

# Multiresolution Mode Decomposition for Adaptive Time Series Analysis

Haizhao Yang

Department of Mathematics, Purdue University, US

August 2019

## Abstract

This paper proposes the *multiresolution mode decomposition* (MMD) as a novel model for adaptive time series analysis. The main conceptual innovation is the introduction of the *multiresolution intrinsic mode function* (MIMF) of the form

$$\sum_{n=-N/2}^{N/2-1} a_n \cos(2\pi n\phi(t)) s_{cn}(2\pi N\phi(t)) + \sum_{n=-N/2}^{N/2-1} b_n \sin(2\pi n\phi(t)) s_{sn}(2\pi N\phi(t))$$

to model nonlinear and non-stationary data with time-dependent amplitudes, frequencies, and waveforms. The multiresolution expansion coefficients  $\{a_n\}$ ,  $\{b_n\}$ , and the shape function series  $\{s_{cn}(t)\}$  and  $\{s_{sn}(t)\}$  provide innovative features for adaptive time series analysis. For complex signals that are a superposition of several MIMFs with well-differentiated phase functions  $\phi(t)$ , a new recursive scheme based on Gauss-Seidel iteration and diffeomorphisms is proposed to identify these MIMFs, their multiresolution expansion coefficients, and shape function series. Numerical examples from synthetic data and natural phenomena are given to demonstrate the power of this new method.

**Keywords.** Multiresolution mode decomposition, multiresolution intrinsic mode function, recursive nonparametric regression, convergence.

**AMS subject classifications:** 42A99 and 65T99.

## 1 Introduction

Extracting useful information from large amounts of oscillatory data is important for a considerable number of real world applications such as medical electrocardiography (ECG) reading [1, 2, 3], atomic crystal images in physics [4, 5], mechanical engineering [6, 7], art investigation [8, 9], geology [10, 11, 12], imaging [13], etc. In order to extract certain features and analyze adaptive components of oscillatory data, it is typical to assume that the signal  $f(t)$  consists of several oscillatory modes like

$$f(t) = \sum_{k=1}^K \alpha_k(t) e^{2\pi i N_k \phi_k(t)} + r(t), \quad (1)$$

for  $t \in [0, 1]$ , where  $\alpha_k(t)$  is the instantaneous amplitude,  $N_k \phi_k(t)$  is the instantaneous phase,  $N_k \phi'_k(t)$  is the instantaneous frequency, and  $r(t)$  is the residual signal. Many methods

have been developed to decompose the signal  $f(t)$  into several modes  $\alpha_k(t)e^{2\pi i N_k \phi_k(t)}$  and to estimate the instantaneous information (the amplitude and phase functions) including the empirical mode decomposition (EMD) approach [14, 15], synchrosqueezed transforms [16, 17], time-frequency reassignment methods [18, 19], adaptive optimization [20, 21], iterative filters [22, 23], etc.

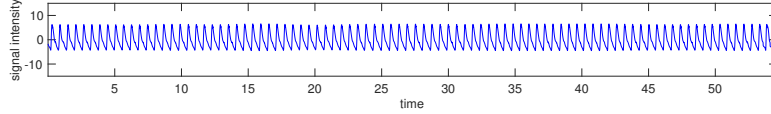
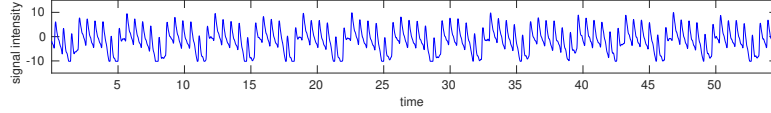
Although modeling oscillation in the form of model (1) is effective in many applications, sinusoidal oscillatory patterns may lose some important physical information when the data contain more complicated features. This observation was raised among EMD approaches and named as the intra-wave phenomenon (see, e.g., [24]). This phenomenon was also modeled with “wave-shape functions” in [25]. Let us use the second terminology with  $\{s_k(t)\}_{1 \leq k \leq K}$  as the shape functions to build a probably better mathematical model for oscillatory data analysis in a form of a superposition of generalized intrinsic mode functions (GIMFs):

$$f(t) = \sum_{k=1}^K \alpha_k(t) s_k(2\pi N_k \phi_k(t)) + r(t) = \sum_{k=1}^K \sum_{n=-\infty}^{\infty} \hat{s}_k(n) \alpha_k(t) e^{2\pi i n N_k \phi_k(t)} + r(t), \quad (2)$$

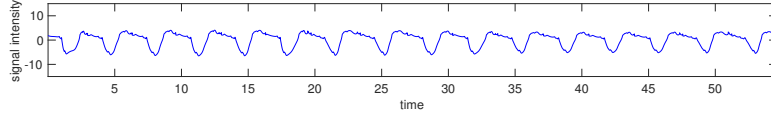
for  $t \in [0, 1]$ , where  $\{s_k(t)\}_{1 \leq k \leq K}$  are  $2\pi$ -periodic shape functions with a unit norm in  $L^2$  and  $\hat{s}(0) = 0$ . The mode decomposition problem with a model in (2) may have different names in the literature; but it is called the generalized mode decomposition (GMD) in this paper. The shape function can reflect complicated evolution patterns of the signal  $f(t)$ . The photoplethysmogram (PPG) signal (see Figure 1) in medical study is one of such complex examples. The PPG signal contains two essential evolution patterns corresponding to the cardiac and respiratory cycles. The shape of the PPG waveform differs from subject to subject and contains valuable information for monitoring the health condition of patients [26]. For more examples, the reader is referred to a detailed survey in [27]. The introduction of shape functions makes it more difficult to solve the decomposition problem and it has been an active research direction to seek its numerical solutions [28, 29, 30, 31].

In spite of considerable successes of analyzing oscillatory time series in the form of mode decomposition in (1) or GMD in (2), these models conflict with the physical intuition that the oscillation pattern of the time series changes in time. For example, the cardiac and respiratory patterns in Figure 1 vary in time. The GMD of the form (2) can only extract average evolution patterns (i.e., time-independent shape functions) to describe the cardiac and respiratory time series, leaving the evolution variance of these patterns (i.e., the deviation from the average evolution pattern) in the residual signal  $r(t)$  (see Panel 4 of Figure 1). However, the evolution variance is more important than the average evolution patterns for detecting diseases and measuring health risk. For example, the electrocardiogram (ECG) is an important tool to examine the functional status of the heart. The ECG waveform consists of three characteristic events (the P, QRS and T-wave as shown in Figure 4) associated with each beat. The detection of abnormal ECG waveforms is important to cardiac disease diagnosis [32, 33] and the abnormality is the deviation of an observed ECG waveform to a standard ECG waveform (e.g. tall R peaks caused by possible thickening of heart muscle wall, wide QRS and wide S waves due to partial or complete right bundle branch block).

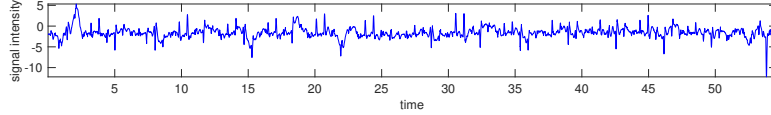
To analyze the fine features of evolution patterns discussed above, this paper proposes the *multiresolution mode decomposition* as a novel model for adaptive time series analysis. The main conceptual innovation is the introduction of the *multiresolution intrinsic mode*



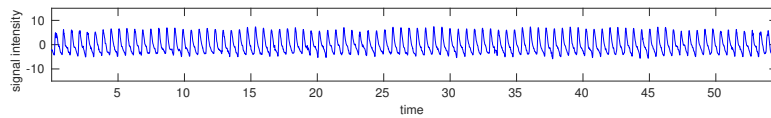
The respiratory mode  $f_2(t)$  by the generalized mode decomposition in (2).



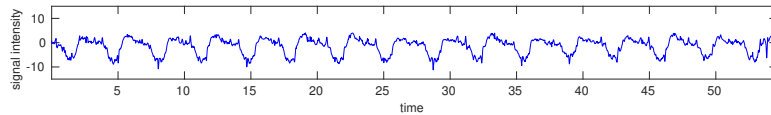
The residual error  $f(t) - f_1(t) - f_2(t)$  of the generalized mode decomposition in (2).



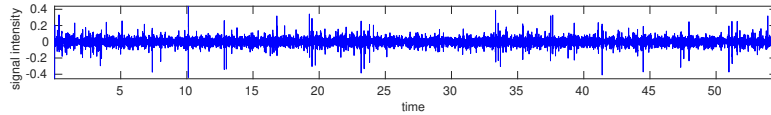
The cardiac mode  $f_1(t)$  by the multiresolution mode decomposition in (6).



The respiratory mode  $f_1(t)$  by the multiresolution mode decomposition in (6).



The residual error  $f(t) - f_1(t) - f_2(t)$  of the multiresolution mode decomposition in (6).



The residual error  $f(t) - f_1(t) - f_2(t)$  of the multiresolution mode decomposition in (6).

Figure 1: Comparison of the GMD (2) and MMD (6) for a photoplethysmogram (PPG) signal. The residual data of the GMD model still contain obvious oscillatory patterns with significant signal intensity, while the residual data of the MMD model in (6) is much weaker and close to i.i.d random noise (see Figure 2 below for a quantitative analysis).

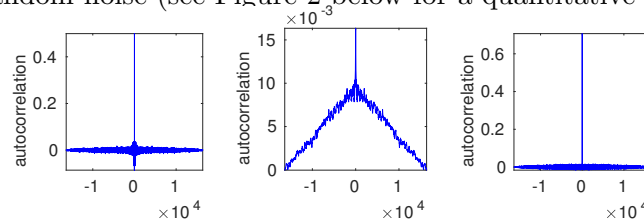


Figure 2: Comparison of the whiteness of the residual signal generated by GMD (2) and MMD (6) for the PPG signal in Figure 1. The autocorrelation of the residual signal by GMD, the residual signal by MMD, and a vector of Gaussian random noise is plotted in the left, middle, and right figures, respectively. Theoretically, the autocorrelation of white noise is an impulse at lag 0. Hence, the results here show that the residual signal by MMD is close to white noise, while the one by GMD still contains correlated oscillation.

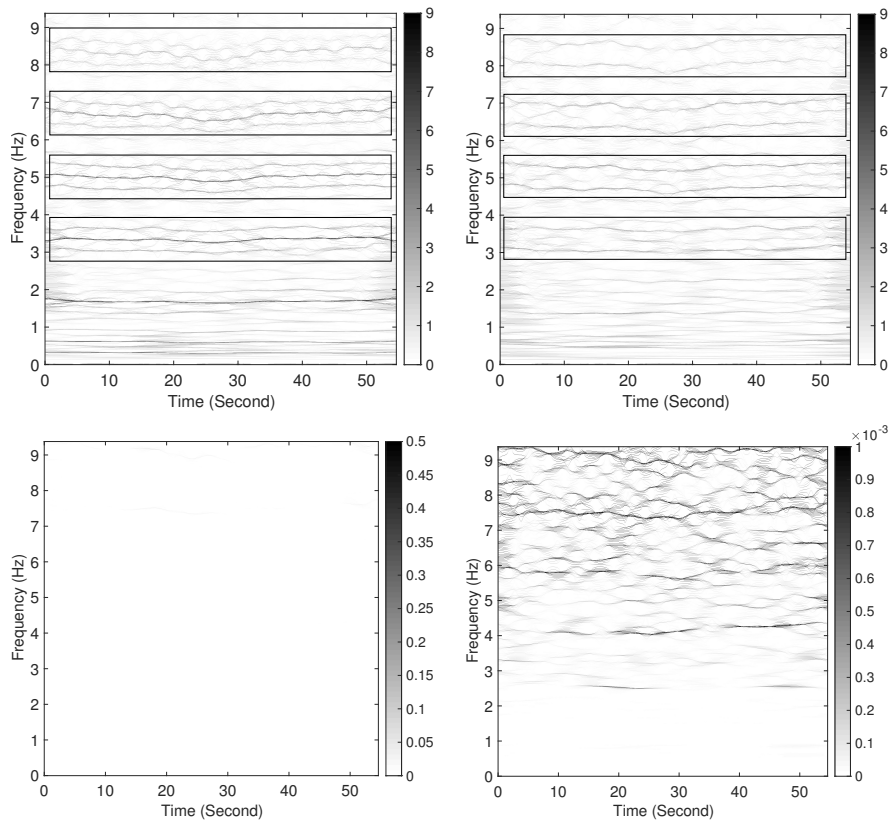


Figure 3: Top-left: the logarithm of the synchrosqueezed transform (SST) of the original PPG signal in Figure 1. Top-right: the logarithm of the SST of the residual signal by GMD (2), i.e., the 4-th signal in Figure 1. There are three major instantaneous frequencies in each rectangle on the left, while there are two in the rectangles on the right. Bottom figures: the logarithm of the SST of the residual signal by MMD (6) with different visualization scales, i.e., the last signal in Figure 1.

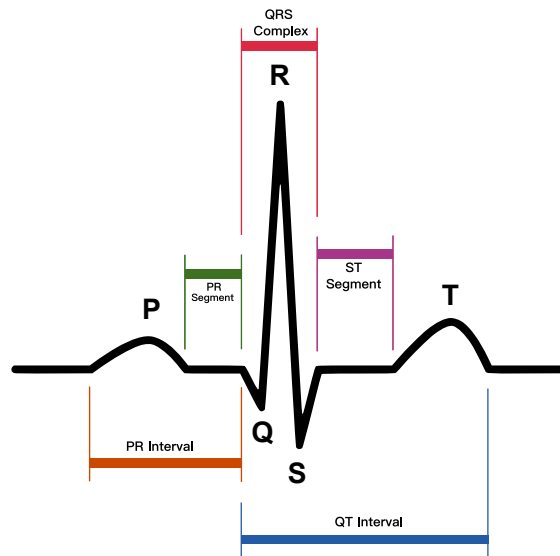


Figure 4: A schematic diagram of normal sinus rhythm for a human heart as seen on ECG [34].

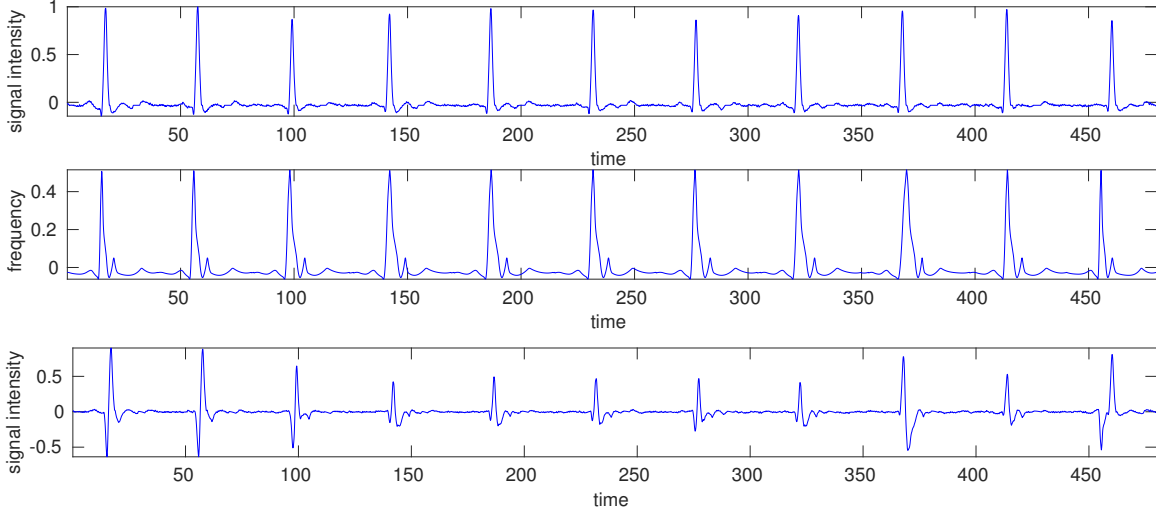


Figure 5: Top: a motion artifact contaminated electrocardiogram (ECG) signal  $f(t)$  modeled by Equation (3). Middle: the 0-band multiresolution approximation  $\mathcal{M}_0(f)(t) = a_0 s_{c0}(2\pi N\phi(t))$  of  $f(t)$ . Bottom:  $f(t) - \mathcal{M}_0(f)(t)$ , the variance of the evolution pattern of  $f(t)$ .

function (MIMF) of the form<sup>1</sup>

$$f(t) = \sum_{n=-N/2}^{N/2-1} a_n \cos(2\pi n\phi(t)) s_{cn}(2\pi N\phi(t)) + \sum_{n=-N/2}^{N/2-1} b_n \sin(2\pi n\phi(t)) s_{sn}(2\pi N\phi(t)) \quad (3)$$

for  $t \in [0, 1]$ , where  $\{a_n\}$  and  $\{b_n\}$  are real numbers,  $\{s_{cn}\}$  and  $\{s_{sn}\}$  are real value functions, to model nonlinear and non-stationary data with time-dependent amplitudes, frequencies, and waveforms. The MIMF is a generalization of the model of GIMF,  $\alpha(t)s(2\pi N\phi(t))$ , in Equation (2) for more accurate data analysis. When  $s_{cn}(t)$  and  $s_{sn}(t)$  in Equation (3) are equal to the same shape function  $s(t)$ , the model in Equation (3) is reduced to  $\alpha(t)s(2\pi N\phi(t))$  once the amplitude function  $\alpha(t)$  is written in the form of its Fourier series expansion. When  $s_{cn}(t)$  and  $s_{sn}(t)$  are different shape functions, the two summations in Equation (3) lead to time-dependent shape functions to describe the nonlinear and non-stationary time series adaptively. A recent paper [27] also tried to address the limitation of GMD in (2) by replacing  $\hat{s}_k(n)\alpha_k(t)$  with a time-varying function, denoted as  $B_{k,n}(t)$ , i.e., introducing more variance to amplitude functions. Our model in (3) emphasizes both the time variance of amplitude and shape functions by introducing multiresolution expansion coefficients and shape function series. As far as we understand, instead of estimating time-varying amplitude and shape functions, [27] proposed an algorithm to eliminate the influence of amplitude and shape functions and estimate phase functions  $N_k\phi_k(t)$ . This algorithm could be a useful tool complimentary to the algorithm proposed in this paper, since we assume phase functions are known and estimate time-varying amplitudes and shapes.

<sup>1</sup>The analytic analog is  $f(t) = \sum_{n=-N/2}^{N/2-1} a_n e^{2\pi i n\phi(t)} s_n(2\pi N\phi(t))$ , where  $\{a_n\}$  are complex numbers and  $\{s_n\}$  are analytic shape functions. Without loss of generality,  $N$  is assumed to be even throughout this paper.

Let  $\mathcal{M}_\ell$  be the operator for computing the  $\ell$ -banded *multiresolution approximation* to a MIMF  $f(t)$  in Equation (3), i.e.,

$$\mathcal{M}_\ell(f)(t) = \sum_{n=-\ell}^{\ell} a_n \cos(2\pi n\phi(t))s_{cn}(2\pi N\phi(t)) + \sum_{n=-\ell}^{\ell} b_n \sin(2\pi n\phi(t))s_{sn}(2\pi N\phi(t)), \quad (4)$$

and  $\mathcal{R}_\ell$  be the operator for the computing the residual sum

$$\mathcal{R}_\ell(f)(t) = f(t) - \mathcal{M}_\ell(f)(t). \quad (5)$$

Then the 0-banded multiresolution approximation  $\mathcal{M}_0(f)(t) = a_0 s_{c0}(2\pi N\phi(t))$  describes the average evolution pattern of the signal, while the rest describe the evolution variance. As shown in Figure 5, if  $f(t)$  is an ECG signal<sup>2</sup>, then it is more obvious to observe the change of the evolution pattern from  $\mathcal{R}_0(f)(t)$  than from  $f(t)$ , e.g., the change of the height of R peaks and the width of QRS and S waves.

