

A FAST ALGORITHM FOR MULTIREOLUTION MODE DECOMPOSITION*

GAO TANG[†] AND HAIZHAO YANG[‡]

Abstract. *Multiresolution mode decomposition* (MMD) is an adaptive tool to analyze a time series $f(t) = \sum_{k=1}^K f_k(t)$, where $f_k(t)$ is a *multiresolution intrinsic mode function* (MIMF) of the form $f_k(t) = \sum_{n=-N/2}^{N/2-1} a_{n,k} \cos(2\pi n\phi_k(t))s_{cn,k}(2\pi N_k\phi_k(t)) + \sum_{n=-N/2}^{N/2-1} b_{n,k} \sin(2\pi n\phi_k(t))s_{sn,k}(2\pi N_k\phi_k(t))$ with time-dependent amplitudes, frequencies, and waveforms. The multiresolution expansion coefficients $\{a_{n,k}\}$, $\{b_{n,k}\}$ and the shape function series $\{s_{cn,k}(t)\}$ and $\{s_{sn,k}(t)\}$ provide innovative features for adaptive time series analysis. The MMD aims at identifying these MIMFs (including their multiresolution expansion coefficients and shape functions series) from their superposition. However, due to the lack of efficient algorithms to solve the MMD problem, the application of MMD for large-scale data science is prohibitive, especially for real-time data analysis. This paper proposes a fast algorithm for solving the MMD problem based on recursive diffeomorphism-based spectral analysis (RDSA). RDSA admits highly efficient numerical implementation via the nonuniform fast Fourier transform; its convergence and accuracy can be guaranteed theoretically. Numerical examples from synthetic data and natural phenomena are given to demonstrate the efficiency of the proposed method.

Key words. mode decomposition, time series, wave shape functions, multiresolution analysis, nonuniform Fast Fourier transform, nonparametric regression

AMS subject classifications. 42A99, 65T99

DOI. 10.1137/18M1220649

1. Introduction. Oscillatory data analysis is important for a considerable number of real world applications such as medical electrocardiography (ECG) reading [1, 25, 30], atomic crystal images in physics [24, 39], mechanical engineering [19, 42], art investigation [10, 38], geology [20, 28, 40], imaging [3], etc. One single record of the data might contain several principal components with different oscillation patterns. The goal is to extract these components and analyze them individually. A typical model in *mode decomposition* is to assume that a signal $f(t)$ defined on $[0, 1]$ consists of several oscillatory modes like

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

where $\alpha_k(t)$ is a smooth, positive, and nonoscillatory instantaneous amplitude, $N_k\phi_k(t)$ is a smooth and strictly increasing instantaneous phase, $N_k\phi'_k(t)$ is the instantaneous frequency, and $r(t)$ is the residual signal. Methods for the mode decomposition problem in (1) include the empirical mode decomposition approach [18, 31],

*Received by the editors October 15, 2018; accepted for publication (in revised form) March 5, 2020; published electronically May 7, 2020.

<https://doi.org/10.1137/18M1220649>

Funding: The second author was supported by the start-up grant from the Department of Mathematic at Purdue University.

[†]Department of Computer Science, University of Illinois at Urbana-Champaign, Champaign, IL (gaotang2@illinois.edu).

[‡]Department of Mathematics, Purdue University, West Lafayette, IN (haizhao@purdue.edu).

synchrosqueezed transforms [4, 11], time-frequency reassignment methods [2, 6], adaptive optimization [12, 16], iterative filters [8, 23], etc.

In complicated applications, sinusoidal oscillatory patterns may lose important physical information [17, 29, 32, 34, 35], which motivates the introduction of shape functions $\{s_k(t)\}_{1 \leq k \leq K}$ and the *generalized mode decomposition* as follows:

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

where $\{s_k(t)\}_{1 \leq k \leq K}$ are 2π -periodic and zero-mean shape functions with a unit norm in $L^2([0, 2\pi])$, $\alpha_k(t)$, and $\phi_k(t)$ are the same functions as in (1). One such example is the photoplethysmogram (PPG) signal (see Figure 1) in medical study. Shape functions reflect complicated evolution patterns of the signal $f(t)$ and contain valuable information for monitoring the health condition of patients [5, 9, 21, 26, 41].

To better analyze time series with time-dependent amplitudes, phases, and shapes, the *multiresolution mode decomposition* (MMD) is proposed in [36] of the form

$$(3) \quad f(t) = \sum_{k=1}^K f_k(t),$$

where each

$$(4) \quad f_k(t) = \sum_{n=-N_k/2}^{N_k/2-1} a_{n,k} \cos(2\pi n \phi_k(t)) s_{c_n,k}(2\pi N_k \phi_k(t)) + \sum_{n=-N_k/2}^{N_k/2-1} b_{n,k} \sin(2\pi n \phi_k(t)) s_{s_n,k}(2\pi N_k \phi_k(t))$$

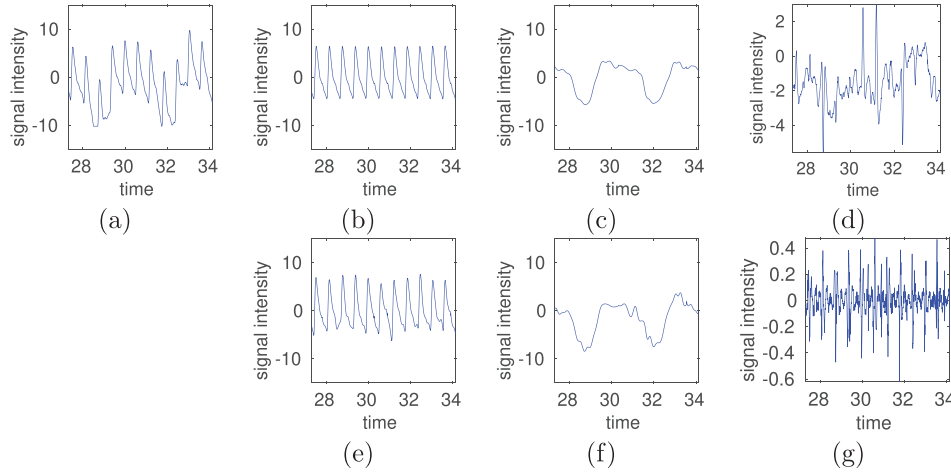


FIG. 1. Decomposition of a real PPG signal in (a). (b), (c), and (d) are the decomposition by the generalized mode decomposition in (2), while (e), (f), and (g) are decomposition by the MMD (3). The decomposed results contain one cardiac mode ((b) and (e)), one respiratory mode ((c) and (f)), and the residual signal ((d) and (g)). The residual signal by MMD is significantly weaker and behaves more like white noise than that by model (2). Figure 2 below visualizes the whiteness of the residual signals.

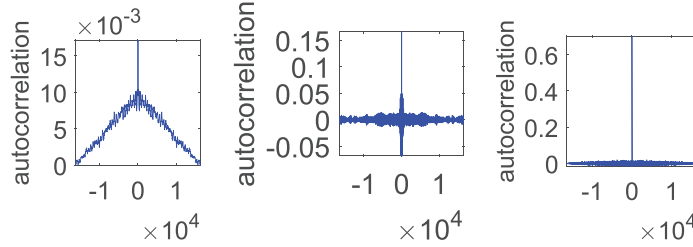


FIG. 2. Comparison of the whiteness of the residual signal generated by (2) and MMD (3) for the PPG signal in Figure 1. The autocorrelation of the residual signal by (2), the residual signal by MMD, and a vector of Gaussian random noise are plotted in the left, middle, and right figures, respectively. 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 that by (2) still contains correlated oscillation.

is a *multiresolution intrinsic mode function* (MIMF) with shape functions and phase functions satisfying the same conditions as in (2). MIMF is a generalization of the model $\alpha_k(t)s_k(2\pi N_k\phi_k(t))$ in (2) for more accurate data analysis (see the comparison of the models (2) and (3) in Figure 1 for the improvement). When $s_{cn,k}(t)$ and $s_{sn,k}(t)$ in (4) are equal to the same shape function $s_k(t)$, the model in (4) is reduced to $\alpha_k(t)s_k(2\pi N_k\phi_k(t))$ once the amplitude function $\alpha_k(t)$ is written in the form of its Fourier series expansion. When $s_{cn,k}(t)$ and $s_{sn,k}(t)$ are different shape functions, the two summations in (4) lead to time-dependent shape functions to describe the nonlinear and nonstationary time series adaption.

A recent paper [22] also tried to address the limitation of model (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 (4) emphasizes both the time variance of amplitude and shape functions by introducing multiresolution expansion coefficients and shape function series.

It was shown in [36] that the MIMF model can capture the evolution variance, which is more important than the average evolution patterns of oscillatory data for detecting diseases and measuring health risk. Let \mathcal{M}_ℓ be the operator for computing the ℓ -banded *multiresolution approximation* to a MIMF $f(t)$, i.e.,

$$(5) \quad \mathcal{M}_\ell(f)(t) = \sum_{n=-\ell}^{\ell} a_n \cos(2\pi n\phi(t))s_{cn}(2\pi N\phi(t)) + \sum_{n=-\ell}^{\ell} b_n \sin(2\pi n\phi(t))s_{sn}(2\pi N\phi(t)),$$

and \mathcal{R}_ℓ be the operator for the computing of the residual sum

$$(6) \quad \mathcal{R}_\ell(f)(t) = f(t) - \mathcal{M}_\ell(f)(t).$$

Then the 0-banded multiresolution approximation $\mathcal{M}_0(f)(t) = a_0s_{c0}(2\pi N\phi(t))$ describes the average evolution pattern of the signal, while the rest describe the evolution variance. Figure 3 shows that, if $f(t)$ is an ECG signal,¹ $\mathcal{R}_0(f)(t)$ visualizes the change of the evolution pattern better than $f(t)$, e.g., the change of the height of R peaks and the width of QRS and S waves.

¹From the PhysiNet <https://physionet.org/>.

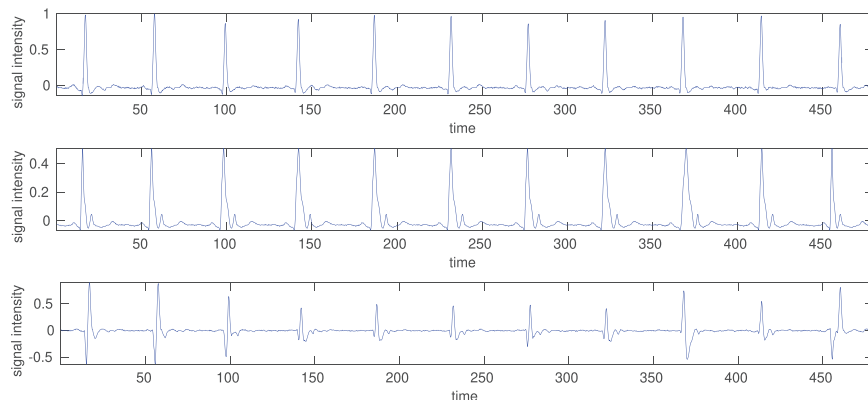


FIG. 3. *Top: a motion artifact contaminated ECG signal $f(t)$ modeled by (4). Middle: the 0-band multiresolution approximation $\mathcal{M}_0(f)(t) = a_0 s_{c0}(2\pi N \phi(t))$ of $f(t)$. Bottom: $f(t) - \mathcal{M}_0(f)(t)$, the variance of the evolution pattern of $f(t)$.*

The MMD problem aims at extracting each MIMF $f_k(t)$, estimating its corresponding multiresolution expansion coefficients $\{a_{n,k}\}$, $\{b_{n,k}\}$, the shape function series $\{s_{cn,k}(t)\}$ and $\{s_{sn,k}(t)\}$, under the assumption that the phase functions $\{\phi_k\}$ are known. Estimating phase functions has been an active research field in mode decomposition, and well-established approaches [2, 4, 6, 8, 11, 12, 16, 18, 22, 23, 31] can be applied to estimate phase functions. Hence, we only focus on the estimation of other quantities in MMD.

Applying the idea of recursive diffeomorphism-based regression (RDBR) [34, 36] has proposed a recursive scheme for decomposing $f(t)$ into several MIMFs, $\{f_k(t)\}$. Due to the repeated application of the expensive diffeomorphism-based regression, the method in [36] is not suitable for analyzing large data sets, especially when real-time analysis is required. Analyzing a single record of a high-resolution ECG or PPG signal with a few minutes of duration could take a whole day. Fast Fourier transform (FFT)-based shape function analysis in [17, 35] is efficient but can be only applied to model (2) and sometimes even for the case when $K = 1$ without any proof of convergence.

This dilemma motivates the design of the recursive diffeomorphism-based spectral analysis (RDSA) in this paper. From the *computational point of view*, RDSA takes only $O(mL \log L)$ operations to solve the MMD problem by taking advantage of the nonuniform Fast Fourier transform (NUFFT), where L is the length of the signal and m is the number of iterations. As we shall see later, the speedup of RDSA over RDBR in [36] can be as large as 1000. From the *theoretical point of view*, RDSA builds the bridge between FFT-based analysis and the RDBR, leading to a complete convergence analysis and filling the gap of theoretical analysis of FFT-based approaches in [17, 35].

A recent paper [33] proposed a complete framework for estimating all instantaneous quantities together in a two-step alternative fitting scheme: (1) fitting shape functions when amplitude and phase estimations are given and (2) fitting amplitude and phase functions when shape estimations are given. RDSA can be applied in this alternative fitting scheme to speed up the convergence. For many challenging numerical examples concerning crossover instantaneous frequencies, close frequencies, and the elimination of numerical errors in both steps via alternative fitting, the reader is referred to [33] for more examples.

We will first introduce RDSA in section 2. The convergence of RDSA is introduced and asymptotically analyzed² in section 3. In section 4, we present some numerical examples to demonstrate the efficiency of RDSA. Finally, we conclude this paper in section 5.

