

## SYNCHROSQUEEZED CURVELET TRANSFORM FOR TWO-DIMENSIONAL MODE DECOMPOSITION\*

HAIZHAO YANG<sup>†</sup> AND LEXING YING<sup>‡</sup>

**Abstract.** This paper introduces the synchrosqueezed curvelet transform as an optimal tool for two-dimensional mode decomposition of wavefronts or banded wave-like components. The synchrosqueezed curvelet transform consists of a generalized curvelet transform with application dependent geometric scaling parameters, and a synchrosqueezing technique for a sharpened phase space representation. In the case of a superposition of banded wave-like components with well-separated wave-vectors, it is proved that the synchrosqueezed curvelet transform is capable of recognizing each component and precisely estimating local wave-vectors. A discrete analogue of the continuous transform and several clustering models for decomposition are proposed in detail. Some numerical examples with synthetic and real data are provided to demonstrate the above properties of the proposed transform.

**Key words.** curvelet transform, synchrosqueezing, banded wave-like components, local wave-vector, phase space representation

**AMS subject classifications.** 42A99, 65T99

**DOI.** 10.1137/130939912

**1. Introduction.** In various applications [2, 12, 23, 24, 25, 26, 28, 30, 34], one is faced with a signal which is a superposition of several components (perhaps non-linear and nonstationary). The frequency or wave-vector of each component is localized in the time-frequency or phase space representation. A natural question would be whether it is possible to separate them according to their localized representations and estimate their local frequencies or wave-vectors. Classical time-frequency or phase space analysis provides several powerful tools for representing and analyzing complex signals. All of these tools essentially fall into two categories: linear or quadratic. As discussed in [9], linear methods have simple and efficient algorithms for forward and inverse transforms, but the resolution is unavoidably limited by the Heisenberg uncertainty principle. Although quadratic methods provide high resolution, the corresponding reconstruction methods are less straightforward and significantly more costly. Furthermore, nonphysical interference between components is more pronounced.

By introducing the synchrosqueezing technique, Daubechies and Maes proposed the synchrosqueezed wavelet transform in [10] and demonstrated that an important class of signals under the assumption of well-separated frequencies could be precisely decomposed. Synchrosqueezing is a reallocation method [1, 6, 7, 10] aiming at a sharpened time-frequency representation by reassigning values of the original representation. Though it has been shown to provide good results for one-dimensional (1D) signals, even with a substantial amount of noise, in higher dimensional space the application of the synchrosqueezed wavelet transform is limited. It cannot distinguish

---

\*Received by the editors October 4, 2013; accepted for publication (in revised form) April 3, 2014; published electronically June 19, 2014.

<http://www.siam.org/journals/sima/46-3/93991.html>

<sup>†</sup>Department of Mathematics, Stanford University, Stanford, CA 94305 (haizhao@math.stanford.edu). This author was partially supported by NSF grant CDI-1027952.

<sup>‡</sup>Department of Mathematics and ICME, Stanford University, Stanford, CA 94305-2125 (lexing@math.stanford.edu). This author was partially supported by NSF grants CAREER DMS-0846501, DMS-1027952, and CDI-1027952.

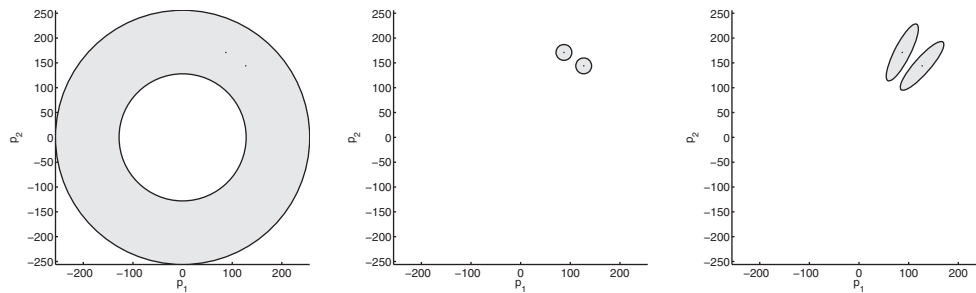


FIG. 1. Comparison of localized supports of continuous wavelets (left), wave packets (middle), and curvelets (right) in the Fourier domain. Two dots in each plot show the support of the Fourier transforms of the superposition of two plane waves  $e^{2\pi i p \cdot x}$  and  $e^{2\pi i q \cdot x}$  with the same wave-number ( $|p| = |q|$ ) but different wave-vectors ( $p \neq q$ ).

two components sharing the same wave-number but having different wave-vectors, because of the isotropic character of the high dimensional wavelet transform. In fact, this is a common phenomenon in many applications of high frequency wave propagation. To specify this problem, let us consider a simple superposition of two plane waves  $e^{2\pi i p \cdot x}$  and  $e^{2\pi i q \cdot x}$  with the same wave-number ( $|p| = |q|$ ) but different wave-vectors ( $p \neq q$ ). In the Fourier domain, the gray region in Figure 1 (left) shows the support of one continuous wavelet. The wavelet cannot distinguish these two plane waves in the sense that the gray region has to cover two dots  $p$  and  $q$  simultaneously, or has to exclude them simultaneously.