In many applications [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13], a signal would be a superposition of several MIMFs, for example, a complex signal

$$f(t) = \sum_{k=1}^K f_k(t), \quad (6)$$

where

$$f_k(t) = \sum_{n=-N/2}^{N/2-1} a_{n,k} \cos(2\pi n\phi_k(t))s_{cn,k}(2\pi N_k\phi_k(t)) + \sum_{n=-N/2}^{N/2-1} b_{n,k} \sin(2\pi n\phi_k(t))s_{sn,k}(2\pi N_k\phi_k(t)).$$

The multiresolution mode decomposition (MMD) problem aims at extracting each MIMF  $f_k(t)$ , estimating its corresponding multiresolution expansion coefficients  $\{a_{n,k}\}$ ,  $\{b_{n,k}\}$ , and the shape function series  $\{s_{cn,k}(t)\}$  and  $\{s_{sn,k}(t)\}$ . As we can see in Figure 1, if MMD is applied to analyze the PPG signal, we can obtain the cardiac and respiratory patterns (Panel 5 and 6 in Figure 1, respectively) with more accurate evolution variance than the model of GMD in (2) in two aspects. In terms of statistical testing, the residual signal of MMD in Panel 7 of Figure 1 contains information close to i.i.d random noise while the residual by (2) still contains correlated oscillation patterns as demonstrated by Figure 2 quantitatively. From the point of view of time-frequency analysis visualized in Figure 3, the synchrosqueezed transform (SST) of the residual by (2) in the top-right panel shares almost the same spectrogram with the SST of the original PPG signal in the top-left panel. Note that the model in (2) seems to capture only partial cardiac and respiratory patterns in the PPG signal and the residual signal still contains significant information resembling the cardiac and respiratory patterns. In fact, there are three major instantaneous frequencies with almost the same geometry in each rectangle on the top-left of Figure 3, while there are two in the rectangles on the top-right, indicating that the model in (2) misses some instantaneous frequencies with similar oscillation patterns as those considered in (2). As a comparison, the SST of the residual by MMD has no obvious spectrogram even if in a much smaller visualization scale (0, 0.5) as shown in the bottom-left panel of Figure 3. In a very small visualization scale like (0, 0.001) in the bottom-right panel, we see that the SST of the residual by MMD indicates no meaningful oscillation pattern. The oscillation

<sup>2</sup>From the PhysiNet <https://physionet.org/>.

patterns missed by (2) and visualized on the top-right panel of Figure 3 have been considered in the MMD model using the summation over different  $n$ 's in (3). The geometry of the instantaneous frequencies of different terms with different  $n$ 's is similar to that of the term when  $n = 0$ .

Although there have been a few algorithms for the GMD in (2) [28, 30, 31], these methods are incapable of either identifying the shape function series  $\{s_{cn,k}(t)\}$  and  $\{s_{sn,k}(t)\}$ , or the multiresolution expansion coefficients  $\{a_{n,k}\}$  and  $\{b_{n,k}\}$  in the multiresolution mode decomposition. Motivated by the recursive diffeomorphism-based regression (RDBR) in [31], this paper proposes a Gauss-Seidel style recursive scheme to solve the multiresolution mode decomposition problem. As we shall see later, the novel recursive scheme has a faster convergence rate than the existing scheme in [31], and more importantly, it is robust to the estimated number of components (i.e., it still returns a meaningful mode decomposition even if the input number of modes is wrong).

To make the presentation of the Gauss-Seidel RDBR for the multiresolution mode decomposition more accessible, we will first introduce the Gauss-Seidel RDBR for the GMD in Section 2. The Gauss-Seidel RDBR for the multiresolution mode decomposition will be introduced in Section 3. In Section 4, we present some numerical examples to demonstrate the efficiency of the proposed RDBR. Finally, we conclude this paper in Section 5.

## 2 Gauss-Seidel recursive scheme for the GMD

In what follows, we introduce the new Gauss-Seidel recursive diffeomorphism-based regression (RDBR) for the GMD. Existing methods [28, 30, 31] for the GMD problem

$$f(t) = \sum_{k=1}^K \alpha_k(t) s_k(2\pi N_k \phi_k(t)) = \sum_{k=1}^K \sum_{n=-\infty}^{\infty} \hat{s}_k(n) \alpha_k(t) e^{2\pi i n N_k \phi_k(t)} \quad (7)$$

generally assume that the instantaneous properties (such as  $\alpha_k(t)$  and  $N_k \phi_k(t)$  in Equation (7)) are available and focus on the estimation of shape functions  $s_k(t)$ . This assumption is based on the observation that: after filtering the signal with a low-pass filter, the GMD problem in Equation (7) becomes the standard mode decomposition problem in (1); afterwards, instantaneous properties can be estimated by well-developed algorithms for the mode decomposition problem [14, 15, 16, 18, 19, 20, 21, 22, 23], especially the deShape SST [27, ?]. Hence, we assume that the instantaneous amplitudes and phases are known in this section.

### 2.1 Algorithm description

As in the Jacobi style RDBR algorithm [31], we assume  $L$  points of measurement  $\{f(t_\ell)\}_{\ell=1,\dots,L}$  with independent and identically distributed (i.i.d.) grid points  $\{t_\ell\}_{\ell=1,\dots,L}$  from a uniform distribution in  $[0, 1]$ . Usually, the grid is deterministic and uniform in  $[0, 1]$ , but an i.i.d. grid enables a better estimation of the shape function with a smaller  $L$ , since it allows the access of the  $2\pi$  periodic shape function  $s(t)$  for every  $t \in [0, 2\pi]$  with a certain probability.

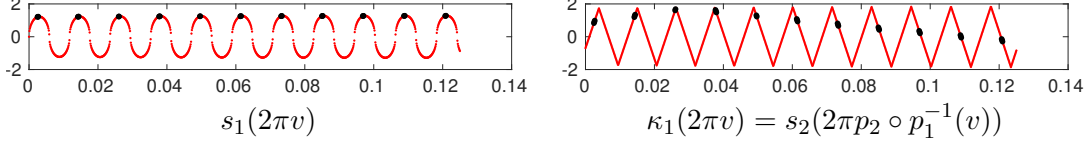


Figure 6: An illustration of the inverse-warping idea in Equation (8) when  $f(t) = s_1(2\pi N_1\phi_1(t)) + s_2(2\pi N_2\phi_2(t))$ . Let  $p_j = N_j\phi_j(t)$  for all  $j$  and inverse-warp  $f(t)$  with  $v = p_1(t)$  into  $h_1(v)$  as in Equation (8), then  $h_1(v)$  contains two main parts: a periodic function,  $s_1(2\pi v)$  with samples  $\{s_1(2\pi v_\ell)\}_{\ell=1,\dots,L}$  (see the left figure), and a non-periodic function,  $\kappa_1(2\pi v)$  with samples  $\{\kappa_1(2\pi v_\ell)\}_{\ell=1,\dots,L}$  (see the right figure).

In the RDBR, we define the inverse-warping data by

$$\begin{aligned}
h_k(v) &= \frac{f \circ p_k^{-1}(v)}{\alpha_k \circ p_k^{-1}(v)} \\
&= s_k(2\pi v) + \sum_{j \neq k} \frac{\alpha_j \circ p_k^{-1}(v)}{\alpha_k \circ p_k^{-1}(v)} s_j(2\pi p_j \circ p_k^{-1}(v)) \\
&:= s_k(2\pi v) + \kappa_k(2\pi v),
\end{aligned} \tag{8}$$

where  $v := p_k(t)$ ,  $p_j(t) = N_j\phi_j(t)$  for all  $j$ , and

$$\kappa_k(2\pi v) := \sum_{j \neq k} \frac{\alpha_j \circ p_k^{-1}(v)}{\alpha_k \circ p_k^{-1}(v)} s_j(2\pi p_j \circ p_k^{-1}(v)).$$

As a consequence, we have a set of measurements of  $h_k(v)$ ,  $\{h_k(v_\ell)\}_{\ell=1,\dots,L}$ , sampled in  $v$  with  $v_\ell = p_k(t_\ell)$  (see Figure 6 for an illustration).

If there was a single mode in  $f(t)$  (e.g.  $f(t) = \alpha_k(t)s_k(2\pi N_k\phi_k(t))$  with  $k = 1$ ), then  $h_k(v)$  is equal to a periodic function  $s_k(2\pi v)$  with period 1. Hence, if we define a folding map  $\tau$  that folds the two-dimensional point set  $\{(v_\ell, h_k(v_\ell))\}_{\ell=1,\dots,L}$  together

$$\tau : (v_\ell, h_k(v_\ell)) \mapsto (\text{mod}(v_\ell, 1), h_k(v_\ell)), \tag{9}$$

then the point set  $\{\tau(v_\ell, s_k(2\pi v_\ell))\}_{\ell=1,\dots,L} \subset \mathbb{R}^2$  is a two-dimensional point set located at the curve  $(v, s_k(2\pi v)) \subset \mathbb{R}^2$  given by the shape function  $s_k(2\pi v)$  with  $v \in [0, 1)$  (see Figure 7 (left) for an example). Using the notations in non-parametric regression, let  $X_k$  be an independent random variable in  $[0, 1)$ ,  $Y_k$  be the response random variable in  $\mathbb{R}$ , and consider  $(x_\ell^{(k)}, y_\ell^{(k)}) = \tau(v_\ell, s_k(2\pi v_\ell))$  as  $L$  i.i.d. samples of the random vector  $(X_k, Y_k)$ , then a simple regression results in the exact shape function as follows. Define

$$s_k^R := \arg \min_{s: \mathbb{R} \rightarrow \mathbb{R}} \mathbb{E}\{|s(2\pi X_k) - Y_k|^2\}, \tag{10}$$

where the superscript  $R$  means the ground truth regression function, then  $s_k = s_k^R$ . If we denote the numerical solution of the above regression problem as  $s_k^P$ , then  $s_k^P \approx s_k^R = s_k$  when  $L$  is sufficiently large.

However, in the case of multiple modes,  $\kappa_k(2\pi v) \neq 0$  and  $(x_\ell^{(k)}, y_\ell^{(k)}) = \tau(v_\ell, s_k(2\pi v_\ell) + \kappa_k(2\pi v_\ell))$  for  $\ell = 1, \dots, L$  can be considered as noise-contaminated i.i.d. samples of a random vector  $(X_k, Y_k)$  (see Figure 7 (right) for an illustration). Hence, the regression in



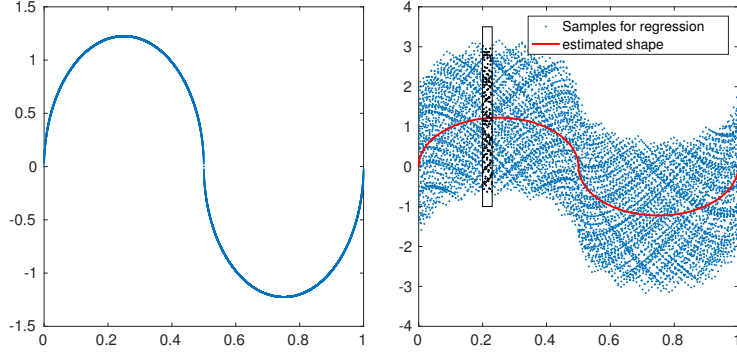


Figure 7: An illustration of the folding idea in Equation (9). Left: the point set  $\{\tau(v_\ell, h_1(v_\ell))\}_{\ell=1, \dots, L} \subset \mathbb{R}^2$  in the case of one mode  $f(t) = s_1(2\pi N_1 \phi_1(t))$ , and the point distribution characterizes the shape function  $s_1(2\pi v)$  exactly. Right: the point set  $\{\tau(v_\ell, h_1(v_\ell))\}_{\ell=1, \dots, L} \subset \mathbb{R}^2$  in the case of two modes  $f(t) = \sum_{k=1}^2 s_k(2\pi N_k \phi_k(t))$ , and the point distribution behaves like noisy observation of  $s_1(2\pi v)$  with an additive noise  $\kappa_1(2\pi v)$ . The partition-based regression computes the regression function (in red) at a point  $v^{bk}$  as the average height of all observations (black points) with sampling locations in a small neighborhood of  $v^{bk}$ .

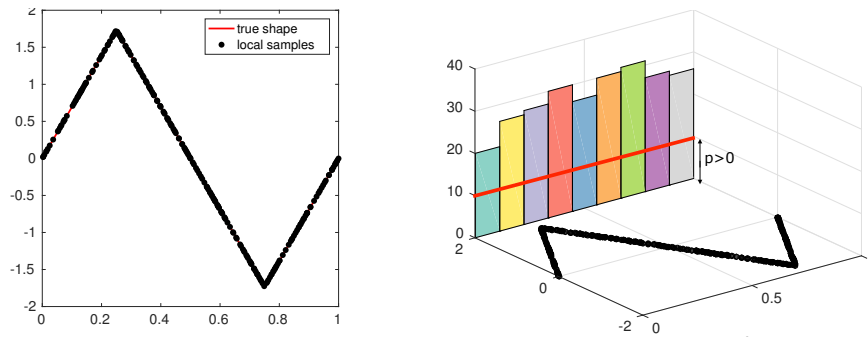


Figure 8: Left: the point set  $\{\text{mod}(z_\ell^{bk}, 1), \kappa_1(2\pi p_1 \circ p_2^{-1}(\text{mod}(z_\ell^{bk}, 1)))\}_\ell = \{\text{mod}(z_\ell^{bk}, 1), s_2(2\pi \text{mod}(z_\ell^{bk}, 1))\}_\ell \subset \mathbb{R}^2$ . Right: a histogram of the points  $\{\text{mod}(z_\ell^{bk}, 1)\}$  in the left figure.

(10) would not estimate the ground true shape function precisely. Let us use the partition-based regression method (or partitioning estimate) in Chapter 4 of [35] as an example. Given a small step side  $h \ll 1$ , the time domain  $[0, 1]$  is uniformly partitioned into  $N^h = \frac{1}{h}$  (assumed to be an integer) parts  $\{[t_n^h, t_{n+1}^h)\}_{n=0, \dots, N^h-1}$ , where  $t_n^h = nh$ . Let  $s_k^P$  denote the estimated regression function by the partition-based regression method with  $L$  samples,  $(x_\ell^{(k)}, y_\ell^{(k)})_{\ell=1, \dots, L}$ . Following Chapter 4 of [35], we define

$$s_k^P(x) := \frac{\sum_{\ell=1}^L \mathcal{X}_{[t_n^h, t_{n+1}^h)}(x_\ell^{(k)}) y_\ell^{(k)}}{\sum_{\ell=1}^L \mathcal{X}_{[t_n^h, t_{n+1}^h)}(x_\ell^{(k)})} \quad (11)$$

as the partition-based regression function, when  $x \in [t_n^h, t_{n+1}^h)$ , where  $\mathcal{X}_{[t_n^h, t_{n+1}^h)}(x)$  is the indicator function supported on  $[t_n^h, t_{n+1}^h)$ . As illustrated by black points in Figure 7 (right), the partition-based regression essentially computes  $s_k^P(x)$  as the average height of all observations  $y_\ell^{(k)}$  with sampling locations  $x_\ell^{(k)}$  in a small neighborhood of  $x$ . Let  $s_k^E := s_k^P - s_k$  be the residual shape function, then the residual error of the GMD after one step of regressions on all modes is

$$r(t) = f(t) - \sum_{k=1}^K \alpha_k(t) s_k^P(2\pi N_k \phi_k(t)) \approx - \sum_{k=1}^K \alpha_k(t) s_k^E(2\pi N_k \phi_k(t)), \quad (12)$$

which might be large, since in general  $s_k \approx s_k^P$  is not true.

The deviation of  $s_k^P$  from  $s_k$  comes from the influence of  $\kappa_k$  in (8). After the folding map in (9), we hope that  $\{\kappa_k(2\pi \bmod(v_\ell, 1))\}_{1 \leq \ell \leq L}$  behave like i.i.d. samples of a mean-zero random noise so that  $s_k^P \approx s_k$ . To understand the behavior of  $\kappa_k(2\pi \bmod(v, 1))$ , let us take the example of  $\kappa_1(2\pi \bmod(v, 1))$  in Figure 7 (right). If we unfold the black samples  $\{h_1(\bmod(v_\ell^{bk}, 1)) = s_1(2\pi \bmod(v_\ell^{bk}, 1)) + \kappa_1(2\pi \bmod(v_\ell^{bk}, 1))\}^3$  (samples with sampling locations  $\bmod(v_\ell^{bk}, 1)$  near the point  $v^{bk}$  in Figure 7 (right)) back to  $h_1(v_\ell^{bk})$ , we see that the black samples  $\kappa_1(2\pi \bmod(v_\ell^{bk}, 1))$  come from the black samples  $\kappa_1(2\pi v_\ell^{bk})$  in Figure 6 (right). If we warp  $\kappa_1(2\pi v_\ell^{bk})$  back to  $\kappa_1(2\pi p_1(t_\ell^{bk}))$ , inverse-warp  $\kappa_1(2\pi p_1(t_\ell^{bk}))$  to  $\kappa_1(2\pi p_1 \circ p_2^{-1}(z_\ell^{bk}))$ , where  $z := p_2(t)$ , and finally fold  $\kappa_1(2\pi p_1 \circ p_2^{-1}(z_\ell^{bk}))$  into one period to obtain

$$\kappa_1(2\pi p_1 \circ p_2^{-1}(\bmod(z_\ell^{bk}, 1))), \quad (13)$$

then from Figure 8 (left) we see that the black samples  $\kappa_1(2\pi p_1 \circ p_2^{-1}(\bmod(z_\ell^{bk}, 1)))$  essentially cover the shape function  $s_2$ . In fact, by the definition of  $\kappa_2$ , we have

$$\kappa_1(2\pi p_1 \circ p_2^{-1}(\bmod(z_\ell^{bk}, 1))) = s_2(2\pi \bmod(z_\ell^{bk}, 1)).$$

Figure 8 (right) shows a histogram of the sampling locations  $\{\bmod(z_\ell^{bk}, 1)\}$ . Note that  $s_2$  has mean zero. Hence, if the point distribution  $\{\bmod(z_\ell^{bk}, 1)\}$  is almost uniform in  $[0, 1]$ ,

<sup>3</sup>Here  $^{bk}$  means samples corresponding to black points in Figure 6 to 8.

then by the partition-based regression formula, we see

$$\begin{aligned}
s_1^P(2\pi v^{bk}) &= s_1(2\pi v^{bk}) + \frac{\sum_{\ell} \kappa_1(2\pi \bmod(v_{\ell}^{bk}, 1))}{\text{the number of samples of } \{ \bmod(v_{\ell}^{bk}, 1) \}} \\
&= s_1(2\pi v^{bk}) + \frac{\sum_{\ell} \kappa_1(2\pi v_{\ell}^{bk})}{\text{the number of samples of } \{v_{\ell}^{bk}\}} \\
&= s_1(2\pi v^{bk}) + \frac{\sum_{\ell} \kappa_1(2\pi p_1(t_{\ell}^{bk}))}{\text{the number of samples of } \{t_{\ell}^{bk}\}} \\
&= s_1(2\pi v^{bk}) + \frac{\sum_{\ell} \kappa_1(2\pi p_1 \circ p_2^{-1}(z_{\ell}^{bk}))}{\text{the number of samples of } \{z_{\ell}^{bk}\}} \\
&= s_1(2\pi v^{bk}) + \frac{\sum_{\ell} \kappa_1(2\pi p_1 \circ p_2^{-1}(\bmod(z_{\ell}^{bk}, 1)))}{\text{the number of samples of } \{ \bmod(z_{\ell}^{bk}, 1) \}} \\
&= s_1(2\pi v^{bk}) + \frac{\sum_{\ell} s_2(2\pi \bmod(z_{\ell}^{bk}, 1))}{\text{the number of samples of } \{ \bmod(z_{\ell}^{bk}, 1) \}} \\
&\approx s_1(2\pi v^{bk}) + \int_0^1 s_2(2\pi z) dz \\
&= s_1(2\pi v^{bk}).
\end{aligned} \tag{14}$$

In the case when the point distribution  $\{ \bmod(z_{\ell}^{bk}, 1) \}$  is far from being uniform in  $[0, 1]$ , we see that  $s_1^P$  is not close to  $s_1$ .

Note that the residual  $r(t)$  in (12) can be viewed as a new superposition of modes with new shape functions  $\{-s_k^E\}$ . This motivates the Jacobi recursive scheme in [31] that repeats the same decomposition procedure to decompose the residual  $r(t)$ , and update the shape function estimation, until the residual is eliminated. It was proved in [31] that if the empirical distribution of  $\{ \bmod(z_{\ell}^{bk}, 1) \}$  corresponding to  $v^{bk}$  for any  $v^{bk}$  is uniformly and strictly positive (e.g., larger than  $p > 0$  as in Figure 8 (right)), then the recursive scheme is able to eliminate the residual  $r(t)$ .

To improve the convergence rate of the Jacobi style recursive scheme, this paper proposes a Gauss-Seidel recursive scheme in Algorithm 1. More importantly, the convergence of the Jacobi recursive scheme is sensitive to the prior information as inputs. For example, if the number of modes  $K$  is not known exactly, the Jacobi recursive scheme may fail to converge, while the Gauss-Seidel recursive scheme would not. In practice, the input prior information is:

1. an estimated number of components  $\bar{K}$ ;
2. a set of estimated phase functions

$$p_k(t) = n_k N_{\tau_k} \phi_{\tau_k}(t) - n_k N_{\tau_k} \phi_{\tau_k}(0), \tag{15}$$

for  $k = 1, \dots, \bar{K}$ , where  $\tau_k \in \{1, \dots, K\}$  and  $n_k \in \mathbb{Z}^+$  are unknown integers;

3. and the corresponding amplitude functions

$$q_k(t) = a_k \alpha_{\tau_k}(t), \tag{16}$$

for  $k = 1, \dots, \bar{K}$ , where  $a_k > 0$  is unknown.

