed into the correlation kernels to compute the output values of the regression. Since better-approximated phase function generates more precise kernels, thus resulting in a smaller difference between the point-wise kernel-based regression output and the original signals, we design the optimization loss as the L_2 distance between regression output and the noisy measurement of the 1D signal. By optimizing this innovative kernel-based loss function, the latent input phase function is retrieved. In this sense, this stage can also be viewed as a latent GP regression problem that aims to recover the latent input of a GP given the output values.

To enhance the performance of the kernel-based optimization, we transform the nonparametric setting to a semi-parametric setup by deploying a divide-and-conquer strategy. We separate the long signals into multiple localized signal chips. These chips are supported on continuous time intervals which can have intersections, as long as the whole time span of the original signal is fully covered. In the phase-updating stage, since each signal chip is time-localized and the variation of the time-instantaneous information is assumed to be sufficiently small, we propose to use low-order polynomials to model the phase and amplitude functions for each of these local chips. By transforming the original nonparametric model to the current semi-parametric one, we can guarantee the local monotonicity and the smoothness of the time-instantaneous information, thus largely enhancing the robustness of the optimization process. Then we summarize the time-instantaneous information for all chips and feed it into the pattern-updating stage to update the oscillatory patterns.

In the pattern-updating stage, state-of-the-art 1D pattern recovery algorithms, *e.g.*, Recursive Diffeomorphism-Based Regression for Shape Functions (RDBR) [24], are applied to update the oscillatory patterns given the renewed time-frequency information and the noisy measurement of the 1D oscillatory signal. This stage allows us to handle oscillatory signals with a broad class of oscillatory patterns as long as a rough initialization is provided, compared to the traditional methods that are mostly limited to trigonometric oscillatory patterns.

The first and second stages of the algorithm are introduced in Sections 2 and 3. Section 4 summarizes PiPs. As we shall see in the numerical examples in Section 5 PiPs works for a wide range of signals in Model (1) while existing methods fail for certain or all aspects. Moreover, for simple signals that can be handled by super-resolution analysis, PiPs achieves better results compared to several state-of-the-art methods with only reasonable initial values.

2 Estimation of Phase Functions

This section explains how to update the phase and amplitude functions in Model (1) when exact or approximated knowledge of the pattern $s_k(t)$ is provided.

To fix our thoughts, we start by introducing the concept of PiPs; these are on-grid auxiliary points that play a role similar to training points in other learning processes. To be specific, we introduce a formal definition as follows.

Definition 1 (PiPs) Let $s(t)$ be a periodic function that satisfies the assumptions in Section 1. We say that the points $(P_i)_{i=0}^l$, with coordinates $(z_i, u_i) \in \mathbb{R}^2$ for all $i \in \{0, \dots, l\}$ are *Pattern-inducing Points (or PiPs)* for $s(t)$, with tolerance h on the interval $[b_0, b] \subset \mathbb{R}$, where $1 \leq b_0 - b < \infty$, if

- a) $z_i = b_0 + \frac{b-b_0}{l}i$
- b) the continuous affine functions $\tilde{s} : [b_0, b] \rightarrow \mathbb{R}$ with breakpoints at the z_i , and such that $\tilde{s}(z_i) = u_i$, $i = 0, \dots, l$, satisfies $\|(\tilde{s} - s)|_{[b_0, b]}\|_\infty \leq h$.

Figure 2(a) shows a cartoon for one component $s_k(\phi_k(t))$ in Model (1). Figure 2(b) shows one corresponding periodic pattern $s_k(t)$, which is also the unwarping result of $s_k(\phi_k(t))$ w.r.t. $\phi_k(t)$. Figure 2(c) is an example of PiPs (green dots) for $s_k(t)$ on $[0, 4]$ with $l = 16$.

Note that for standard signals with trigonometric patterns, the PiPs is considered to be already known. In practise, we set $b_0 = 0$ without loss of generality. The pattern resolution l should be large enough to

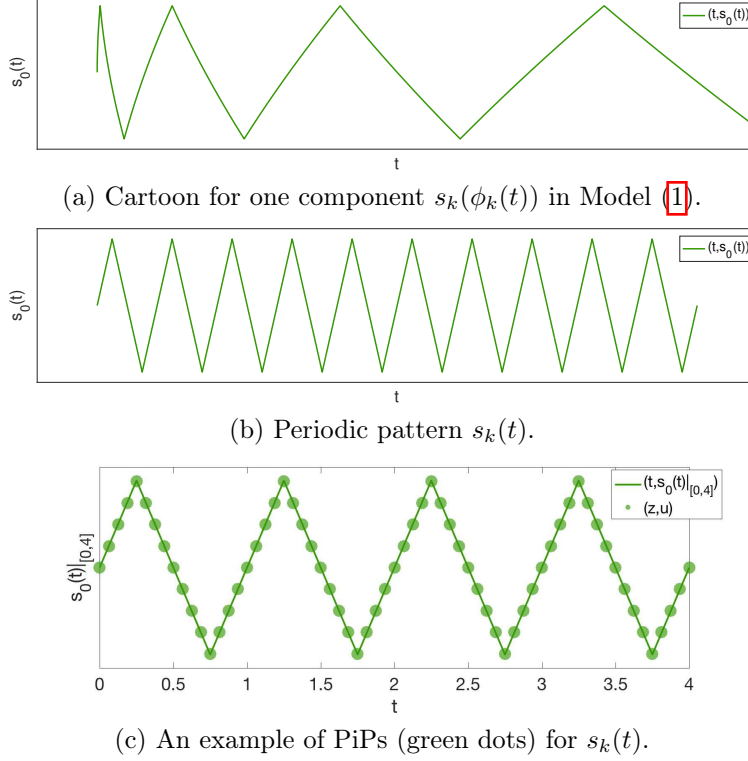


Figure 2: The original oscillatory pattern and two non-oscillatory patterns with no approximation error.

describe the details of each pattern. Moreover, the left-hand bound b should be larger than the upper bound of $\phi_k(t)$ to guarantee the optimization performance.

Next by fixing the PiPs as the training points, a nonparametric-regression-based optimization algorithm is designed to retrieve the latent phase and amplitude functions by maximizing the posterior distribution of the observed samples.

2.1 GP Regression

Suppose $\mathbf{y} \in \mathbb{R}^N$ are observations¹ of

$$y(t) = \sum_k^K y_k(t), \quad \text{where} \quad y_k(t) = f_k(t) + ns_k(t),$$

sampled at time points $\mathbf{t} = [t_1, \dots, t_N] \in \mathbb{R}^N$. Here we consider $ns_k(t)$ as GP with mean 0 and fixed point-wise variance σ_k^2 . Thus $y(t)$, being the sum of K independent GP's, can also be modeled as a GP and tackled with respective tools. In the rest of this section, we illustrate the key idea of this work, *i.e.*, using PiPs ($y(\phi)$) to formulate this nonparametric regression problem, rather than modeling $y(t)$ directly.

For each mode $f_k(t) = a_k(t)s_k(\phi_k(t))$, denote $\mathbf{a}_k = a_k(\mathbf{t})$, $\boldsymbol{\phi}_k = \phi_k(\mathbf{t})$ and $\mathbf{f}_k = \mathbf{a}_k \odot s_k(\mathbf{t})$ as the respective discretization² of $a_k(t)$, $\phi_k(t)$ and $f_k(t)$ at time samples $\mathbf{t}$. It's easy to see that if we set $a_k(t)$ to a constant say $a_k(t) = 1$, the re-arranged mode points $(\boldsymbol{\phi}_k, \mathbf{f}_k) = (\boldsymbol{\phi}_k, s_k(\boldsymbol{\phi}_k))$ should lie exactly on the oscillatory pattern $s_k(t)$. On the other hand, since $\phi_k(t)$ is a strictly monotonic function, for any $\phi(\mathbf{t})$ that deviates from $\phi_k(\mathbf{t})$ module 1, *i.e.*, $\phi(\mathbf{t}) \neq \phi_k(\mathbf{t}) \pmod{1}$, $(\boldsymbol{\phi}, \mathbf{f}_k)$ should deviate from pattern $s_k(t)$. This is illustrated in Figure 3 (a)(b) respectively, where $s_k(t)$ is has a triangular pattern and $\phi_k(t) = 2t$, $t \in [0, 1]$.