To overcome this inherent limitation of the synchrosqueezed wavelet transform in high dimensional space, the synchrosqueezed wave packet transform (SSWPT) was developed in [33], inspired by the localized support of wave packets in the Fourier domain. The finer supports result in the better resolution for wave-number separation and, more importantly, the anisotropic supports contribute to the angular separation of wave-vectors. As shown in Figure 1 (middle), in the Fourier domain, the supports of  $e^{2\pi i p \cdot x}$  and  $e^{2\pi i q \cdot x}$  are in the supports of two different wave packets, as long as  $p$  and  $q$  are well-separated. Yang and Ying [33] proved that SSWPT could identify different nonlinear and nonstationary high frequency wave-like components with different wave-vectors in high dimensional space in a general case, even with severe noise.

In some applications such as wave field separation problems [28, 30] and ground roll removal problems [2, 12, 34] in geophysics, it is required to separate overlapping wavefronts or banded wave-like components. In this case, the boundary of these components gives rise to many nonzero coefficients of wave packet transform, which results in unexpected interferential synchrosqueezed energy distribution (see Figure 2 top-right). This would dramatically reduce the accuracy of local wave-vector estimation, because the locations of nonzero energy provide estimation of local wave-vectors. As shown in Figure 2 (top-right), there exists misleading local wave-vector estimates at the location where the signal is negligible. Even if at the location where the signal is relevant, the relative error is still unacceptable. To solve this problem, an empirical idea is that good basis elements in the synchrosqueezed transform should look like the components, i.e., they should appear in a needle-like shape. An optimal solution is curvelets. The curvelet transform is anisotropic (as shown in Figure 1 right) and is designed for optimally representing curved edges [4, 29] and banded wavefronts [3]. This motivates the design of the synchrosqueezed curvelet transform (SSCT) as an optimal tool to estimate local wave-vectors of wavefronts or banded wave-like components in