Note that the curves  $\{p'_k(t) = n_k N_{\tau_k} \phi'_{\tau_k}(t)\}$  naturally belong to a few groups, each of which corresponds to the multiple of a fundamental instantaneous frequency  $N_k \phi'_k(t)$  for some  $k$ . Following the curve classification idea in Algorithm 3.7 and Theorem 3.9 in [28], we are able to classify the curves  $\{p'_k(t) = n_k N_{\tau_k} \phi'_{\tau_k}(t)\}$  and identify the corresponding fundamental instantaneous frequencies  $\{N_{\tau_k} \phi'_{\tau_k}(t)\}$ , which give the fundamental instantaneous phases up to an unknown initial phase  $\{N_{\tau_k} \phi_{\tau_k}(t) - N_{\tau_k} \phi_{\tau_k}(0)\}$ . Therefore, we can assume that  $\{\tau_k\}_{1 \leq k \leq \bar{K}}$  are distinct and  $n_k = 1$  for all  $k$  in Equation (15) and (16) for clean data (i.e. the prior information only contains the fundamental instantaneous frequencies without their multiples, but the prior information may still miss some instantaneous frequencies). The reader is referred to [28] for more detail. However, in the presence of noise, the instantaneous frequency estimations may have large errors leading the failure of the curve classification. In this case, the prior information may contain instantaneous frequencies that are multiples of the fundamental instantaneous frequencies (i.e., there may be some  $n_k \neq 1$ ). Even if the prior information above misses some fundamental frequencies or contains the multiples of fundamental frequencies, the Gauss-Seidel recursive scheme below

can still recover reasonably accurate modes from their superposition.

1	Input: $L$ points of i.i.d. measurement $\{f(t_\ell)\}_{\ell=1,\dots,L}$ with $t_\ell \in [0, 1]$ , estimated instantaneous phases $\{p_k\}_{k=1,\dots,\bar{K}}$ , amplitudes $\{q_k\}_{k=1,\dots,\bar{K}}$ , an accuracy parameter $\epsilon < 1$ , and the maximum iteration number $J$ .
2	Output: the estimated shape functions $\{\bar{s}_k\}_{k=1,\dots,\bar{K}}$ and the estimated modes $\{\bar{f}_k(t)\}_{k=1,\dots,\bar{K}} = \{q_k(t)\bar{s}_k(2\pi p_k(t))\}_{k=1,\dots,\bar{K}}$ at the sampling grid points $\{t_\ell\}_{1 \leq \ell \leq L}$ .
3	Initialize: let $r_1^{(0)} = f$ , $\epsilon_1 = \epsilon_2 = 1$ , $\epsilon_0 = 2$ , the iteration number $j = 0$ , $\dot{s}_k^{(0)} = 0$ , and $\bar{s}_k^{(0)} = 0$ for all $k = 1, \dots, \bar{K}$ .
4	Compute $N_k$ as the integer nearest to the average of $p'_k(t)$ for $k = 1, \dots, \bar{K}$ .
5	Sort $\{N_k\}_{1 \leq k \leq \bar{K}}$ in an ascending order and reorder the amplitude and phase functions accordingly.
6	<b>while</b> $j < J$ , $\epsilon_1 > \epsilon$ , $\epsilon_2 > \epsilon$ , and $ \epsilon_1 - \epsilon_0  > \epsilon$ <b>do</b>
7	<b>for</b> $k = 1, \dots, \bar{K}$ <b>do</b>
8	Define
	$h_k^{(j)} = \frac{r_k^{(j)} \circ p_k^{-1}}{q_k \circ p_k^{-1}},$
	and we know it is sampled on grid points $v_\ell = p_k(t_\ell)$ .
9	Observe that $\{\tau(v_\ell, h_k^{(j)}(v_\ell))\}_{\ell=1,\dots,L}$ behaves like a sequence of i.i.d. samples
	of a certain random vector $(X_k, Y_k^{(j)})$ with $X_k \in [0, 1]$ .
10	Solve the distribution-free regression problem
	$\dot{s}_k^{(j+1)} \approx s_k^{R,(j+1)} = \arg \min_{s: \mathbb{R} \rightarrow \mathbb{R}} \mathbb{E}\left\{\left s(2\pi X_k) - Y_k^{(j)}\right ^2\right\}, \quad (17)$
	where $\dot{s}_k^{(j+1)}$ denotes the numerical solution approximating the ground truth solution $s_k^{R,(j+1)}$ .
11	Update $\dot{s}_k^{(j+1)} = \dot{s}_k^{(j+1)} - \frac{1}{2\pi} \int_0^{2\pi} \dot{s}_k^{(j+1)}(t) dt$ for all $k$ .
12	Let $\bar{s}_k^{(j+1)} = \bar{s}_k^{(j)} + \dot{s}_k^{(j+1)}$ for all $k$ .
13	If $k < \bar{K}$ , then let $r_{k+1}^{(j)} = r_k^{(j)} - q_k(t)\dot{s}_k^{(j+1)}(2\pi p_k(t))$ ; otherwise, let
	$r_1^{(j+1)} = r_k^{(j)} - q_k(t)\dot{s}_k^{(j+1)}(2\pi p_k(t))$ .
14	Update $\epsilon_0 = \epsilon_1$ , $\epsilon_1 = \ r_1^{(j+1)}\ _{L^2}$ , $\epsilon_2 = \max_k \{\ \dot{s}_k^{(j+1)}\ _{L^2}\}$ .
15	Set $j = j + 1$ .
16	Let $\bar{s}_k = \bar{s}_k^{(j+1)}$ for all $k$ .

**Algorithm 1:** Gauss-Seidel recursive diffeomorphism-based regression (RDBR).

Remark: as pointed out in the Jacobi style RDBR [31], the unknown shifts  $\{n_k N_{\tau_k} \phi_{\tau_k}(0)\}$  in Equation (15) and the unknown prefactors  $\{a_k\}$  in Equation (16) have been absorbed in the estimation of shape functions and the reconstruction of modes. Hence, it is not necessary to know them a priori. In the convergence analysis, instead of Equation (15) and (16), we assume that

$$p_k(t) = n_k N_{\tau_k} \phi_{\tau_k}(t), \quad (18)$$

and

$$q_k(t) = \alpha_{\tau_k}(t). \quad (19)$$

## 2.2 Convergence analysis

In this section, an asymptotic analysis on the convergence of the Gauss-Seidel recursive diffeomorphism-based regression (RDBR) in Algorithm 1 is provided for a class of time series that is a superposition of several generalized intrinsic mode functions as follows. To make the analysis self-contained, a few definitions in [31] will be repeated.

**Definition 2.1.** *Generalized shape functions:* The generalized shape function class  $\mathcal{S}_M$  consists of  $2\pi$ -periodic functions  $s(t)$  in the Wiener Algebra with a unit  $L^2([0, 2\pi])$ -norm and a  $L^\infty$ -norm bounded by  $M$  satisfying the following spectral conditions:

1. The Fourier series of  $s(t)$  is uniformly convergent;
2.  $\sum_{n=-\infty}^{\infty} |\widehat{s}(n)| \leq M$  and  $\widehat{s}(0) = 0$ ;
3. Let  $\Lambda$  be the set of integers  $\{|n| : \widehat{s}(n) \neq 0\}$ . The greatest common divisor  $\gcd(s)$  of all the elements in  $\Lambda$  is 1.

**Definition 2.2.** A function  $f(t) = \alpha(t)s(2\pi N\phi(t))$  for  $t \in [0, 1]$  is a generalized intrinsic mode function (GIMF) of type  $(M, N)$ , if  $s(t) \in \mathcal{S}_M$  and  $\alpha(t)$  and  $\phi(t)$  satisfy the conditions<sup>4</sup> below:

$$\alpha(t) \in C^\infty, \quad 1/M \leq \alpha \leq M, \quad \phi(t) \in C^\infty, \quad 1/M \leq |\phi'| \leq M.$$

In the analysis of the Gauss-Seidel RDBR, the partition-based regression method (or partitioning estimate) in Chapter 4 of [35] will be adopted. Recall the definition of this regression introduced in (11):

$$s^R(x) \approx s^P(x) := \frac{\sum_{\ell=1}^L \mathcal{X}_{[t_n^h, t_{n+1}^h)}(x_\ell) y_\ell}{\sum_{\ell=1}^L \mathcal{X}_{[t_n^h, t_{n+1}^h)}(x_\ell)},$$

when  $x \in [t_n^h, t_{n+1}^h)$ . The following theorem given in Chapter 4 in [35] estimates the  $L_2$  risk of the approximation  $s^P \approx s^R$  as follows.

**Theorem 2.3.** *For the uniform partition with a step side  $h$  in  $[0, 1)$  as defined just above, assume that*

$$\begin{aligned} \text{Var}(Y|X = x) &\leq \sigma^2, \quad x \in \mathbb{R}, \\ |s^R(x) - s^R(z)| &\leq C|x - z|, \quad x, z \in \mathbb{R}, \end{aligned}$$

*$X$  has a compact support  $[0, 1)$ , and there are  $L$  i.i.d. samples of  $(X, Y)$ . Then the partition-based regression method provides an estimated regression function  $s^P$  to approximate the ground truth regression function  $s^R$ , where*

$$s^R = \arg \min_{s: \mathbb{R} \rightarrow \mathbb{R}} \mathbb{E}\{|s(2\pi X) - Y|^2\},$$

---

<sup>4</sup> In the analysis of the synchrosqueezed transform in [28], a GIMF of type  $(M, N)$  requires stronger conditions as follows:

$$\begin{aligned} \alpha(t) &\in C^\infty, \quad |\alpha'| \leq M, \quad 1/M \leq \alpha \leq M \\ \phi(t) &\in C^\infty, \quad 1/M \leq |\phi'| \leq M, \quad |\phi''| \leq M. \end{aligned}$$

However, the RDBR requires much weaker conditions.

with an  $L^2$  risk bounded by

$$\mathbb{E}\|s^P - s^R\|^2 \leq c_0 \frac{\sigma^2 + \|s^R\|_{L^\infty}^2}{Lh} + C^2 h^2,$$

where  $c_0$  is a constant independent of the number of samples  $L$ , the regression function  $s$ , the step side  $h$ , and the Lipschitz continuity constant  $C$ .

As we shall see later in the assumption of Theorem 2.5, when the small step size  $h$  in the partition-based regression is fixed, we require the number of samples  $L$  to be sufficiently large such that the  $L_2$  risk of the approximation  $s^P \approx s^R$  is small enough. To simplify notations, let  $\mathcal{L}^C$  be the class of functions that are Lipschitz continuous with a constant  $C$ . Denote the set of sampling grid points  $\{t_\ell\}_{\ell=1,\dots,L}$  in Algorithm 1 as  $\mathcal{T}$ . To estimate the regression function using the partition-based regression method,  $\mathcal{T}$  is divided into several subsets as follows. For  $i, j = 1, \dots, K$ ,  $i \neq j$ ,  $m, n = 0, \dots, N^h - 1$ , let

$$\mathcal{T}_h^{ij}(m, n) = \left\{ t \in \mathcal{T} : \text{mod}(p_i(t), 1) \in [t_m^h, t_m^h + h), \text{mod}(p_j(t), 1) \in [t_n^h, t_n^h + h) \right\},$$

and

$$\mathcal{T}_h^i(m) = \left\{ t \in \mathcal{T} : \text{mod}(p_i(t), 1) \in [t_m^h, t_m^h + h) \right\},$$

then  $\mathcal{T} = \cup_{m=0}^{N^h-1} \mathcal{T}_h^i(m) = \cup_{m=0}^{N^h-1} \cup_{n=0}^{N^h-1} \mathcal{T}_h^{ij}(m, n)$ . Let

$$D_h^{ij}(m, n) \quad \text{and} \quad D_h^i(m) \tag{20}$$

denote the number of points in  $\mathcal{T}_h^{ij}(m, n)$  and  $\mathcal{T}_h^i(m)$ , respectively.

**Definition 2.4.** Suppose phase functions  $p_k(t) = N_k \phi_k(t)$  for  $t \in [0, 1]$ , and  $k = 1, \dots, K$ , where  $\phi_k(t)$  satisfies<sup>5</sup>

$$\phi_k(t) \in C^\infty, \quad 1/M \leq |\phi_k'| \leq M.$$

Then the collection of phase functions  $\{p_k(t)\}_{1 \leq k \leq K}$  is said to be  $(M, N, K, h, \beta, \gamma)$ -well-differentiated and denoted as  $\{p_k(t)\}_{1 \leq k \leq K} \subset \mathcal{WD}(M, N, K, h, \beta, \gamma)$ , if the following conditions are satisfied:

1.  $N_k \geq N$  for  $k = 1, \dots, K$ ;
2.  $\gamma := \min_{m, n, i \neq j} D_h^{ij}(m, n)$  satisfies  $\gamma > 0$ , where  $D_h^{ij}(m, n)$  (and  $D_h^i(m)$  below) is defined in (20);
3. Let

$$\beta_{i,j} := \left( \sum_{m=0}^{N^h-1} \frac{1}{D_h^i(m)} \left( \sum_{n=0}^{N^h-1} (D_h^{ij}(m, n) - \gamma)^2 \right) \right)^{1/2}$$

for all  $i \neq j$ , then  $\beta := \max\{\beta_{i,j} : i \neq j\}$  satisfies  $M^2(K-1)\beta < 1$ .

<sup>5</sup> In the analysis of the Jacobi style RDBR in [31], stronger conditions for the well-differentiated phase functions were required as follows:

$$\phi_k(t) \in C^\infty, \quad 1/M \leq |\phi_k'| \leq M, \quad |\phi_k''| \leq M.$$

In fact, these conditions can be further weakened when the phase functions are assumed to be known and the synchrosqueezed transform is not involved in the analysis of RDBR.

In the above definition,  $\gamma$  quantifies the dissimilarity between phase functions. The larger  $\gamma$  is, the more dissimilarity phase functions have. If two phases are very similar, there might be some nearly empty sets  $\mathcal{T}_h^{ij}(m, n)$  and hence  $\gamma$  is small. If  $\gamma$  is larger, the numbers  $\{D_h^{ij}(m, n)\}_{m, n}$  are closer and  $\beta$  would be smaller. To guarantee a large  $\gamma$ ,  $N$  and  $L$  should be sufficiently large. To give more intuition of the well-differentiation, let us revisit the example in Figure 6 to 8. Note that: 1) the second condition  $\gamma := \min_{m, n, i \neq j} D_h^{ij}(m, n) > 0$  requires that the empirical distribution of  $\{\text{mod}(z_\ell^{bk}, 1)\}$  in (13) is uniformly and strictly positive for all cases in the partition-based regression; 2) the number  $\beta$  in the third condition quantifies how uniform the distribution of  $\{\text{mod}(z_\ell^{bk}, 1)\}$  is in the partition-based regression. The fact that these distributions are close to being uniform is equivalent to the fact that  $\beta \approx 0$ . As we have seen in the example in (15), if the distribution of  $\{\text{mod}(z_\ell^{bk}, 1)\}$  is close to being uniform, which can be guaranteed if  $\gamma$  is large and  $\beta$  is close to zero, then one-step of diffeomorphism-based regression is already very accurate. Repeatedly applying the diffeomorphism-based regression can quickly eliminate the residual error.

With these notations defined, we are ready to present the main analysis of the Gauss-Seidel RDBR. Let's recall that in each iteration of Algorithm 1, if we denote the target shape function as  $s_k^{(j)}$  then the given data is

$$r_k^{(j)}(t) = \sum_{\ell=1}^{k-1} \alpha_\ell(t) s_\ell^{(j+1)}(2\pi p_\ell(t)) + \sum_{\ell=k}^K \alpha_\ell(t) s_\ell^{(j)}(2\pi p_\ell(t)). \quad (21)$$

By reformulating the regression problem

$$s_k^{R, (j+1)} = \arg \min_{s: \mathbb{R} \rightarrow \mathbb{R}} \mathbb{E} \left\{ \left| s(2\pi X_k) - Y_k^{(j)} \right|^2 \right\} \quad (22)$$

$$= \arg \min_{s: \mathbb{R} \rightarrow \mathbb{R}} \mathbb{E} \left\{ \left| s(2\pi X_k) - (Y_k^{(j)} - s_k^{(j)}(2\pi X_k)) \right|^2 \right\} - s_k^{(j)}, \quad (23)$$

we see that

$$s_k^{R, (j+1)} = s_k^{(j)} + s_k^{E, (j)},$$

where

$$s_k^{E, (j)}(2\pi x) := \mathbb{E} \{ Y_k^{(j)} - s_k^{(j)}(2\pi X_k) | X_k = x \} \neq 0 \quad (24)$$

due to the perturbation caused by other modes. In the next iteration, the target shape function  $s_k^{(j+1)} = -s_k^{E, (j)}$ . Hence, the key convergence analysis is to show that  $s_k^{E, (j)}$  decays as  $j \rightarrow \infty$ .

In what follows, we assume that an accuracy parameter  $\epsilon$  is fixed. Furthermore, suppose GIMF's  $f_k(t) = \alpha_k(t) s_k(2\pi N_k \phi_k(t))$ ,  $k = 1, \dots, K$ , have phases in  $\mathcal{WD}(M, N, K, h, \beta, \gamma)$ , all generalized shape functions and amplitude functions are in the space  $\mathcal{L}^C$ . Under these conditions, all regression functions  $s_k^{(j)} \in \mathcal{L}^C$  and have bounded  $L^\infty$  norm linearly depending only on  $M$  and  $K$ . By Line 8 in Algorithm 1, we have the nice and key condition that  $\int_0^1 s_k^{(j)}(2\pi t) dt = 0$  at each iteration for all  $k$  and  $j$ . Note that  $\text{Var}(Y_k^{(j)} | X_k = x)$  is bounded by a constant linearly depending only on  $M$  and  $K$  as well. For the fixed  $\epsilon$  and  $C$ , there exists  $h_0(\epsilon, C)$  such that  $C^2 h^2 < \epsilon^2$  if  $0 < h < h_0$ . By the abuse of notation,  $O(\epsilon)$  is used instead of  $Ch$  later. By Theorem 2.3, for the fixed  $\epsilon$ ,  $M$ ,  $K$ ,  $C$ , and  $h$ , there exists  $L_0(\epsilon, M, K, C, h)$  such that the  $L^2$  error of the partition-based regression is bounded by  $\epsilon^2$ . In what follows,  $h$  is smaller than  $h_0$ ,  $L$  is larger than  $L_0$ , and hence all estimated



regression functions approximate the ground truth regression function with an  $L^2$  error of order  $\epsilon$ . Under these conditions and assumptions, as long as  $N$  and  $L$  are large enough,  $s_k^{E,(j)}$  is shown to decay to  $O(\epsilon)$  as  $j \rightarrow \infty$ , and the decay rate will be estimated.

**Theorem 2.5.** (Convergence of the Gauss-Seidel RDBR when  $\bar{K} = K$ ) Suppose  $\bar{K} = K$ ,  $\{1, \dots, K\} = \{\tau_1, \dots, \tau_K\}$ ,  $n_k = 1$  for all  $k$  in Equation (18) and (19). Under the conditions listed in the paragraph immediately preceding this theorem, we have

$$\|s_k^{E,(j)}\|_{L^2} \leq O(c_0\epsilon + (M^2(K-1)\beta)^{c(j,k)})$$

and

$$\|r_k^{(j+1)}\|_{L^2} \leq O(c_0\epsilon + (M^2(K-1)\beta)^{c(j,1)})$$

for all  $j \geq 0$  and  $1 \leq k \leq K$ , where  $c(j,k) = \left\lceil \frac{jK+k}{K-1} \right\rceil$ , the smallest integer larger than or equal to  $\frac{jK+k}{K-1}$ ,  $c_0 = \frac{1}{1-M^2(K-1)\beta}$  is a constant number,  $s_k^{E,(j)}$  is defined in Equation (24) and  $r_k^{(j)}$  is defined in Equation (21).

*Proof.* First, we start with the case when  $K = 2$  and  $\alpha_k(t) = 1$  for all  $t$  and  $k$ .