As we know, there are four essential components for nonparametric regression: the input and output of the training points, and the input and output of the testing points. Based on the aforementioned observation,

¹We will use bold font for vectors and $(\cdot)_i$ for the i th element of the respective vector.

² $\odot$ is entry-wise product between vectors or matrix.

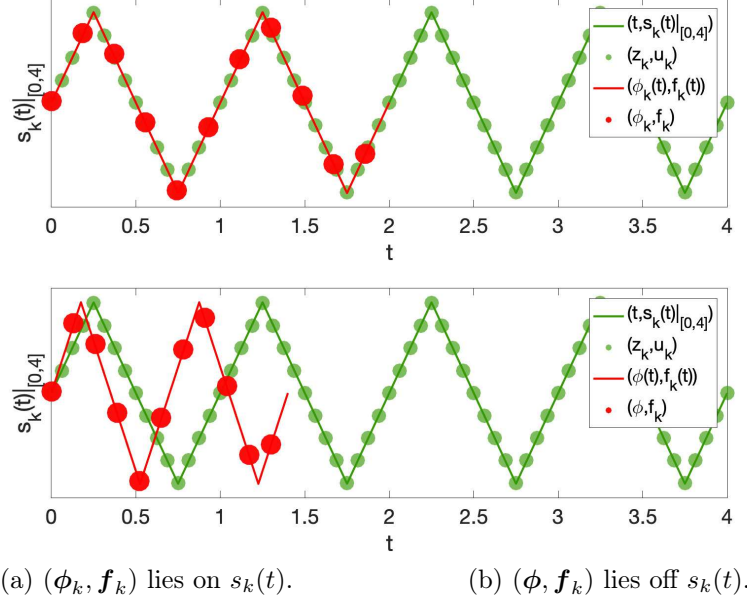


Figure 3: Phase-signal pairs (red points) that lies on or off the $s_k(t)$ (green curve). (a) When the phase is set to the ground truth value ϕ_k , the signal points $\mathbf{f}_k$ lie on $s_k(t)$. Hence non-parametric regression can be used to estimate the signal value (red points) from the PiPs $(\mathbf{z}_k, \mathbf{u}_k)$ (green points). (b) When the phase deviates from ground truth, non-parametric regression can't be used to estimate signal value (red points) given the PiPs (green points).

if we set $(\mathbf{z}_k, \mathbf{u}_k)$'s as PiPs for $f_k(t)$, $[\min \phi_k, \max \phi_k] \in [b_0, b]$, and treat them as training points of some nonparametric regression, a plenty of existing non-parametric approaches can be applied to get a reasonable approximation of $\mathbf{f}_k$ (testing output) at ground truth locations ϕ_k (testing input). Figure 3(a) shows that if $\phi_k = 2t$ is correctly estimated, then $(\phi_k, \mathbf{f}_k)$ should lie on $s_k(t)$. In this case, non-parametric regression can be applied to infer the signal density $\mathbf{f}_k$ at phase position ϕ_k (red dots) using the PiPs (green curve). Figure 3(b) shows that if $\phi = 1.5t \neq \phi_k$, $(\phi, \mathbf{f}_k)$ deviates from $s_k(t)$. In this case, standard nonparametric methods fail to estimate $\mathbf{f}_k$ at ϕ (red dots) using PiPs (green curve) with a large marginal loss.

We use the GP regression with squared exponential (SE) kernel to implement this nonparametric regression step directly on $y_k(\phi_k)$, as it shows robust performance under heavy noise for the stochastic formulation of $y(t)$. We remark here that identical formulation can be derived by performing nonparametric kernel estimator on $f_k(t)$ and treating $ns_k(t)$ as white noise sampled on a continuous interval.

The SE kernel for GP is defined as

$$\mathbf{k}(t, t') = \beta^{\text{SE}} \exp \left(-\frac{1}{2} \alpha^{\text{SE}} (t - t')^2 \right), \quad (2)$$

with kernel parameters β^{SE} and α^{SE} , and the posterior of $\mathbf{y}_k$ given $\phi_k, (\mathbf{z}_k, \mathbf{u}_k)$ can be written as

$$p(\mathbf{y}_k | \phi_k, \mathbf{z}_k, \mathbf{u}_k) = \mathcal{N}(\mathbf{f}_k, \sigma_k^2 I_N) = \mathcal{N}(K_{NM,k} K_{MM,k}^{-1} \mathbf{u}_k, \sigma_k^2 I_N), \quad (3)$$

according to [25], where the third equality is obtained by computing the marginal distribution of $\mathbf{y}_k$. In Eq. (3), $K_{NN,k}$ is a $N \times N$ covariance matrix, $(K_{NN,k})_{ij} = k((\phi_k)_i, (\phi_k)_j)$ ³ for $\phi_k \in \mathbb{R}^N$, and similarly $(K_{MM,k})_{ij} = k((\mathbf{z}_k)_i, (\mathbf{z}_k)_j)$ for $\mathbf{z}_k \in \mathbb{R}^{bl}$, $(K_{MN,k})_{ij} = k((\mathbf{z}_k)_i, (\phi_k)_j) = (K_{NM,k})_{ji}$.

2.2 Optimization Loss

$p(\mathbf{y}_k | \mathbf{u}_k, \phi_k)$ is higher when ϕ_k are more accurately provided. Based on this fact and independence between different modes, a loss function measuring the differences between the joint probability model $\prod_k p(\mathbf{y}_k | \mathbf{u}_k, \phi_k, \mathbf{z}_k)$

³Denote $(\cdot)_i$ as the i th entry of a vector, and $(\cdot)_{ij}$ as the ij th entry of a matrix.

and the ground truth signal $\mathbf{y}$ is derived to update the latent phase functions. To be specific, when $a_k(t) \equiv 1$, we have

$$\begin{aligned} p(\mathbf{y}|\phi_1, \mathbf{z}_1, \mathbf{u}_1, \dots, \phi_K, \mathbf{z}_K, \mathbf{u}_K) &= p(\mathbf{y}|\mathbf{y}_1, \dots, \mathbf{y}_K) \prod_{k=1}^K p(\mathbf{y}_k|\phi_k, \mathbf{z}_k, \mathbf{u}_k) \\ &= \mathcal{N}\left(\sum_{k=1}^K K_{NM,k} K_{MM,k}^{-1} \mathbf{u}_k, \sum_{k=1}^K \sigma_k^2 I_N\right). \end{aligned} \quad (4)$$

We recover the phase function $\phi(t)$ by maximizing the conditional probability of the signal distribution defined in Eq. (4). When $a_k(t)$ is not a constant, Eq. (4) can be further modified as following,

$$\begin{aligned} \mathcal{L}(\mathbf{y}; \mathbf{z}_1, \mathbf{u}_1, \dots, \mathbf{z}_K, \mathbf{u}_K; \phi_1, \mathbf{a}_1, \dots, \phi_K, \mathbf{a}_K) \\ = \log p(\mathbf{y}|\phi_1, \mathbf{z}_1, \mathbf{u}_1, \dots, \phi_K, \mathbf{z}_K, \mathbf{u}_K) = \log \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \end{aligned} \quad (5)$$

where

$$\begin{aligned} \boldsymbol{\mu} &= \sum_{k=1}^K \mathbf{a}_k \odot K_{NM,k} K_{MM,k}^{-1} \mathbf{u}_k, \\ \boldsymbol{\Sigma} &= \sum_{k=1}^K \sigma_k^2 I_N. \end{aligned}$$

We expect to update $\phi_k(t)$ and amplitudes $a_k(t)$ by maximizing $\mathcal{L}$ given in Eq. (5).