2. RDSA.

2.1. Diffeomorphism-based spectral analysis (DSA) for a single MIMF.

We first introduce the DSA for a single MIMF defined as follows.

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

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 Λ be the set of integers $\{|n| : \widehat{s}(n) \neq 0\}$. The greatest common divisor $\text{gcd}(s)$ of all the elements in Λ is 1.*

DEFINITION 2.2.

(7)

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

is a MIMF of type (M_0, M, N, ϵ) 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

$$\sum_{n=-N/2}^{N/2-1} |a_n| \leq M, \quad \sum_{n=-N/2}^{N/2-1} |b_n| \leq M, \quad \sum_{n=-N/2}^{N/2-1} |a_n| - \sum_{n=-M_0}^{M_0-1} |a_n| \leq \epsilon,$$

$$\sum_{n=-N/2}^{N/2-1} |b_n| - \sum_{n=-M_0}^{M_0-1} |b_n| \leq \epsilon;$$

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

As usual, we assume that the phase function $N\phi(t)$ and N are available; these quantities can be estimated using time-frequency concentration methods [2, 4, 6, 11], or adaptive optimization [7, 12, 16]. With the abuse of notations, we will use the same notation for continuous and discrete functions or transforms for simplicity. Without loss of generality, we assume that the signal $f(t)$ is uniformly sampled on $[0, 1]$ with L grid points

$$(8) \quad \left\{ t_\ell := \frac{\ell}{L} : 0 \leq \ell \leq L - 1, \ell \in \mathbb{Z} \right\};$$

the discrete Fourier transform of f denoted as $\widehat{f}(\xi)$ (or $\mathcal{F}(f)$) is defined on $\{\xi \in \mathbb{Z} : -\frac{L}{2} \leq \xi \leq \frac{L}{2} - 1\}$.

²Notations in the asymptotic analysis: we shall use the $O(\epsilon)$ notation, as well as the related notations \lesssim and \gtrsim ; in particular, we write $F = O(\epsilon)G$ if there exists a constant C (which we will not specify further) such that $|F| \leq C\epsilon|G|$; here C may depend on some general parameters as detailed just before Theorem 3.4.

We further assume that $f(t)$ satisfies a periodic boundary condition only in theoretical analysis for simplicity; in the case of nonperiodic boundary condition, the proposed method still works. In all of our numerical examples, signals are nonperiodic.

When $f(t)$ is a MIMF, the smooth function $\phi(t)$ serves as a diffeomorphism mapping $\cos(2\pi m\phi(t))f(t)$ to

$$(9) \quad \begin{aligned} h(t) &= (\cos(2\pi m\phi) f) \circ \phi^{-1}(t) \\ &= \sum_{n=-N/2}^{N/2-1} \frac{a_n}{2} (\cos(2\pi(m+n)t) + \cos(2\pi(m-n)t)) s_{cn}(2\pi Nt) \\ &\quad + \sum_{n=-N/2}^{N/2-1} \frac{b_n}{2} (\sin(2\pi(n+m)t) + \sin(2\pi(n-m)t)) s_{sn}(2\pi Nt). \end{aligned}$$

Let us define a scaling operator T_N mapping a function $g(\xi)$ to a function $T_N(g)$ via $T_N(g)(\xi) := g(N\xi)$. In the discrete case, this is equivalent to subsampling the function $g(\xi)$ at the grid points $\{N\xi\}_{\xi \in \mathbb{Z}}$. After Fourier transform and subsampling, we have

$$(10) \quad T_N(\widehat{h}(\xi)) = 2^{1-|\text{sgn}(m)|} \frac{a_m}{2} \sum_{\ell} \widehat{s}_{cm}(\ell) \delta(\xi - \ell),$$

where $\delta(\cdot)$ denotes the Dirac delta function. In practice, $\widehat{h}(\xi)$ can be evaluated via the NUFFT of $\cos(2\pi m\phi(t))f(t)$ on nonuniform grids $\{\psi_\ell := \phi(\frac{\ell}{L}) : 0 \leq \ell \leq L-1, \ell \in \mathbb{Z}\}$.

Equations (9) and (10) result in

$$(11) \quad a_m s_{cm}(2\pi t) = 2^{|\text{sgn}(m)|} \mathcal{F}^{-1} (T_N (\mathcal{F} ((\cos(2\pi m\phi) f) \circ \phi^{-1}))) (t),$$

and similarly we have

$$(12) \quad b_m s_{sm}(2\pi t) = 2^{|\text{sgn}(m)|} \mathcal{F}^{-1} (T_N (\mathcal{F} ((\sin(2\pi m\phi) f) \circ \phi^{-1}))) (t)$$

for $t \in [0, 1]$. Since all shape functions have a unit $L^2([0, 2\pi])$ -norm,³ we have

$$(13) \quad a_m = \sqrt{2\pi} \|2^{|\text{sgn}(m)|} \mathcal{F}^{-1} (T_N (\mathcal{F} ((\cos(2\pi m\phi) f) \circ \phi^{-1}))) \|_{L^2([0,1])}$$

and

$$(14) \quad b_m = \sqrt{2\pi} \|2^{|\text{sgn}(m)|} \mathcal{F}^{-1} (T_N (\mathcal{F} ((\sin(2\pi m\phi) f) \circ \phi^{-1}))) \|_{L^2([0,1])},$$

where the prefactor $\sqrt{2\pi}$ comes from changing the integral domain from $[0, 2\pi]$ to $[0, 1]$. Hence,

$$(15) \quad s_{cm}(2\pi t) = \begin{cases} \frac{1}{a_m} 2^{|\text{sgn}(m)|} \mathcal{F}^{-1} (T_N (\mathcal{F} ((\cos(2\pi m\phi) f) \circ \phi^{-1}))) (t), & a_m \neq 0, \\ 0, & a_m = 0, \end{cases}$$

and

$$(16) \quad s_{sm}(2\pi t) = \begin{cases} \frac{1}{b_m} 2^{|\text{sgn}(m)|} \mathcal{F}^{-1} (T_N (\mathcal{F} ((\sin(2\pi m\phi) f) \circ \phi^{-1}))) (t), & b_m \neq 0, \\ 0, & b_m = 0. \end{cases}$$

³In numerical implementation, we have a bandwidth parameter L_s for shape functions, i.e., only consider the Fourier series coefficient vector $\widehat{s} \in \mathbb{C}^{L_s}$ with entries $\widehat{s}(m)$ for $-\frac{L_s}{2} \leq m \leq \frac{L_s}{2} - 1$ in the reconstruction of a shape function vector $\vec{s} \in \mathbb{R}^{L_s}$ with entries $s(2\pi t)$ sampled on the grid points $\{t = \frac{k}{L_s} : 0 \leq k \leq L_s - 1, k \in \mathbb{Z}\}$. The discrete analog of the L^2 -norm $\|s\|_{L^2([0,2\pi])}$ of a function $s(t)$ is defined as $\sqrt{\frac{2\pi}{L_s}} \|\widehat{s}\|_{\ell^2}$.

Algorithm 1. DSA for shape functions and expansion coefficients with $O(mL \log L)$ operation complexity). The operation complexity comes from the fact that all routines are dominated by a NUFFT of size at most L and pointwise summation of vectors of size at most L .

- 1 Input: A single MIMF $f(t)$ in (7) and the phase function $p(t) = N\phi(t)$ sampled over t_ℓ , $\ell = 0, \dots, L-1$, frequency parameters N , a bandwidth parameter L_s , and a set of scale indices $\mathfrak{S} = \{n_1, \dots, n_m\}$.
- 2 Output: The shape functions s_{cn} and s_{sn} , the expansion coefficients a_n and b_n , and the partial summation $f_c(t)$ and $f_s(t)$.
- 3 Compute the expansion coefficients a_n and b_n for $n \in \mathfrak{S}$ according to (13) and (14).⁴
- 4 Evaluate the shape functions $s_{cn}(2\pi t)$ and $s_{sn}(2\pi t)$ for $n \in \mathfrak{S}$ according to (15) and (16) on uniform grids $\{t = \frac{k}{L_s} : 0 \leq k \leq L_s - 1, k \in \mathbb{Z}\}$.
- 5 Evaluate $a_n \cos(2\pi n\phi(t))s_{cn}(2\pi N\phi(t))$ and $b_n \sin(2\pi n\phi(t))s_{sn}(2\pi N\phi(t))$ for $n \in \mathfrak{S}$ based on interpolating the shape functions from uniform grids to nonuniform grids $\{t = \phi(\frac{k}{L}) : 0 \leq k \leq L-1, k \in \mathbb{Z}\}$.
- 6 Compute the partial summation $f_c(t) = \sum_{n \in \mathfrak{S}} a_n \cos(2\pi n\phi(t))s_{cn}(2\pi N\phi(t))$ and $f_s(t) = \sum_{n \in \mathfrak{S}} b_n \sin(2\pi n\phi(t))s_{sn}(2\pi N\phi(t))$.

The above discussion can be summarized in Algorithm 1 for estimating shape functions and expansion coefficients from a single MIMF in (7).

2.2. RDSA for multiple MIMFs. Next, in the case of a superposition of several MIMFs,

$$(17) \quad f(t) = \sum_{k=1}^K f_k(t),$$

where each

$$(18) \quad f_k(t) = \sum_{n=-N/2}^{N/2-1} a_{n,k} \cos(2\pi n\phi_k(t))s_{cn,k}(2\pi N_k\phi_k(t)) \\ + \sum_{n=-N/2}^{N/2-1} b_{n,k} \sin(2\pi n\phi_k(t))s_{sn,k}(2\pi N_k\phi_k(t)),$$

we propose the RDSA to extract each MIMF, estimate its corresponding multiresolution expansion coefficients and the shape function series from the superposition.

Due to the interference between different MIMFs and directly applying Algorithm 1 with an input signal in (17) and a phase function $N_k\phi_k(t)$ would not lead to accurate estimation of the multiresolution expansion coefficients, denoted as $\hat{a}_{n,k}$ and $\hat{b}_{n,k}$, and shape function series of $f_k(t)$, denoted as $\hat{s}_{cn,k}$ and $\hat{s}_{sn,k}$. This motivates the application of Algorithm 1 combined with the recursive scheme proposed in [36]. The intuition of the recursive scheme can be summarized as follows. We can get a rough estimation of $f_k(t)$ via Algorithm 1, denoted as

⁴Note that in Definition 2.1 all shape functions have zero-mean. Hence, in the numerical implementation of $T_N(\hat{h})(\xi)$ in (10), we will manually make $T_N(\hat{h})(0) = 0$ if $T_N(\hat{h})(0) \neq 0$ due to noise perturbation in the signal $f(t)$.

$$\begin{aligned} \dot{f}_k(t) &= \sum_{n=-N/2}^{N/2-1} \dot{a}_{n,k} \cos(2\pi n\phi_k(t)) \dot{s}_{cn,k}(2\pi N_k\phi_k(t)) \\ &\quad + \sum_{n=-N/2}^{N/2-1} \dot{b}_{n,k} \sin(2\pi n\phi_k(t)) \dot{s}_{sn,k}(2\pi N_k\phi_k(t)). \end{aligned}$$

Hence, the residual signal $r(t) := f(t) - \dot{f}(t)$ is again a new superposition of MIMFs. The recursive scheme applies Algorithm 1 again to $r(t)$ to estimate new multiresolution expansion coefficients and shape function series. We hope that the new estimations can correct the estimation error in the previous step; if this correction idea is applied repeatedly, we hope that the residual signal will decay and the estimation error will approach to zero. In more particular, RDSA can be summarized in Algorithm 2. In the pseudocode in Algorithm 2, the input and output of Algorithm 1 is denoted as

$$[\{s_{cn}\}_{n \in \mathfrak{S}}, \{s_{sn}\}_{n \in \mathfrak{S}}, \{a_n\}_{n \in \mathfrak{S}}, \{b_n\}_{n \in \mathfrak{S}}, f_s, f_c] = DSA(f, p, N, L_s, \mathfrak{S}).$$

When the input M_1 of Algorithm 2 is set to be empty, Algorithm 2 returns the M_0 -banded multiresolution approximation to each MIMF $f_k(t)$, its corresponding multiresolution expansion coefficients, and shape function series.

In fact, we have two for-loops to apply Algorithm 1 repeatedly to correct the estimation error: (1) one for-loop for the scale index n in Algorithm 1; (2) another one for the MIMF component index k in Algorithm 2. Note that in the case of a superposition of several MIMFs, the estimation provided by line 12 in Algorithm 2 is not accurate: the estimation error of a larger $|n|$ is much larger than that of a smaller $|n|$ because $|a_{n,k}|$ and $|b_{n,k}|$ usually decay quickly in $|n|$. As the iteration goes on, the multiresolution expansion coefficients with a small scale index n in the residual signal will decay since previous estimation steps try to eliminate them in the residual signal; only after a sufficiently large number of iterations in j can line 12 in Algorithm 2 can give accurate estimations for multiresolution expansion coefficients with a large $|n|$. Hence, to make Algorithm 2 converge, a large number of iteration number J_1 might be required.

To reduce the number of iterations J_1 in Algorithm 2, it might be better to put the k -for-loop inside the n -for-loop as in Algorithm 3, i.e., eliminating the multiresolution expansion coefficients with a small $|n|$ in the residual signal first before estimating those coefficients with a large $|n|$. It is still unclear which algorithm is faster since it relies on the decay rate of multiresolution expansion coefficients in $|n|$. Hence, a block size parameter \mathfrak{b} is used to make a balance: when $J_2 = 1$ and $\mathfrak{b} = M_0 + 1$ in Algorithm 3, Algorithm 3 essentially becomes Algorithm 2; when $\mathfrak{b} = 1$ in Algorithm 3, Algorithm 3 only computes the multiresolution expansion coefficients and shape functions for two scale indices per iteration in ℓ in line 5 of Algorithm 3. In the pseudocode in Algorithm 3, the input and output of Algorithm 2 are denoted as

$$\begin{aligned} &[\mathfrak{S}, \{f_k^{est}(t)\}_{1 \leq k \leq K}, \{s_{cn,k}\}_{1 \leq k \leq K, n \in \mathfrak{S}}, \{s_{sn,k}\}_{1 \leq k \leq K, n \in \mathfrak{S}}, \{a_{n,k}\}_{1 \leq k \leq K, n \in \mathfrak{S}}, \\ &\quad \{b_{n,k}\}_{1 \leq k \leq K, n \in \mathfrak{S}}] \\ &= RDSA_1(f, \{p_k\}_{1 \leq k \leq K}, \epsilon, J_1, L_s, M_0, M_1). \end{aligned}$$

3. Convergence analysis. Although the RDSAs in Algorithm 2 and 3 are mainly based on Fourier analysis, it can be proved that they are equivalent to the RDBR in [36], which leads to the theory of the convergence of RDSA. Since Algorithm 2 is a special case of Algorithm 3, we will only focus on the convergence analysis of Algorithm 3. Without loss of generality, we assume $\mathfrak{b} = 1$ in the analysis.