Recall that  $p_k(t)$  can be considered as a diffeomorphism from  $\mathbb{R}$  to  $\mathbb{R}$  transforming data in the  $t$  domain to the  $p_k(t)$  domain. We have introduced the inverse-warping data

$$\begin{aligned} h_k^{(j)}(v) &= r_k^{(j)} \circ p_k^{-1}(v) \\ &= s_k^{(j)}(2\pi v) + \sum_{\ell=1}^{k-1} s_\ell^{(j+1)}(2\pi p_\ell \circ p_k^{-1}(v)) + \sum_{\ell=k+1}^K s_\ell^{(j)}(2\pi p_\ell \circ p_k^{-1}(v)) \\ &:= s_k^{(j)}(2\pi v) + \kappa_k^{(j)}(2\pi v), \end{aligned}$$

where  $v = p_k(t)$ . After the folding map

$$\tau : (v, h_k(v)) \mapsto (\text{mod}(v, 1), h_k^{(j)}(v)),$$

we have  $(x_\ell, y_\ell) = \tau(v_\ell, s_k^{(j)}(2\pi v_\ell) + \kappa_k^{(j)}(2\pi v_\ell))$  for  $\ell = 1, \dots, L$  as  $L$  i.i.d. samples of a random vector  $(X_k, Y_k^{(j)})$ , where  $X_k \in [0, 1]$ . We can assume the target shape functions  $s_k^{(j)}$  for all  $k$  at the  $j$ th step are known in the analysis, although they are not known in practice. The partition-based regression method is applied (not necessary to know the distribution of the random vector  $(X_k, Y_k^{(j)})$ ) to solve the following regression problem approximately

$$\arg \min_{s: \mathbb{R} \rightarrow \mathbb{R}} \mathbb{E} \left\{ \left| s(2\pi X_k) - (Y_k^{(j)} - s_k^{(j)}(2\pi X_k)) \right|^2 \right\}, \quad (25)$$

and the solution is denoted as  $s_k^{P,(j)}$ . We would like to emphasize that  $s_k^{P,(j)}$  is only used in the analysis and it is not computed in Algorithm 1. Recall notations in Definition 3.2. By the partition-based regression method, when  $x \in [t_m^h, t_m^h + h)$ ,

$$s_1^{P,(j)}(2\pi x) = \frac{\sum_{n=0}^{N^h-1} \left( s_2^{(j)}(2\pi t_n^h) + O(\epsilon) \right) D_h^{12}(m, n)}{D_h^1(m)},$$

and

$$s_2^{P,(j)}(2\pi x) = \frac{\sum_{n=0}^{N^h-1} \left( s_1^{(j+1)}(2\pi t_n^h) + O(\epsilon) \right) D_h^{21}(m, n)}{D_h^2(m)}, \quad (26)$$

where  $O(\epsilon)$  comes from the approximation of the  $\mathcal{L}^C$  function  $s_i$  using the values on grid points  $t_n^h$ . Note that in the case of the Jacobi style RDBR in [31], the term  $s_1^{(j+1)}$  in Equation (26) is replaced with  $s_1^{(j)}$ , since the Jacobi style iteration doesn't use the latest estimations of shape functions. The following argument is similar to that for Lemma 3.3 in [31]. It is easy to check that

$$|s_1^{P,(j)}(2\pi x)| \leq O(\epsilon) + \frac{\sum_{n=0}^{N^h-1} s_2^{(j)}(2\pi t_n^h) (D_h^{12}(m, n) - \gamma)}{D_h^1(m)}$$

and

$$|s_2^{P,(j)}(2\pi x)| \leq O(\epsilon) + \frac{\sum_{n=0}^{N^h-1} s_1^{(j+1)}(2\pi t_n^h) (D_h^{21}(m, n) - \gamma)}{D_h^2(m)},$$

which imply that

$$\|s_1^{P,(j)}\|_{L^2} = O(\epsilon) + \left( \sum_{n=0}^{N^h-1} \left( s_2^{(j)}(2\pi t_n^h) \right)^2 h \right)^{1/2} \left( \sum_{m=0}^{N^h-1} \left( \sum_{n=0}^{N^h-1} \left( \frac{D_h^{12}(m, n) - \gamma}{D_h^1(m)} \right)^2 \right) \right)^{1/2}$$

and

$$\|s_2^{P,(j)}\|_{L^2} = O(\epsilon) + \left( \sum_{n=0}^{N^h-1} \left( s_1^{(j+1)}(2\pi t_n^h) \right)^2 h \right)^{1/2} \left( \sum_{m=0}^{N^h-1} \left( \sum_{n=0}^{N^h-1} \left( \frac{D_h^{21}(m, n) - \gamma}{D_h^2(m)} \right)^2 \right) \right)^{1/2}$$

by the triangle inequality and Hölder's inequality. Since  $s_2^{(j)}$  and  $s_1^{(j+1)}$  are in  $\mathcal{L}^C$ , we have

$$\left( \sum_{n=0}^{N^h-1} \left( s_2^{(j)}(2\pi t_n^h) \right)^2 h \right)^{1/2} = \|s_2^{(j)}\|_{L^2} + O(\epsilon)$$

and

$$\left( \sum_{n=0}^{N^h-1} \left( s_1^{(j+1)}(2\pi t_n^h) \right)^2 h \right)^{1/2} = \|s_1^{(j+1)}\|_{L^2} + O(\epsilon).$$

Since phase functions are in  $\mathcal{WD}(M, N, h, \beta, \gamma)$ ,

$$\left( \sum_{m=0}^{N^h-1} \left( \sum_{n=0}^{N^h-1} \left( \frac{D_h^{ki}(m, n) - \gamma}{D_h^k(m)} \right)^2 \right) \right)^{1/2} \leq \beta < 1.$$

Hence,

$$\|s_1^{P,(j)}\|_{L^2} \leq O(\epsilon) + \beta \|s_2^{(j)}\|_{L^2}^2,$$

and

$$\|s_2^{P,(j)}\|_{L^2} \leq O(\epsilon) + \beta \|s_1^{(j+1)}\|_{L^2}^2.$$

By the conditions just listed immediately before Theorem 2.5, and the definition in Equation (24), we have

$$\|s_1^{E,(j)}\|_{L^2} \leq O(\epsilon) + \|s_1^{P,(j)}\|_{L^2} \leq O(\epsilon) + \beta \|s_2^{(j)}\|_{L^2}^2, \quad (27)$$

and

$$\|s_2^{E,(j)}\|_{L^2} \leq O(\epsilon) + \|s_2^{P,(j)}\|_{L^2} \leq O(\epsilon) + \beta \|s_1^{(j+1)}\|_{L^2}^2.$$

Note that  $s_1^{E,(j)} = s_1^{(j+1)}$ , it holds that

$$\|s_2^{E,(j)}\|_{L^2} \leq O(\epsilon) + \beta \|s_1^{(j+1)}\|_{L^2}^2 = O(\epsilon) + \beta \|s_1^{E,(j)}\|_{L^2}^2 \leq O(\epsilon) + \beta^2 \|s_2^{(j)}\|_{L^2}^2. \quad (28)$$

By Equation (27), (28), and the mathematical induction similar to that for Theorem 3.5 in [31], it is easy to show that

$$\|s_k^{E,(j)}\|_{L^2} \leq O(c_0\epsilon + \beta^{c(j,k)})$$

where  $c(j,k) = 2j + k$  and  $c_0 = \frac{1}{1-\beta}$  coming from the geometric sequence due to the summation of the  $O(\epsilon)$  term for all  $j \geq 0$ .

To care the general case, we need to extend the argument to  $K > 2$  and non-constant  $\alpha_k$ . We shall do this in two steps: first  $K > 2$  but  $\alpha_k \equiv 1$  for all  $k$ , and then, finally,  $K > 2$  and varying  $\alpha_k$ . Rather than repeating the earlier argument in full detail, adapted to these more general situations, we indicate simply, for both steps, what extra estimates need to be taken into account. This may not give the sharpest estimate, but this is not a concern for now.

Next, we prove the case when  $K > 2$  and  $\alpha_k(t) = 1$  for all  $t$  and  $k$ . Similarly, by the definition of the partition-based regression and the triangle inequality, we have

$$\begin{aligned} |s_k^{P,(j)}(2\pi x)| &\leq O(K\epsilon) + \sum_{i=1}^{k-1} \frac{\sum_{n=0}^{N^h-1} s_i^{(j+1)}(2\pi t_n^h) (D_h^{ki}(m, n) - \gamma)}{D_h^k(m)} \\ &\quad + \sum_{i=k+1}^K \frac{\sum_{n=0}^{N^h-1} s_i^{(j)}(2\pi t_n^h) (D_h^{ki}(m, n) - \gamma)}{D_h^k(m)}. \end{aligned}$$

Hence, by the triangle inequality and the Hölder inequality again, it holds that

$$\begin{aligned} \|s_k^{P,(j)}\|_{L^2} &\leq O(K\epsilon) + \sum_{i=1}^{k-1} \left( \sum_{m=0}^{N^h-1} \left( \frac{\sum_{n=0}^{N^h-1} s_i^{(j+1)}(2\pi t_n^h) (D_h^{ki}(m, n) - \gamma)}{D_h^k(m)} \right)^2 h \right)^{1/2} \\ &\quad + \sum_{i=k+1}^K \left( \sum_{m=0}^{N^h-1} \left( \frac{\sum_{n=0}^{N^h-1} s_i^{(j)}(2\pi t_n^h) (D_h^{ki}(m, n) - \gamma)}{D_h^k(m)} \right)^2 h \right)^{1/2} \\ &\leq O(K\epsilon) + \sum_{i=1}^{k-1} \left( \sum_{n=0}^{N^h-1} (s_i^{(j+1)}(2\pi t_n^h))^2 h \right)^{1/2} \left( \sum_{m=0}^{N^h-1} \left( \sum_{n=0}^{N^h-1} \left( \frac{D_h^{ki}(m, n) - \gamma}{D_h^k(m)} \right)^2 \right) \right)^{1/2} \\ &\quad + \sum_{i=k+1}^K \left( \sum_{n=0}^{N^h-1} (s_i^{(j)}(2\pi t_n^h))^2 h \right)^{1/2} \left( \sum_{m=0}^{N^h-1} \left( \sum_{n=0}^{N^h-1} \left( \frac{D_h^{ki}(m, n) - \gamma}{D_h^k(m)} \right)^2 \right) \right)^{1/2} \\ &\leq O(\epsilon) + \sum_{i=1}^{k-1} \beta \|s_i^{(j+1)}\|_{L^2} + \sum_{i=k+1}^K \beta \|s_i^{(j)}\|_{L^2}. \end{aligned}$$

Since  $s_k^{E,(j)} - s_k^{P,(j)} = O(\epsilon)$ , we know

$$\|s_k^{E,(j)}\|_{L^2} \leq O(\epsilon) + \sum_{i=1}^{k-1} \beta \|s_i^{(j+1)}\|_{L^2} + \sum_{i=k+1}^K \beta \|s_i^{(j)}\|_{L^2}.$$

Similar to the case of two components, by the equation just above and mathematical induction, one can show that

$$\|s_k^{E,(j)}\|_{L^2} \leq O(c_0\epsilon + \beta^{c(j,k)})$$

where  $c_0 = \frac{1}{1-(K-1)\beta}$  coming from the geometric sequence due to the summation of the  $O(\epsilon)$  term for all  $j \geq 0$ ,

$$c(j,k) = \left\lceil \frac{jK+k}{K-1} \right\rceil,$$

and  $\lceil \cdot \rceil$  is the ceiling operator.

Finally, we prove the case when amplitude functions are smooth functions but not a constant 1. If the instantaneous frequencies are sufficiently large, depending on  $\epsilon$ ,  $M$ ,  $K$ , and  $C$ , amplitude functions are nearly constant up to an approximation error of order  $\epsilon$ . The time domain  $[0, 1]$  is divided into sufficiently small intervals such that amplitude functions are nearly constant inside each interval. Accordingly, the samples  $(x_\ell, y_\ell) = \tau(v_\ell, s_k(2\pi v_\ell) + \kappa_k(v_\ell))$  for  $\ell = 1, \dots, L$  of the random vector  $(X_k, Y_k^{(j)})$  is divided into groups and the partition-based regression method is applied to estimate the regression function for each group. This is similar to data splitting in nonparametric regression. The bound of  $|s_k^{P,(j)}(x)|$  is a weighted average of the bound given by each group, and the weight comes the number of points in each group over the total number of samples. Note that  $\|\alpha_k\|_{L^\infty} \leq M$ . By repeating the analysis above, it is simple to show

$$\|s_k^{E,(j)}\|_{L^2} \leq O(\epsilon) + M^2 \sum_{i=1}^{k-1} \beta \|s_i^{(j+1)}\|_{L^2} + M^2 \sum_{i=k+1}^K \beta \|s_i^{(j)}\|_{L^2}.$$

where  $M^2$  comes from

$$\frac{\alpha_i \circ p_k^{-1}(v)}{\alpha_k \circ p_k^{-1}(v)}$$

in  $\kappa_k(v)$  after warping. By mathematical induction similar to that for Theorem 3.5 in [31] again, it holds that

$$\|s_k^{E,(j)}\|_{L^2} \leq O(c_0\epsilon + (M^2(K-1)\beta)^{c(j,k)}) \quad (29)$$

where  $c_0 = \frac{1}{1-M^2(K-1)\beta}$  coming from the geometric sequence due to the summation of the  $O(\epsilon)$  term for all  $j \geq 0$ , and

$$c(j,k) = \left\lceil \frac{jK+k}{K-1} \right\rceil.$$

This finishes the proof of the first part of Theorem 2.5.

By the definition of  $r_k^{(j)}$  in Equation (21) and the inequality in Equation (29), it holds that

$$\|r_k^{(j+1)}\|_{L^2} \leq O(c_0\epsilon + (M^2(K-1)\beta)^{c(j,1)}),$$

which completes the proof of Theorem 2.5.  $\square$

Theorem 2.5 shows that the regression function in each step of Algorithm 1 decays, if  $M^2(K-1)\beta < 1$ , in the  $L^2$  sense up to a fixed accuracy parameter as the iteration number becomes large. Hence, the recovered shape function converges and the residual decays up to a fixed accuracy parameter, if  $M^2(K-1)\beta < 1$ . When the iteration number is sufficiently large, the accuracy of the RDBR in Theorem 2.5 is as good as a single step of regression

in Theorem 2.3. Compared to the convergence theorem of the Jacobi RDBR in [31], both the Jacobi style and the Gauss-Seidel style recursive scheme have linear convergence, but the Gauss-Seidel style has a smaller rate of convergence; especially in the case of two components, the rate of convergence of the Gauss-Seidel style is  $\beta^2$  while the one of the Jacobi style is  $\beta$ .

**Theorem 2.6.** (*Noise robustness of the Gauss-Seidel RDBR*) Let  $f_k(t) = \alpha_k(t)s_k(2\pi N_k\phi_k(t))$ ,  $k = 1, \dots, K$ , be  $K$  GIMF's and  $f(t) = \sum_{k=1}^K f_k(t) + n(t)$ , where  $n(t)$  is a random noise with a bounded variance  $\sigma^2$ . Under the other conditions introduced in Theorem 2.5, for the given  $\epsilon$ ,  $\exists L_0(\epsilon, M, K, C, h, \sigma)$ , if  $L > L_0$ , then

$$\|s_k^{E,(j)}\|_{L^2} \leq O(c_0\epsilon + (M^2(K-1)\beta)^{c(j,k)})$$

and

$$\|r_k^{(j+1)}\|_{L^2} \leq O(c_0\epsilon + (M^2(K-1)\beta)^{c(j,1)}),$$

for all  $j \geq 0$  and  $1 \leq k \leq K$ , where  $c(j,k) = \left\lceil \frac{jK+k}{K-1} \right\rceil$ ,  $c_0 = \frac{1}{1-M^2(K-1)\beta}$  is a constant number,  $s_k^{E,(j)}$  is defined in Equation (24) and  $r_k^{(j)}$  is defined in Equation (21).

Theorem 2.6 is an immediate result of Theorem 2.3 and 2.5. It shows that as soon as the number of sampling points  $L$  is large enough, the noise effect will be negligible.

Next, we discuss the case when the estimated number of components is inexact, i.e.,  $\bar{K} \neq K$  in Algorithm 1. Two main situations are concerned here: (1)  $\bar{K} > K$  and  $\{p_k\}_{1 \leq k \leq \bar{K}}$  consists of all the fundamental phase functions and their multiples; (2)  $\bar{K} < K$  and  $\{p_k\}_{1 \leq k \leq \bar{K}}$  are fundamental phase functions. The convergence analysis of other situations can be generalized from these two situations.

In the first situation when  $\bar{K} > K$ ,  $\{p'_k(t) = n_k N_{\tau_k} \phi'_{\tau_k}(t)\}_{1 \leq k \leq \bar{K}}$  naturally belong to  $K$  groups  $\{\mathcal{G}_k\}_{1 \leq k \leq K}$ , each of which corresponds to the multiple of a fundamental instantaneous frequency  $p'_k(t)$ . For each  $p_k$ , the RDBR tries to identify a shape function and the result depends on the order of  $p_k$  in the set  $\{p_k\}$ . To make sure that the shape functions corresponding to fundamental phases are approximately the ground truth shape functions, there is a sorting procedure in Line 4 and 5 of Algorithm 1. By the similar analysis in Theorem 2.5, one can show that the summation of the shape function estimations corresponding to the phase functions within each group  $\mathcal{G}_k$  converges to the  $k$ th ground truth shape function  $s_k$ ; i.e., let

$$\tilde{s}_k^{(j)} = \sum_{\tau \in \mathcal{G}_k} \bar{s}_\tau^{(j)}, \quad (30)$$

then  $\lim_{j \rightarrow \infty} \tilde{s}_k^{(j)} = O(\epsilon) + s_k$ . In particular, Theorem 2.7 characterizes its convergence rate as follows.

**Theorem 2.7.** (*Convergence of the Gauss-Seidel RDBR when  $\bar{K} > K$* ) Suppose  $\bar{K} > K$  and  $\{p_k\}_{1 \leq k \leq \bar{K}}$  in Algorithm 1 contains all the fundamental phase functions. Under the conditions listed in the paragraph immediately preceding Theorem 2.5, we have

$$\|\tilde{s}_k^{(j)} - s_k\|_{L^2} \leq O(c_0\epsilon + (M^2(K-1)\beta)^{c(j,k)})$$

and

$$\|\tilde{r}_k^{(j+1)}\|_{L^2} \leq O(c_0\epsilon + (M^2(K-1)\beta)^{c(j,1)})$$

for all  $j \geq 0$  and  $1 \leq k \leq K$ , where  $c(j, k) = \left\lceil \frac{jK+k}{K-1} \right\rceil$ ,  $c_0 = \frac{1}{1-M^2(K-1)\beta}$  is a constant number,  $\tilde{s}_k^{(j)}$  is defined in Equation (30) and  $\tilde{r}_k^{(j)}$  is defined as

$$\tilde{r}_k^{(j)} = r_{\tau_k}^{(j)},$$

where  $\tau_k \in \mathcal{G}_k$  satisfies  $n_{\tau_k} \geq n_\nu$  for all  $\nu \in \mathcal{G}_k$ , and the notation  $n_\nu$  is introduced in Equation (18).

*Proof.* The proof of Theorem 2.7 can be generalized from Theorem 2.5. The phase functions  $\{p_\nu\}_{\nu \in \mathcal{G}_k}$  compete with each other to obtain a larger  $L^2$ -norm in their own shape function estimation  $\bar{s}_\nu^{(j)}$ . This competition does not have negative effects on the convergence of Algorithm 1; in fact, the more members in  $\mathcal{G}_k$ , the faster convergence of the shape function estimation, because the estimation error  $\|\bar{s}_k^{(j)} - s_k\|_{L^2}$  can be reduced by the extra regressions due to more than one member in  $\mathcal{G}_k$ .  $\square$

It is worth pointing out that the Gauss-Seidel RDBR converges in the sense of Theorem 2.7 when  $\bar{K} > K$ ; however, the Jacobi RDBR diverges when  $\bar{K} > K$ .

In the second situation when  $\bar{K} < K$ , it is assumed that  $\{p_k\}_{1 \leq k \leq \bar{K}}$  are all fundamental phase functions. Since not all the fundamental phase functions are used in Algorithm 1, the components corresponding to the missing fundamental phase functions always remain in the residual signal  $r_k^{(j)}$ . Hence, the shape function estimation corresponding to the known phase functions cannot be improved by recursive regression, and Algorithm 1 stops iteration quickly since the stopping criteria  $|\epsilon_1 - \epsilon_0| \leq \epsilon$  is soon satisfied. After Algorithm 1 stops, by the arguments in Theorem 2.5, it is easy to see that the estimated shape function  $\bar{s}_k$  satisfies

$$\|\bar{s}_k - s_k\|_{L^2} \leq O(\epsilon + M^2\beta \sum_{\ell \neq k} \|s_\ell\|_{L^2})$$

for  $1 \leq k \leq \bar{K}$ .