2.3 Semi-parametric setting: Parameterization of Phases and Amplitudes

However, directly updating phase and amplitude functions by maximizing the loss function in Eq. (5) is highly unstable when no prior information is considered, *e.g.*, $\phi_k(t)$ is smooth and monotone. Thus, we learn the phase function $\phi_k(t)$ and amplitude function $\mathbf{a}_k(t)$ as sample points from a function, instead of treating ϕ_k and $\mathbf{a}_k$ as independent variable. Among all possible choices, representing $\phi_k(t)$ as low-degree polynomials with order D is effective in practice:

$$\phi_k = \sum_{d=0}^D (B)_{kd} \mathbf{t}^d, \quad \text{and} \quad \mathbf{a}_k = \sum_{d=0}^{D_c} (C)_{kd} \mathbf{t}^d. \quad (6)$$

Here $\mathbf{t}^j$ is the j th power of $\mathbf{t}$. B and C are $N \times (D+1)$ and $N \times (D_c+1)$ real matrices, with the k th row ($\mathbf{b}_k$ and $\mathbf{c}_k$) representing the polynomial coefficients of $\phi_k(t)$ and $\mathbf{a}_k(t)$, respectively. Substituting the parameterized forms of ϕ_k and $\mathbf{a}_k$ into Eq. (5), we can get

$$\mathcal{L}(\mathbf{y}, \mathbf{z}_1, \mathbf{u}_1, \dots, \mathbf{z}_K, \mathbf{u}_K; B, C) = \log \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \quad (7)$$

where

$$\begin{aligned} \boldsymbol{\mu} &= \sum_{k=1}^K \left(\sum_{d=0}^{D_c} (C)_{kd} \mathbf{t}^d \right) \odot K_{NM,k} K_{MM,k}^{-1} \mathbf{u}_k, \\ \boldsymbol{\Sigma} &= \sum_{k=1}^K \sigma_k^2 I_N. \end{aligned}$$

Note that parameters in ϕ_k are implicitly included in the kernel matrices $K_{NM,k}$ and $K_{NN,k}$.

The nonparametric setting (Eq. (5)) thus transforms to a semi-parametric setting as Eq. (7). As a result, the monotonicity and smoothness of ϕ_k is guaranteed, along with a more stabilized performance for this highly non-convex optimization problem. Ignoring the constant terms, we end up with an equivalent MSE loss between $\mathbf{y}$ and $\boldsymbol{\mu}$

$$\mathcal{L}_0(\mathbf{y}; B, C, \mathbf{z}_1, \mathbf{u}_1, \dots, \mathbf{z}_K, \mathbf{u}_K) = \|\mathbf{y} - \boldsymbol{\mu}\|_2, \quad (8)$$

which will be directly optimized by gradient descent methods.

2.4 Divide-and-Conquer Strategy

In practice, setting $D = 2$ provides a reasonable approximation to signals localized in time, because both $\phi_k(t)$ and $a_k(t)$ vary slowly. For signals that can not be well approximated via lower degree polynomials, a *divide-and-conquer* strategy is applied to obtain a global point estimation for the phase (and amplitude) functions. This divide-and-conquer strategy involves three steps. (i) We separate the long signals into multiple localized signal chips. These chips are supported on continuous time intervals which can have intersections, as long as the whole time span of the original signal is fully covered. (ii) We update $\phi_k(t)$ (and $a_k(t)$) for each short chip using polynomial estimation as Eq. (7) or Eq. (8). (iii) A robust curve fitting algorithm [26, 27] is applied to obtain the final global estimation from the previous steps.

If the oscillatory patterns are unknown and need to be updated, to improve the pattern estimate result, Step (iii) can be replaced by some more time-consuming variants. This is detailed in the following section.

3 Estimation of Oscillatory Patterns

In this section, we introduce the way of approximating the oscillating patterns $s_k(t)$. The phase and amplitude functions ϕ_k and a_k estimated from the previous sections are fixed in this step. We do not aim at closed formulas for $s_k(t)$. As introduced in Section 2 it is sufficient to estimate the pattern inducing points (z_k, u_k) to represent the non-oscillating pattern $\hat{s}_k(t)|_{[a,b]}$. When amplitude and phase functions are given, shape function estimation has been studied thoroughly in previous works, like [28, 24, 29, 30]. There are no quantitative criteria to measure how well the shape function estimate performs when the amplitude and phase function estimate is not very good. These methods achieve good performance when the inferred amplitude and phase functions are close to the ground truth.

3.1 Summarize Output from Localized Chips

In this section, we illustrate two variants in Step (iii) of Section 2.4 that can improve the accuracy of pattern estimation.

First, since different signal chips generate instantaneous information estimates with distinct qualities and these qualities of instantaneous information can be partially manifested by the final loss of Eq. (7) or Eq. (8), we propose a loss-selective variant of Step (iii). To be specific, instead of feeding all the chips' output into a smooth curve fitting module to generate the pattern estimate, we select chips with relatively low loss value for the pattern estimation. For this end, two constants are prefixed: a threshold value τ_1 to admit all chips that satisfies $\mathcal{L} < \tau_1$; if all chips has loss larger than τ_1 , then we use another quantile constant τ_2 to select a portion of chips with the lower loss value. As a result, we can enhance the possibility that better estimated instantaneous information is applied, while instantaneous information with lower qualities is discarded, to update the oscillatory patterns.

However, empirical experiments show that lower loss value from Eq. (7) or Eq. (8) does not always guarantee a better pattern recovery. We further propose a more expensive loss-selective variant of Step (iii) by computing and summarizing the distribution of the final losses generated by the chips in the next iteration step for each chip in the current iteration step. In practice, we set the average of final losses in the next iteration as our fresh chip quality indicator and then apply the identical chip selective scheme work as illustrated in the aforementioned paragraph. Although it's not always the case that this fresh indicator guarantees a better guideline for pattern updating, this variant can guarantee a lower instantaneous information loss in the next round of iteration. We note that this variant involves extensive computing of the trial-and-errors for selecting ideal chips, thus is highly time-consuming compared to the first variant.

4 Overview of PiPs

In this section, an overview of the whole algorithm is presented. PiPs repeatedly applies alternatively updates between spectral information and oscillatory patterns until convergence. The overall loss function of PiPs can

be written as:

$$\mathcal{L}(\mathbf{y}; \mathbf{B}, \mathbf{C}, \mathbf{z}_1, s_1(\cdot), \dots, \mathbf{z}_K, s_K(\cdot)) = \log \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \quad (9)$$

where $\boldsymbol{\mu} = \sum_{k=1}^K \left(\sum_{d=0}^{D_c} (C)_{dk} \mathbf{t}^d \right) \odot K_{NM,k} K_{MM,k}^{-1} s_k(\mathbf{z}_k)$, and $\boldsymbol{\Sigma} = \sum_{k=1}^K \sigma_k^2 \mathbf{I}$.