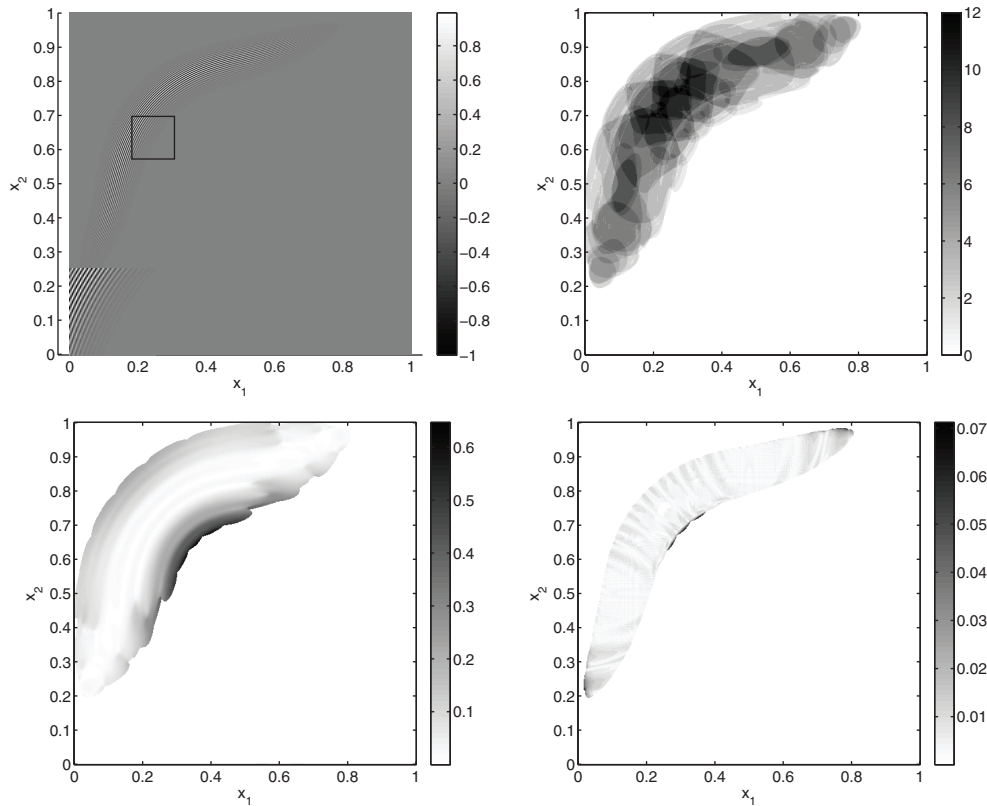


FIG. 2. *Top-left:* A banded deformed plane wave,  $f(x) = e^{-\frac{(\phi(x)-0.7)^2}{(4/N)^2}} e^{2\pi i N \phi(x)}$ , where  $N = 135$  and  $\phi(x) = x_1 + (1 - x_2) + 0.1 \sin(2\pi x_1) + 0.1 \sin(2\pi(1 - x_2))$ . *Top-right:* Number of nonzero discrete synchrosqueezed energy of SSWPT at each grid point of space domain. *Bottom-left:* Relative error between the mean local wave-vector estimate (defined in [33]) and the exact local wave-vector using SSWPT. *Bottom-right:* Relative error between the mean local wave-vector estimate and the exact local wave-vector using SSCT.

this paper. The estimate of local wave-vectors provided by SSCT is much better than that by SSWPT, as shown in Figure 2 (bottom). As a particular interest in theory, we will explore the limit of the bandwidth of these banded wave-like components that can be analyzed by the SSCT.

**1.1. SSCT.** Below is a brief introduction to the general curvelet transform with a radial scaling parameter  $t < 1$  and an angular scaling parameter  $s \in (\frac{1}{2}, t)$ . Similar to the discussion in [33], it is crucial to assume  $\frac{1}{2} < s < t < 1$ , so as to obtain accurate estimates of local wave-vectors for reasonable large wavenumbers. It is proved in the next section that  $s < t$  guarantees precise estimates in the case of banded wave-like components. Here is some notation for the general curvelet transform:

1. The scaling matrix

$$A_a = \begin{pmatrix} a^t & 0 \\ 0 & a^s \end{pmatrix},$$

where  $a$  is the distance from the center of one curvelet to the origin of Fourier domain.

- The rotation angle  $\theta$  and rotation matrix

$$R_\theta = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}.$$

- The unit vector  $e_\theta = (\cos \theta, \sin \theta)^T$  of rotation angle  $\theta$  and  $T$  denotes a transpose.
- $\theta_\alpha$  represents the argument of given vector  $\alpha$ .
- $w(x)$  of  $x \in \mathbb{R}^2$  denotes the mother curvelet, which is in the Schwartz class and has a nonnegative, radial, real-valued, smooth Fourier transform  $\widehat{w}(\xi)$  with support equal to the unit ball  $B_1(0)$  in the Fourier domain. The mother curvelet is required to obey the admissibility condition:  $\exists 0 < c_1 < c_2 < \infty$  such that

$$c_1 \leq \int_0^{2\pi} \int_1^\infty a^{-(t+s)} |\widehat{w}(A_a^{-1} R_\theta^{-1}(\xi - a \cdot e_\theta))|^2 a da d\theta \leq c_2$$

for any  $|\xi| \geq 1$ .

With the notation above, we are ready to define a family of curvelets through scaling, modulation, and translation as follows, controlled by the geometric parameter  $s$  and  $t$ .

DEFINITION 1.1. For  $\frac{1}{2} < s < t < 1$ , define  $\widehat{w_{a\theta b}}(\xi) = \widehat{w}(A_a^{-1} R_\theta^{-1}(\xi - a \cdot e_\theta)) e^{-2\pi i b \cdot \xi} a^{-\frac{t+s}{2}}$  as a general curvelet in the Fourier domain. Equivalently, in the space domain, the corresponding general curvelet is

$$\begin{aligned} w_{a\theta b}(x) &= \int_{\mathbb{R}^2} \widehat{w}(A_a^{-1} R_\theta^{-1}(\xi - a \cdot e_\theta)) e^{-2\pi i b \cdot \xi} e^{2\pi i \xi \cdot x} a^{-\frac{t+s}{2}} d\xi \\ &= a^{\frac{t+s}{2}} \int_{\mathbb{R}^2} \widehat{w}(y) e^{-2\pi i b \cdot (R_\theta A_a y + a \cdot e_\theta)} e^{2\pi i x \cdot (R_\theta A_a y + a \cdot e_\theta)} dy \\ &= a^{\frac{t+s}{2}} e^{2\pi i a(x-b) \cdot e_\theta} w(A_a R_\theta^{-1}(x - b)). \end{aligned}$$

In such a way, a family of curvelets  $\{w_{a\theta b}(x), a \in [1, \infty), \theta \in [0, 2\pi), b \in \mathbb{R}^2\}$  is constructed.