Algorithm 2. (The first RDSA for MMD). The operation complexity of this algorithm is $O(\max\{M_0, M_1\}J_1KL \log L)$ since the essential cost is the application of the DSA in Algorithm 1 for J_1K times.

```

1 Input:  $L$  points of measurement  $\{f(t_\ell)\}_{\ell=0,\dots,L-1}$  with  $t_\ell \in [0, 1]$ , estimated
   instantaneous phases  $\{p_k\}_{k=1,\dots,K}$ , an accuracy parameter  $\epsilon$ , the maximum
   iteration number  $J_1$ , and bandwidth parameters  $L_s$ ,  $M_0$ , and  $M_1$ .
2 Output: A scale index set  $\mathfrak{S}$ ,  $f_k^{est}(t)$  at the sampling grid points  $\{t_\ell\}_{0 \leq \ell \leq L-1}$ ,
   its multiresolution expansion coefficients  $\{a_{n,k}\}_{n \in \mathfrak{S}}$  and  $\{b_{n,k}\}_{n \in \mathfrak{S}}$ , and its
   shape function series  $\{s_{cn,k}\}_{n \in \mathfrak{S}}$  and  $\{s_{sn,k}\}_{n \in \mathfrak{S}}$  for  $1 \leq k \leq K$ .
3 if  $M_1$  has not been specified then
4    $\lfloor$  Define the scale index set  $\mathfrak{S} = \{-M_0, -M_0 + 1, \dots, M_0\}$ .
5 else
6    $\lfloor$  Define  $\mathfrak{S} = \{-M_1 + 1, -M_1 + 2, \dots, -M_0, \} \cup \{M_0, M_0 + 1, \dots, M_1 - 1, \}$ .
7 Initialize: let  $a_{n,k} = 0$ ,  $b_{n,k} = 0$ ,  $s_{cn,k} = 0$ ,  $s_{sn,k} = 0$ ,  $f_k^{est}(t) = 0$  for all  $k$  and
    $n \in \mathfrak{S}$ ; let  $c = \|f\|_{L^2}$ ; let  $e = 1$ ; let  $r^{(0)} = f$ .
8 Compute  $N_k$  as the integer nearest to the average of  $p'_k(t)$  for  $k = 1, \dots, K$ .
9 Sort  $\{N_k\}_{1 \leq k \leq K}$  in an ascending order and reorder the phase functions
   accordingly.
10 for  $j = 1, 2, \dots, J_1$ , do
11   for  $k = 1, \dots, K$  do
12      $\lfloor$   $\{\bar{s}_{cn}\}_{n \in \mathfrak{S}}, \{\bar{s}_{sn}\}_{n \in \mathfrak{S}}, \{\bar{a}_n\}_{n \in \mathfrak{S}}, \{\bar{b}_n\}_{n \in \mathfrak{S}}, \bar{f}_s, \bar{f}_c =$ 
        $DSA(r^{(j-1)}, p_k, N_k, L_s, \mathfrak{S})$ .
13      $s_{cn,k} \leftarrow s_{cn,k} + \bar{s}_{cn}$  and  $s_{sn,k} \leftarrow s_{sn,k} + \bar{s}_{sn}$  for  $n \in \mathfrak{S}$ .
14     Update  $f_k^{est}(t) \leftarrow f_k^{est}(t) + \bar{f}_c + \bar{f}_s$ .
15     if  $k < K$  then
16        $\lfloor$  Update  $r^{(j-1)} \leftarrow r^{(j-1)} - \bar{f}_c - \bar{f}_s$ .
17     else
18        $\lfloor$  Compute  $r^{(j)} = r^{(j-1)} - \bar{f}_c - \bar{f}_s$ .
19   if  $\|r^{(j)}\|_{L^2}/c \leq \epsilon$ , then break the for-loop.
20   if  $\|r^{(j)}\|_{L^2}/c \geq e - \epsilon$  then
21      $\lfloor$  Break the for loop.
22   else
23      $\lfloor$   $e = \|r^{(j)}\|_{L^2}/c$ .
24 Let  $a_{n,k} = \|s_{cn,k}\|_{L^2}$  and  $s_{cn,k} = s_{cn,k}/a_{n,k}$  for all  $k$  and  $n \in \mathfrak{S}$ .
25 Let  $b_{n,k} = \|s_{sn,k}\|_{L^2}$  and  $s_{sn,k} = s_{sn,k}/b_{n,k}$  for all  $k$  and  $n \in \mathfrak{S}$ .

```

3.1. Preliminaries. Before presenting the theory for RDSA, let us revisit RDBR for MMD in [36]. In RDBR, if $f(t) = a_n s_{cn}(2\pi N\phi(t))$, we define the inverse-warping data by $h(v) = f \circ p^{-1}(v) = a_n s_{cn}(2\pi v)$, where $v = p(t) = N\phi(t)$. As a consequence, we have a set of measurements of $h(v)$ sampled on $\{h(v_\ell)\}_{\ell=0,\dots,L-1}$ with $v_\ell = p(t_\ell)$. Note that $h(v)$ is a periodic function with period 1. Hence, if we define a folding map τ that folds the two-dimensional point set $\{(v_\ell, h(v_\ell))\}_{\ell=0,\dots,L-1}$ together,

$$(19) \quad \tau : (v_\ell, h(v_\ell)) \mapsto (\text{mod}(v_\ell, 1), h(v_\ell)),$$

Algorithm 3. The second RDSA for MMD. For the purpose of simplicity, we assume that $M_0 + 1$ is a multiple of \mathbf{b} . The operation complexity is bounded by $O(\max\{M_0, M_1\}J_1KL \log L)$ since this is a faster algorithm than Algorithm 3.

1	Input: L points of measurement $\{f(t_\ell)\}_{\ell=0, \dots, L-1}$ with $t_\ell \in [0, 1]$, estimated instantaneous phases $\{p_k\}_{k=1, \dots, K}$, accuracy parameters ϵ_1 and ϵ_2 , the maximum iteration numbers J_1 and J_2 , bandwidth parameters M_0 and L_s , and a block-size parameter \mathbf{b} .
2	Output: $\mathcal{M}_{M_0}(f_k)(t)$ at the sampling grid points $\{t_\ell\}_{0 \leq \ell \leq L-1}$, 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 $1 \leq k \leq K$.
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 $j = 1, 2, \dots, J_2$, do
5	for $\ell = 1, \dots, (M_0 + 1)/\mathbf{b}$ do
6	Compute $m = (\ell - 1)\mathbf{b}$ and apply
	$\mathfrak{S}, \{f_k^{est}(t)\}_{1 \leq k \leq K}, \{\bar{s}_{cn,k}\}_{1 \leq k \leq K, n \in \mathfrak{S}}, \{\bar{s}_{sn,k}\}_{1 \leq k \leq K, n \in \mathfrak{S}}, \{a_{n,k}\}_{1 \leq k \leq K, n \in \mathfrak{S}},$ $\{b_{n,k}\}_{1 \leq k \leq K, n \in \mathfrak{S}} = RDSA_1(r^{(j-1)}, \{p_k\}_{1 \leq k \leq K}, \epsilon_2, J_1, L_s, m, m + \mathbf{b}).$
7	for $k = 1, \dots, K$ do
8	$s_{cn,k} \leftarrow s_{cn,k} + \bar{s}_{cn,k}$ and $s_{sn,k} \leftarrow s_{sn,k} + \bar{s}_{sn,k}$.
9	Update $\mathcal{M}_{M_0}(f_k)(t) \leftarrow \mathcal{M}_{M_0}(f_k)(t) + \bar{f}_k^{est}$.
10	Compute $r^{(j-1)} \leftarrow r^{(j-1)} - \bar{f}_k^{est}$.
11	$r^{(j)} = r^{(j-1)}$.
12	If $\ r^{(j)}\ _{L^2}/c \leq \epsilon_1$, then break the for loop.
13	if $\ r^{(j)}\ _{L^2}/c \geq e - \epsilon_1$ then
14	Break the for loop.
15	else
16	$e = \ r^{(j)}\ _{L^2}/c$.
17	Let $a_{n,k} = \ s_{cn,k}\ _{L^2}$ and $s_{cn,k} = s_{cn,k}/a_{n,k}$ for all $k = 1, \dots, K$ and $n = -M_0, \dots, M_0$.
18	Let $b_{n,k} = \ s_{sn,k}\ _{L^2}$ and $s_{sn,k} = s_{sn,k}/b_{n,k}$ for all $k = 1, \dots, K$ and $n = -M_0, \dots, M_0$.

then the point set $\{\tau(v_\ell, a_n s_{cn}(2\pi v_\ell))\}_{\ell=0, \dots, L-1} \subset \mathbb{R}^2$ is a two-dimensional point set located at the curve $(v, a_n s_{cn}(2\pi v)) \subset \mathbb{R}^2$ given by the shape function $a_n s_{cn}(2\pi v)$ with $v \in [0, 1]$. Using the notations in nonparametric regression, let X be an independent random variable in $[0, 1]$, let Y be the response random variable in \mathbb{R} , and consider $(x_\ell, y_\ell) = \tau(v_\ell, a_n s_{cn}(2\pi v_\ell))$ as L samples of the random vector (X, Y) ; then a simple regression results in the shape function

$$(20) \quad a_n s_{cn} = s^R := \arg \min_{s: \mathbb{R} \rightarrow \mathbb{R}} \mathbb{E} \left\{ |s(2\pi X) - Y|^2 \right\},$$

where the superscript R means the ground truth regression function.

RDBR applies the partition-based regression method (or partitioning estimate) in Chapter 4 of [15] to solve the above regression problem. Given a small step size $h \ll 1$, the time domain $[0, 1]$ is uniformly partitioned into $N^h = \frac{1}{h}$ (assumed to be an integer) parts $\{[t_k^h, t_{k+1}^h)\}_{k=0, \dots, N^h-1}$, where $t_k^h = kh$. Let s^P denote the estimated regression function by the partition-based regression method with L samples. Following the definition in Chapter 4 of [15], we have a piecewise function

$$(21) \quad s^P(2\pi x) := \frac{\sum_{\ell=0}^{L-1} \mathcal{X}_{[t_k^h, t_{k+1}^h)}(x_\ell) y_\ell}{\sum_{\ell=0}^{L-1} \mathcal{X}_{[t_k^h, t_{k+1}^h)}(x_\ell)} = \frac{\sum_{\ell=0}^{L-1} \mathcal{X}_{[t_k^h, t_{k+1}^h)}(\text{mod}(v_\ell, 1)) a_n s_{cn}(2\pi v_\ell)}{\sum_{\ell=0}^{L-1} \mathcal{X}_{[t_k^h, t_{k+1}^h)}(\text{mod}(v_\ell, 1))} \\ = \frac{\sum_{\ell=0}^{L-1} \mathcal{X}_{[t_k^h, t_{k+1}^h)}(\text{mod}(N\phi(t_\ell), 1)) f(t_\ell)}{\sum_{\ell=0}^{L-1} \mathcal{X}_{[t_k^h, t_{k+1}^h)}(\text{mod}(N\phi(t_\ell), 1))}$$

when $x \in [t_k^h, t_{k+1}^h)$, where $\mathcal{X}_{[t_k^h, t_{k+1}^h)}(x)$ is the indicator function supported on $[t_k^h, t_{k+1}^h)$. When L is sufficiently large,

$$(22) \quad a_n s_{cn}(2\pi x) = s^R(2\pi x) \approx s^P(2\pi x) = \frac{\sum_{\ell=0}^{L-1} \mathcal{X}_{[t_k^h, t_{k+1}^h)}(\text{mod}(N\phi(t_\ell), 1)) f(t_\ell)}{\sum_{\ell=0}^{L-1} \mathcal{X}_{[t_k^h, t_{k+1}^h)}(\text{mod}(N\phi(t_\ell), 1))}$$

when $x \in [t_k^h, t_{k+1}^h)$, and the approximation is robust against noise perturbation (Chapter 4 of [15]).

More rigorously, the following theorem given in Chapter 4 in [15] estimates the L_2 risk of the approximation $s^P \approx s^R$ as follows.