The goal is to maximize the conditional probability of $p(\mathbf{y}|\mathbf{z}_1, \dots, \mathbf{z}_K)$ with respect to the parameters in phase, amplitude and shape functions, i.e.,

$$\sup_{s_k(\cdot) \in \mathcal{S}} \max_{\mathbf{B}, \mathbf{C}} \mathcal{L}(\mathbf{y}; \mathbf{B}, \mathbf{C}, \mathbf{z}_1, s_1(\cdot), \dots, \mathbf{z}_K, s_K(\cdot)), \quad (10)$$

where $\mathcal{S}$ indicates the set of predefined non-oscillatory patterns supported on $[0, 1]$. $\mathbf{B}$ and $\mathbf{C}$ are the matrix form of phase and amplitude functions in Eq. (6). The pseudo-code of the proposed algorithm is given in Algorithm 1. We only put the most basic update procedure in the pseudo-code. We implement the gradient descent using Adam [31] with learning rate set to 0.005.

For each outer loop, we first update the parameterized phase and amplitude functions for each component using the GP based gradient descent method as introduced in Section 2. Secondly, the oscillation patterns are updated with RBDR as illustrated in Section 3. For sparsely sampled signals (less than 1000 sample points as with the sparse PPG example), PiPs usually converges with two or three outer loop iterations with phase updating part converges within 3000 steps. An illustrative convergence pattern is shown in Figure 4 whereas we can observe the pattern updating stage is effective and crucial to the overall optimization process. Theoretical analysis of this alternative approach will be treated as a future work of this approach.

Algorithm 1 (PiPs) Note that an alternative is to initialize the phase $\{\phi_k\}$ and amplitude $\{a_k\}$ as fixed, while updating the oscillatory patterns first. The order of updating which component first depends on the signal and prior knowledge. Note that if the initialization of oscillatory patterns is better than those of amplitude and phase functions, we update oscillatory patterns first. Since the problem is not convex, the global convergence analysis of the proposed algorithm would be interesting future work.

Input: N measurements of $(\mathbf{t}, \mathbf{y})$, the number of components K , the polynomial degrees D_c and D , input of PiPs $\mathbf{z}_k$ ($k=1, \dots, K$) and the accuracy parameter ϵ .

Initialization: Initialize the estimates of oscillatory patterns $s_k(\cdot)$, the phase and amplitude parametrized matrices $\mathbf{B}$ and $\mathbf{C}$, and set $iter = 0$. For long signals, divide the long signal into short chips (Section 3).

while $iter < MaxIter$ and model not converge **do**

Compute $\mathbf{u}_k = s(\mathbf{z}_k)$ ($k = 1, \dots, K$) for each component.

for each chip do

for $i = 1$ to $Iter_{\phi, a}$ **do**

Fix $s_k(\cdot)$, compute $\Delta \mathbf{b}_k = \frac{\nabla \mathcal{L}}{\nabla \mathbf{a}_k} \frac{\nabla \mathbf{a}_k}{\nabla \mathbf{b}_k}$, $\Delta \mathbf{c}_k = \frac{\nabla \mathcal{L}}{\nabla \phi_k} \frac{\nabla \phi_k}{\nabla \mathbf{c}_k}$ for $k = 1, \dots, K$. (Section 2.3).

Update $\mathbf{b}_k = \mathbf{b}_k + \gamma \Delta \mathbf{b}_k$, and $\mathbf{c}_k = \mathbf{c}_k + \gamma \Delta \mathbf{c}_k$.

Compute $\mathbf{a}_k = \sum_{d=0}^D B_{dk} \mathbf{t}^d$ and $\phi_k = \sum_{d=0}^D C_{dk} \mathbf{t}^d$.

Update oscillatory patterns $s_k(\cdot)$ for $k = 1, \dots, K$ with selected chips. (Section 3).

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.

In many applications, e.g. ECG and PPG data analysis, heuristic properties of the physical system are often available and we know the rough range of instantaneous frequencies $\omega_k(t)$. Hence, we can apply a band-pass filter to $f(t) = \sum_{k=1}^K a_k(t) s_k(\phi_k(t)) = \sum_{k=1}^K \sum_n \hat{s}_k(n) a_k(t) e^{jn\phi_k(t)}$ with Fourier expansion on the shape function. Then, we estimate amplitude and phase of $\hat{s}_k(1) a_k(t) e^{j\phi_k(t)}$ in a certain frequency band using traditional time-frequency analysis methods [17, 18]. Generally, $\hat{s}_k(1)$ is much larger than the rest Fourier coefficients. Thus, by setting K to 1, Synchrosqueezed-based methods [32] can be directly applied without a band filter. Another initialization method is to directly set $\mathbf{u}_k = \sin(\mathbf{z}_k)$. Since we adopt local patch segmentation in Section 2, components decomposed by Fourier series expansion become approximately orthogonal to each other in a short time period ($\{\hat{s}_k(1) a_k(t) e^{j\phi_k(t)}\}_{k=1}^K$). Hence, Algorithm 1 can recover the amplitude and phase functions corresponding to $\{\hat{s}_k(1)\}_{k=1}^K$ since they usually have the largest K magnitude.

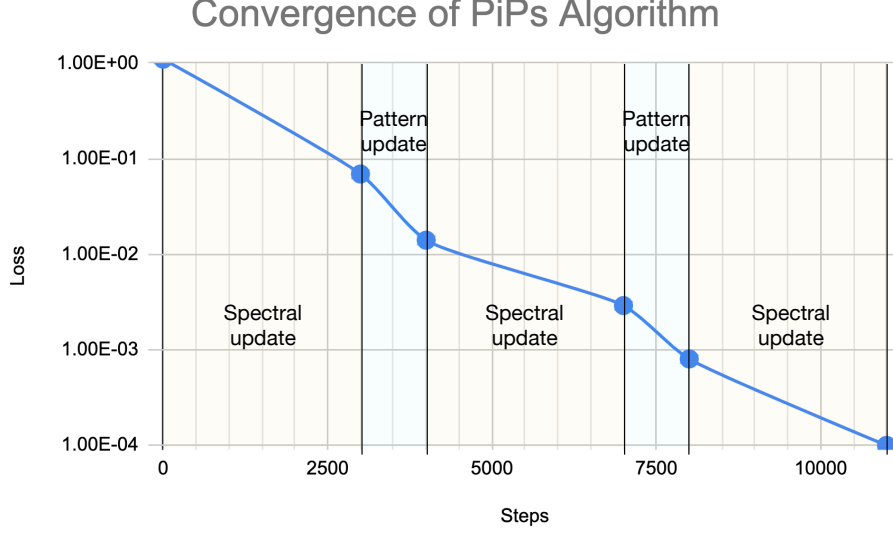


Figure 4: Illustration for the convergence process of Algorithm 1 for the PPG sparse dataset.

5 Experiments

In this section, we provide numerical examples to demonstrate the performance of PiPs⁴, especially in the case of super-resolution and adaptive time-frequency analysis. Optimization problems in all examples are solved by Adam [33] aiming at better local minimizers. We choose degree-1 (or degree-2 when specified) polynomials to approximate local amplitude and phase functions in these optimization problems. The hyperparameters of PiPs 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 to use 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. We also created a set of oscillation patterns with non-trigonometric shapes to facilitate testing. Whenever appeared, we refer

$$s_1^{\text{eg}}(t) = \begin{cases} \frac{16}{\pi}t - \frac{4}{\pi}, & \text{when } 0 \leq t < 0.5, \\ \frac{12}{\pi} - \frac{16}{\pi}t, & \text{when } 0.5 \leq t < 1, \end{cases}$$

and

$$s_2^{\text{eg}}(t) = \begin{cases} -8t + 3, & \text{when } 0.25 \leq t < 0.5, \\ 8t - 7, & \text{when } 0.75 \leq t < 1, \\ 1, & \text{otherwise.} \end{cases}$$

These shape functions have been visualized in Fig. 9(b) and (c).