By definition, the Fourier transform  $\widehat{w_{a\theta b}}(\xi)$  has an ellipse-like support  $\{\xi : |A_a^{-1} R_\theta^{-1}(\xi - a \cdot e_\theta)| \leq 1\}$  centered at  $a \cdot e_\theta$  with a major radius  $a^t$  and a minor radius  $a^s$ . It is natural to require  $a \geq 1$  in order to keep the consideration regarding the shape of curvelets valid. Meanwhile,  $w_{a\theta b}(x)$  is centered in space at  $b$  with an essential support of length  $O(a^{-s})$  and width  $O(a^{-t})$ . By this appropriate construction, each curvelet is scaled to have the same  $L^2$  norm with the mother curvelet  $w(x)$ . Notice that if  $s = \frac{1}{2}$  and  $t = 1$ , these functions would be qualitatively similar to standard two-dimensional (2D) curvelets. When  $s = t$ , these functions would become general wave packets in [33]. As  $s = t$  approaching 1 or  $\frac{1}{2}$ , they are getting close to wavelets or wave atoms [11], respectively.

Similar to the classical curvelet transform, the general curvelet transform is defined to be the inner product of a given signal and each curvelet as follows.

DEFINITION 1.2. The general curvelet transform of a function  $f(x)$  is a function

$$\begin{aligned} W_f(a, \theta, b) &= \langle w_{a\theta b}, f \rangle = \int_{\mathbb{R}^2} \overline{w_{a\theta b}(x)} f(x) dx \\ &= \langle \widehat{w_{a\theta b}}, \widehat{f} \rangle = \int_{\mathbb{R}^2} \overline{\widehat{w_{a\theta b}}(\xi)} \widehat{f}(\xi) d\xi \end{aligned}$$

for  $a \in [1, \infty)$ ,  $\theta \in [0, 2\pi)$ ,  $b \in \mathbb{R}^2$ .

If the Fourier transform  $\widehat{f}(\xi)$  vanishes for  $|\xi| < 1$ , one can check the following  $L^2$  norms equivalence up to a uniform constant factor following the proof of Theorem 1 in [5], i.e.,

$$c_1 \int |f(x)|^2 dx \leq \int |W_f(a, \theta, b)|^2 a da d\theta db \leq c_2 \int |f(x)|^2 dx.$$

Below is a simple example to show how the synchrosqueezing technique estimates local wave-vectors. Let us consider a plane wave function

$$f(x) = \alpha e^{2\pi i N \beta \cdot x},$$

where  $\alpha$  and  $\beta$  are nonzero constants of order  $O(1)$  and  $N$  is a sufficiently large constant. The general curvelet transform of  $f(x)$  is

$$\begin{aligned} W_f(a, \theta, b) &= \int_{\mathbb{R}^2} \alpha e^{2\pi i N \beta \cdot x} a^{\frac{s+t}{2}} w(A_a R_\theta^{-1}(x-b)) e^{-2\pi i a(x-b) \cdot e_\theta} dx \\ &= a^{-\frac{s+t}{2}} \alpha \int_{\mathbb{R}^2} e^{2\pi i N \beta \cdot (b + R_\theta A_a^{-1} y)} w(y) e^{-2\pi i a^{1-t} y_1} dy \\ &= a^{-\frac{s+t}{2}} \alpha e^{2\pi i N \beta \cdot b} \overline{\widehat{w}(A_a^{-1} R_\theta^{-1}(a \cdot e_\theta - N\beta))}. \end{aligned}$$

Notice that  $\widehat{w}(\xi)$  is compactly supported in the unit ball, and  $W_f(a, \theta, b)$  is able to provide a preliminary estimate of the local wave-vector  $N\beta$ , since the nonzero  $W_f(a, \theta, b)$  is located in the regime

$$|A_a^{-1} R_\theta^{-1}(a \cdot e_\theta - N\beta)| \leq 1.$$

This implies that, for each  $b$ ,  $W_f(a, \theta, b)$  has a support of length  $O(|N\beta|^t)$  and width  $O(|N\beta|^s)$  around the wave-vector  $N\beta$  in the variable  $a$  and  $\theta$ . Nevertheless, the resolution of this estimate is too low. Further observation tells us that the oscillation of  $W_f(a, \theta, b)$  in the  $b$  variable in fact uncovers  $N\beta$  by

$$\begin{aligned} \nabla_b W_f(a, \theta, b) &= 2\pi i N \beta a^{-\frac{s+t}{2}} \alpha e^{2\pi i N \beta \cdot b} \overline{\widehat{w}(A_a^{-1} R_\theta^{-1}(a \cdot e_\theta - N\beta))} \\ &= (2\pi i W_f(a, \theta, b)) N \beta, \end{aligned}$$

where  $\nabla_b = (\partial_{b_1}, \partial_{b_2})^T$  denotes a partial derivative operator in the variable  $b$ . This motivates the definition of the local wave-vector estimation for a general function  $f(x)$  as follows.