THEOREM 3.1. *For the uniform partition with a step size h in $[0, 1)$ as defined just above, assume that*

$$\text{Var}(Y|X = x) \leq \sigma^2, \quad x \in \mathbb{R},$$

$$|s^R(x) - s^R(z)| \leq C|x - z|, \quad x, z \in \mathbb{R},$$

X has a compact support $[0, 1)$, and there are L independently and identically distributed 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

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

with an L^2 risk bounded by

$$(23) \quad \mathbb{E} \|s^P - s^R\|^2 \leq c_0 \frac{\sigma^2 + \|s^R\|_{L^\infty}^2}{Lh} + C^2 h^2,$$

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

If $f(t)$ is a MIMF, i.e.,

$$(24) \quad f(t) = \sum_{n=-N/2}^{N/2-1} a_n \cos(2\pi n\phi(t)) s_{cn}(2\pi N\phi(t)) + \sum_{n=-N/2}^{N/2-1} b_n \sin(2\pi n\phi(t)) s_{sn}(2\pi N\phi(t)),$$

under the same condition as in (22), it was shown in [36] that

$$(25) \quad a_n s_{cn}(2\pi x) \approx \frac{2^{|\text{sgn}(n)|} \sum_{\ell=0}^{L-1} \mathcal{X}_{[t_k^h, t_{k+1}^h)}(\text{mod}(N\phi(t_\ell), 1)) \cos(2\pi n\phi(t_\ell)) f(t_\ell)}{\sum_{\ell=0}^{L-1} \mathcal{X}_{[t_k^h, t_{k+1}^h)}(\text{mod}(N\phi(t_\ell), 1))}$$

and

$$(26) \quad b_n s_{sn}(2\pi x) \approx \frac{2^{|\text{sgn}(n)|} \sum_{\ell=0}^{L-1} \mathcal{X}_{[t_k^h, t_{k+1}^h)}(\text{mod}(N\phi(t_\ell), 1)) \sin(2\pi n\phi(t_\ell)) f(t_\ell)}{\sum_{\ell=0}^{L-1} \mathcal{X}_{[t_k^h, t_{k+1}^h)}(\text{mod}(N\phi(t_\ell), 1))}$$

when $x \in [t_k^h, t_{k+1}^h)$, by a similar argument as in (22) and the fact that the oscillation in amplitude functions $\cos(2\pi n\phi(t))$ and $\sin(2\pi n\phi(t))$ removes the influence of other terms in (24) on the estimation of $a_n s_{cn}(2\pi x)$ and $b_n s_{sn}(2\pi x)$, respectively.

In practice, in the case of a superposition of MIMFs, RDBR uses the same recursive algorithm as in Algorithm 3 (when $\mathfrak{b} = 1$) to solve the MMD problem. Unlike RDSA that uses the DSA in Algorithm 1⁵ to estimate shape functions, RDBR applies (25) and (26). Even though in each iteration (25) and (26) cannot give exact estimation, [36] proves that the estimation error can be corrected recursively as long as the MIMFs are well-differentiated. The well-differentiation of MIMFs relies on the well-differentiation of phase functions. Denote the set of sampling grid points $\{t_\ell\}_{\ell=0, \dots, L-1}$ in (8) as \mathcal{T} . \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

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

and

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

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

$$(27) \quad D_h^{ij}(m, n) \quad \text{and} \quad D_h^i(m)$$

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

DEFINITION 3.2. *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*

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

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 (27);
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 $M^2(K-1)\beta < 1$.

⁵The DSA is called in line 12 in Algorithm 2, which is called in line 6 in Algorithm 3.

DEFINITION 3.3. *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, ϵ) 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 $\mathcal{WS}(M_0, M, N, K, h, \beta, \gamma, \epsilon)$.

We recall again that in the case of $\mathbf{b} = 1$, RDBR replaces DSA in line 12 of Algorithm 2, which is used in line 6 in Algorithm 3, with (25) and (26) to estimate shape functions. Under the well-differentiation condition introduced just above, [36] proves that the estimation error of (25) and (26) (denoted as $s_{cn,k}^{E,(j)}$ and $s_{sn,k}^{E,(j)}$, respectively) in each iteration of the for-loop for J_2 in Algorithm 3 can be corrected recursively: the estimation errors of shape functions in the j th step become the target shape function to be estimated in the $(j+1)$ th step; to show the convergence of RDBR, it is sufficient to show that $s_{cn,k}^{E,(j)}$ and $s_{sn,k}^{E,(j)}$ decay as $j \rightarrow \infty$. Theorem 3.4 below (see the proof of Theorem 3.3 in [36]) shows that the estimation error decays to $O(\epsilon)$ as the iteration number goes to infinity.

Recall that, when we write $O(\cdot)$, \lesssim , or \gtrsim , the implicit constants may depend on M_0, M, K, C , and no other parameters.

THEOREM 3.4 (convergence of RDBR for MMD). *Suppose all shape functions are in the space of Lipschitz continuous functions with a constant C and ϵ is an accuracy parameter. Assume that $J_1 = 1$, (25), and (26) are used to estimate shape functions instead of DSA in Algorithm 3. For fixed ϵ, M_0, M, K , and C , there exists $h_0(\epsilon, C)$ such that for all $h < h_0(\epsilon, C)$, there exist $L_0(\epsilon, M_0, M, K, C, h)$ and $N_0(\epsilon, M, K, C, h)$ such that, when $L > L_0, N > N_0$, and $f(t) \in \mathcal{WS}(M_0, M, N, K, h, \beta, \gamma, \epsilon)$, we have*

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

⁶The estimation error of a shape function at step j , $s_{cn,k}^{E,(j)}$ is defined as the difference of the ground truth regression function of the regression problem at step j and the target shape function at step j , $a_{n,k}^{(j-1)} s_{cn,k}^{(j-1)}$, i.e., $s_{cn,k}^{E,(j)} = \arg \min s : \mathbb{R} \rightarrow \mathbb{R} \quad \mathbb{E}\{|s(2\pi X_{cn,k}^{(j-1)}) - Y_{cn,k}^{(j-1)}|^2\} - a_{n,k}^{(j-1)} s_{cn,k}^{(j-1)}$, where $(X_{cn,k}^{(j-1)}, Y_{cn,k}^{(j-1)})$ has samples $(\text{mod}(N_k\phi_k(t_\ell), 1), 2^{|\text{sgn}(n)|} \cos(2\pi n\phi_k(t_\ell))r^{(j-1)}(t_\ell))$ for t_ℓ from (8), where $r^{(j-1)}$ is the MIMF at step j as used in line 6 in Algorithm 3. Similarly, we define the estimation error $s_{sn,k}^{E,(j)}$ as $s_{sn,k}^{E,(j)} = \arg \min s : \mathbb{R} \rightarrow \mathbb{R} \quad \mathbb{E}\{|s(2\pi X_{sn,k}^{(j-1)}) - Y_{sn,k}^{(j-1)}|^2\} - b_{n,k}^{(j-1)} s_{sn,k}^{(j-1)}$, where $(X_{sn,k}^{(j-1)}, Y_{sn,k}^{(j-1)})$ has samples $(\text{mod}(N_k\phi_k(t_\ell), 1), 2^{|\text{sgn}(n)|} \sin(2\pi n\phi_k(t_\ell))r^{(j-1)}(t_\ell))$.

and

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

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 and $s_{cn,k}^{E,(j)}$ and $s_{sn,k}^{E,(j)}$ are defined just before this theorem.

As shown in [36], in each regression step, the variation of noise perturbation, which comes from the interference of other components, is bounded by a constant depending only on M_0 , M , and K . For the fixed ϵ and C , there exists $h_0(\epsilon, C)$ such that $C^2 h^2 < \epsilon^2$ if $0 < h < h_0$. For the fixed ϵ , M_0 , M , K , C , and h , there exists $L_0(\epsilon, M_0, M, K, C, h)$ such that, if $L > L_0$, then the L^2 error of the partition-based regression is bounded by ϵ^2 according to Theorem 3.1. Under these conditions, one can prove Theorem 3.4 using classical inequalities like the triangle inequality, Hölder's inequality, and Taylor expansion, following the steps in [36, Theorem 3.3] and the ideas in [34, Theorem 3.5].

3.2. Theory of RDSA. With the theory of RDBR introduced in the previous section, we are ready to prove the convergence of RDSA in this section. The main idea is to prove that RDSA is a special kind of RDBR by the downsampling theorem (aliasing theorem).

DEFINITION 3.5. Suppose L , N , and $\frac{L}{N}$ are integers. A downsampling operator, denoted as $\mathcal{D}_{N,L}$, with a factor N is a map from $x \in \mathbb{C}^L$ to $y = \mathcal{D}_{N,L}(x) \in \mathbb{C}^{\frac{L}{N}}$ such that

$$y[n] = x[nN]$$

for $n = 0, \dots, \frac{L}{N} - 1$.

DEFINITION 3.6. Suppose L , N , and $\frac{L}{N}$ are integers. An aliasing operator, denoted as $\mathcal{A}_{N,L}$, with a factor N is a map from $x \in \mathbb{C}^L$ to $y = \mathcal{A}_{N,L}(x) \in \mathbb{C}^{\frac{L}{N}}$ such that

$$y[n] = \sum_{j=0}^{N-1} x \left[n + j \frac{L}{N} \right]$$

for $n = 0, \dots, \frac{L}{N} - 1$.

THEOREM 3.7 (downsampling theorem). Suppose L , N , and $\frac{L}{N}$ are integers. For all $x \in \mathbb{C}^L$, it holds that

$$\mathcal{F}(\mathcal{D}_{N,L}(x)) = \frac{1}{N} \mathcal{A}_{N,L}(\mathcal{F}(x)),$$

where \mathcal{F} denotes the discrete Fourier transform.

The reader is referred to [27] for the proof of Theorem 3.7. An immediate result of Theorem 3.7 is the following convergence theorem for RDSA.

THEOREM 3.8 (convergence of RDSA for MMD). Suppose all shape functions are in the space of Lipschitz continuous functions with a constant C and ϵ is an accuracy parameter. Assume that $J_1 = 1$ and Algorithm 1 is used to estimate shape functions in Algorithm 3. For fixed ϵ , M_0 , M , K , and C , there exists $h_0(\epsilon, C)$ such that for all $h < h_0(\epsilon, C)$, there exist $L_0(\epsilon, M_0, M, K, C, h)$ and $\bar{N}_0(\epsilon, M_0, M, K, C, h)$ such that, when $L > L_0$, $N > \bar{N}_0$, $\frac{N}{L} < h_0$, and $f(t) \in \mathcal{WS}(M_0, M, N, K, \frac{N}{L}, \beta, \gamma, \epsilon)$, we have

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

and

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

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 and $s_{cn,k}^{E,(j)}$ and $s_{sn,k}^{E,(j)}$ are defined just before this theorem.

Proof. In the first part of the proof, we show that (11) and (12) are equivalent to partition-based regression with a step size $\frac{N}{L}$ up to an approximation error due to the NUFFT. Since the approximation error of the NUFFT can be controlled within arbitrary accuracy [13], we assume that this approximation error is $O(\epsilon)$.

For a MIMF $f(t)$ as defined in (7), let us define a uniform grid

$$\left\{ \psi_\ell = \phi(0) + \frac{(\phi(1) - \phi(0))\ell}{L} : 0 \leq \ell \leq L - 1, \ell \in \mathbb{Z} \right\}.$$

Define a vector $\vec{c}f \in \mathbb{R}^L$ associated with the function $f(t)$ such that the ℓ th entry is $\vec{c}f[\ell] = \cos(2\pi n\psi_\ell)(f \circ \phi^{-1}(\psi_\ell))$. Define a vector $\vec{s}_{cn} \in \mathbb{R}^L$ associated with the function $a_n s_{cn}(2\pi t)$ such that the ℓ th entry is $\vec{s}_{cn}[\ell] = a_n s_{cn}(2\pi t_\ell)$, where t_ℓ is from the uniform grid

$$\left\{ t_\ell = \frac{\ell}{L} : 0 \leq \ell \leq L - 1, \ell \in \mathbb{Z} \right\}.$$

By Theorem 3.7, we know

$$\mathcal{F}^{-1}(\mathcal{D}_{N,L}(\mathcal{F}(\vec{c}f))) = \frac{1}{N} \mathcal{A}_{N,L}(\vec{c}f).$$

By the definition of $\mathcal{D}_{N,L}$ and T_N and the fact that the right \mathcal{F} in (11) is carried out via the NUFFT, we see that (11) is equivalent to

$$\vec{s}_{cn} = O(\epsilon) + \frac{2^{|\text{sgn}(n)|}}{N} \mathcal{A}_{N,L}(\vec{c}f),$$

i.e.,

$$\begin{aligned} \vec{s}_{cn}[k] &= O(\epsilon) + \frac{2^{|\text{sgn}(n)|}}{N} \sum_{j=0}^{N-1} \vec{c}f \left[k + j \frac{L}{N} \right] \\ &= O(\epsilon) + \frac{2^{|\text{sgn}(n)|}}{N} \sum_{j=0}^{N-1} \cos \left(2\pi n \psi_{k+j\frac{L}{N}} \right) \left(f \circ \phi^{-1} \left(\psi_{k+j\frac{L}{N}} \right) \right). \end{aligned}$$

If we write the above equation in the terminology of partition-based regression, then the above equation (and hence (11)) is equivalent to

(28)

$$\begin{aligned} &a_n s_{cn}(2\pi x) \\ &= O(\epsilon) + \frac{2^{|\text{sgn}(n)|} \sum_{\ell=0}^{L-1} \mathcal{X}_{[t_k^{N/L}, t_{k+1}^{N/L})} \left(\frac{\text{mod}(N(\psi_\ell - \psi_0), \phi(1) - \phi(0))}{\phi(1) - \phi(0)} \right) \cos(2\pi n\psi_\ell) (f \circ \phi^{-1}(\psi_\ell))}{\sum_{\ell=0}^{L-1} \mathcal{X}_{[t_k^{N/L}, t_{k+1}^{N/L})} \left(\frac{\text{mod}(N(\psi_\ell - \psi_0), \phi(1) - \phi(0))}{\phi(1) - \phi(0)} \right)} \end{aligned}$$

when $x \in [t_k^{N/L}, t_{k+1}^{N/L})$, where $t_k^{N/L} = k \frac{N}{L}$, for $k = 0, \dots, \frac{L}{N} - 1$. In this special case of partition-based regression, the samples are

$$\left\{ \frac{\text{mod}(N(\psi_\ell - \psi_0), \phi(1) - \phi(0))}{\phi(1) - \phi(0)}, \cos(2\pi n \psi_\ell) (f \circ \phi^{-1}(\psi_\ell)) \right\}_{\ell=0, \dots, L-1},$$

where $\frac{\text{mod}(N(\psi_\ell - \psi_0), \phi(1) - \phi(0))}{\phi(1) - \phi(0)} = t_{\text{mod}(\ell, L/N)}^{N/L} \in [0, 1]$ is always on the partition grid points (with a step size $\frac{N}{L}$) of the partition-based regression. In fact, these samples are uniformly distributed on the partition grid points and each grid point has N samples.