### 3 RDBR for the multiresolution mode decomposition

#### 3.1 Algorithm description

In what follows, we modify the Gauss-Seidel RDBR in Algorithm 1 to solve the multiresolution mode decomposition. The multiresolution mode decomposition problem aims at extracting each multiresolution intrinsic mode function  $f_k(t)$  from their superposition  $f(t) = \sum_{k=1}^K f_k(t)$ , estimating their corresponding multiresolution expansion coefficients and the shape function series defined as follows.

**Definition 3.1.** A function

$$f(t) = \sum_{n=-N/2}^{N/2-1} a_n \cos(2\pi n\phi(t)) s_{cn}(2\pi N\phi(t)) + \sum_{n=-N/2}^{N/2-1} b_n \sin(2\pi n\phi(t)) s_{sn}(2\pi N\phi(t)) \quad (31)$$

is a multiresolution intrinsic mode function (MIMF) of type  $(M_0, M, N, \epsilon)$  defined on  $[0, 1]$ , if the conditions below are satisfied:

- the shape function series  $\{s_{cn}(t)\}$  and  $\{s_{sn}(t)\}$  are in  $\mathcal{S}_M$ ;

- the multiresolution expansion coefficients  $\{a_n\}$  and  $\{b_n\}$  satisfy

$$\begin{aligned} \sum_{n=-N/2}^{N/2-1} |a_n| &\leq M, & \sum_{n=-N/2}^{N/2-1} |a_n| - \sum_{n=-M_0}^{M_0-1} |a_n| &\leq \epsilon, \\ \sum_{n=-N/2}^{N/2-1} |b_n| &\leq M, & \sum_{n=-N/2}^{N/2-1} |b_n| - \sum_{n=-M_0}^{M_0-1} |b_n| &\leq \epsilon; \end{aligned}$$

- $\phi(t)$  satisfies

$$\phi(t) \in C^\infty, \quad 1/M \leq |\phi'| \leq M.$$

A MIMF is a generalization of the GIMF in Equation (2) for time-dependent shape functions to describe the nonlinear and non-stationary time series adaptively, as we shall see in numerical examples in Section 4. A single MIMF itself is a superposition of GIMF's that share the same phase function; identifying its multiresolution expansion coefficients and shape function series requires separating these GIMF's.

The only difference to distinguish different GIMF's in a MIMF is the frequency of the oscillation in  $\cos(2\pi n\phi(t))$  and  $\sin(2\pi n\phi(t))$  (see Equation (31)). Hence, if  $\phi(t)$  and  $N$  are known exactly, Fourier analysis in the coordinate of  $\phi(t)$  (instead of  $t$ ) can estimate the multiresolution expansion coefficients  $\{a_n\}$ ,  $\{b_n\}$ , and the shape function series  $\{s_{cn}\}$  and  $\{s_{sn}\}$ . However, in practice  $\phi(t)$  and  $N$  are only known approximately and the estimation error in  $\phi(t)$  can be amplified by a factor  $O(Nm)$ , where  $m$  is the frequency bandwidth of shape functions, if the Fourier analysis in the  $\phi(t)$  coordinate is applied. This instability makes the Fourier approach less attractive in analyzing a MIMF.

An immediate question is whether the diffeomorphism-based regression idea in Section 3 can estimate the GIMF's in a MIMF. The main concerns are: when regression is applied to estimate one GIMF, other GIMF's acting as additive perturbation seem to cause a large estimation error; when  $n \neq 0$ ,  $\cos(2\pi n\phi(t))$  and  $\sin(2\pi n\phi(t))$  as the amplitudes of GIMF's are occasionally zero, making it numerically infeasible to apply the diffeomorphism-based regression directly. Fortunately, for the first concern, the oscillatory functions  $\cos(2\pi n\phi(t))$  and  $\sin(2\pi n\phi(t))$  have zero mean in the coordinate of  $\phi(t)$ , exactly cancelling out the noise perturbation. As for the second concern, by the formulas of trigonometric functions,

$$\begin{aligned} \cos(2\pi m\phi(t))f(t) &= \sum_{n=-N/2}^{N/2-1} \frac{a_n}{2} (\cos(2\pi(m+n)\phi(t)) + \cos(2\pi(m-n)\phi(t))) s_{cn}(2\pi N\phi(t)) \\ &+ \sum_{n=-N/2}^{N/2-1} \frac{b_n}{2} (\sin(2\pi(n+m)\phi(t)) + \sin(2\pi(n-m)\phi(t))) s_{sn}(2\pi N\phi(t)), \end{aligned}$$

where there is only one term with a non-zero-mean amplitude,  $\frac{a_m}{2} s_{cm}(2\pi N\phi(t))$ , implying that the diffeomorphism-based regression could be able to estimate  $a_m$  and  $s_{cm}$  from  $\cos(2\pi m\phi(t))f(t)$ . Similarly, we can estimate  $b_m$  and  $s_{sm}$  from  $\sin(2\pi m\phi(t))f(t)$ . Hence, by applying the diffeomorphism-based regression to  $\cos(2\pi n\phi(t))f(t)$  and  $\sin(2\pi n\phi(t))f(t)$  for  $n = -N/2, \dots, N/2 - 1$ , we can estimate all the multiresolution expansion coefficients  $\{a_n\}$ ,  $\{b_n\}$ , and the shape function series  $\{s_{cn}\}$  and  $\{s_{sn}\}$ .

However, in the presence of a superposition of several MIMFs, a sequence of diffeomorphism-based analysis discussed just above cannot estimate all the multiresolution expansion coefficients and the shape function series. The reason for this inaccuracy is similar to the motivation of the RDBR for GMD in Section 2. Hence, we expect that a modified version of the RDBR can solve the multiresolution mode decomposition problem under the condition of well-differentiated phase functions defined as follows.

**Definition 3.2.** *Suppose*

$$f_k(t) = \sum_{n=-N_k/2}^{N_k/2-1} a_{n,k} \cos(2\pi n\phi_k(t)) s_{cn,k}(2\pi N_k\phi_k(t)) + \sum_{n=-N_k/2}^{N_k/2-1} b_{n,k} \sin(2\pi n\phi_k(t)) s_{sn,k}(2\pi N_k\phi_k(t)).$$

is a MIMF of type  $(M_0, M, N_k, \epsilon)$  for  $t \in [0, 1]$ ,  $k = 1, \dots, K$ , and

$$\{p_k(t) = N_k\phi_k(t)\}_{1 \leq k \leq K} \subset \mathcal{WD}(M, N, K, h, \beta, \gamma),$$

then  $f(t) = \sum_{k=1}^K f_k(t)$  is said to be a well-differentiated superposition of MIMFs of type  $(M_0, M, N, K, h, \beta, \gamma, \epsilon)$ . Denote the set of all these functions  $f(t)$  as  $\mathcal{WS}(M_0, M, N, K, h, \beta, \gamma, \epsilon)$ .

Recall that  $\mathcal{M}_\ell$  is the operator for computing the  $\ell$ -banded multiresolution approximation to a MIMF  $f(t)$  in Equation (31), i.e.,

$$\mathcal{M}_\ell(f)(t) = \sum_{n=-\ell}^{\ell} a_n \cos(2\pi n\phi(t)) s_{cn}(2\pi N\phi(t)) + \sum_{n=-\ell}^{\ell} b_n \sin(2\pi n\phi(t)) s_{sn}(2\pi N\phi(t)). \quad (32)$$

Ideally, we hope to apply the modified RDBR algorithm in Algorithm 2 to  $f(t) - \mathcal{M}_{n-1}(f)(t)$  for computing the multiresolution expansion coefficients  $\{a_{n,k}\}_{1 \leq k \leq K}$ ,  $\{b_{n,k}\}_{1 \leq k \leq K}$ , and the shape functions  $\{s_{cn,k}\}_{1 \leq k \leq K}$  and  $\{s_{sn,k}\}_{1 \leq k \leq K}$ , where  $\mathcal{M}_{n-1}(f)(t)$  is available from previous computation. However, the modified RDBR can only return estimations approximately and  $\mathcal{M}_{n-1}(f)(t)$  is only available approximately. This motivates us to repeatedly apply the same idea to refine the estimations. In summary, Algorithm 3 below identifies  $\mathcal{M}_{M_0}(f_k)(t)$ , its multiresolution expansion coefficients  $\{a_{n,k}\}$  and  $\{b_{n,k}\}$ , and its shape function series  $\{s_{cn,k}\}$  and  $\{s_{sn,k}\}$  from the superposition  $f(t) \in \mathcal{WS}(M_0, M, N, K, h, \beta, \gamma, \epsilon)$  based on repeating the modified RDBR in Algorithm 2. In the pseudo-code in Algorithm 3, the input and output of Algorithm 2 is denoted as

$$[\{\bar{s}_k\}_{1 \leq k \leq \bar{K}}, \{\bar{f}_k\}_{1 \leq k \leq \bar{K}}] = \text{RDBR}(f, \{p_k\}_{1 \leq k \leq \bar{K}}, n, tp, \epsilon, J).$$

The MMD algorithm in Algorithm 3 is essentially  $J_1 J_2 M_0$  iterations of recursive regression, while the GMD solved by Algorithm 1 requires  $J_1$  iterations when we set  $J = J_1$ . The main computational cost in each iteration in both algorithms is the same and depends on the cost of regression. Hence, the computational cost of the MMD over the one of GMD is always  $O(J_2 M_0)$  independent of data and computational environment. In the current version of code, both GMD and MMD are implemented with a knot-free spline regression in [36], which might not be efficient enough. In a parallel paper concerning the efficient numerical implementation of the MMD (including GMD) algorithm, we have proposed fast algorithms to solve the MMD problem [?].



```

1 Input:  $L$  points of i.i.d. measurement  $\{f(t_\ell)\}_{\ell=1,\dots,L}$  with  $t_\ell \in [0, 1]$ , estimated
instantaneous phases  $\{p_k\}_{k=1,\dots,\bar{K}}$ , an accuracy parameter  $\epsilon < 1$ , the maximum
iteration number  $J$ , frequency  $n$ , the type of amplitude  $tp$ .
2 Output: the estimated shape functions  $\{\bar{s}_k\}_{k=1,\dots,\bar{K}}$ , and the estimated modes
 $\{\bar{f}_k(t)\}_{k=1,\dots,\bar{K}}$  at the sampling grid points  $\{t_\ell\}_{1 \leq \ell \leq L}$ .
3 Initialize: let  $r_1^{(0)} = f$ ,  $\epsilon_1 = \epsilon_2 = 1$ ,  $\epsilon_0 = 2$ , the iteration number  $j = 0$ ,  $\dot{s}_k^{(0)} = 0$ ,
 $\bar{s}_k^{(0)} = 0$ , and  $\bar{f}_k = 0$  for all  $k = 1, \dots, \bar{K}$ .
4 Compute  $N_k$  as the integer nearest to the average of  $p'_k(t)$  for  $k = 1, \dots, \bar{K}$ .
5 Sort  $\{N_k\}_{1 \leq k \leq \bar{K}}$  in an ascending order and reorder the amplitude and phase
functions accordingly.
6 while  $j < J$ ,  $\epsilon_1 > \epsilon$ ,  $\epsilon_2 > \epsilon$ , and  $|\epsilon_1 - \epsilon_0| > \epsilon$  do
7   for  $k = 1, \dots, \bar{K}$  do
8     if  $tp = 1$  then
9        $\lfloor$  Let  $g(t) = \cos(2\pi \frac{n}{N_k} p_k(t))$  and evaluate  $h_k^{(j)}(t) = g(p_k^{-1}(t))r_k^{(j)}(p_k^{-1}(t))$ .
10      else
11         $\lfloor$  Let  $g(t) = \sin(2\pi \frac{n}{N_k} p_k(t))$  and evaluate  $h_k^{(j)}(t) = g(p_k^{-1}(t))r_k^{(j)}(p_k^{-1}(t))$ .
12      Repeat Line 9 to 11 in Algorithm 1 to compute  $\dot{s}_k^{(j+1)}$ .
13      if  $m = 0$  then
14         $\lfloor$  Let  $\dot{f}_k^{(j+1)}(t) = \dot{s}_k^{(j+1)}(2\pi p_k(t))$ .
15      else
16         $\lfloor$  Let  $\dot{f}_k^{(j+1)}(t) = 2g(t)\dot{s}_k^{(j+1)}(2\pi p_k(t))$  and  $\dot{s}_k^{(j+1)} = 2\dot{s}_k^{(j+1)}$ .
17      Let  $\bar{s}_k^{(j+1)} = \bar{s}_k^{(j)} + \dot{s}_k^{(j+1)}$ .
18      Update  $\bar{f}_k(t) \leftarrow \bar{f}_k(t) + \dot{f}_k^{(j+1)}(t)$ .
19      if  $k < \bar{K}$  then
20         $\lfloor$  Let  $r_{k+1}^{(j)} = r_k^{(j)} - \dot{f}_k^{(j+1)}(t)$ .
21      else
22         $\lfloor$  Let  $r_1^{(j+1)} = r_k^{(j)} - \dot{f}_k^{(j+1)}(t)$ .
23      Update  $\epsilon_0 = \epsilon_1$ ,  $\epsilon_1 = \|r_1^{(j+1)}\|_{L^2}$ ,  $\epsilon_2 = \max_k \{\|\dot{s}_k^{(j+1)}\|_{L^2}\}$ .
24      Set  $j = j + 1$ .

```

**Algorithm 2:** A modified RDBR for the multiresolution mode decomposition.

```

1 Input:  $L$  points of i.i.d. measurement  $\{f(t_\ell)\}_{\ell=1,\dots,L}$  with  $t_\ell \in [0, 1]$ , estimated
instantaneous phases  $\{p_k\}_{k=1,\dots,\bar{K}}$ , accuracy parameters  $\epsilon_1$  and  $\epsilon_2$ , the maximum
iteration numbers  $J_1$  and  $J_2$ , and the band-width parameter  $M_0$ .
2 Output:  $\mathcal{M}_{M_0}(f_k)(t)$  at the sampling grid points  $\{t_\ell\}_{1 \leq \ell \leq L}$ , its multiresolution
expansion coefficients  $\{a_{n,k}\}_{n=-M_0,\dots,M_0}$  and  $\{b_{n,k}\}_{n=-M_0,\dots,M_0}$ , and its shape
function series  $\{s_{cn,k}\}_{n=-M_0,\dots,M_0}$  and  $\{s_{sn,k}\}_{n=-M_0,\dots,M_0}$  for  $1 \leq k \leq \bar{K}$ .
3 Initialize: let  $a_{n,k} = 0$ ,  $b_{n,k} = 0$ ,  $s_{cn,k} = 0$ ,  $s_{sn,k} = 0$ ,  $\mathcal{M}_{M_0}(f_k) = 0$  for all  $k$  and  $n$ ; let
 $c = \|f\|_{L^2}$ ; let  $e = 1$ ; let  $r^{(0)} = f$ .
4 for  $j = 1, 2, \dots, J_1$ , do
5   for  $n = 0, 1, -1, \dots, M_0, -M_0$  do
6      $[\{\bar{s}_k\}_{1 \leq k \leq \bar{K}}, \{\bar{f}_k\}_{1 \leq k \leq \bar{K}}] = RDBR(r^{(j-1)}, \{p_k\}_{1 \leq k \leq \bar{K}}, n, 1, \epsilon_2, J_2)$ .
7     for  $k = 1, \dots, \bar{K}$  do
8        $s_{cn,k} \leftarrow s_{cn,k} + \bar{s}_k$ .
9       Update  $\mathcal{M}_{M_0}(f_k)(t) \leftarrow \mathcal{M}_{M_0}(f_k)(t) + \bar{f}_k$ .
10      Compute  $r^{(j)} = r^{(j-1)} - \bar{f}_k$ .
11     if  $|n| > 0$  then
12        $[\{\bar{s}_k\}_{1 \leq k \leq \bar{K}}, \{\bar{f}_k\}_{1 \leq k \leq \bar{K}}] = RDBR(r^{(j)}, \{p_k\}_{1 \leq k \leq \bar{K}}, n, 0, \epsilon_2, J_2)$ .
13       for  $k = 1, \dots, \bar{K}$  do
14          $s_{sn,k} \leftarrow s_{sn,k} + \bar{s}_k$ .
15         Update  $\mathcal{M}_{M_0}(f_k)(t) \leftarrow \mathcal{M}_{M_0}(f_k)(t) + \bar{f}_k$ .
16         Update  $r^{(j)} \leftarrow r^{(j)} - \bar{f}_k$ .
17     If  $\|r^{(j)}\|_{L^2}/c \leq \epsilon_1$ , then break the for loop.
18     if  $\|r^{(j)}\|_{L^2}/c \geq e - \epsilon_1$  then
19       Break the for loop.
20     else
21        $e = \|r^{(j)}\|_{L^2}/c$ .
22 Let  $a_{n,k} = \|s_{cn,k}\|_{L^2}$  and  $s_{cn,k} = s_{cn,k}/a_{n,k}$  for all  $k$  and  $n$ .
23 Let  $b_{n,k} = \|s_{sn,k}\|_{L^2}$  and  $s_{sn,k} = s_{sn,k}/b_{n,k}$  for all  $k$  and  $n$ .

```

**Algorithm 3:** Multiresolution mode decomposition.

### 3.2 Convergence analysis

In this section, an asymptotic analysis on the convergence of the multiresolution mode decomposition in Algorithm 3 is introduced. We assume that the number of MIMFs is known exactly, i.e.,  $\bar{K} = K$ ; we also focus on the case when  $J_2 = 1$  in Algorithm 3; the analysis for other cases can be directly generalized. Similar to the theory in Section 2, the main analysis is to prove that the for-loop in Line 5 in Algorithm 3 is able to approximately identify the multiresolution expansion coefficients  $a_{n,k}$  and  $b_{n,k}$ , and estimate the shape function series  $s_{cn,k}$  and  $s_{sn,k}$  for  $1 \leq k \leq K$  and  $-M_0 \leq n \leq M_0$ . Though the Gauss-Seidel iteration is used in Algorithm 3 for faster convergence, we will show a looser bound of the convergence by using the Jacobi style iteration for the purpose of simplicity.

In the  $(j+1)$ th iteration in Line 4, the residual function is

$$\begin{aligned} r^{(j)}(t) &= \sum_{k=1}^K \sum_{n=-M_0}^{M_0} \cos(2\pi n\phi_k(t)) s_{cn,k}^{(j)}(2\pi N_k\phi_k(t)) \\ &\quad + \sum_{k=1}^K \sum_{n=-M_0}^{M_0} \sin(2\pi n\phi_k(t)) s_{sn,k}^{(j)}(2\pi N_k\phi_k(t)) + O(\epsilon), \end{aligned} \quad (33)$$

where the superscript  $(j)$  indicates the new target shape function series in the  $(j+1)$ th iteration of Line 4 in Algorithm 3, and the multiresolution expansion coefficients have been absorbed in the shape functions. Since we adopt the Jacobi style iteration in the analysis, the residual function in (33) remains the same throughout the  $j$ th iteration, i.e. Line 9, 10, 15, and 16 were postponed until the end of the for-loop in Line 5. For the  $k$ th component and frequency  $n$ , the diffeomorphism-based analysis with a constant amplitude function and the phase function  $N_k\phi_k(t)$  is applied to  $2 \cos(2\pi n\phi_k(t))r^{(j)}(t)$  and  $2 \sin(2\pi n\phi_k(t))r^{(j)}(t)$  to estimate the target shape functions  $s_{cn,k}^{(j)}$  and  $s_{sn,k}^{(j)}$ , respectively.