DEFINITION 1.3. *The local wave-vector estimation of a function  $f(x)$  at  $(a, \theta, b)$  is*

$$(1) \quad v_f(a, \theta, b) = \frac{\nabla_b W_f(a, \theta, b)}{2\pi i W_f(a, \theta, b)}$$

for  $a \in [1, \infty)$ ,  $\theta \in [0, 2\pi)$ ,  $b \in \mathbb{R}^2$  such that  $W_f(a, \theta, b) \neq 0$ .

It is remarkable that  $v_f(a, \theta, b)$  estimates the local wave-vectors independently of the amplitude  $\alpha$  or the position  $b$ . Hence, if the coefficients with the same  $v_f$  are reallocated together, then there would be a sharpened phase space representation of  $f(x)$ , a clear picture of nonzero energy concentrating around local wave-vectors. Mathematically speaking, the synchrosqueezed energy distribution is defined as follows.

DEFINITION 1.4. *Given  $f(x)$ ,  $W_f(a, \theta, b)$ , and  $v_f(a, \theta, b)$ , the synchrosqueezed energy distribution  $T_f(v, b)$  is*

$$(2) \quad T_f(v, b) = \int |W_f(a, \theta, b)|^2 \delta(\Re v_f(a, \theta, b) - v) a da d\theta$$

for  $v \in \mathbb{R}^2$ ,  $b \in \mathbb{R}^2$ .

Here  $\delta$  denotes the Dirac delta function and  $\Re v_f(a, \theta, b)$  means the real part of a complex number  $v_f(a, \theta, b)$ . For  $f(x)$  with Fourier transform vanishing for  $|\xi| < 1$ , the norm equivalence

$$\int T_f(v, b) dv db = \int |W_f(a, \theta, b)|^2 da d\theta db \approx \|f\|_2^2$$

holds as a consequence of the  $L^2$  norm equivalence between  $W_f(a, \theta, b)$  and  $f(x)$ .

Equipped with the definitions above, let us consider now a general function of the form

$$f(x) = e^{-(\phi(x)-c)^2/\sigma^2} \alpha(x) e^{2\pi i N \phi(x)}$$

with a smooth amplitude  $\alpha(x)$ , a smooth phase  $\phi(x)$ , a banded parameter  $\sigma = \Theta(N^{-\eta})$  ( $\eta < t$ ), and a sufficiently large  $N$ . It will be shown that the general curvelet transform  $W_f(a, \theta, b)$  for each  $b$  is essentially supported in the following set:

$$(3) \quad \{(a, \theta) : |A_a^{-1} R_\theta^{-1}(a \cdot e_\theta - N \nabla \phi(b))| \leq 1\}.$$

In the meantime,  $v_f(a, \theta, b)$  is an accurate estimation of the local wave-vector  $N \nabla \phi$  independent of  $a$  and  $\theta$ , which implies that the essential support of the synchrosqueezed energy distribution  $T_f(v, b)$  in  $v$  is concentrating around  $N \nabla \phi$  at each location.

**1.2. Mode decomposition.** In the previous subsection, the property of the SSCT, that it concentrates the energy of a banded wave-like component around its wave-vectors, has been informally discussed. In what follows, the procedure of the mode decomposition after synchrosqueezing will be presented. For simplicity, let

$$f(x) = e^{-(\phi_1(x)-c_1)^2/\sigma_1^2} \alpha_1(x) e^{2\pi i N \phi_1(x)} + e^{-(\phi_2(x)-c_2)^2/\sigma_2^2} \alpha_2(x) e^{2\pi i N \phi_2(x)}$$

with smooth amplitudes  $\alpha_1(x)$  and  $\alpha_2(x)$ , banded parameters  $\sigma_1$  and  $\sigma_2$  of order  $\Theta(N^{-\eta})$  ( $\eta < t$ ), and smooth phases  $N \phi_1(x)$  and  $N \phi_2(x)$  for a sufficiently large  $N$ . Let us assume that at each position the local wave-vectors  $N \nabla \phi_1(x)$  and  $N \nabla \phi_2(x)$  are sufficiently large and well-separated from each other.