Similarly, we see that (12) is equivalent to

$$\begin{aligned} & (29) \\ & b_n s_{sn}(2\pi x) \\ & = O(\epsilon) + \frac{2^{|\text{sgn}(n)|} \sum_{\ell=0}^{L-1} \mathcal{X}_{[t_k^{N/L}, t_{k+1}^{N/L})} \left(\frac{\text{mod}(N(\psi_\ell - \psi_0), \phi(1) - \phi(0))}{\phi(1) - \phi(0)} \right) \sin(2\pi n \psi_\ell) (f \circ \phi^{-1}(\psi_\ell))}{\sum_{\ell=0}^{L-1} \mathcal{X}_{[t_k^{N/L}, t_{k+1}^{N/L})} \left(\frac{\text{mod}(N(\psi_\ell - \psi_0), \phi(1) - \phi(0))}{\phi(1) - \phi(0)} \right)} \end{aligned}$$

when $x \in [t_k^{N/L}, t_{k+1}^{N/L})$. (29) is again a special case of partition-based regression with samples

$$\left\{ \frac{\text{mod}(N(\psi_\ell - \psi_0), \phi(1) - \phi(0))}{\phi(1) - \phi(0)}, \sin(2\pi n \psi_\ell) (f \circ \phi^{-1}(\psi_\ell)) \right\}_{\ell=0, \dots, L-1}.$$

The step size of the sampling domain $[0, 1]$ is $\frac{N}{L}$.

Recall that RDBR uses formulas (25) and (26) to estimate shape functions, and these formulas come from partition-based regression with sampling points

$$\{\text{mod}(N\phi(t_\ell), 1), \cos(2\pi n\phi(t_\ell))f(t_\ell)\}_{0 \leq \ell \leq L-1}$$

and

$$\{\text{mod}(N\phi(t_\ell), 1), \sin(2\pi n\phi(t_\ell))f(t_\ell)\}_{0 \leq \ell \leq L-1},$$

respectively. The step size of the sampling domain $[0, 1]$ is a fixed parameter h .

By Theorem 3.4, we see that, if RDBR was used to estimate shape functions (i.e., formulas (25) and (26) were used), then for fixed ϵ , M_0 , M , K , and C , there exists $h_0(\epsilon, C)$ such that for all $h < h_0(\epsilon, C)$, there exist $L_0(\epsilon, M_0, M, K, C, h)$ and $N_0(\epsilon, M_0, M, K, C, h)$ such that, when $L > L_0(\epsilon, M_0, M, K, C, h)$, $N > N_0(\epsilon, M_0, M, K, C, h)$, and $f(t) \in \mathcal{WS}(M_0, M, N, K, h, \beta, \gamma, \epsilon)$, we have

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

and

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

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 and $s_{cn,k}^{E,(j)}$ and $s_{sn,k}^{E,(j)}$ are defined just before this theorem.

Hence, in the second part of the proof of Theorem 3.8 for RDSA, we only need to clarify the conditions under which the estimations by (25) and (26) are almost the same as those by (28) and (29), respectively, up to a small difference $O(\epsilon)$.

Under the conditions of Theorem 3.4, as mentioned right after Theorem 3.4 in this paper, in each step of regression in (25) and (26), the estimated regression function only differs from the ground truth regression function up to an L^2 error bounded by ϵ^2 . More particularly, Theorem 3.1 gives the error bound as follows:

$$O\left(\frac{\sigma^2 + \|s^R\|_{L^\infty}^2}{L \cdot h} + C^2 h^2\right),$$

where σ^2 is the variation of noise perturbation (coming from the interference between different components) and σ^2 is bounded by a constant depending only on M_0 , M , and K ; s^R denotes the ground truth regression function for the regression problem, and it has an L^∞ -norm depending on M_0 , M , and K as well; C is the Lipschitz continuity constant; L is the number of samples; and h is the step size of the partition-based regression. Hence, there exists $h_0(\epsilon, C)$ such that for all $h < h_0(\epsilon, C)$, there exists $L_0(\epsilon, M_0, M, K, C, h)$ such that, when $L > L_0(\epsilon, M_0, M, K, C, h)$ we have

$$O\left(\frac{\sigma^2 + \|s^R\|_{L^\infty}^2}{L \cdot h} + C^2 h^2\right) \lesssim O(\epsilon^2).$$

Similarly by Theorem 3.1, we see that the estimated regression function by (28) and (29) has an L^2 error bounded by

$$O\left(\frac{\sigma^2 + \|s^R\|_{L^\infty}^2}{L \cdot \frac{N}{L}} + C^2 \frac{N^2}{L^2}\right) = O\left(\frac{1}{N} + \frac{N^2}{L^2}\right).$$

Hence, there exists $\dot{N}_0(\epsilon, M_0, M, K, C)$ such that, when $L > L_0(\epsilon, M_0, M, K, C, h)$, $N > \dot{N}_0(\epsilon, M_0, M, K, C)$, $\frac{N}{L} < h_0(\epsilon, C)$, we have

$$O\left(\frac{\sigma^2 + \|s^R\|_{L^\infty}^2}{L \cdot \frac{N}{L}} + C^2 \frac{N^2}{L^2}\right) \lesssim O(\epsilon^2).$$

Hence, following the proof of Theorem 3.3 in [36], we can prove that for fixed ϵ , M_0 , M , K , and C , there exists $h_0(\epsilon, C)$ such that for all $h < h_0(\epsilon, C)$, there exist $L_0(\epsilon, M_0, M, K, C, h)$ and

$$\bar{N}_0(\epsilon, M_0, M, K, C, h) := \max\{N_0(\epsilon, M_0, M, K, C, h), \dot{N}_0(\epsilon, M_0, M, K, C)\}$$

such that, when $L > L_0(\epsilon, M_0, M, K, C, h)$, $N > \bar{N}_0(\epsilon, M_0, M, K, C, h)$, $\frac{N}{L} < h_0(\epsilon, C)$, and $f(t) \in \mathcal{WS}(M_0, M, N, K, \frac{N}{L}, \beta, \gamma, \epsilon)$, we have

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

and

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

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 and $s_{cn,k}^{E,(j)}$ and $s_{sn,k}^{E,(j)}$ are defined just before this theorem.

The reason for requiring

$$f(t) \in \mathcal{WS}\left(M_0, M, N, K, \frac{N}{L}, \beta, \gamma, \epsilon\right)$$

instead of

$$f(t) \in \mathcal{WS}(M_0, M, N, K, h, \beta, \gamma, \epsilon)$$

is that the partition-based regression in (28) and (29) has a step size $\frac{N}{L}$ instead of h . \square

4. Numerical examples. In this section, some numerical examples of synthetic and real data are provided to demonstrate the proposed properties of RDSA. In all synthetic examples, we assume the instantaneous phases and amplitudes are known and only focus on verifying the RDSA in section 2 and its convergence theory in section 3. In most examples, we apply the one-dimensional highly redundant synchrosqueezed wave packet transform (SSWPT) [35, 37] to estimate instantaneous phases and amplitudes as inputs of RDSA. The implementation of SSWPT is publicly available in SynLab.⁷ Some more packages for estimating instantaneous frequencies can be found in [14]. The code for the RDSA is available online as well in a MATLAB package named DeCom.⁸

Let us summarize the main parameters in the above packages and in Algorithm 3. 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;
- ϵ_{sst} : a threshold for the wave packet coefficients.

In Algorithm 3, main parameters are

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

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

4.1. Convergence of RDSA. In this section, we provide numerical examples to verify the convergence theory of RDSA in section 3. For a fixed accuracy parameter ϵ , Theorem 3.8 shows that as long as instantaneous frequencies are sufficiently high and the number of samples is large enough, RDSA is able to estimate shape functions from a class of superpositions of MIMFs. The residual error in the iterative scheme linearly converges to a quantity of order ϵ . Since it is difficult to specify the relation

TABLE 1

Parameters in SynLab and Algorithm 3. The notation “–” means the corresponding parameter is not used or will be specified later in the example.

Figure	s	rad	red	ϵ_{sst}	J_1	J_2	M_0	ϵ	L_s	L
4	–	–	–	–	10	200	20	1e-13	2000	–
5, 6, 7	0.5	1.5	8	1e-3	10	200	–	1e-6	5000	2^{16}
8, 9, 10	0.5	1	8	1e-3	10	200	–	1e-6	5000	4000
11, 12, 13	–	–	–	–	10	200	10	1e-6	2000	2^{16}
14 (left 2), 15	0.75	1	8	1e-3	10	200	10	1e-6	2000	2^{16}
14 (right 2), 16	0.6	1	16	1e-2	10	200	10	1e-6	2000	2^{16}
17, 18, 19	–	–	–	–	10	200	10	1e-6	2000	2^{16}
20, 21, 22	0.5	1.5	8	1e-3	10	200	40	1e-6	1000	2^{16}

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

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

of the rate of convergence and other parameters explicitly in the analysis, we provide numerical examples to study this rate quantitatively.

In all examples in this section, we consider a simple case when the signal has two components with piecewise linear and continuous shapes. This makes it easier to verify the convergence analysis. For example, we consider a signal of the form

$$(30) \quad f(t) = f_1(t) + f_2(t),$$

where

$$f_1(t) = s_1(2\pi N\phi_1(t)) = s_1(2\pi N(t + 0.006 \sin(2\pi t)))$$

and

$$f_2(t) = s_2(2\pi N\phi_2(t)) = s_2(2\pi N(t + 0.006 \cos(2\pi t)));$$

$s_1(2\pi t)$ and $s_2(2\pi t)$ are shape functions defined on $[0, 1]$ as shown in Figure 4 (left). Here f_1 and f_2 can be considered as two intrinsic mode functions as well as two MIMFs by definition.

First, we fix the number of samples $L = 2^{19}$, vary the parameter N in (30), and estimate the convergence rate numerically. By Theorem 3.8 (adapted to the example in this section), the residual norm ϵ_1 in Algorithm 3 converges to $O(\epsilon)$ as follows:

$$\epsilon_1^{(j)} = O(\epsilon) + \beta^j O(1).$$

Hence, if we define a sequence $\{\mu_j\}$ by

$$\mu_j = \log \left(\left| \epsilon_1^{(j-1)} - \epsilon_1^{(j)} \right| \right)$$

and a sequence $\{\eta_j\}$ by

$$\eta_j = \mu_j - \mu_{j+1},$$

then η_j approximately quantifies the convergence in the j th iteration and should be nearly a constant close to $-\log(\beta)$. Figure 4(c) visualizes the sequences $\{\eta_j(\beta)\}$ generated from different signals with various N 's. It shows that when N is sufficiently large, $\{\eta_j\}$ are approximately a constant for all j and hence the convergence is linear; when N is small, RDSA converges sublinearly since $\eta_j > 0$ for all j and $\{\eta_j\}$ decays as j becomes large. After a few iterations, the residual error has been small enough. Hence, we do not show the results when the iteration number is larger than 7.

Second, we fix $N = 100$, vary the number of samples $L = 2^m$ with $m = 9, 10, \dots, 18$, and show the accuracy of RDSA after it converges. To obtain results with an accuracy as high as possible, we let $J_2 = 200$ and $\epsilon = 1e - 13$. Figure 4(d) shows that the final residual norm ϵ_1 after RDSA converged essentially decays in L .

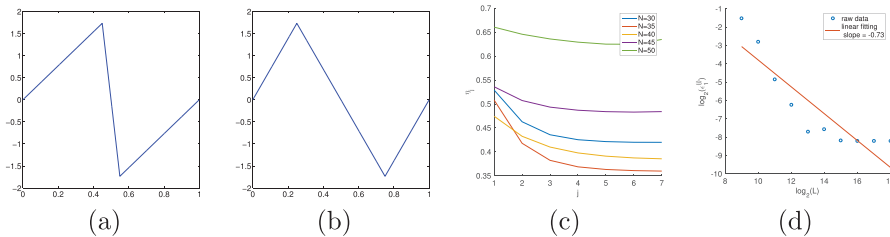


FIG. 4. (a) Shape function s_1 in (30). (b) Shape function s_2 in (30). (c) Estimated convergence rates β in different iteration steps when different values are assigned to N in (30). (d) The relation of the final residual norm $\epsilon^{(j)}$ (after the RDSA has been terminated) and the number of samples L .

TABLE 2

The speedup of RDSA against RDBR, i.e., t_{RDBR}/t_{RDSA} , for estimating shape functions in one iteration of the recursive scheme. The above table shows the speedup for various numbers of samples L and essential frequencies N .

$N \setminus L$	2^{10}	2^{11}	2^{12}	2^{13}	2^{14}	2^{15}	2^{16}	2^{17}	2^{18}	2^{19}
50	59	154	287	713	217	900	946	1246	1396	1278
70	525	682	723	1364	771	916	1136	1293	1645	1650
90	482	436	593	645	1184	611	945	1451	1501	2086
110	382	505	685	700	1056	1026	966	1253	1416	1298

4.2. The speedup of RDSA against RDBR. In this section, we compare the computational efficiency of RDSA proposed in this paper and RDBR in [36]. In this comparison, we still adopt the simple example in (30) and only compare the computational time of one iteration in RDSA and RDBR, i.e., the time (denoted as t_{RDBR}) for performing Algorithm 2 in [36] with $J = 1$ and the time (denoted as t_{RDSA}) for performing Algorithm 1 in this paper for computing only one shape function. The speedup of RDSA against RDBR, i.e., t_{RDBR}/t_{RDSA} , is shown in Table 2 for various L 's and N 's. Since the main computational cost for RDSA is the NUFFT, which has a computational complexity $O(L \log L)$ for a problem of size L , the RDSA is highly efficient. The speedup of RDSA against RDBR is much more prominent as L increases. Hence, RDSA is a more practical algorithm for MMD than RDBR when the problem size is large.