In particular,

$$\begin{aligned} &2 \cos(2\pi n\phi_k(t))r^{(j)}(t) \\ &= \sum_{k=1}^K \sum_{m=-M_0}^{M_0} 2 \cos(2\pi n\phi_k(t)) \cos(2\pi m\phi_k(t)) s_{cm,k}^{(j)}(2\pi N_k\phi_k(t)) \\ &\quad + \sum_{k=1}^K \sum_{m=-M_0}^{M_0} 2 \cos(2\pi n\phi_k(t)) \sin(2\pi m\phi_k(t)) s_{sm,k}^{(j)}(2\pi N_k\phi_k(t)) + O(\epsilon), \end{aligned} \quad (34)$$

where the term  $2 \cos(2\pi n\phi_k(t)) \cos(2\pi m\phi_k(t)) s_{cm,k}^{(j)}(2\pi N_k\phi_k(t))$  results in

$$s_{cn,k}^{(j)}(2\pi N_k\phi_k(t)) + \cos(4\pi n\phi_k(t)) s_{cn,k}^{(j)}(2\pi N_k\phi_k(t)).$$

$s_{cn,k}^{(j)}(2\pi N_k\phi_k(t))$  is the only term with a non-zero-mean amplitude function in 34. Following the same notations as in Section 2.2, let

$$\begin{aligned} h_{cn,k}^{(j)}(v) &= 2 \left( \cos(2\pi n\phi_k(t))r^{(j)} \right) \circ p_k^{-1}(v) \\ &= s_{cn,k}^{(j)}(2\pi v) + \kappa_{cn,k}^{(j)}(2\pi v) + O(\epsilon), \end{aligned}$$

where  $v = p_k(t) = N_k\phi_k(t)$  and  $\kappa_{cn,k}^{(j)}$  comes from other terms in (34). After the folding map

$$\tau : \left( v, h_{cn,k}^{(j)}(v) \right) \mapsto \left( \text{mod}(v, 1), h_{cn,k}^{(j)}(v) \right),$$

we have  $(x_\ell, y_\ell) = \tau(v_\ell, s_{cn,k}^{(j)}(2\pi v_\ell) + \kappa_{cn,k}^{(j)}(2\pi v_\ell) + O(\epsilon))$  for  $\ell = 1, \dots, L$  as  $L$  i.i.d. samples of a random vector  $(X_{cn,k}, Y_{cn,k}^{(j)})$ , where  $X_{cn,k} \in [0, 1]$ . Hence, the regression problem in (34) can be written as

$$\begin{aligned} s_{cn,k}^{R,(j+1)} &= \arg \min_{s: \mathbb{R} \rightarrow \mathbb{R}} \mathbb{E} \left\{ \left| s(2\pi X_{n,k}) - Y_{cn,k}^{(j)} \right|^2 \right\} \\ &= \arg \min_{s: \mathbb{R} \rightarrow \mathbb{R}} \mathbb{E} \left\{ \left| s(2\pi X_{n,k}) - (Y_{cn,k}^{(j)} - s_{cn,k}^{(j)}(2\pi X_{cn,k})) \right|^2 \right\} - s_{cn,k}^{(j)}, \end{aligned} \quad (35)$$

we have

$$s_{cn,k}^{R,(j+1)} = s_{cn,k}^{(j)} + s_{cn,k}^{E,(j)},$$

where

$$s_{cn,k}^{E,(j)}(2\pi x) := \mathbb{E} \{ Y_{cn,k}^{(j)} - s_{cn,k}^{(j)}(2\pi X_{cn,k}) | X_{cn,k} = x \} \neq 0 \quad (36)$$

due to the perturbation caused by  $\kappa_{cn,k}^{(j)}$ . In the next iteration, the target shape function  $s_{cn,k}^{(j+1)} = -s_{cn,k}^{E,(j)}$ . Hence, the key convergence analysis is to show that  $s_{cn,k}^{E,(j)}$  decays as  $j \rightarrow \infty$ .

Similarly, when the diffeomorphism-based regression is applied to  $s_{sn,k}^{(j)}$ , we are able to estimate  $s_{sn,k}^{(j)}$  approximately and the estimation error  $s_{sn,k}^{E,(j)}$  is the estimation target  $s_{sn,k}^{(j+1)} = -s_{sn,k}^{E,(j)}$  in the next iteration.

In what follows, we assume that an accuracy parameter  $\epsilon$  is fixed. Furthermore, suppose  $f(t) \in \mathcal{WS}(M_0, M, N, K, h, \beta, \gamma, \epsilon)$ , and all shape functions are in the space  $\mathcal{L}^C$ . Under these conditions, all regression functions  $s_{cn,k}^{(j)} \in \mathcal{L}^C$ ,  $s_{sn,k}^{(j)} \in \mathcal{L}^C$ , and have bounded  $L^\infty$  norm depending only on  $M$  and  $K$ . By Algorithm 2, we have the nice and key conditions that  $\int_0^1 s_{cn,k}^{(j)}(2\pi t) dt = 0$  and  $\int_0^1 s_{sn,k}^{(j)}(2\pi t) dt = 0$  at each iteration for all  $n, k$  and  $j$ . Note that  $\text{Var}(Y_{cn,k}^{(j)} | X_{cn,k} = x)$  and  $\text{Var}(Y_{sn,k}^{(j)} | X_{sn,k} = x)$  are bounded by a constant depending only on  $M$  and  $K$  as well. For the fixed  $\epsilon$  and  $C$ , there exists  $h_0(\epsilon, C)$  such that  $C^2 h^2 < \epsilon^2$  if  $0 < h < h_0$ . By the abuse of notation,  $O(\epsilon)$  is used instead of  $Ch$  later. By Theorem 2.3, for the fixed  $\epsilon, M_0, M, K, C$ , and  $h$ , there exists  $L_0(\epsilon, M_0, M, K, C, h)$  such that the  $L^2$  error of the partition-based regression is bounded by  $\epsilon^2$ . In what follows,  $h$  is smaller than  $h_0$ ,  $L$  is larger than  $L_0$ , and hence all estimated regression functions approximate the ground truth regression function with an  $L^2$  error of order  $\epsilon$ . Under these conditions and assumptions,  $s_{cn,k}^{E,(j)}$  and  $s_{sn,k}^{E,(j)}$  are shown to decay to  $O(\epsilon)$  as  $j \rightarrow \infty$ , as long as  $L$  and  $N$  are sufficiently large, and the decay rate will be estimated.

**Theorem 3.3.** *(Convergence of the multiresolution mode decomposition) Under the conditions listed in the paragraph immediately preceding this theorem, suppose  $N_k$  and  $\phi_k(t)$  are known for all  $k$ , and  $J_2 = 1$  in Algorithm 3, as long as  $L$  and  $N$  are sufficiently large, we have*

$$\|s_{cn,k}^{E,(j)}\|_{L^2} \leq O(c_0 \epsilon + (\beta(2M_0 + 1)(K - 1))^j),$$

and

$$\|s_{sn,k}^{E,(j)}\|_{L^2} \leq O(c_0 \epsilon + (\beta(2M_0 + 1)(K - 1))^j)$$

for all  $j \geq 0$  and  $1 \leq k \leq K$ , where  $c_0 = \frac{1}{1 - \beta(2M_0 + 1)(K - 1)}$  is a constant number,  $s_{cn,k}^{E,(j)}$  is defined in Equation (36) and  $s_{sn,k}^{E,(j)}$  is defined similarly.

*Proof.* We only prove the case when  $K = 2$ ; the proof for a general case follows a similar discussion of the proof of Theorem 2.5.

Recall notations in Definition 3.2. When the partition-based regression method is applied to solve the regression problem in (35) to obtain the approximate regression function  $s_{cn,1}^{P,(j)}(2\pi x)$ , we have

$$s_{cn,1}^{P,(j)}(2\pi x) = \frac{\sum_{n=0}^{N^h-1} \left( \kappa_{cn,1}^{(j)}(2\pi t_n^h) + O(\epsilon) \right) D_h^{12}(m, n)}{D_h^1(m)},$$

for  $x \in [t_m^h, t_m^h + h)$ , where  $O(\epsilon)$  comes from the approximation of  $\mathcal{L}^C$  functions using the values on grid points  $t_n^h$ , and the  $O(\epsilon)$  term in the residual function in (33). The following argument is similar to that for Lemma 3.3 in [31]. It is easy to check that

$$|s_{cn,1}^{P,(j)}(2\pi x)| \leq O(\epsilon) + \frac{\sum_{n=0}^{N^h-1} \bar{\kappa}_{cn,1}^{(j)}(2\pi t_n^h) (D_h^{12}(m, n) - \gamma)}{D_h^1(m)} + \gamma \frac{\sum_{n=0}^{N^h-1} \bar{\kappa}_{cn,1}^{(j)}(2\pi t_n^h)}{D_h^1(m)}, \quad (37)$$

where

$$\bar{\kappa}_{cn,k}^{(j)}(2\pi v) := \kappa_{cn,k}^{(j)}(2\pi v) - \dot{\kappa}_{cn,k}^{(j)}(2\pi v),$$

and

$$\begin{aligned} \dot{\kappa}_{cn,k}^{(j)}(2\pi v) &:= \sum_{m=-M_0}^{M_0} 2 \cos(2\pi \frac{n}{N_k} v) \cos(2\pi \frac{m}{N_k} v) s_{cm,k}^{(j)}(2\pi v) \\ &+ \sum_{m=-M_0}^{M_0} 2 \cos(2\pi \frac{n}{N_k} v) \sin(2\pi \frac{m}{N_k} v) s_{sm,k}^{(j)}(2\pi v) - s_{cn,k}^{(j)}(2\pi v) \\ &= \sum_{m=-M_0}^{M_0} \left( \cos(2\pi \frac{m+n}{N_k} v) + \cos(2\pi \frac{n-m}{N_k} v) \right) s_{cn,k}^{(j)}(2\pi v) \\ &+ \sum_{m=-M_0}^{M_0} \left( \sin(2\pi \frac{n+m}{N_k} v) + \sin(2\pi \frac{m-n}{N_k} v) \right) s_{sn,k}^{(j)}(2\pi v) - s_{cn,k}^{(j)}(2\pi v), \end{aligned}$$

since

$$\frac{\sum_{n=0}^{N^h-1} \dot{\kappa}_{cn,1}^{(j)}(2\pi t_n^h) (D_h^{12}(m, n))}{D_h^1(m)} = O(\epsilon), \quad (38)$$

which is due to the fact that the oscillation in the smooth amplitudes ( $\cos(2\pi \frac{\pm n \pm m}{N_k} v)$  and  $\sin(2\pi \frac{\pm n \pm m}{N_k} v)$ ) can cancel out the summation in 38.

The inequalities in (37) implies that

$$\begin{aligned} \|s_{cn,1}^{P,(j)}\|_{L^2} &\leq O(\epsilon) + \left( \sum_{n=0}^{N^h-1} \left( \bar{\kappa}_{cn,1}^{(j)}(2\pi t_n^h) \right)^2 h \right)^{1/2} \left( \sum_{m=0}^{N^h-1} \left( \sum_{n=0}^{N^h-1} \left( \frac{D_h^{12}(m, n) - \gamma}{D_h^1(m)} \right)^2 \right) \right)^{1/2} \\ &+ \gamma \frac{\sum_{n=0}^{N^h-1} \bar{\kappa}_{cn,1}^{(j)}(2\pi t_n^h)}{D_h^1(m)} \end{aligned} \quad (39)$$

by the triangle inequality and Hölder's inequality. Note that

$$\begin{aligned} \left( \sum_{n=0}^{N^h-1} \left( \bar{\kappa}_{cn,k}^{(j)}(2\pi t_n^h) \right)^2 h \right)^{1/2} &\leq \sum_{\ell \neq k} \sum_{m=-M_0}^{M_0} \left( \sum_{n=0}^{N^h-1} \left( s_{cm,\ell}^{(j)}(2\pi t_n^h) \right)^2 h \right)^{1/2} \\ &\quad + \sum_{\ell \neq k} \sum_{m=-M_0}^{M_0} \left( \sum_{n=0}^{N^h-1} \left( s_{sm,\ell}^{(j)}(2\pi t_n^h) \right)^2 h \right)^{1/2} \end{aligned}$$

by the triangle inequality. Since all shape functions are in  $\mathcal{L}^C$ , we have

$$\left( \sum_{n=0}^{N^h-1} \left( s_{cm,\ell}^{(j)}(2\pi t_n^h) \right)^2 h \right)^{1/2} = \|s_{cm,\ell}^{(j)}\|_{L^2} + O(\epsilon),$$

and

$$\left( \sum_{n=0}^{N^h-1} \left( s_{sm,\ell}^{(j)}(2\pi t_n^h) \right)^2 h \right)^{1/2} = \|s_{sm,\ell}^{(j)}\|_{L^2} + O(\epsilon)$$

for all  $\ell$  and  $m$ . Hence,

$$\begin{aligned} &\left( \sum_{n=0}^{N^h-1} \left( \bar{\kappa}_{cn,k}^{(j)}(2\pi t_n^h) \right)^2 h \right)^{1/2} \\ &\leq O(\epsilon) + (2M_0 + 1)(K - 1) \max_{1 \leq k \leq K, -M_0 \leq m \leq M_0} \left( \|s_{cm,k}^{(j)}\|_{L^2} + \|s_{sm,k}^{(j)}\|_{L^2} \right). \end{aligned} \quad (40)$$

Since phase functions are in  $\mathcal{WD}(M, N, h, \beta, \gamma)$ ,

$$\left( \sum_{m=0}^{N^h-1} \left( \sum_{n=0}^{N^h-1} \left( \frac{D_h^{ki}(m, n) - \gamma}{D_h^k(m)} \right)^2 \right) \right)^{1/2} \leq \beta < 1. \quad (41)$$

Hence, by (39), (40), and (41), it holds that

$$\begin{aligned} \|s_{cn,1}^{P,(j)}\|_{L^2} &\leq O(\epsilon) + \gamma \frac{\sum_{n=0}^{N^h-1} \bar{\kappa}_{cn,1}^{(j)}(2\pi t_n^h)}{D_h^1(m)} \\ &\quad + \beta(2M_0 + 1)(K - 1) \max_{1 \leq k \leq K, -M_0 \leq m \leq M_0} \left( \|s_{cm,k}^{(j)}\|_{L^2} + \|s_{sm,k}^{(j)}\|_{L^2} \right). \end{aligned} \quad (42)$$

Note that  $D_h^k(m) \geq N^h \gamma = \gamma/h$ . Hence,  $\frac{\gamma}{D_h^k(m)} \leq h$  and

$$\gamma \frac{\sum_{n=0}^{N^h-1} \bar{\kappa}_{cn,1}^{(j)}(2\pi t_n^h)}{D_h^1(m)} \leq h \sum_{n=0}^{N^h-1} \bar{\kappa}_{cn,1}^{(j)}(2\pi t_n^h). \quad (43)$$

Since  $s_{cn,k}^{(j)} \in \mathcal{L}^C$ ,  $s_{sn,k}^{(j)} \in \mathcal{L}^C$ ,  $\int_0^1 s_{cn,k}^{(j)}(2\pi t) dt = 0$ , and  $\int_0^1 s_{sn,k}^{(j)}(2\pi t) dt = 0$  for all  $n$  and  $k$ , by the stationary phase approximation, we have

$$h \sum_{n=0}^{N^h-1} \bar{\kappa}_{cn,1}^{(j)}(2\pi t_n^h) \leq O(\epsilon), \quad (44)$$

as long as the number of grid points,  $L$ , and the lower bound of the oscillation frequency,  $N$ , are sufficiently large. In sum, by (42), (43), and (44), we have

$$\begin{aligned} \|s_{cn,1}^{P,(j)}\|_{L^2} &\leq O(\epsilon) \\ &+\beta(2M_0+1)(K-1)\max_{1\leq k\leq K,-M_0\leq m\leq M_0}\left(\|s_{cm,k}^{(j)}\|_{L^2}+\|s_{sm,k}^{(j)}\|_{L^2}\right). \end{aligned} \quad (45)$$

In fact, following the arguments for (45), we can bound  $\|s_{cn,k}^{P,(j)}\|_{L^2}$  and  $\|s_{sn,k}^{P,(j)}\|_{L^2}$  for all  $k$  with the same bound as in (45).

By the conditions just listed immediately before Theorem 3.3, and the definition in Equation (36), we have

$$\begin{aligned} \|s_{cn,k}^{E,(j)}\|_{L^2} &\leq O(\epsilon)+\|s_{cn,k}^{P,(j)}\|_{L^2}\leq \\ O(\epsilon)+\beta(2M_0+1)(K-1)\max_{1\leq k\leq K,-M_0\leq m\leq M_0}\left(\|s_{cm,k}^{(j)}\|_{L^2}+\|s_{sm,k}^{(j)}\|_{L^2}\right), \end{aligned} \quad (46)$$

and

$$\begin{aligned} \|s_{sn,k}^{E,(j)}\|_{L^2} &\leq O(\epsilon)+\|s_{sn,k}^{P,(j)}\|_{L^2}\leq \\ O(\epsilon)+\beta(2M_0+1)(K-1)\max_{1\leq k\leq K,-M_0\leq m\leq M_0}\left(\|s_{cm,k}^{(j)}\|_{L^2}+\|s_{sm,k}^{(j)}\|_{L^2}\right). \end{aligned} \quad (47)$$

By Equation (46), (47), and mathematical induction, it is easy to show that

$$\|s_{cn,k}^{E,(j)}\|_{L^2}\leq O(c_0\epsilon+(\beta(2M_0+1)(K-1))^j),$$

and

$$\|s_{sn,k}^{E,(j)}\|_{L^2}\leq O(c_0\epsilon+(\beta(2M_0+1)(K-1))^j)$$

for all  $j\geq 0$  and  $1\leq k\leq K$ , where  $c_0=\frac{1}{1-\beta(2M_0+1)(K-1)}$  is a constant number coming from the geometric sequence due to the summation of the  $O(\epsilon)$  term for all  $j\geq 0$ .

The above proof is just for  $K=2$ . To care the general case for  $K>2$ , the only difference is the presence of more terms in the inequality estimations.  $\square$

Theorem 3.3 shows that the regression function in each iteration of Line 4 in Algorithm 3 decays, if  $J_2=1$  and  $\beta(2M_0+1)(K-1)<1$ , in the  $L^2$  sense up to a fixed accuracy parameter as the iteration number becomes large. Hence, the recovered shape function converges and the residual decays up to a fixed accuracy parameter. For a general case when  $J_2>1$ , the convergence of Algorithm 3 is obvious following Theorem 3.3. It is tedious to compare the convergence rate for different  $J_2$ , since it depends on how fast the multiresolution expansion coefficients  $a_{cn,k}$ ,  $a_{sn,k}$ ,  $b_{cn,k}$ , and  $b_{sn,k}$  decay in  $n$ . In general, the faster the coefficients decay in  $n$ , the larger  $J_2$  should be used. Another immediate result of Theorem 3.3 is that, the correlation of MIMF's with well-differentiated phase functions is asymptotically zero in the sense of recursive diffeomorphism-based regression. In other words, when the MIMF and MMD model is applied to analyze a time series, the resulting representation by MIMF's is unique. The robustness of Algorithm 3.3 is an immediate result following the proof of Theorem 2.6 and 2.7.

## 4 Numerical Example

In this section, some numerical examples of synthetic and real data are provided to support the multiresolution mode decomposition (MMD) model and Algorithm 3. We apply the least squares spline regression method with free knots in [36] to solve all the regression problems in this paper. The implementation of the regression method is available online<sup>6</sup>. In all synthetic examples, we assume the fundamental instantaneous phases are known and only focus on verifying the theory in Section 3. In real examples, we apply the one-dimensional synchrosqueezed wave packet transform (SSWPT) to estimate instantaneous phases as inputs of the multiresolution mode decomposition in Algorithm 3. The implementation of the SSWPT is available in SynLab<sup>7</sup>, while the code for the MMD is available online in a MATLAB package named DeCom<sup>8</sup>.

Before presenting results, we would like to summarize the main parameters in the above packages and in Algorithm 3. In the spline regression with free knots, main parameters are

- $nk$ : the number of free knots;
- $krf$ : the knot removal factor, a number quantifying how likely a free knot would be removed;
- $ord$ : the highest degree of spline polynomials.

In SynLab, main parameters are

- $s$ : a geometric scaling parameter;
- $rad$ : the support size of the mother wave packet in the Fourier domain;
- $red$ : a redundancy parameter, the number of frames in the wave packet transform;
- $\epsilon_{sst}$ : a threshold for the wave packet coefficients.

In Algorithm 3, main parameters are

- $J_1$ : the maximum number of iterations allowed in Algorithm 3;
- $J_2$ : the maximum number of iterations allowed in Algorithm 2;
- $M_0$ : the bandwidth parameter;
- $\epsilon_1 = \epsilon_2 = \epsilon$ : the accuracy parameter.

For the purpose of convenience, the synthetic data is defined in  $[0, 1]$  and sampled on a uniform grid. All these parameters in different examples are summarized in Table 1.

In the noisy synthetic examples, Gaussian random noise with a distribution  $\mathcal{N}(0, \sigma^2)$  is used. The signal-to-noise ration (SNR) of  $f(t)$  is defined as

$$\text{SNR}[dB] = 10 \log_{10} \left( \frac{\|f\|_{L^2}}{\sigma^2} \right) \quad (48)$$

where  $\sigma^2$  is the variation of the noise.

<sup>6</sup> Available at <https://www.mathworks.com/matlabcentral/fileexchange/25872-free-knot-spline-approximation>.