The decomposition relies on four steps, summarized below:

1. By (3), the essential supports of  $W_{f_1}(a, \theta, b)$  and  $W_{f_2}(a, \theta, b)$  are contained in the sets

$$P_1 = \{(a, \theta, b) : |A_a^{-1} R_\theta^{-1}(a \cdot e_\theta - N \nabla \phi_1(b))| \leq 1\},$$

$$P_2 = \{(a, \theta, b) : |A_a^{-1} R_\theta^{-1}(a \cdot e_\theta - N \nabla \phi_2(b))| \leq 1\}.$$

Because both  $|N \nabla \phi_1(b)|$  and  $|N \nabla \phi_2(b)|$  are large, and  $N \nabla \phi_1(x)$  and  $N \nabla \phi_2(x)$  are sufficiently well-separated, these two sets are essentially disjoint. Hence, the essential support of  $W_f(a, \theta, b)$  is separated into two essentially disjoint sets, each of which corresponds to one component in  $f(x)$ .

2. The separation in step 1 implies that for each  $b$

$$v_f(a, \theta, b) = v_{f_1}(a, \theta, b) \approx N \nabla \phi_1 \text{ in } P_1$$

and

$$v_f(a, \theta, b) = v_{f_2}(a, \theta, b) \approx N \nabla \phi_2 \text{ in } P_2.$$

Though  $v_f(a, \theta, b)$  is defined wherever  $W_f(a, \theta, b) \neq 0$ , it is only relevant when  $|W_f(a, \theta, b)|$  is above a significant level, as will be shown in Theorem 2.3, (5).

Hence, it is sufficient to compute  $v_f(a, \theta, b)$  in these disjoint essential supports  $P_1$  and  $P_2$  to estimate local wave-vectors of each component.

3. The separation in step 2 shows that  $T_f(v, b)$  is essentially concentrating around two well-separated 2D surfaces

$$S_1 = \{(N|\nabla\phi_1(b)|, \theta_{\nabla\phi_1(b)}, b) : b \in \mathbb{R}^2\}$$

and

$$S_2 = \{(N|\nabla\phi_2(b)|, \theta_{\nabla\phi_2(b)}, b) : b \in \mathbb{R}^2\}.$$

Hence, the essential support of  $T_f(v, b)$  separates into two well disjoint sets  $U_1$  and  $U_2$ .

4. Notice that  $T_f(v, b) = T_{f_1}(v, b)$  in  $U_1$  and, respectively,  $T_f(v, b) = T_{f_2}(v, b)$  in  $U_2$ . Once  $U_1$  and  $U_2$  are identified by some clustering technique, each component of  $f(x)$  can be recovered by

$$\begin{aligned} f_1(x) &= \int_{\mathbb{R}v_f(a, \theta, b) \in U_1} \tilde{w}_{a\theta b}(x) W_f(a, \theta, b) da d\theta db, \\ f_2(x) &= \int_{\mathbb{R}v_f(a, \theta, b) \in U_2} \tilde{w}_{a\theta b}(x) W_f(a, \theta, b) da d\theta db, \end{aligned}$$

where the set of functions  $\{\tilde{w}_{a\theta b}(x), a \in [1, \infty), \theta \in [0, 2\pi), b \in \mathbb{R}^2\}$  is the dual frame of  $\{w_{a\theta b}(x), a \in [1, \infty), \theta \in [0, 2\pi), b \in \mathbb{R}^2\}$ .

The synchrosqueezing steps 2 and 3 are indispensable, because they improve the resolution of original results significantly so that clustering is possible for decomposition. In step 4, the reconstruction is based on the Calderon-type reconstruction formula for the reason that curvelet transforms, unlike wavelet transforms in [9], do not have a reconstruction formula that integrates their coefficients over the scale parameter with a proper weight. In effect, numerical examples in [9] are based on the Calderon-type reconstruction formula, since it works more robustly in noisy cases.

**1.3. Related work.** There is another interesting line of work for mode decomposition, which is the empirical mode decomposition (EMD) initiated and refined by Huang et al. in [20, 21]. Starting from the most oscillatory mode, the EMD method decomposes a signal into a collection of intrinsic mode functions and estimates instantaneous frequencies via the Hilbert transform. However, the dependence on local extrema limits its applications in noisy cases. To address the robustness problem, some variants were proposed in [18, 31]. Following the idea of EMD, there are two existing methods for high dimensional mode decomposition. The first is based on high dimensional interpolation [26, 27, 23, 24] and the second applies a 1D decomposition to each dimension and then combines the results with a proper combination strategy [19, 25, 32]. In spite of their considerable success, these existing methods in this research line are not suitable to separate two modes with similar wave-numbers but different wave-vectors due to the lack of an anisotropic angular separation, as illustrated in detail in [33].