4.3. Analysis of MIMFs in real data. In this section, we apply RDSA to analyze MIMFs in real applications. We adopt the same numerical examples in [36] to compare the performance of RDSA and RDBR. To save space, only the results of RDSA will be provided. The reader is referred to [36] for the results of RDBR as comparison. The first example is an ECG record from a normal subject, and the second example is a motion-contaminated ECG record. More details about the ECG data can be found in <https://www.physionet.org/physiobank/database/>. We compute the band-limited multiresolution approximations of the first example and visualize them in Figures 5, 6, and 7; the band-limited multiresolution approximations of the second example are plotted in Figures 8, 9, and 10. Note that when the bandwidth of the multiresolution approximation increases, the approximation error decreases, and finer variation of the time series can be captured. Figures 7 and 10 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 shapes of an ECG signal we see in the raw data are 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. The results by RDSA validate the MIMF model again. Compared to RDBR, the residual $f(t) - \mathcal{M}_{40}(f)(t)$ by RDSA in Figures 6 and 9 is smaller, which implies that RDSA is better to handle fine details of the signal than RDBR.

4.4. MMD for synthetic data. In this section, two synthetic examples of MMD are provided to demonstrate the effectiveness of RDSA. We consider a simple case when the signal has two MIMFs with ECG shape functions. In particular, we let the shape function series of each MIMF contain the same ECG shape function. This makes it easier to verify Algorithm 3.

First synthetic example. In the first example, we consider a signal of the form

$$(31) \quad f(t) = f_1(t) + f_2(t),$$

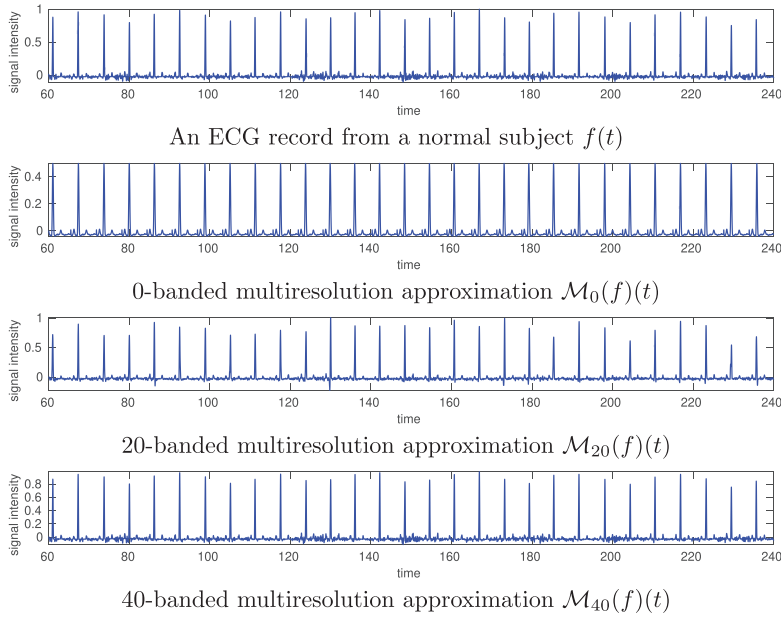


FIG. 5. Multiresolution approximations of an ECG record from a normal subject.

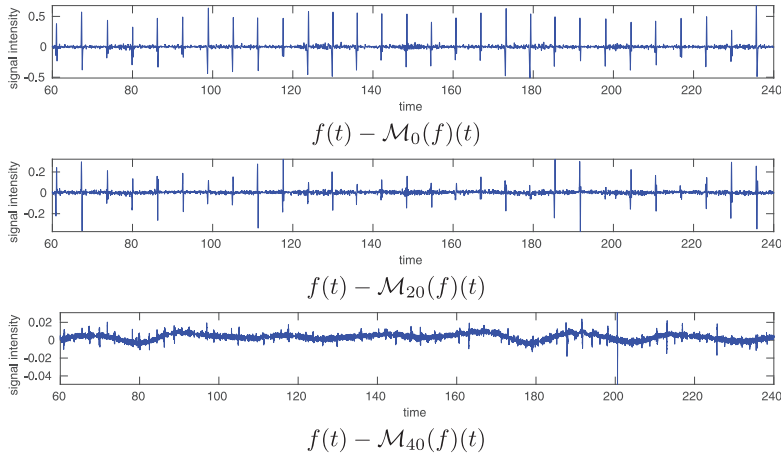


FIG. 6. The residual of the multiresolution approximations of an ECG record from a normal subject in Figure 5.

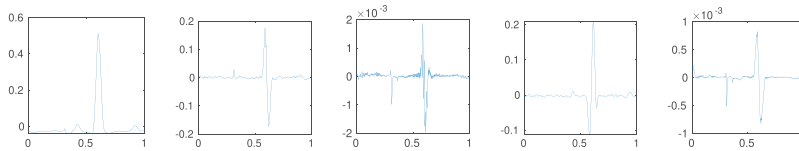


FIG. 7. Estimated shape functions $a_0 s_{c0}(t)$, $a_1 s_{c1}(t)$, $a_{-1} s_{c-1}(t)$, $b_1 s_{s1}(t)$, and $b_{-1} s_{s-1}(t)$ for the ECG signal in Figure 5.

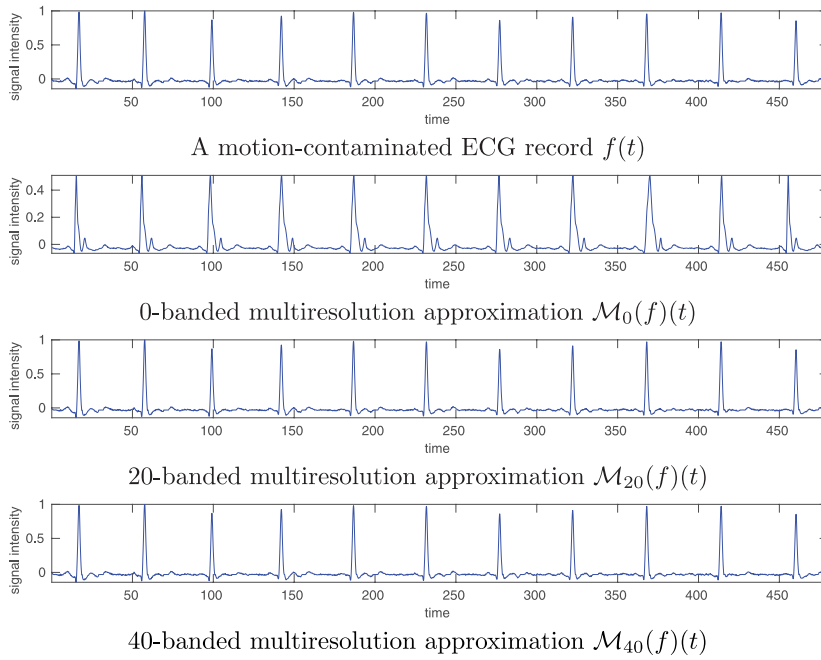


FIG. 8. *Multiresolution approximations of a motion-contaminated ECG record.*

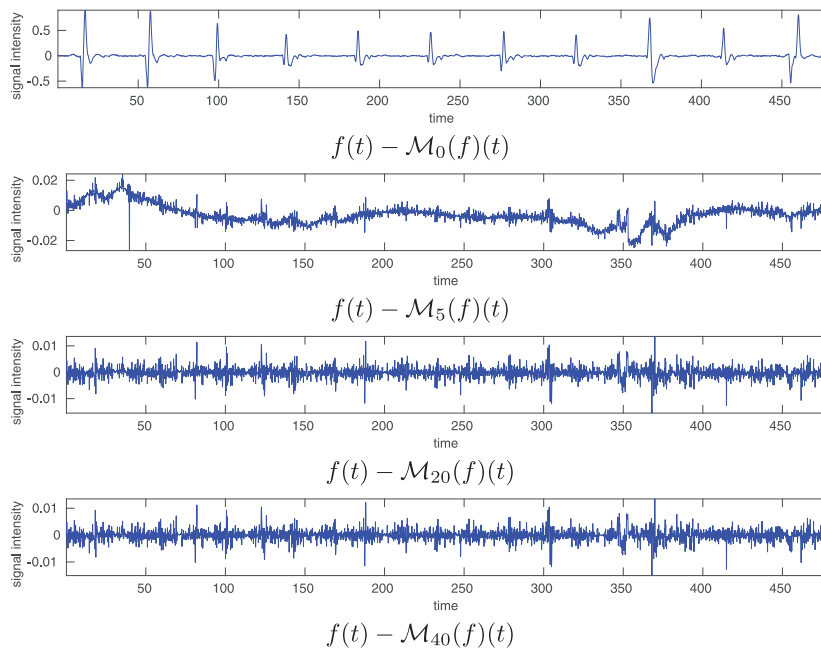


FIG. 9. *The residual of the multiresolution approximations of a motion-contaminated ECG record in Figure 8.*

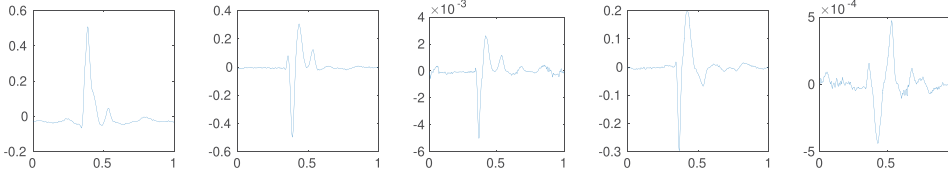


FIG. 10. Estimated shape functions $a_0s_{c0}(t)$, $a_1s_{c1}(t)$, $a_{-1}s_{c-1}(t)$, $b_1s_{s1}(t)$, and $b_{-1}s_{s-1}(t)$ for the ECG signal in Figure 8.

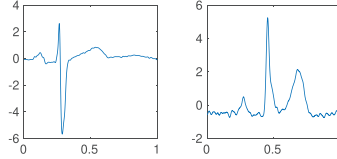


FIG. 11. Shape function $s_1(2\pi t)$ in (32) and $s_2(2\pi t)$ in (33).

where

$$(32) \quad f_1(t) = \alpha_1(\phi_1(t))s_1(300\pi\phi_1(t)),$$

$$(33) \quad f_2(t) = \alpha_2(\phi_2(t))s_2(440\pi\phi_2(t)),$$

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

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

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

and

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

$s_1(2\pi t)$ and $s_2(2\pi t)$ are generalized shape functions defined on $[0, 1]$ as shown in Figure 11.

First, we apply Algorithm 3 with the known instantaneous phases just above to estimate the multiresolution expansion coefficients, the shape functions series, and two components in (32) and (33). The product of the multiresolution expansion coefficient and its corresponding shape function is shown in Figure 12. To quantify the estimation error for the multiresolution expansion coefficient and its shape function, we compute the L^2 -norms of the estimation error and list them in the caption of Figure 12. The accuracy is as high as $1e-3$. Identified components are shown in Figure 13. The estimation errors are very small; the estimated results and the ground truth are almost indistinguishable.

Second, we apply the SSWPT to estimate instantaneous frequencies of the signal in (31) and then apply Algorithm 3 with the estimated instantaneous frequencies to estimate two components in (32) and (33). The estimated instantaneous frequencies and components are shown in Figures 14 (left two) and 15. The estimation errors are very small; the estimated results and the ground truth are almost indistinguishable.

Finally, we repeat the experiment as in the last paragraph with a noisy version of the signal in (31) with Gaussian random noise $\mathcal{N}(0, 0.25)$ to demonstrate the

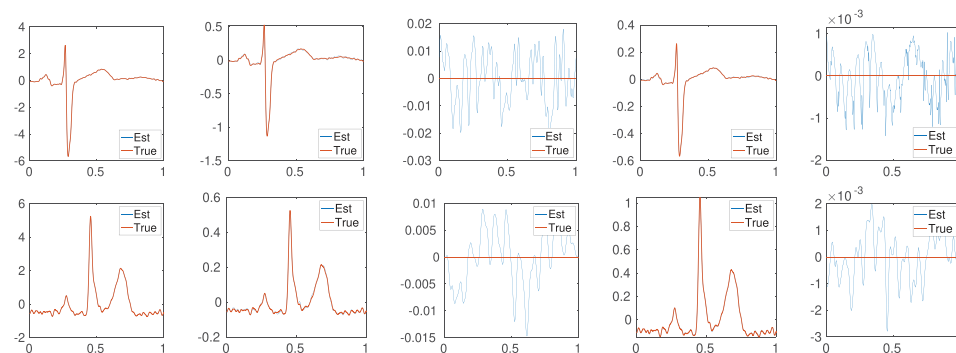


FIG. 12. *Top: estimated shape functions $a_{0,1}s_{c0,1}(2\pi t)$, $a_{1,1}s_{c1,1}(2\pi t)$, $a_{-1,1}s_{c-1,1}(2\pi t)$, $b_{1,1}s_{s1,1}(2\pi t)$, and $b_{-1,1}s_{s-1,1}(2\pi t)$ of $f_1(t)$ in (32). Bottom: estimated shape functions $a_{0,2}s_{c0,2}(2\pi t)$, $a_{1,2}s_{c1,2}(2\pi t)$, $a_{-1,2}s_{c-1,2}(2\pi t)$, $b_{1,2}s_{s1,2}(2\pi t)$, and $b_{-1,2}s_{s-1,2}(2\pi t)$ of $f_2(t)$ in (33). The L^2 -norm of the estimation errors in these 10 figures is $3.4e-3$, $9.1e-3$, $2.8e-3$, $9.1e-3$, $5.6e-4$, $6.2e-3$, $5.6e-3$, $4.1e-3$, $5.3e-3$, and $9.0e-4$, respectively.*

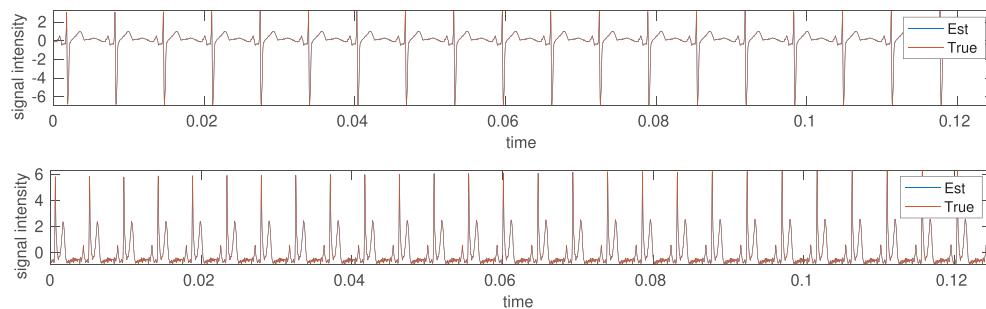


FIG. 13. *Recovered components by the RDSA with ground true instantaneous frequencies as compared to the ground true signals in (32) and (33).*

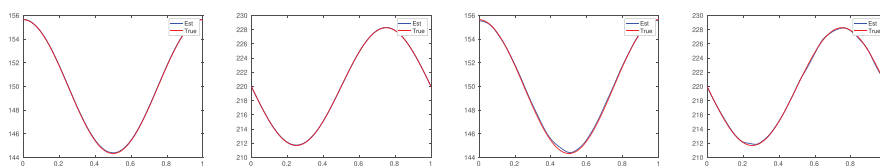


FIG. 14. *Estimated instantaneous frequencies of the example in (32) and (33). Left two figures: estimation when the signal is clean. Right two figures: estimation when the signal is noisy.*

robustness of the proposed method. The estimated instantaneous frequencies and components are shown in Figures 14 (right two) and 16. The estimation errors are small, and the proposed method works well for reasonably noisy signals.