<sup>7</sup> Available at <https://github.com/HaizhaoYang/SynLab>.

<sup>8</sup> Available at <https://github.com/HaizhaoYang/DeCom>.



figure	$nk$	$krf$	$ord$	$s$	$rad$	$red$	$\epsilon_{sst}$	$J_1$	$J_2$	$M_0$	$\epsilon$	$L$
7, 8, 9	20	1.0001	3	0.5	1.5	8	1e-3	200	10	20	1e-6	$2^{16}$
10, 11, 12	20	1.0001	3	0.5	1	8	1e-3	200	10	20	1e-6	4000
14, 15	20	1.0001	3	–	–	–	–	200	10	10	1e-6	$2^{15}$
16, 17, 18	20	1.0001	3	0.5	1.5	8	1e-3	200	10	40	1e-6	$2^{16}$

Table 1: Parameters in the spline regression, SynLab, and Algorithm 3. The notation “–” means the corresponding parameter is not used.

#### 4.1 Synthetic multiresolution mode decomposition

In this section, clean and noisy synthetic time series are provided to demonstrate the effectiveness and the robustness of the multiresolution mode decomposition in Algorithm 3.

**The first synthetic example.** We consider a simple case when the signal has two MIMFs with ECG shape functions. In particular, we generate MIMFs such that  $s_{sn,k} = s_{cn,k} = s_k$  for all  $n \geq 0$  and  $k$ , and  $s_{sn,k} = s_{cn,k} = 0$  for all  $n < 0$  and  $k$ , resulting in an example that can be either considered as a GMD or an MMD problem. Through this example we see that Algorithm 3 for MMD can also be applied to solve the GMD problem. For example, we consider a signal of the form

$$f(t) = f_1(t) + f_2(t) + ns, \quad (49)$$

where  $ns$  denotes Gaussian random noise with a distribution  $\mathcal{N}(0, \sigma^2)$ ,

$$f_1(t) = \alpha_1(\phi_1(t))s_1(300\pi\phi_1(t)), \quad (50)$$

$$f_2(t) = \alpha_2(\phi_2(t))s_2(440\pi\phi_2(t)), \quad (51)$$

$$\alpha_1(t) = 1 + 0.2 \cos(2\pi t) + 0.1 \sin(2\pi t),$$

$$\alpha_2(t) = 1 + 0.1 \cos(2\pi t) + 0.2 \sin(2\pi t),$$

$$\phi_1(t) = t + 0.006 \sin(2\pi t),$$

and

$$\phi_2(t) = t + 0.006 \cos(2\pi t).$$

$s_1(2\pi t)$  and  $s_2(2\pi t)$  are generalized shape functions defined in  $[0, 1]$  as shown in Figure 9. In the noiseless example when  $ns = 0$ , we apply the MMD in Algorithm 3 with the known instantaneous phases mentioned just above to estimate the multiresolution expansion coefficients and the shape functions series. Note that it is sufficient to show the estimation accuracy of the product of the multiresolution expansion coefficient and its corresponding shape function; the products of the first five leading expansion coefficients and shape functions are shown in Figure 10. The estimation errors are very small; the estimated results and the ground truth are almost indistinguishable. The numerical results of a noisy signal with  $ns = \mathcal{N}(0, 2.25)$  are shown in Figure 11. Even if the SNR for the leading terms  $\mathcal{M}_0(f_1)(t)$  and  $\mathcal{M}_0(f_2)(t)$  are near  $-10$ , the estimation  $a_{0,1}s_{c0,1}(t)$  is almost equal to the ground truth  $s_1(t)$ , and the estimation  $a_{0,2}s_{c0,2}(t)$  is almost equal to the ground truth  $s_2(t)$ . The SNR for the other terms are even much smaller (less than  $-20$ ). However, as shown in Figure 11, the estimation of other terms are still reasonable, capturing the main trend of the shape functions. Therefore, Algorithm 3 is quite robust against noise perturbation.

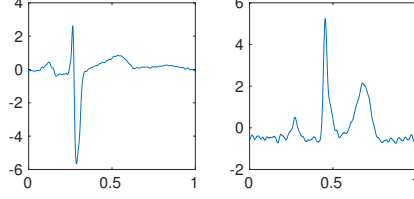


Figure 9: Shape function  $s_1(2\pi t)$  in (50) and  $s_2(2\pi t)$  in (51).

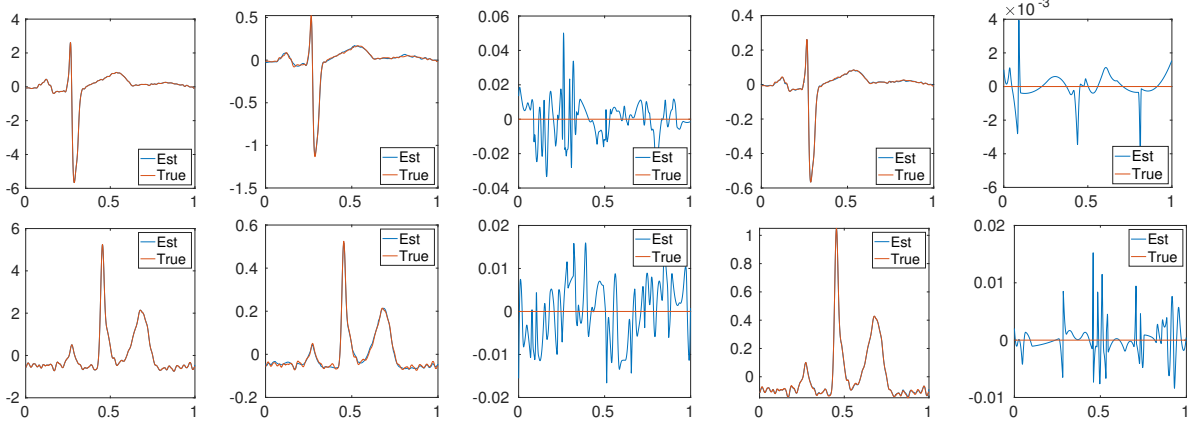


Figure 10: Top row: estimated shape functions  $a_{0,1}s_{c0,1}(2\pi t)$ ,  $a_{1,1}s_{c1,1}(2\pi t)$ ,  $a_{-1,1}s_{c-1,1}(2\pi t)$ ,  $b_{1,1}s_{s1,1}(2\pi t)$ , and  $b_{-1,1}s_{s-1,1}(2\pi t)$  of  $f_1(t)$  in (50). Bottom row: estimated shape functions  $a_{0,2}s_{c0,2}(2\pi t)$ ,  $a_{1,2}s_{c1,2}(2\pi t)$ ,  $a_{-1,2}s_{c-1,2}(2\pi t)$ ,  $b_{1,2}s_{s1,2}(2\pi t)$ , and  $b_{-1,2}s_{s-1,2}(2\pi t)$  of  $f_2(t)$  in (51).

## 4.2 Multiresolution intrinsic mode functions in real data

In this section, time series in real application are provided to support the model of multiresolution intrinsic mode functions (MIMFs). The first example is an ECG record from a normal subject and the second example is a motion-contaminated ECG record. The reader is referred to <https://www.physionet.org/physiobank/database/> for more details about the ECG data. We compute the band-limited multiresolution approximations of the first example and visualize them in Figure 12, 13, and 14; the band-limited multiresolution approximations of the second example are plotted in Figure 15, 16, and 17. Note that when the bandwidth of the multiresolution approximation increases, the approximation error decreases, and finer variation of the time series can be captured. These observations support the model of MIMF as a superposition of several oscillatory components. In particular, the results in Figure 13 shows that, as the bandwidth increases, the variation of the time-varying amplitude of the signal has been captured in the high frequency components of the MIMF. Figure 14 and 17 show the first five shape functions of these two examples, respectively; all shape functions vary a lot at different level of resolution. The actual time-varying shape of an ECG signals we see in the raw data is not exactly any single shape function in the shape function series; they are actually the results of all shape functions in the shape function series. This completes the validation of the proposed MIMF model.

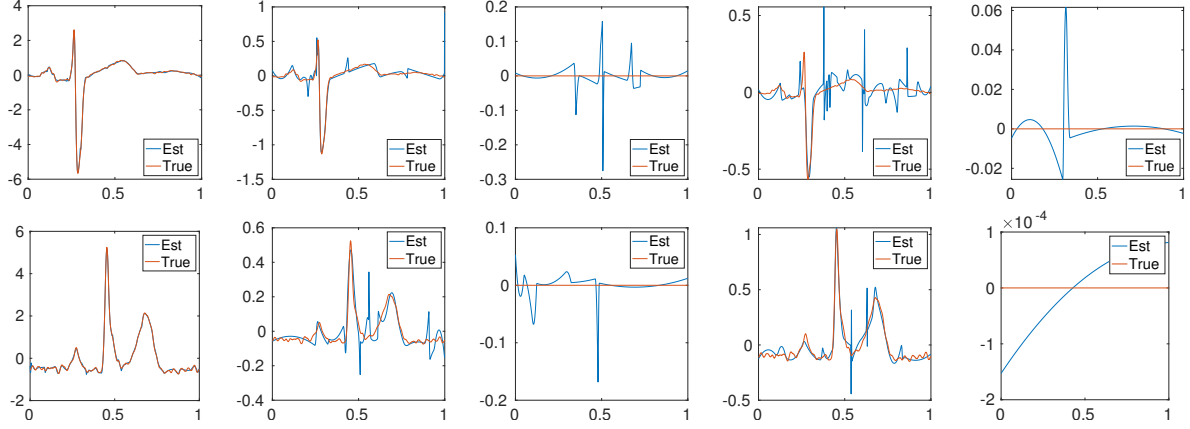


Figure 11: Top row: estimated shape functions  $a_{0,1}s_{c0,1}(2\pi t)$ ,  $a_{1,1}s_{c1,1}(2\pi t)$ ,  $a_{-1,1}s_{c-1,1}(2\pi t)$ ,  $b_{1,1}s_{s1,1}(2\pi t)$ , and  $b_{-1,1}s_{s-1,1}(2\pi t)$  of  $f_1(t) + ns$ , where  $f_1(t)$  is given in (50) and  $ns$  is Gaussian random noise with variance 2.25. Bottom row: estimated shape functions  $a_{0,2}s_{c0,2}(2\pi t)$ ,  $a_{1,2}s_{c1,2}(2\pi t)$ ,  $a_{-1,2}s_{c-1,2}(2\pi t)$ ,  $b_{1,2}s_{s1,2}(2\pi t)$ , and  $b_{-1,2}s_{s-1,2}(2\pi t)$  of  $f_2(t) + ns$ , where  $f_2(t)$  is given in (51) and  $ns$  is Gaussian random noise with variance 2.25.

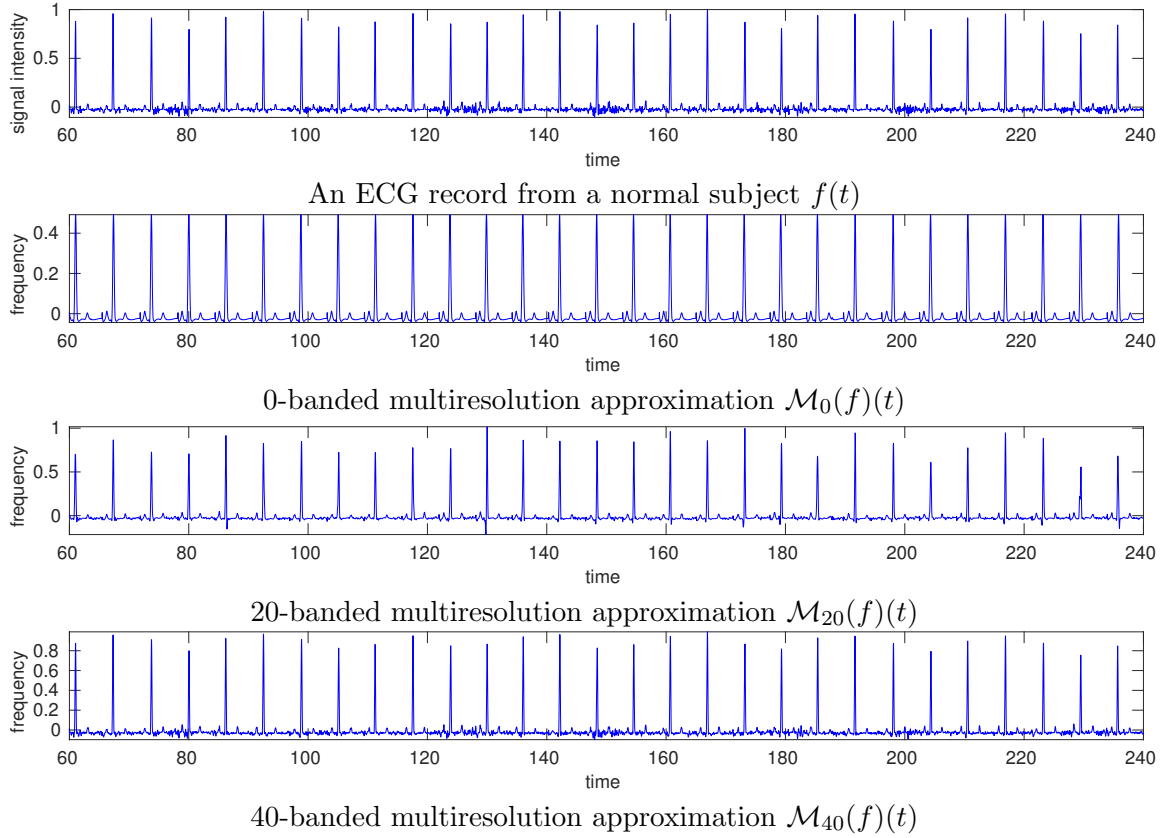


Figure 12: Multiresolution approximations of an ECG record from a normal subject.

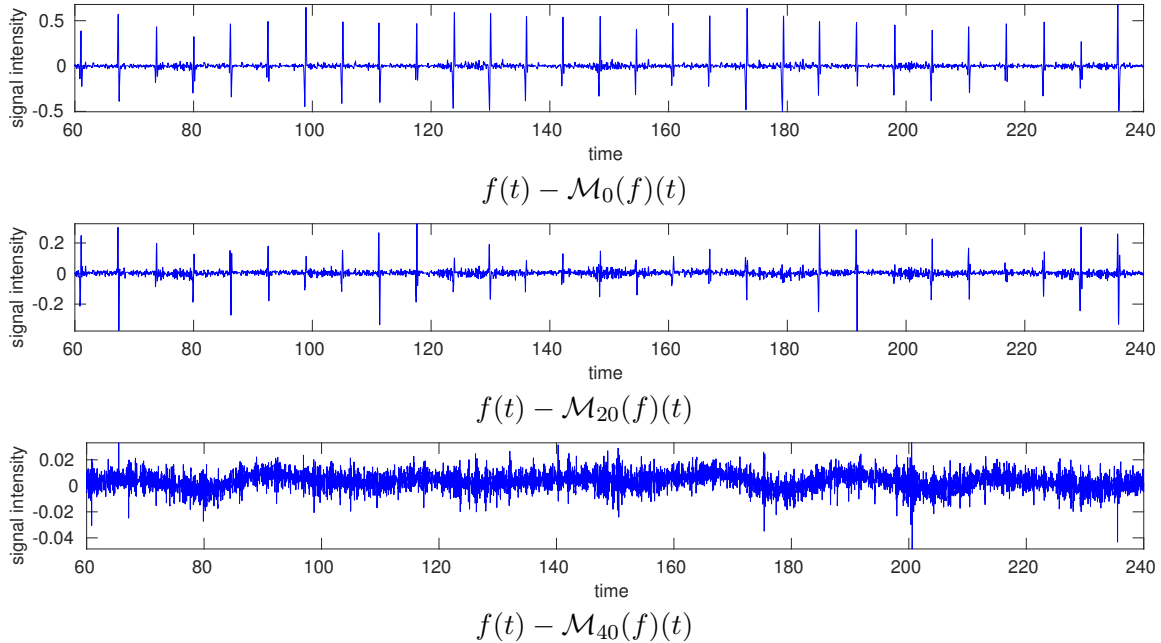


Figure 13: The residual of the multiresolution approximations of an ECG record from a normal subject in Figure 12.

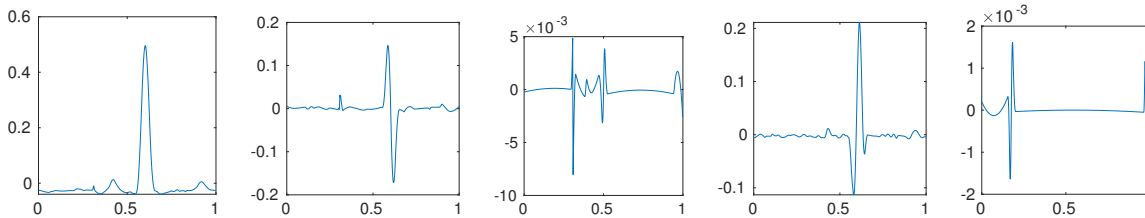


Figure 14: Estimated shape functions  $a_0s_{c0}(t)$ ,  $a_1s_{c1}(t)$ ,  $a_{-1}s_{c-1}(t)$ ,  $b_1s_{s1}(t)$ , and  $b_{-1}s_{s-1}(t)$  for the ECG signal in Figure 12.

### 4.3 Real example of multiresolution mode decomposition

In this section, we provide a real example to demonstrate the application of the multiresolution mode decomposition. This is an example of photoplethysmography (PPG)<sup>9</sup> that contains the hemodynamical information as well as the respiration information. Hence, the PPG signal essentially contains two MIMFs. In this example, the instantaneous frequencies and phases are not known and they are estimated via the synchrosqueezed transform in [28]. Figure 18 shows the estimated fundamental instantaneous frequencies of the respiratory and cardiac cycles. Inputting their corresponding instantaneous phases into the MMD algorithm, the PPG signal is separated into a respiratory MIMF and a cardiac MIMF as shown in Figure 19; their leading multiresolution shape functions are shown in Figure 20.

The last two panels of Figure 19 shows that the PPG signal has been completely separated into two MIMFs; the residual signal only contains noise, a smooth trend, and some

<sup>9</sup>From <http://www.capnobase.org>.

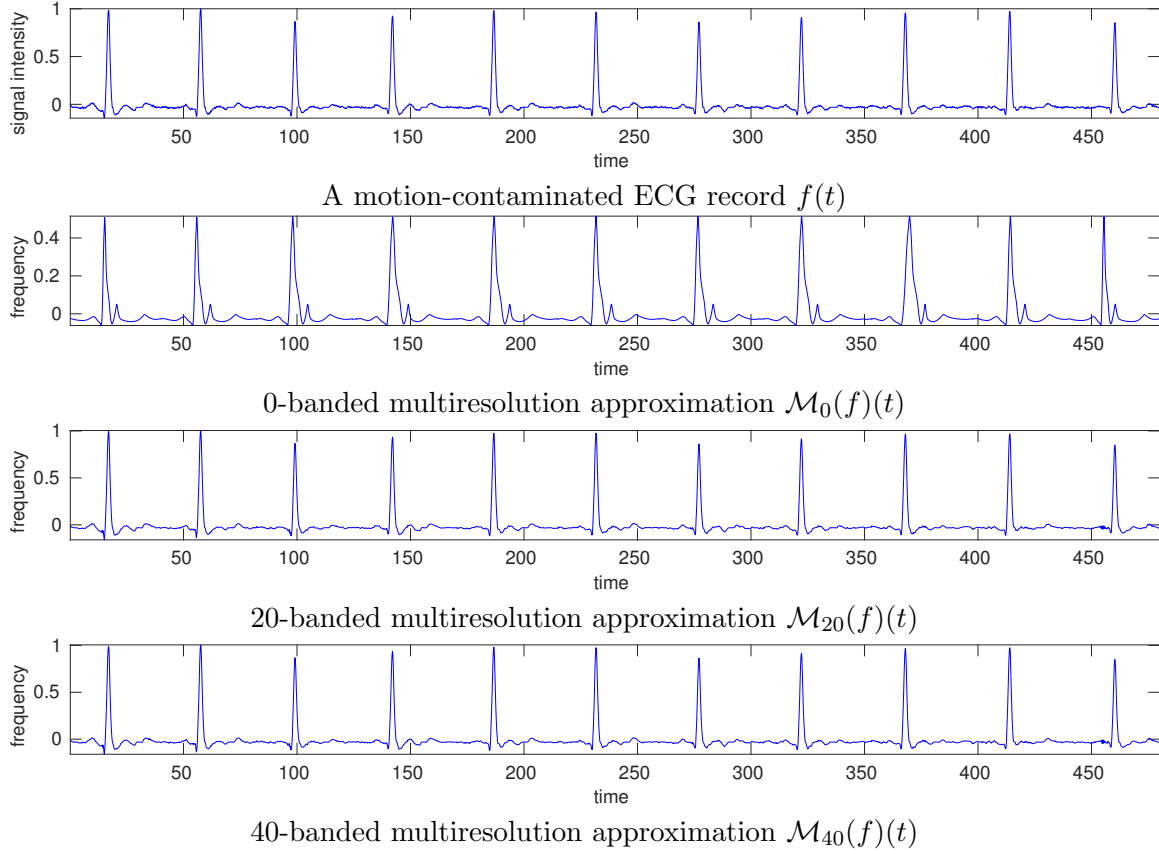


Figure 15: Multiresolution approximations of a motion-contaminated ECG record.

sharp changes that are not correlated to the oscillation in MIMFs. It is worth to exploring the application of MIMFs and MMD for fault detection in health data. For example, the detection of abnormal ECG waveforms is important to cardiac disease diagnosis [32, 33]; the abnormality can be identified by detecting the sharp changes in the residual signal that are not correlated to the normal oscillation pattern. The second panel shows that the MIMF model can characterize time-varying shape functions, while the third panel shows that the MIMF model can capture the time-varying amplitude functions.