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 [34], ME [14, 35, 36], and ESPRIT [15]. Hence, we will compare PiPs with these methods⁵ to show the advantages of PiPs. Although the Fourier transform usually fails [34] to identify $\{a_k\}$ and $\{\omega_k\}$, we use its results as the initialization for PiPs.

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 $s_1(\cdot)$ and $s_2(\cdot)$ are chosen as follows:

⁴Code is available on <https://github.com/JierenXu/PiPs>

⁵Code from <http://people.ece.umn.edu/~georgiou/files/HRTSA/SpecAn.html>

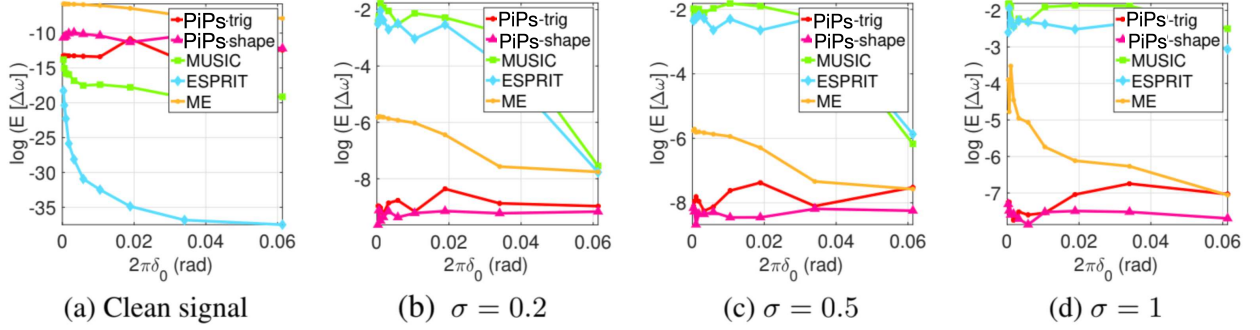


Figure 5: Frequency estimate (absolute) error of $f^{\{1\}}(t) = \cos(2\pi\omega_1 t) + \sin(2\pi\omega_2 t)$, where $\omega_1 = 38.8/1024$ and $\omega_2 = (38.8 + \delta_0)/1024$ with different δ_0 and white noise $\mathcal{N}(0, \sigma^2)$.

- (i) $s_1(t) = \cos(t)$ and $s_2(t) = \sin(t)$ for standard super resolution comparison to other baseline methods. The estimation results by PiPs are denoted as PiPs-trig and are visualized as red lines in Fig. 5
- (ii) $s_1(t) = s_1^{eg}(t)$ and $s_2(t) = s_2^{eg}(t)$ for super resolution comparison with special oscillation patterns. The estimation results by PiPs are denoted as PiPs-shape and are visualized as pink lines in Fig. 5

In these two examples, $\omega_1 = 38.8/1024$ and $\omega_2 = (38.8 + \delta_0)/1024$; δ_0 varies from $0.05/1024$ to $10/1024$; and the noise variance is $\sigma^2 = 10^{-1.6}$. The sampling rate is 1 Hz and the number of samples is $N = 100$ in this example. Fig. 5 shows the frequency estimation accuracy of PiPs, MUSIC, ESPRIT, and ME.

As we can see, PiPs 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). All baseline methods can not directly apply to instantiation (ii) with non-trigonometric oscillation patterns besides PiPs.

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. 6 summarizes the results of frequency estimates in this experiment. ESPRIT and MUSIC lose accuracy in all tests. PiPs and ME achieve high accuracy when the number of samples is large and PiPs 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 PiPs for estimating close and crossover time-varying instantaneous frequencies. An adaptive time-frequency analysis algorithm, ConceFT [37], is used as a comparison. And local approximation degree is set to $d = 2$ in this section.

Close frequencies and phase estimation error We use $f^{\{3\}}(t) = s_1(2\pi(10/10.24t + 230/10.24t^2)) + s_2(2\pi((10/10.24 + \delta_0)t + 250/10.24t^2))$, where the two instantiations of $s_1(\cdot)/s_2(\cdot)$ are (i) $s_1 = \cos/s_2 = \sin$ (Fig. 7(b)) and (ii) $s_1 = s_1^{eg}/s_2 = s_2^{eg}$ as in Fig. 7(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 [38] 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, PiPs still can identify instantaneous frequencies and phases with reasonably good accuracy. The result is summarized in Fig. 7

Fig. 7(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

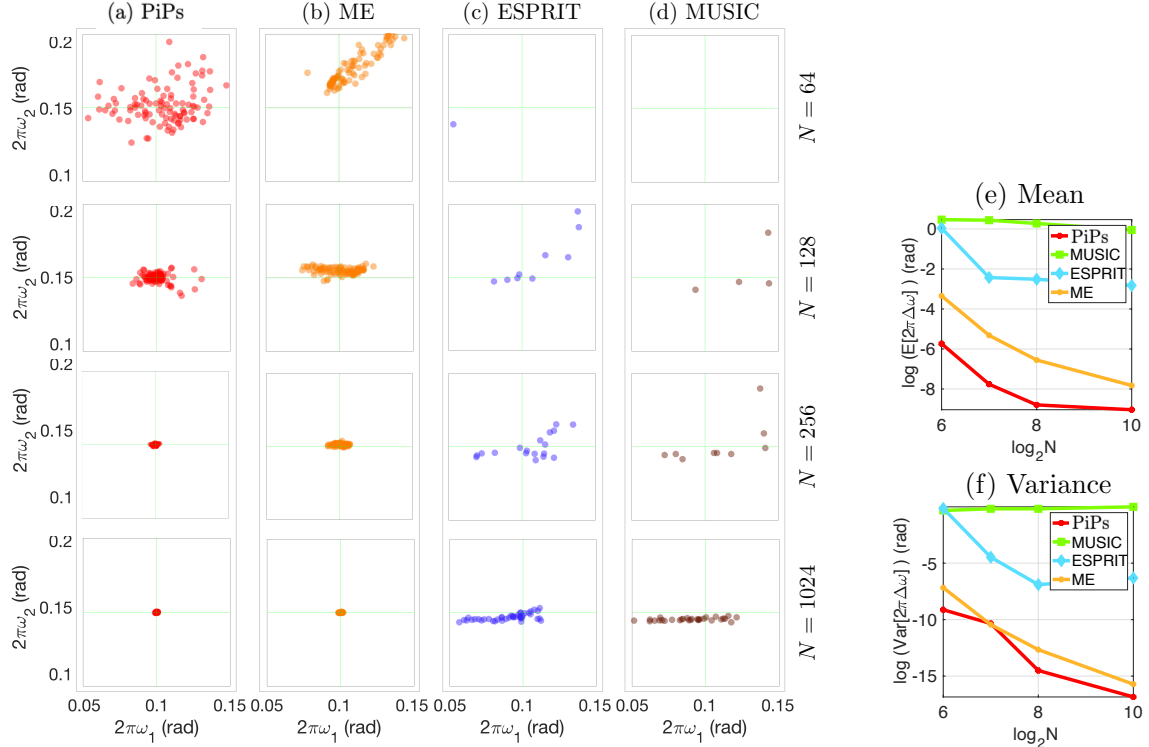


Figure 6: Left: results of $f^{\{2\}}$ with different number of samples $N = 64, 128, 256$, and 1024 from top to bottom, and by different methods in an order of PiPs, ME, ESPRIT, and MUSIC from left to right. The ground truth frequencies are $(2\pi\omega_1, 2\pi\omega_2) = (0.1, 0.15)$. 100 tests with 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.

the first row of Fig. 7 (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 a large noise case with $\sigma_0 = 1$, PiPs 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, PiPs 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. 8 visualizes the ground truth instantaneous frequencies, the time-frequency distribution by ConceFT, the initialization, and the estimation results of PiPs. 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 PiPs. Although the initialization is very poor, PiPs is still able to estimate the instantaneous frequencies with a reasonably good accuracy no matter in clean or noisy cases. When the number of components K is known, we generally can average out the energy band to obtain one instantaneous frequency function and initialize all instantaneous frequencies in PiPs using this function from empirical observations. Similar initialization strategy is used in the following examples.