Second synthetic example. In the second example, we investigate the performance of the RDSA in the case of intrawave phenomenon together with multiple crossover frequencies. This is a very challenging case of interest in the community of empirical mode decomposition. We consider a signal of the form

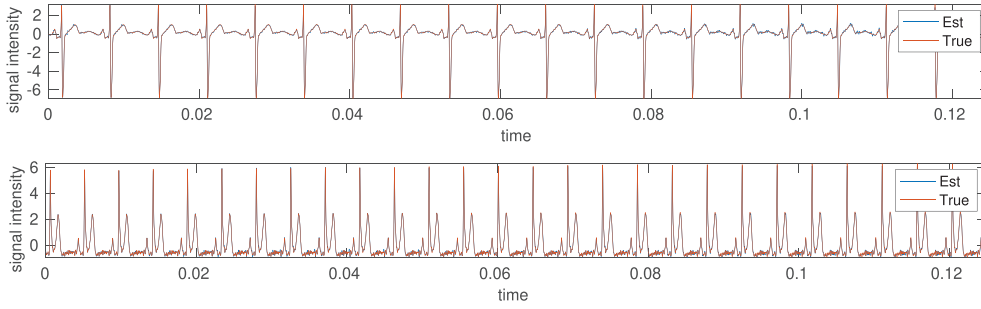


FIG. 15. Recovered components by the RDSA with estimated instantaneous frequencies from SSWPT as compared to the ground true signals in (32) and (33) when the signal f in (31) is clean.

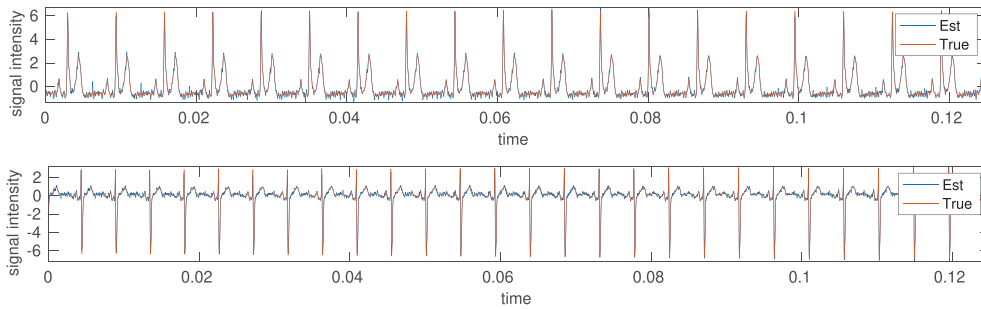


FIG. 16. Recovered components by the RDSA with estimated instantaneous frequencies from SSWPT as compared to the ground true signals in (32) and (33) when the signal f in (31) is noisy.

$$(34) \quad f(t) = f_1(t) + f_2(t),$$

where

$$(35) \quad f_1(t) = \alpha_1(\phi_1(t))s_1(300\pi\phi_1(t)),$$

$$(36) \quad f_2(t) = \alpha_2(\phi_2(t))s_2(300\pi\phi_2(t)),$$

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

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

$$\phi_1(t) = x + 0.01 \sin(20\pi t),$$

and

$$\phi_2(t) = x + 0.01 \cos(20\pi t).$$

$s_1(2\pi t)$ and $s_2(2\pi t)$ are generalized shape functions defined on $[0, 1]$ as shown in Figure 11.

We apply Algorithm 3 with the known instantaneous phases just above (see the corresponding instantaneous frequencies in Figure 17) to estimate the multiresolution expansion coefficients, the shape functions series, and two components in (35) and (36). The product of the multiresolution expansion coefficient and its corresponding shape function is shown in Figure 18. To quantify the estimation error for the multiresolution expansion coefficient and its shape function, we compute the L^2 -norms of the estimation error and list them in the caption of Figure 18. The accuracy is a high

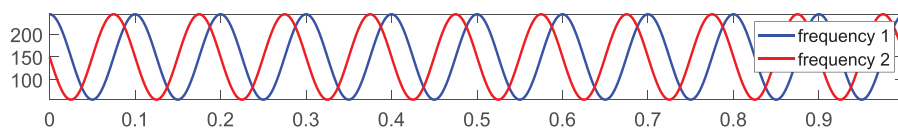


FIG. 17. The instantaneous frequencies of the example in (35) and (36) with intrawave and crossover phenomenon.

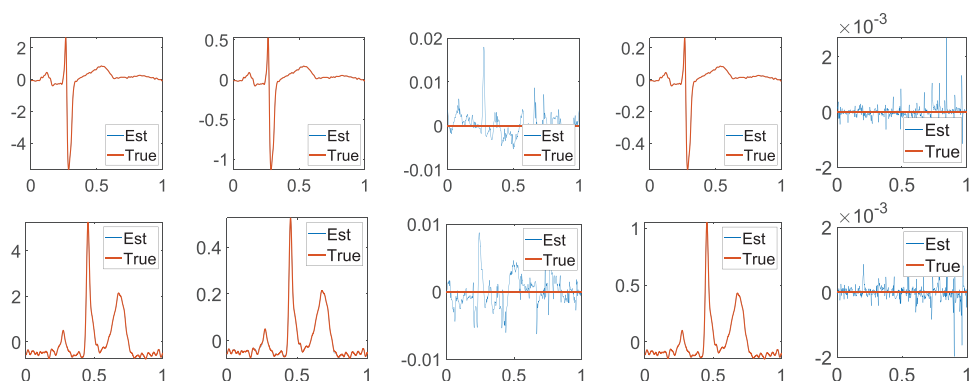


FIG. 18. Top: estimated shape functions $a_{0,1}s_{c0,1}(2\pi t)$, $a_{1,1}s_{c1,1}(2\pi t)$, $a_{-1,1}s_{c-1,1}(2\pi t)$, $b_{1,1}s_{s1,1}(2\pi t)$, and $b_{-1,1}s_{s-1,1}(2\pi t)$ of $f_1(t)$ in (35). Bottom: estimated shape functions $a_{0,2}s_{c0,2}(2\pi t)$, $a_{1,2}s_{c1,2}(2\pi t)$, $a_{-1,2}s_{c-1,2}(2\pi t)$, $b_{1,2}s_{s1,2}(2\pi t)$, and $b_{-1,2}s_{s-1,2}(2\pi t)$ of $f_2(t)$ in (36). The L^2 -norm of the estimation errors in these 10 figures is $2.0e-3$, $2.8e-3$, $4.6e-4$, $2.9e-3$, $2.9e-4$, $5.6e-3$, $2.4e-3$, $1.1e-3$, $2.3e-3$, and $2.4e-4$, respectively.

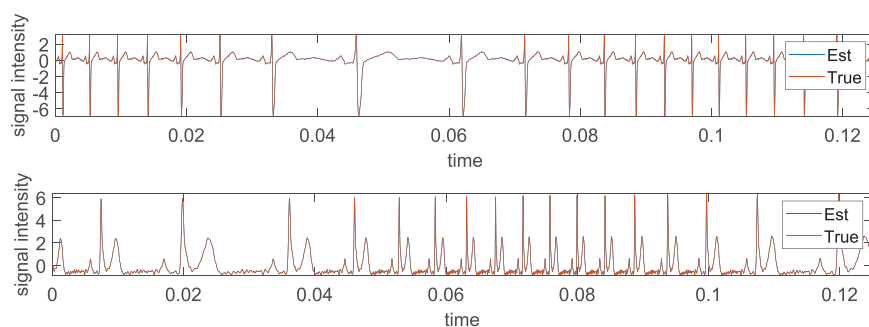


FIG. 19. Recovered components by the RDSA with ground true instantaneous frequencies as compared to the ground true signals in (35) and (36).

as $1e-3$. Identified components are shown in Figure 19. The estimation errors are very small; the estimated results and the ground truth are almost indistinguishable. After the completion of this manuscript, we developed a new framework for estimating instantaneous information in the case of intrawave and crossover frequencies in [33]. The reader is referred to [33] for the performance of RDSA applied with estimated instantaneous information.

4.5. MMD for real data. This is an example of PPG⁹ that contains a hemodynamical MIMF and a respiration MIMF. The instantaneous frequencies and phases

⁹From <http://www.capnabase.org>.

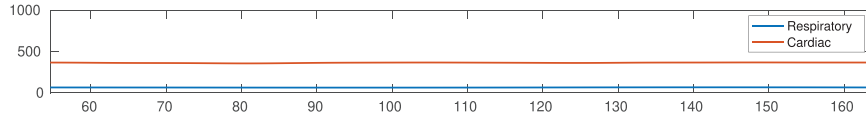


FIG. 20. Estimated fundamental instantaneous frequencies of the real PPG signal in the first panel of Figure 21 by the synchrosqueezed transform.

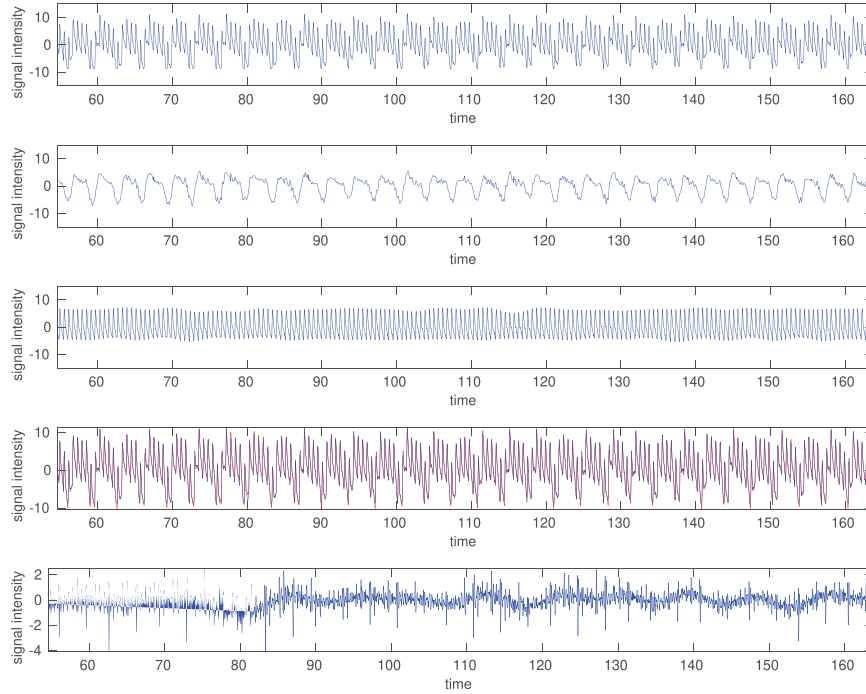


FIG. 21. First panel: the raw PPG signal $f(t)$. Second panel: the respiratory MIMF $f_1(t)$. Third panel: the cardiac MIMF $f_2(t)$. Fourth panel: the summation of the respiratory and cardiac MIMF's $f_1(t) + f_2(t)$ (red) compared to the raw PPG signal $f(t)$ (blue). The fifth panel: the residual signal $f(t) - f_1(t) - f_2(t)$.

are not known, and they are estimated via the synchrosqueezed transform in [35]. Figure 20 shows the estimated instantaneous frequencies of the respiratory and cardiac cycles. Inputting their corresponding instantaneous phases into RDSA, the PPG signal is separated into a respiratory MIMF and a cardiac MIMF as shown in Figure 21; their leading multiresolution shape functions are shown in Figure 22.

The last two panels of Figure 21 shows that the PPG signal has been completely separated into two MIMFs; the residual signal only contains noise, a smooth trend, and some sharp changes that are not correlated to the oscillation in MIMFs. 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.