Following the same methodology of extracting modes one by one from the most oscillatory one, Hou and Shi proposed an optimization scheme for mode decomposition in [16, 17]. Inspired by recent developments of compressive sensing, the first paper, [16], is based on total variations, while the second one, [17], is based on the sparse representation in a data-driven time-frequency dictionary. The convergence of the

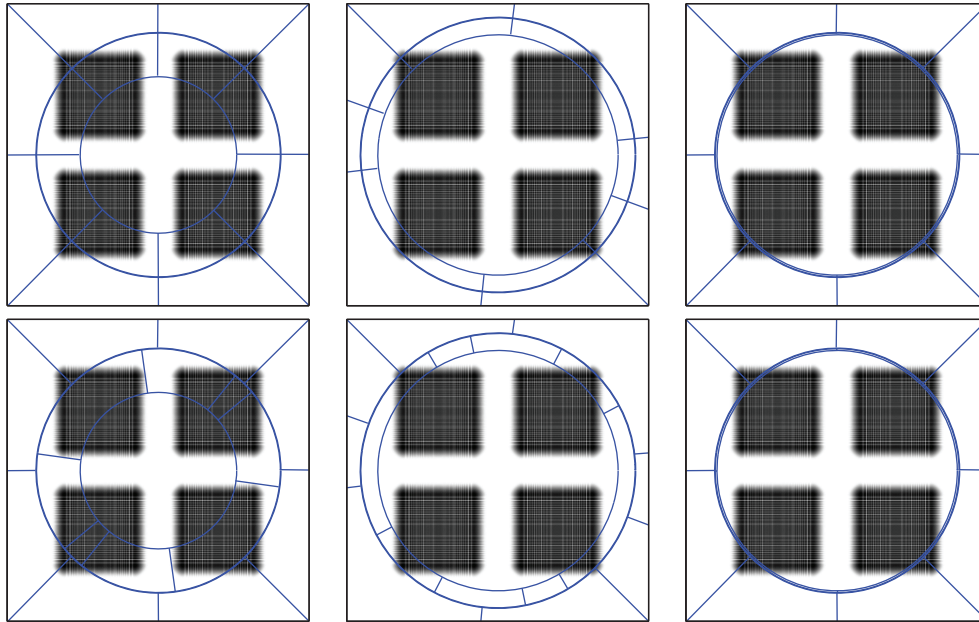


FIG. 3. Detected Fourier boundaries provided by the 2D empirical curvelet transform for the superposition  $f(x)$  in (4). Top: Results using the curvelet EWT-I approach. From left to right, we apply the “tophat,” the “morpho,” and the “poly” boundary detection method, respectively. Bottom: Results using the curvelet EWT-II approach. From left to right, we apply the “tophat,” the “morpho,” and the “poly” boundary detection method, respectively. Each figure shows the 2D Fourier power spectrum of  $f(x)$  and the detected Fourier boundaries (in blue) using a certain approach.

data-driven time-frequency analysis method under a certain sparsity assumption was proved recently in [15]. However, the analysis of high dimensional case is still under active research.

Recently, 1D empirical transforms were proposed in [13] with an application in mode decomposition using adaptive time frequency representations. The idea is generalized to 2D in [14] with three main algorithms. Algorithm 1 is based on the Fourier spectra of 1D data slices and, hence, lacks the anisotropic angular separation for the same reason as the 2D EMD methods using 1D data slices. Algorithm 2 is based on a 2D isotropic wavelet transform. Hence, it cannot distinguish two different modes if they share the same wave number. Algorithm 3 generalizes some anisotropic transforms, such as the ridgelet transform and the curvelet transform. However, the Fourier boundaries detection method in this algorithm may not be robust and sometimes leads to misleading tilings in the Fourier domain for these transforms. For example, let us consider a superposition of two deformed plane waves given by

$$(4) \quad f(x) = e^{2\pi i N \phi_1(x)} + e^{2\pi i N \phi_2(x)},$$

where

$$\begin{aligned} \phi_1(x) &= \phi_1(x_1, x_2) = x_1 + x_2 + \beta \sin(2\pi x_1) + \beta \sin(2\pi x_2), \\ \phi_2(x) &= \phi_2(x_1, x_2) = -x_1 + x_2 - \beta \sin(2\pi x_1) + \beta \sin(2\pi x_2) \end{aligned}$$

with  $N = 100$  and  $\beta = 0.1$ . Figure 3 shows the detected Fourier boundaries provided by different Fourier boundaries detection methods for the example  $f(x)$  in (4). These



detected Fourier boundaries cannot separate the supports of the wave-like components in  $f(x)$  in the Fourier domain, which will lead to misleading decomposition results. Hence, the 2D Empirical transforms may not be suitable to address the problems discussed in this paper.