5.3 Estimation of amplitudes, phases, and shapes simultaneously

Finally, we apply PiPs 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. 9. 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.

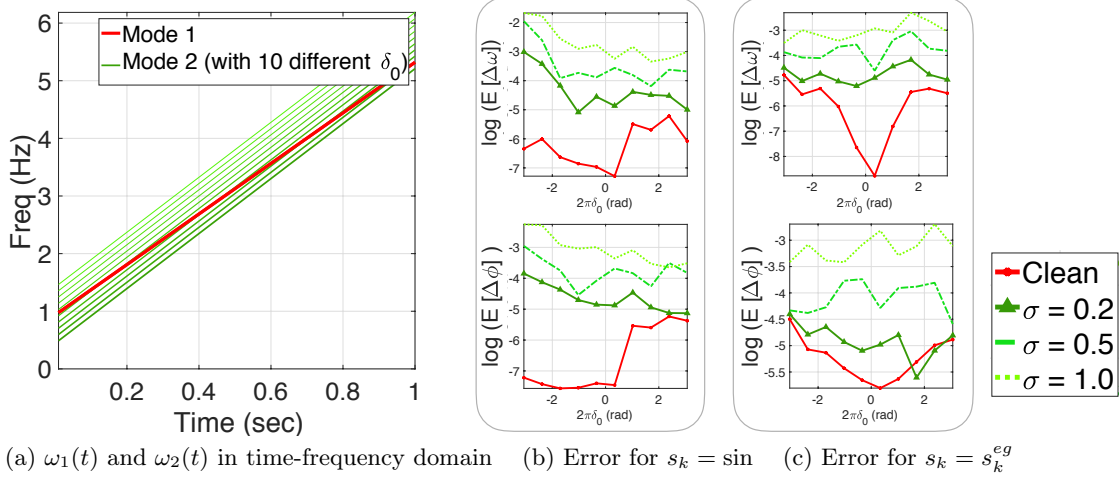


Figure 7: Short signal $f^{\{3\}}(N = 100)$ with linear frequencies. (a) visualizes all the instantaneous frequencies of our synthetic components as the spectral gap parameter δ_0 takes the values $(i - 5)/10.24$ for $i = 0, 1, \dots, 9$. (b) is the estimation error for frequency (top) and phase (bottom) estimates when $s_k = \sin$; (c) is for $s_k = s_k^{eg}$.

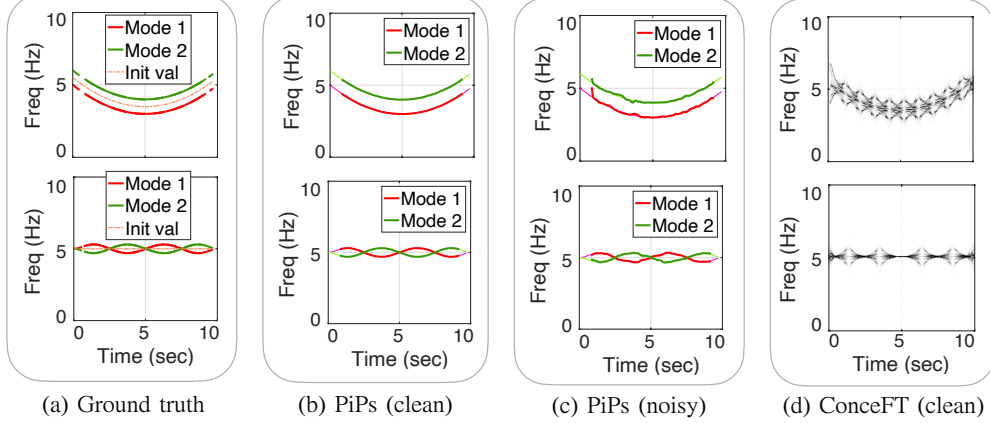


Figure 8: Instantaneous frequency estimates for signals with close and crossover frequencies. (a) ground truth instantaneous frequencies and initialization of PiPs. (b) estimated instantaneous frequencies for clean signals. (c) estimated instantaneous frequencies for noisy signals. (d) time-frequency distribution by ConceFT.

The frequency estimates are initialized as *one* constant centered in the peak spectrogram by ConceFT (see Fig. 9 (a)). As we can see in Fig. 9 (b) and (c), PiPs is able to estimate shape functions with a reasonably good accuracy and the reconstructed components match the ground truth components very well.

In Fig. 10 and 11, a similar initialization strategy is applied to two more examples, $f^{\{7\}}(t) = \sum_{k=1}^2 s_k^{eg}(\omega_k^{smooth}(t)) + \mathcal{N}(0, 0.2)$ and $f^{\{8\}}(t) = \sum_{k=1}^2 s_k^{eg}(\omega_k^{near}(t)) + \mathcal{N}(0, 0.2)$, respectively. Here

$$\begin{aligned} \omega_1^{smooth}(t) &= 0.08t(t - 10), \quad \omega_2^{smooth}(t) = 0.08t(t - 10) + 1, \\ \omega_1^{near}(t) &= t + 4, \quad \text{and} \quad \omega_2^{near}(t) = -t + 6. \end{aligned}$$

In other words, signal $f^{\{7\}}(t)$ and $f^{\{8\}}(t)$ both have more difficult frequency time-frequency representation to resolve, one with non-linearity and one with contact frequency curves. The results in Fig. 10 and 11 show that PiPs can still obtain sharp mode decomposition results for both clean and noisy cases with sparsely sampled data points in these challenging examples.

In the last example, we apply PiPs to a real signal from photoplethysmogram (PPG) (see Fig. 1). 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 lungs in the human body, where Fig. 1(b) shows the mode decomposition result. As can be seen, two modes with

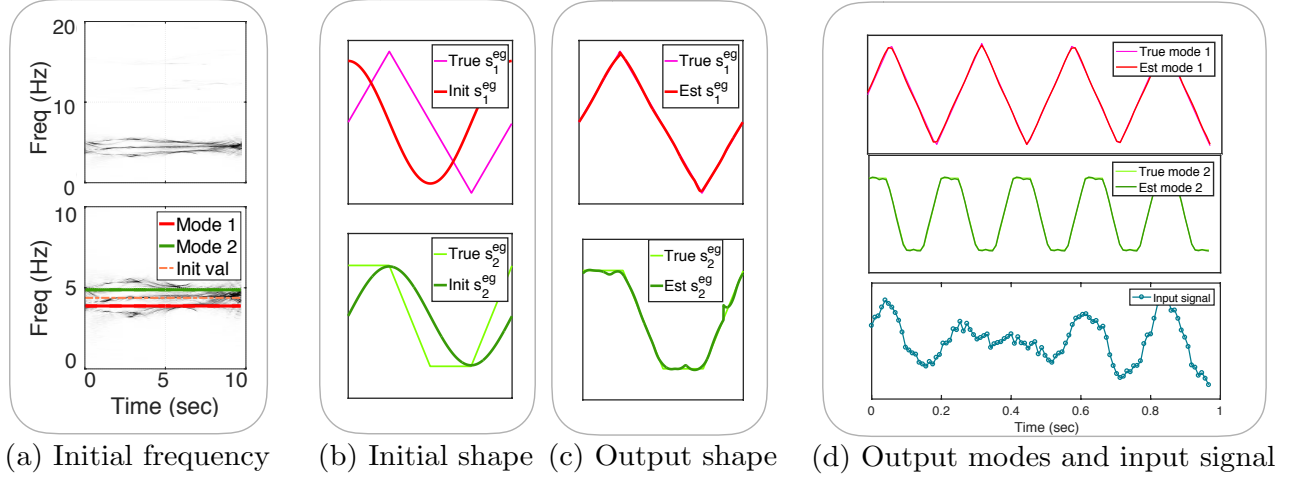


Figure 9: PiPs is applied to estimate the amplitude, phase, and shapes of a synthetic signal $f^{\{6\}}(t)$ consisting of two components. (a) the time-frequency distribution of $f^{\{6\}}(t)$ by ConceFT in two different frequency ranges. ConceFT cannot reveal the ground truth instantaneous frequencies (in red and green). But we can initialize PiPs 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 PiPs.

highly domain-specific patterns are accurately recovered under the naive cosine and sine oscillation pattern initialization from these 100 samples.