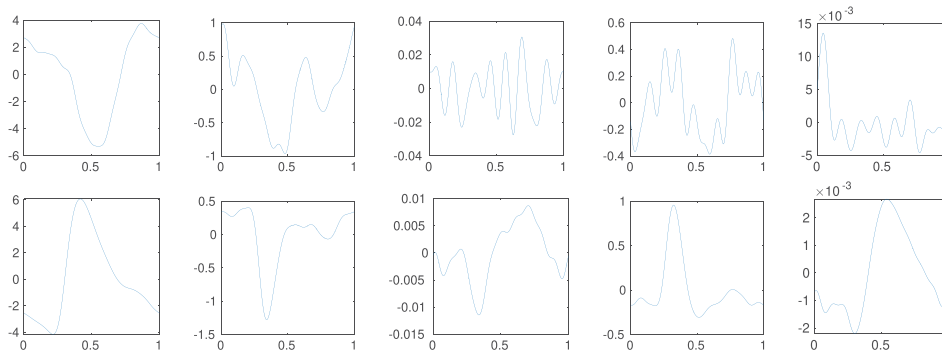


FIG. 22. *Top: estimated shape functions $a_{0,1}s_{c0,1}(2\pi t)$, $a_{1,1}s_{c1,1}(2\pi t)$, $a_{-1,1}s_{c-1,1}(2\pi t)$, $b_{1,1}s_{s1,1}(2\pi t)$, and $b_{-1,1}s_{s-1,1}(2\pi t)$ for the respiratory MIMF. Bottom: estimated shape functions $a_{0,2}s_{c0,2}(2\pi t)$, $a_{1,2}s_{c1,2}(2\pi t)$, $a_{-1,2}s_{c-1,2}(2\pi t)$, $b_{1,2}s_{s1,2}(2\pi t)$, and $b_{-1,2}s_{s-1,2}(2\pi t)$ for the cardiac MIMF.*

5. Conclusion. This paper proposed the RDSA for the MMD. The convergence of RDSA has been theoretically and numerically proved. The computational efficiency is significantly better than the RDBR in [36]. RDSA analyzes oscillatory time series by providing their multiresolution expansion coefficients and shape function series; these features would be more meaningful than those by traditional Fourier analysis and wavelet analysis. As we have seen in numerical examples, these features visualize important variation of signals, which are important for abnormality detection in oscillatory time series. The computational efficiency of RDSA makes the MMD a practical model for large-scale time series analysis and online data analysis, e.g., real-time monitoring systems for heart condition.

The fast algorithms proposed in this paper can be naturally extended to higher dimensional spaces for the applications like atomic crystal images in physics [24, 39], art investigation [10, 38], geology [20, 28, 40], imaging [3], etc. In higher dimensional spaces, the computational efficiency is a crucial issue. Hence, the extension of RDSA to higher dimensional spaces would be very important.

Acknowledgment. H. Y. thanks Ingrid Daubechies for discussion.

REFERENCES

- [1] E. ALONSO, E. ARAMENDI, D. GONZÁLEZ-OTERO, U. AYALA, M. DAYA, AND J. K. RUSSELL, *Empirical mode decomposition for chest compression and ventilation detection in cardiac arrest*, *Comput. Cardiology*, 41 (2014), pp. 17–20.
- [2] F. AUGER AND P. FLANDRIN, *Improving the readability of time-frequency and time-scale representations by the reassignment method*, *IEEE Trans. Signal Process.*, 43 (1995), pp. 1068–1089.
- [3] X. BAI, M. XING, F. ZHOU, G. LU, AND Z. BAO, *Imaging of micromotion targets with rotating parts based on empirical-mode decomposition*, *IEEE Trans. Geosci. Remote Sensing*, 46 (2008), pp. 3514–3523, <https://doi.org/10.1109/TGRS.2008.2002322>.
- [4] R. BEHERA, S. MEIGNEN, AND T. OBERLIN, *Theoretical analysis of the second-order synchrosqueezing transform*, *Appl. Comput. Harmon. Anal.*, 45 (2018), pp. 379–404, <https://doi.org/10.1016/j.acha.2016.11.001>.
- [5] G. S. H. CHAN, P. M. MIDDLETON, N. H. LOVELL, AND B. G. CELLER, *Extraction of photoplethysmographic waveform variability by lowpass filtering*, in *Proceedings of the 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*, pp. 5568–5571, <https://doi.org/10.1109/IEMBS.2005.1615746>.

- [6] E. CHASSANDE-MOTTIN, F. AUGER, AND P. FLANDRIN, *Time-frequency/time-scale reassignment*, in *Wavelets and Signal Processing*, L. Debnath, ed., Birkhäuser Boston, MA, 2003, pp. 233–267.
- [7] S. CHEN, X. DONG, Z. PENG, W. ZHANG, AND G. MENG, *Nonlinear chirp mode decomposition: A variational method*, *IEEE Trans. Signal Process.*, 65 (2017), pp. 6024–6037, <https://doi.org/10.1109/TSP.2017.2731300>.
- [8] A. CICONE, J. LIU, AND H. ZHOU, *Adaptive local iterative filtering for signal decomposition and instantaneous frequency analysis*, *Appl. Comput. Harmon. Anal.*, 41 (2016), pp. 384–411, <https://doi.org/10.1016/j.acha.2016.03.001>.
- [9] A. CICONE AND H.-T. WU, *How nonlinear-type time-frequency analysis can help in sensing instantaneous heart rate and instantaneous respiratory rate from photoplethysmography in a reliable way*, *Front. Physiology*, 8 (2017), 701, <https://doi.org/10.3389/fphys.2017.00701>.
- [10] B. CORNELIS, H. YANG, A. GOODFRIEND, N. OCON, J. LU, AND I. DAUBECHIES, *Removal of canvas patterns in digital acquisitions of paintings*, *IEEE Trans. Image Process.*, 26 (2017), pp. 160–171, <https://doi.org/10.1109/TIP.2016.2621413>.
- [11] I. DAUBECHIES, J. LU, AND H.-T. WU, *Synchrosqueezed wavelet transforms: An empirical mode decomposition-like tool*, *Appl. Comput. Harmon. Anal.*, 30 (2011), pp. 243–261, <https://doi.org/10.1016/j.acha.2010.08.002>.
- [12] K. DRAGOMIRETSKIY AND D. ZOSSO, *Variational mode decomposition*, *IEEE Trans. Signal Process.*, 62 (2014), pp. 531–544, <https://doi.org/10.1109/TSP.2013.2288675>.
- [13] A. DUTT AND V. ROKHLIN, *Fast Fourier transforms for nonequispaced data*, *SIAM J. Sci. Comput.*, 14 (1993), pp. 1368–1393, <https://doi.org/10.1137/0914081>.
- [14] D. FOURER, J. HARMOUCHE, J. SCHMITT, T. OBERLIN, S. MEIGNEN, F. AUGER, AND P. FLANDRIN, *The ASTRES toolbox for mode extraction of non-stationary multicomponent signals*, in *Proceedings of the 25th European Signal Processing Conference (EUSIPCO)*, 2017, pp. 1130–1134, <https://doi.org/10.23919/EUSIPCO.2017.8081384>.
- [15] L. GYÖRFI, M. KOHLER, A. KRZYŻAK, AND H. WALK, *A Distribution-Free Theory of Nonparametric Regression*, Springer Ser. Statist., Springer, New York, 2002.
- [16] T. Y. HOU AND Z. SHI, *Data-driven time-frequency analysis*, *Appl. Comput. Harmon. Anal.*, 35 (2013), pp. 284–308, <https://doi.org/10.1016/j.acha.2012.10.001>.
- [17] T. Y. HOU AND Z. SHI, *Extracting a shape function for a signal with intra-wave frequency modulation*, *Philos. Trans. Roy. Soc. A*, 374 (2016), 20150194 <https://doi.org/10.1098/rsta.2015.0194>.
- [18] N. E. HUANG, Z. SHEN, S. R. LONG, M. C. WU, H. H. SHIH, Q. ZHENG, N.-C. YEN, C. C. TUNG, AND H. H. LIU, *The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis*, *Proc. A*, 454 (1998), pp. 903–995, <https://doi.org/10.1098/rspa.1998.0193>.
- [19] W. HUANG, Z. SHEN, N. E. HUANG, AND Y. C. FUNG, *Engineering analysis of biological variables: An example of blood pressure over 1 day*, *Proc. Natl. Acad. Sci. USA*, 95 (1998), pp. 4816–4821.
- [20] Y. HUANYIN, G. HUADONG, H. CHUNMING, L. XINWU, AND W. CHANGLIN, *A SAR interferogram filter based on the empirical mode decomposition method*, in *Proceedings of the IEEE 2001 International Geoscience and Remote Sensing Symposium*, pp. 2061–2063, <https://doi.org/10.1109/IGARSS.2001.977903>.
- [21] D. LI, H. ZHAO, AND S. DOU, *A new signal decomposition to estimate breathing rate and heart rate from photoplethysmography signal*, *Biomed. Signal Process. Control*, 19 (2015), pp. 89–95, <https://doi.org/10.1016/j.bspc.2015.03.008>.
- [22] C.-Y. LIN, S. LI, AND H.-T. WU, *Wave-Shape Function Analysis – When Cepstrum Meets Time-Frequency Analysis*, e-print, arXiv:1605.01805 [physics.data-an], (2016).
- [23] L. LIN, Y. WANG, AND H. ZHOU, *Iterative filtering as an alternative algorithm for empirical mode decomposition*, *Adv. Adaptive Data Anal.*, 1 (2009), pp. 543–560, <https://doi.org/10.1142/S179353690900028X>.
- [24] J. LU, B. WIRTH, AND H. YANG, *Combining 2D synchrosqueezed wave packet transform with optimization for crystal image analysis*, *J. Mech. Phys. Solids*, 89 (2016), pp. 194–210, <https://doi.org/10.1016/j.jmps.2016.01.002>.
- [25] E. PINHEIRO, O. POSTOLACHE, AND P. GIRÃO, *Empirical mode decomposition and principal component analysis implementation in processing non-invasive cardiovascular signals*, *Measurement*, 45 (2012), pp. 175–181, <https://doi.org/https://doi.org/10.1016/j.measurement.2011.03.022>.
- [26] A. REISNER, M.D., P. A. SHALTIS, PH.D., D. MCCOMBIE, AND H. H. ASADA, PH.D., *Utility of the photoplethysmogram in circulatory monitoring*, *Anesthesiology*, 108 (2008), pp. 950–958, <https://doi.org/10.1097/ALN.0b013e31816c89e1>.

- [27] J. O. SMITH, *Mathematics of the Discrete Fourier Transform (DFT)*, W3K Publishing, <http://www.w3k.org/books/>, 2007.
- [28] J. B. TARY, R. H. HERRERA, J. HAN, AND M. VAN DER BAAN, *Spectral estimation-What is new? What is next?*, Rev. Geophys., 52 (2014), pp. 723–749, <https://doi.org/10.1002/2014rg000461>.
- [29] H.-T. WU, *Instantaneous frequency and wave shape functions (I)*, Appl. Comput. Harmon. Anal., 35 (2013), pp. 181–199, <https://doi.org/10.1016/j.acha.2012.08.008>.
- [30] H.-T. WU, Y.-H. CHAN, Y.-T. LIN, AND Y.-H. YEH, *Using synchrosqueezing transform to discover breathing dynamics from ECG signals*, Appl. Comput. Harmon. Anal., 36 (2014), pp. 354–359, <https://doi.org/10.1016/j.acha.2013.07.003>.
- [31] Z. WU AND N. E. HUANG, *Ensemble empirical mode decomposition: A noise-assisted data analysis method*, Adv. Adaptive Data Anal., 1 (2009), pp. 1–41, <https://doi.org/10.1142/S1793536909000047>.
- [32] Z. WU, N. E. HUANG, AND X. CHEN, *Some considerations on physical analysis of data.*, Adv. Adaptive Data Anal., 3 (2011), pp. 95–113.
- [33] J. XU, Y. LI, D. DUNSON, I. DAUBECHIES, AND H. YANG, *Non-Oscillatory Pattern Learning for Non-Stationary Signals*, e-prints, arXiv:1805.08102 [stat.ML], 2018.
- [34] J. XU, H. YANG, AND I. DAUBECHIES, *Recursive diffeomorphism-based regression for shape functions*, SIAM J. Math. Anal., 50 (2018), pp. 5–32, <https://doi.org/10.1137/16M1097535>.
- [35] H. YANG, *Synchrosqueezed wave packet transforms and diffeomorphism based spectral analysis for 1D general mode decompositions*, Appl. Comput. Harmon. Anal., 39 (2015), pp. 33–66, <https://doi.org/10.1016/j.acha.2014.08.004>.
- [36] H. YANG, *Multiresolution mode decomposition for adaptive time series analysis*, Appl. Comput. Harmon. Anal., to appear, <https://doi.org/10.1016/j.acha.2019.09.006>.
- [37] H. YANG, *Statistical analysis of synchrosqueezed transforms*, Appl. Comput. Harmon. Anal., 45 (2018), pp. 526–550, <https://doi.org/10.1016/j.acha.2017.01.001>.
- [38] H. YANG, J. LU, W. P. BROWN, I. DAUBECHIES, AND L. YING, *Quantitative canvas weave analysis using 2-D synchrosqueezed transforms: Application of time-frequency analysis to art investigation*, IEEE Signal Process. Mag., 32 (2015), pp. 55–63, <https://doi.org/10.1109/MSP.2015.2406882>.
- [39] H. YANG, J. LU, AND L. YING, *Crystal image analysis using 2D synchrosqueezed transforms*, Multiscale Model. Simul., 13 (2015), pp. 1542–1572, <https://doi.org/10.1137/140955872>.
- [40] H. YANG AND L. YING, *Synchrosqueezed curvelet transform for two-dimensional mode decomposition*, SIAM J. Math. Anal., 46 (2014), pp. 2052–2083, <https://doi.org/10.1137/130939912>.
- [41] Y. YE, Y. CHENG, W. HE, M. HOU, AND Z. ZHANG, *Combining nonlinear adaptive filtering and signal decomposition for motion artifact removal in wearable photoplethysmography*, IEEE Sensors J., 16 (2016), pp. 7133–7141, <https://doi.org/10.1109/JSEN.2016.2597265>.
- [42] C. ZHANG, Z. LI, C. HU, S. CHEN, J. WANG, AND X. ZHANG, *An optimized ensemble local mean decomposition method for fault detection of mechanical components*, Measurement Sci. Tech., 28 (2017), 035102.