## 5 Conclusion

This paper proposed the multiresolution intrinsic mode function (MIMF) as a new model to simulate oscillatory time series with time-varying amplitudes and shapes. In the case of a superposition of several MIMFs, a novel multiresolution decomposition algorithm based on the idea of recursive diffeomorphism-based regression is proposed to separate the signal into individual MIMFs. The convergence and the robustness of recursive scheme has been theoretically and numerically proved. The application of MIMFs and MMD is not limited to decomposing signals into several components with well-differentiated instantaneous phase functions; they can also be used for time series denoising due to the robustness of the MMD algorithm; the multiresolution expansion coefficients and shape function series can also provide better features for adaptive time series analysis than traditional Fourier analysis

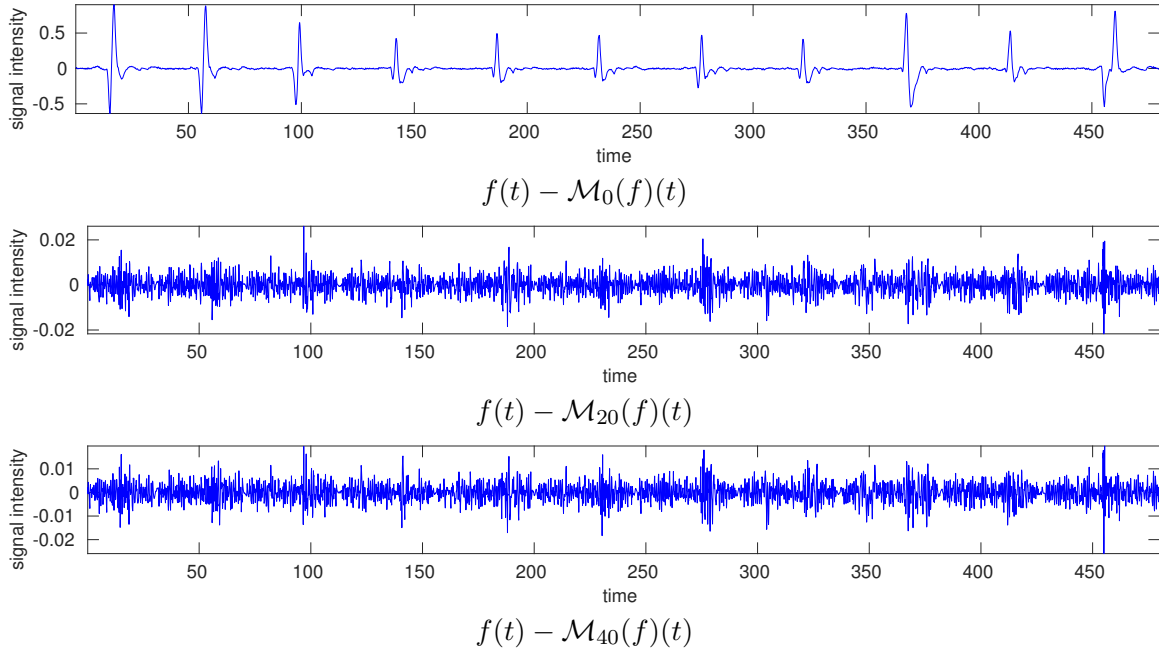


Figure 16: The residual of the multiresolution approximations of a motion-contaminated ECG record in Figure 15.

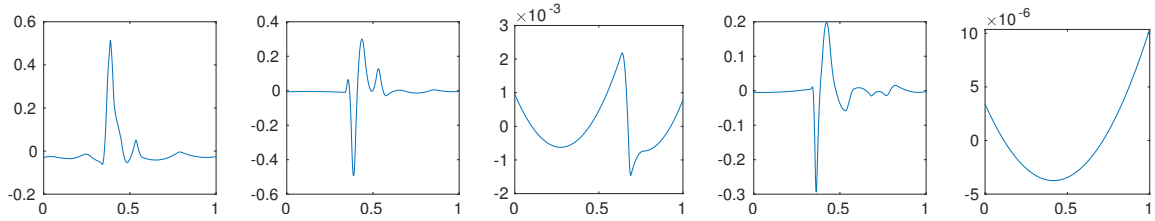


Figure 17: Estimated shape functions  $a_0 s_{c0}(t)$ ,  $a_1 s_{c1}(t)$ ,  $a_{-1} s_{c-1}(t)$ ,  $b_1 s_{s1}(t)$ , and  $b_{-1} s_{s-1}(t)$  for the ECG signal in Figure 15.

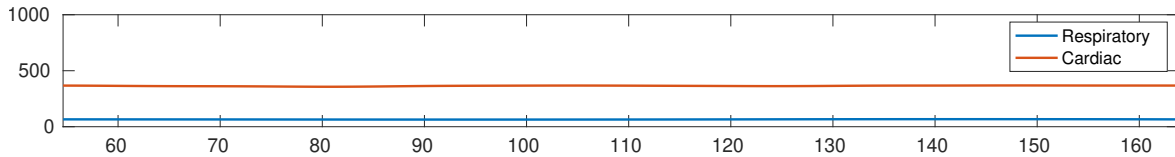


Figure 18: Estimated fundamental instantaneous frequencies of the real PPG signal in the first panel of Figure 19 by the synchrosqueezed transform.

and wavelet analysis.

The MIMF model is inspired by the visual observation of nonlinear and non-stationary oscillatory time series with time-dependent amplitudes and shape functions. Theoretically it has been proved that the correlation of MIMF's with well-differentiated phase functions is asymptotically zero in the sense of recursive diffeomorphism-based regression. In other

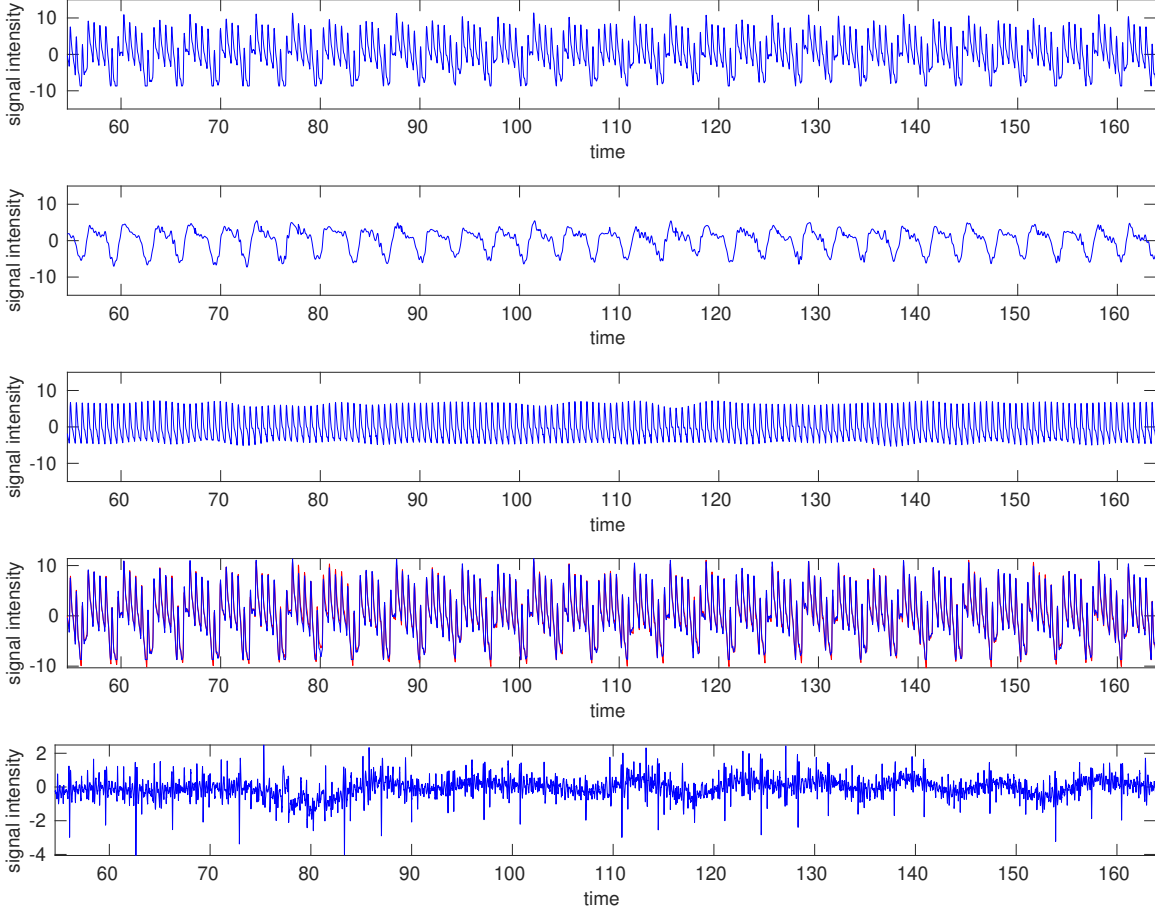


Figure 19: First panel: the raw PPG signal  $f(t)$ . Second panel: the respiratory MIMF  $f_1(t)$ . Third panel: the cardiac MIMF  $f_2(t)$ . Fourth panel: the summation of the respiratory and cardiac MIMFs  $f_1(t) + f_2(t)$  (red) compared to the raw PPG signal  $f(t)$  (blue). The fifth panel: the residual signal  $f(t) - f_1(t) - f_2(t)$ .

words, when the MIMF and MMD model is applied to analyze a time series, the resulting representation by MIMF's is unique. Numerically, it is shown that MIMF's match real oscillatory time series well, especially those from health data. To theoretically validate the MIMF model for oscillatory time series, a more fundamental but challenging approach is to understand the governing dynamics that generate the time series. For example, despite several attempts [37] over decades still it has not been successful to establish dynamical models faithfully describing real ECG signals – not to mention the variability among different individuals. The MIMF model may serve as a prototype that separates the variability of complicated signals into simpler components, the generator of whose dynamics may be more accessible.

Though adaptive time-series analysis is the main motivation of MIMF and MMD discussed in this paper, these models and algorithms can be naturally extended to higher dimensional spaces. Applications include atomic crystal images in physics [4, 5], art investigation [8, 9], geology [10, 11, 12], imaging [13], etc. In higher dimensional spaces, the computational efficiency is a crucial issue. There are mainly two directions for future

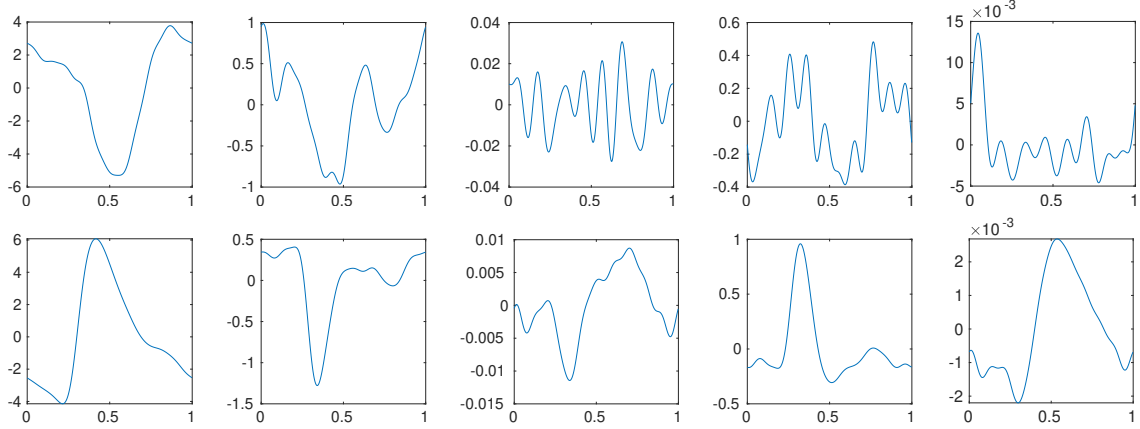


Figure 20: Top row: estimated shape functions  $a_{0,1}s_{c0,1}(2\pi t)$ ,  $a_{1,1}s_{c1,1}(2\pi t)$ ,  $a_{-1,1}s_{c-1,1}(2\pi t)$ ,  $b_{1,1}s_{s1,1}(2\pi t)$ , and  $b_{-1,1}s_{s-1,1}(2\pi t)$  for the respiratory MIMF. Bottom row: estimated shape functions  $a_{0,2}s_{c0,2}(2\pi t)$ ,  $a_{1,2}s_{c1,2}(2\pi t)$ ,  $a_{-1,2}s_{c-1,2}(2\pi t)$ ,  $b_{1,2}s_{s1,2}(2\pi t)$ , and  $b_{-1,2}s_{s-1,2}(2\pi t)$  for the cardiac MIMF.

works for fast algorithms for MMD. A natural idea is to develop fast regression techniques to reduce the time for each iteration in the recursive scheme in MMD; a more challenging question is to develop new recursive schemes to update all the multiresolution expansion coefficients and shape function series simultaneously, instead of updating them one by one in each iteration in Algorithm 2 and 3.

**Acknowledgments.** H.Y. thanks Ingrid Daubechies for her fruitful discussion.

## References

- [1] 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.
- [2] Eduardo Pinheiro, Octavian Postolache, and Pedro Giro. Empirical mode decomposition and principal component analysis implementation in processing non-invasive cardiovascular signals. *Measurement*, 45(2):175 – 181, 2012. Special issue on Electrical Instruments.
- [3] Erik Alonso, Elisabete Aramendi, Digna González-Otero, Unai Ayala, Mohamud Daya, and James K. Russell. Empirical mode decomposition for chest compression and ventilation detection in cardiac arrest. In *Computing in Cardiology 2014*, pages 17–20, Sept 2014.
- [4] Haizhao Yang, Jianfeng Lu, and Lexing Ying. Crystal image analysis using 2D synchrosqueezed transforms. *Multiscale Modeling & Simulation*, 13(4):1542–1572, 2015.
- [5] Jianfeng Lu, Benedikt Wirth, and Haizhao Yang. Combining 2D synchrosqueezed wave packet transform with optimization for crystal image analysis. *Journal of the Mechanics and Physics of Solids*, pages –, 2016.



- [6] 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.
- [7] 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.
- [8] 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.
- [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, Jan 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] Haizhao Yang and Lexing Ying. Synchrosqueezed curvelet transform for two-dimensional mode decomposition. *SIAM Journal on Mathematical Analysis*, 46(3):2052–2083, 2014.
- [12] Yue Huanyin, Guo Huadong, Han Chunming, Li Xinwu, and Wang 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, pages 2061–2063 vol.5, 2001.
- [13] Xueru Bai, Mengdao Xing, Feng Zhou, Guangyue Lu, and Zheng Bao. Imaging of micromotion targets with rotating parts based on empirical-mode decomposition. *IEEE Transactions on Geoscience and Remote Sensing*, 46(11):3514–3523, Nov 2008.
- [14] 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.
- [15] Zhaohua Wu and Norden E. Huang. Ensemble empirical mode decomposition: A noise-assisted data analysis method. *Advances in Adaptive Data Analysis*, 01(01):1–41, 2009.
- [16] Ingrid Daubechies, Jianfeng Lu, and Hau-Tieng Wu. Synchrosqueezed wavelet transforms: an empirical mode decomposition-like tool. *Appl. Comput. Harmon. Anal.*, 30(2):243–261, 2011.
- [17] Ratikanta Behera, Sylvain Meignen, and Thomas Oberlin. Theoretical Analysis of the Second-order Synchrosqueezing Transform. working paper or preprint, December 2015.
- [18] François Auger and Patrick Flandrin. Improving the readability of time-frequency and time-scale representations by the reassignment method. *Signal Processing, IEEE Transactions on*, 43(5):1068–1089, 1995.

- [19] Eric Chassande-Mottin, Francois Auger, and Patrick Flandrin. Time-frequency/time-scale reassignment. In *Wavelets and signal processing*, Appl. Numer. Harmon. Anal., pages 233–267. Birkhäuser Boston, Boston, MA, 2003.
- [20] Konstantin Dragomiretskiy and Dominique Zosso. Variational mode decomposition. *Signal Processing, IEEE Transactions on*, 62(3):531–544, Feb 2014.
- [21] Thomas Y. Hou and Zuoqiang Shi. Data-driven timefrequency analysis. *Applied and Computational Harmonic Analysis*, 35(2):284 – 308, 2013.
- [22] Luan Lin, Yang Wang, and Haomin Zhou. Iterative filtering as an alternative algorithm for empirical mode decomposition. *Advances in Adaptive Data Analysis*, 1(4):543–560, 10 2009.
- [23] Antonio Cicone, Jingfang Liu, and Haomin Zhou. Adaptive local iterative filtering for signal decomposition and instantaneous frequency analysis. *Applied and Computational Harmonic Analysis*, 41(2):384 – 411, 2016. Sparse Representations with Applications in Imaging Science, Data Analysis, and Beyond, Part IISI: {ICCHAS} Outgrowth, part 2.
- [24] Zhaohua Wu, Norden E. Huang, and Xianyao Chen. Some considerations on physical analysis of data. *Advances in Adaptive Data Analysis*, 3(1-2):95–113, 2011.
- [25] Hau-Tieng Wu. Instantaneous frequency and wave shape functions (i). *Applied and Computational Harmonic Analysis*, 35(2):181 – 199, 2013.
- [26] Andrew Reisner, M.D., Phillip A. Shaltis, Ph.D., Devin McCombie, and H Harry Asada, Ph.D. Utility of the photoplethysmogram in circulatory monitoring. *Anesthesiology*, 108(5):950–958, 2008.
- [27] Chen-Yun Lin, Su Li, and Hau-Tieng Wu. Wave-shape function analysis – when cepstrum meets time-frequency analysis. *arXiv:1605.01805[physics.data-an]*, 2016.
- [28] 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.
- [29] Thomas Y. Hou and Zuoqiang Shi. Extracting a shape function for a signal with intra-wave frequency modulation. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 374(2065), 2016.
- [30] Charles K. Chui, Yu-Ting Lin, and Hau-Tieng Wu. Real-time dynamics acquisition from irregular samples with application to anesthesia evaluation. *Analysis and Applications*, 14(04):537–590, 2016.
- [31] Jieren Xu, Haizhao Yang, and Ingrid Daubechies. Recursive Diffeomorphism-Based Regression for Shape Functions. 2016.
- [32] Lippincott Williams & Wilkins. *ECG Interpretation Made Incredibly Easy!* Made Incredibly Easy. Wolters Kluwer/Lippincott Williams & Wilkins Health, 2011.
- [33] Steve Herndon, Benjamin Holland, and Ajeet Vinayak. *What is the Role and Impact of the ECG in the Patient with Hyperkalemia?*, pages 362–370. Wiley-Blackwell, 2009.

- [34] Wikimedia Commons. File:sinusrhythmlabels.svg — wikimedia commons, the free media repository, 2017. [Online; accessed 24-June-2017].
- [35] László Györfi, Micael Kohler, Adam Krzyżak, and Harro Walk. *A distribution-free theory of nonparametric regression*. Springer series in statistics. Springer, New York, Berlin, Paris, 2002. Autre(s) tirage(s) : 2010.
- [36] David L. B. Jupp. Approximation to data by splines with free knots. *SIAM Journal on Numerical Analysis*, 15(2):328–343, 1978.
- [37] Saptarshi Das and Koushik Maharatna. Fractional dynamical model for the generation of eeg like signals from filtered coupled van-der pol oscillators. *Computer Methods and Programs in Biomedicine*, 112(3):490 – 507, 2013.