6 Conclusion

This paper proposed a novel alternatively learning scheme (PiPs) between spectral information and periodical patterns to address several oscillatory data analysis problems, including signal decomposition, super-resolution, and signal sub-sampling. The method achieves state-of-the-art results for noisy and sparsely sampled cases on several datasets, and demonstrates its potentials in real world applications. Though numerical convergence of the proposed method has been observed, an interesting future direction is to analyze the convergence theoretically, especially statistical analysis in the presence of noise.

Acknowledgments. H. Y. was partially supported by the NSF Award DMS-2244988 and DMS-2206333, and the Office of Naval Research Award N00014-23-1-2007.

References

- [1] Albena Dimitrova Veltcheva. Wave and group transformation by a hilbert spectrum. *Coastal Engineering Journal*, 44(4), 2002.
- [2] Wei Huang, Zheng Shen, Norden E. Huang, and Yuan Cheng Fung. Engineering analysis of biological variables: An example of blood pressure over 1 day. *Proc. Natl. Acad. Sci.*, 95, 1998.
- [3] Haizhao Yang and Lexing Ying. Synchrosqueezed curvelet transform for two-dimensional mode decomposition. *SIAM Journal on Mathematical Analysis*, 46(3):2052–2083, 2014.
- [4] Haizhao Yang, Jianfeng Lu, W.P. Brown, I. Daubechies, and Lexing Ying. Quantitative canvas weave analysis using 2-D synchrosqueezed transforms: Application of time-frequency analysis to art investigation. *Signal Processing Magazine, IEEE*, 32(4):55–63, July 2015.
- [5] Hau-Tieng Wu, Yi-Hsin Chan, Yu-Ting Lin, and Yung-Hsin Yeh. Using synchrosqueezing transform to discover breathing dynamics from ECG signals. *Applied and Computational Harmonic Analysis*, 36(2):354 – 359, 2014.

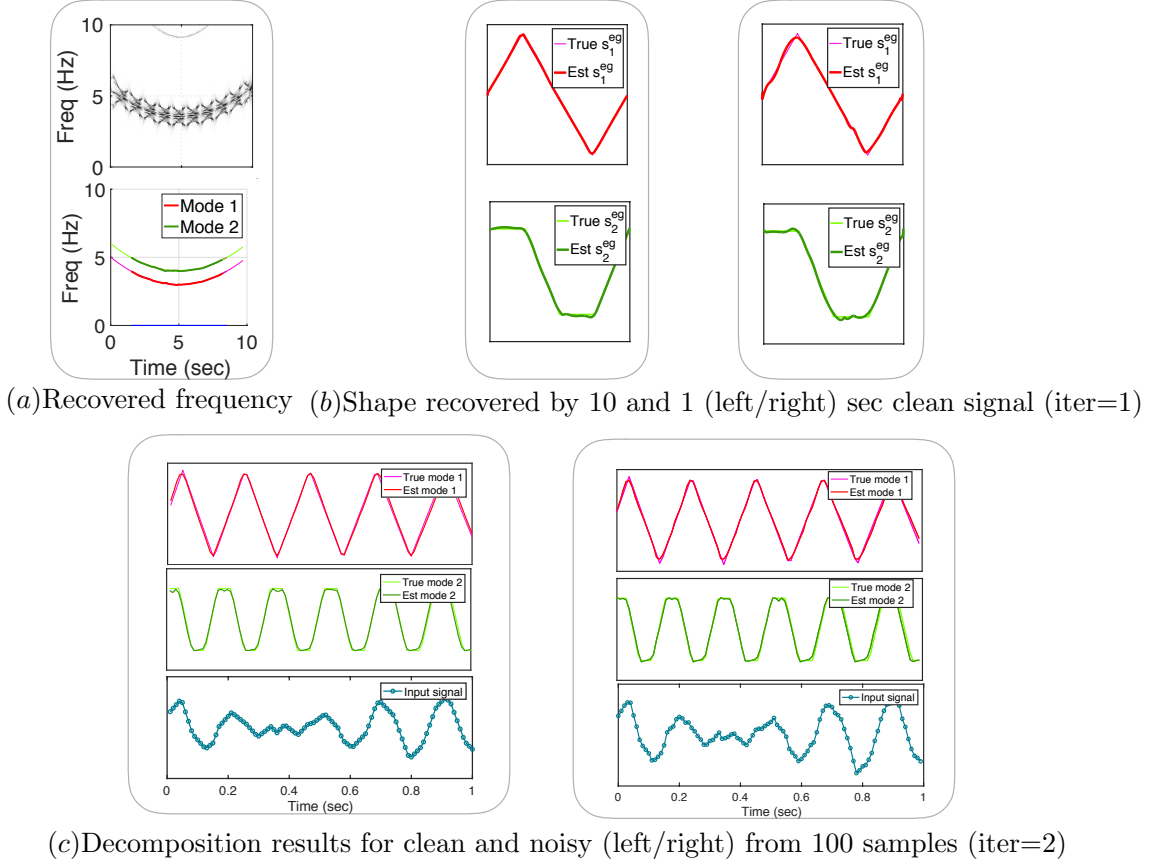


Figure 10: PiPs is applied to estimate the amplitude, phase, and shapes of a synthetic signal $f^{\{7\}}(t)$ consisting of two components. (a) the time-frequency distribution of $f^{\{7\}}(t)$ by ConceFT in two different frequency ranges. ConceFT cannot reveal the ground truth instantaneous frequencies (in red and green). But we can initialize PiPs by averaging out the distribution (see the dash pink line). (b) the ground truth shape functions and their estimates. In this case we apply the divide-and-conquer trick due to the non-linearity of frequency functions and take 10 signal chips whereas each for 1 second. We observed that when more chips are applied, the shape are more accurately estimated. (c) the noisy signal $f^{\{7\}}(t)$ and the reconstructed components by PiPs.

- [6] Eduardo Pinheiro, Octavian Postolache, and Pedro Girão. Empirical mode decomposition and principal component analysis implementation in processing non-invasive cardiovascular signals. *Measurement*, 45(2):175–181, 2012.
- [7] Haizhao Yang, Jianfeng Lu, and Lexing Ying. Crystal image analysis using 2D synchrosqueezed transforms. *Multiscale Modeling & Simulation*, 13(4):1542–1572, 2015.
- [8] Chao Zhang, Zhixiong Li, Chao Hu, Shuai Chen, Jianguo Wang, and Xiaogang Zhang. An optimized ensemble local mean decomposition method for fault detection of mechanical components. *Measurement Science and Technology*, 28(3):035102, 2017.
- [9] Bruno Cornelis, Haizhao Yang, Alex Goodfriend, Noelle Ocon, Jianfeng Lu, and Ingrid Daubechies. Removal of canvas patterns in digital acquisitions of paintings. *IEEE Transactions on Image Processing*, 26(1):160–171, 2017.
- [10] Jean B. Tary, Roberto H. Herrera, Jiajun Han, and Mirko van der Baan. Spectral estimation-What is new? What is next? *Rev. Geophys.*, 52(4):723–749, December 2014.
- [11] Wenjing Liao and Albert Fannjiang. Music for single-snapshot spectral estimation: Stability and super-resolution. *Applied and Computational Harmonic Analysis*, 40(1):33–67, 2016.

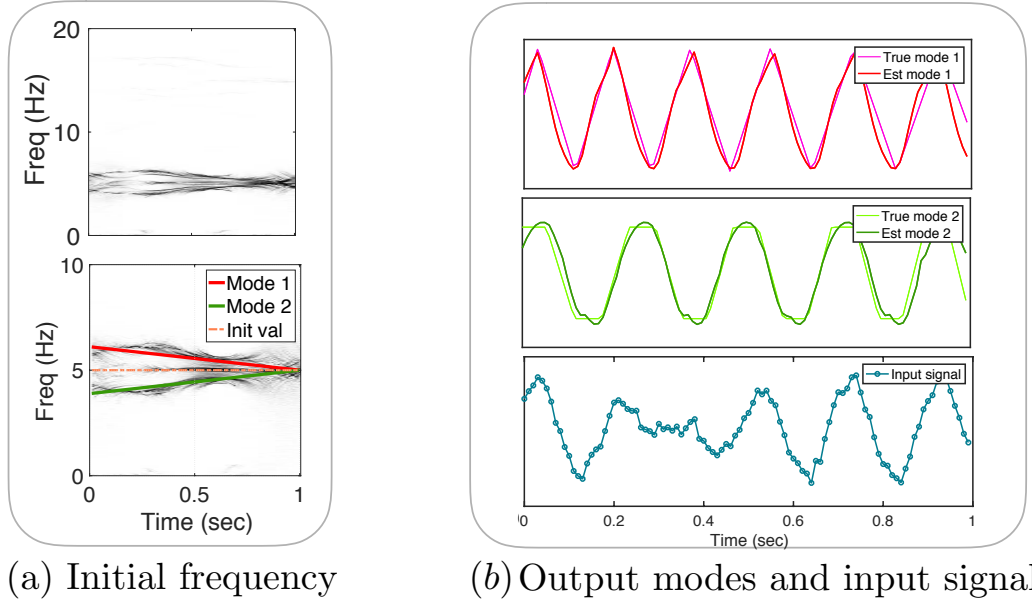


Figure 11: PiPs is applied to estimate the amplitude, phase, and shapes of a synthetic signal $f^{\{8\}}(t)$ consisting of two components. (a) the time-frequency distribution of $f^{\{8\}}(t)$ by ConceFT in two different frequency ranges. ConceFT cannot reveal the ground truth instantaneous frequencies (in red and green). But we can initialize PiPs by averaging out the distribution (see the dash pink line). (b) the noisy signal $f^{\{8\}}(t)$ and the reconstructed components by PiPs.

- [12] Emmanuel J Candès and Carlos Fernandez-Granda. Towards a mathematical theory of super-resolution. *Communications on Pure and Applied Mathematics*, 67(6):906–956, 2014.
- [13] Marvin HJ Gruber. Statistical digital signal processing and modeling, 1997.
- [14] John Parker Burg. The relationship between maximum entropy spectra and maximum likelihood spectra. *Geophysics*, 37(2):375–376, 1972.
- [15] Richard Roy and Thomas Kailath. Esprit-estimation of signal parameters via rotational invariance techniques. *IEEE Transactions on acoustics, speech, and signal processing*, 37(7):984–995, 1989.
- [16] François Auger and Patrick Flandrin. Improving the readability of time-frequency and time-scale representations by the reassignment method. *IEEE Transactions on signal processing*, 43(5):1068–1089, 1995.
- [17] I. Daubechies and S. Maes. A nonlinear squeezing of the continuous wavelet transform based on auditory nerve models. In *Wavelets in Medicine and Biology*, pages 527–546. CRC Press, 1996.
- [18] Haizhao Yang. Statistical analysis of synchrosqueezed transforms. *Applied and Computational Harmonic Analysis*, 45(3):526–550, 2018.
- [19] 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. *R. Soc. Lond. Proc. Ser. A Math. Phys. Eng. Sci.*, 454(1971):903–995, 1998.
- [20] Zhaohua Wu, Norden E. Huang, and Xianyao Chen. The multi-dimensional ensemble empirical mode decomposition method. *Adv. Adapt. Data Anal.*, 1(3):339–372, 2009.
- [21] Zhaohua Wu and Norden E Huang. Ensemble empirical mode decomposition: a noise-assisted data analysis method. *Advances in Adaptive Data Analysis*, 01(01):1, 2009.
- [22] Bin Zhu and David B Dunson. Locally adaptive bayes nonparametric regression via nested gaussian processes. *Journal of the American Statistical Association*, 108(504):1445–1456, 2013.

- [23] László Györfi, Michael Kohler, Adam Krzyzak, Harro Walk, et al. *A distribution-free theory of nonparametric regression*, volume 1. Springer, 2002.
- [24] Jieren Xu, Haizhao Yang, and Ingrid Daubechies. Recursive diffeomorphism-based regression for shape functions. *SIAM Journal on Mathematical Analysis*, 50(1):5–32, 2018.
- [25] Carl Edward Rasmussen and Christopher KI Williams. *Gaussian processes for machine learning*, volume 1. MIT press Cambridge, 2006.
- [26] Damien Garcia. Robust smoothing of gridded data in one and higher dimensions with missing values. *Computational statistics & data analysis*, 54(4):1167–1178, 2010.
- [27] Damien Garcia. A fast all-in-one method for automated post-processing of piv data. *Experiments in fluids*, 50(5):1247–1259, 2011.
- [28] Gao Tang and Haizhao Yang. A fast algorithm for multiresolution mode decomposition. *arXiv preprint arXiv:1712.09338*, 2017.
- [29] Haizhao Yang. Synchrosqueezed wave packet transforms and diffeomorphism based spectral analysis for 1d general mode decompositions. *Applied and Computational Harmonic Analysis*, 39(1):33–66, 2015.
- [30] Haizhao Yang. Multiresolution mode decomposition for adaptive time series analysis. *Applied and Computational Harmonic Analysis*, 52:25–62, 2021.
- [31] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- [32] Ingrid Daubechies, Jianfeng Lu, and Hau-Tieng Wu. Synchrosqueezed wavelet transforms: An empirical mode decomposition-like tool. *Applied and computational harmonic analysis*, 30(2):243–261, 2011.
- [33] Ian Goodfellow, Yoshua Bengio, Aaron Courville, and Yoshua Bengio. *Deep learning*, volume 1. MIT press Cambridge, 2016.
- [34] Ralph Schmidt. Multiple emitter location and signal parameter estimation. *IEEE transactions on antennas and propagation*, 34(3):276–280, 1986.
- [35] Tryphon T Georgiou. Spectral estimation via selective harmonic amplification. *IEEE Transactions on Automatic Control*, 46(1):29–42, 2001.
- [36] Tryphon T Georgiou. Spectral analysis based on the state covariance: the maximum entropy spectrum and linear fractional parametrization. *IEEE transactions on Automatic Control*, 47(11):1811–1823, 2002.
- [37] Ingrid Daubechies, Yi (Grace) Wang, and Hau-tieng Wu. Conceft: concentration of frequency and time via a multitapered synchrosqueezed transform. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 374(2065), 2016.
- [38] Daniel Griffin and Jae Lim. Signal estimation from modified short-time fourier transform. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 32(2):236–243, 1984.