The rest of the paper is organized as follows. The main theoretical results of SSCT are presented in section 2. We prove that SSCT is able to estimate the local wave-vectors under some well-separation condition of the local wave-vectors of multiple highly oscillatory components. In section 3, a discrete analog of SSCT and some clustering methods in the phase space are introduced. Section 4 compares several numerical examples on local wave-vector estimation using SSWPT and SSCT, and provides decomposition examples with synthetic and real data to demonstrate the proposed properties of SSCT. Finally, this article will end with some discussions in section 5.

**2. Analysis of the transform.** In this section, we define a class of superpositions of multiple banded components with well-separated local wave-vectors and prove that the SSCT is able to estimate these local wave-vectors accurately. Throughout the analysis, the scaling parameters  $s$  and  $t$  are fixed such that  $\frac{1}{2} < s < t < 1$  and  $\eta < t$ .

**DEFINITION 2.1.** For any  $c \in \mathbb{R}$ ,  $N > 0$ , and  $M > 0$ , a function  $f(x) = e^{-(\phi(x)-c)^2/\sigma^2} \alpha(x) e^{2\pi i N \phi(x)}$  is a banded intrinsic mode function of type  $(M, N, \eta)$  if  $\sigma \geq N^{-\eta}$ ,  $\alpha(x)$  and  $\phi(x)$  satisfy

$$\begin{aligned} \alpha(x) &\in C^\infty, \quad |\nabla \alpha| \leq M, \quad 1/M \leq \alpha \leq M, \\ \phi(x) &\in C^\infty, \quad 1/M \leq |\nabla \phi| \leq M, \quad |\nabla^2 \phi| \leq M. \end{aligned}$$

The banded term  $e^{-(\phi(x)-c)^2/\sigma^2}$  is introduced to study the limit of the bandwidth of these components. In the space domain, a general curvelet at the scale  $a = O(N)$  has a width  $O(N^t)$ .  $\sigma \geq N^\eta$  with  $\eta < t$  indicates that the bandwidth  $\sigma$  of  $e^{-(\phi(x)-c)^2/\sigma^2} \alpha(x) e^{2\pi i N \phi(x)}$  can be almost as narrow as the width of a general curvelet that is sharing the same wave number  $O(N)$ . If  $\eta$  tends to  $-\infty$ , the banded intrinsic mode function will become the one discussed in [33]. So, the model in this article is more general.

**DEFINITION 2.2.** A function  $f(x)$  is a well-separated superposition of type  $(M, N, \eta, K)$  if

$$f(x) = \sum_{k=1}^K f_k(x),$$

where each  $f_k(x) = e^{-(\phi_k(x)-c_k)^2/\sigma_k^2} \alpha_k(x) e^{2\pi i N \phi_k(x)}$  is a banded intrinsic mode function of type  $(M, N, \eta)$  and they satisfy the separation condition:  $\forall a \in [1, \infty)$  and  $\forall \theta \in [0, 2\pi)$ , there is at most one banded intrinsic mode function  $f_k$  satisfying that

$$|A_a^{-1} R_\theta^{-1} (a \cdot e_\theta - N \nabla \phi_k(b))| \leq 1.$$

We denote by  $F(M, N, \eta, K)$  the set of all such functions.

Recall that  $W_f(a, \theta, b)$  is the general curvelet transform of a function  $f(x)$  with geometric scaling parameter  $\frac{1}{2} < s < t < 1$ , and  $v_f(a, \theta, b)$  is the local wave-vector estimation. The following theorem is the main theoretical result for the SSCT. It studies the dependence of  $N$  in the class of well-separated superposition of type  $F(M, N, \eta, K)$  on an accuracy parameter  $\epsilon$  in the estimate of local wave vectors provided by the SSCT.

**THEOREM 2.3.** For a function  $f(x)$ , which is a well-separated superposition of some type  $(M, N, \eta, K)$ , and any  $\varepsilon > 0$ , define

$$(5) \quad R_{f,\varepsilon} = \left\{ (a, \theta, b) : |W_f(a, \theta, b)| \geq a^{-\frac{s+t}{2}} \sqrt{\varepsilon} \right\}$$

and

$$Z_{f,k} = \left\{ (a, \theta, b) : |A_a^{-1} R_\theta^{-1} (a \cdot e_\theta - N \nabla \phi_k(b))| \leq 1 \right\}$$

for  $1 \leq k \leq K$ . For fixed  $M, s, t, \eta$ , and  $\varepsilon$ , there exists  $N_0(M, s, t, \eta, \varepsilon) > 0$  such that for any  $N > N_0(M, s, t, \eta, \varepsilon)$  and  $f(x) \in F(M, N, \eta, K)$  the following statements hold:

- (i)  $\{Z_{f,k} : 1 \leq k \leq K\}$  are disjoint and  $R_{f,\varepsilon} \subset \bigcup_{1 \leq k \leq K} Z_{f,k}$ .
- (ii) For any  $(a, \theta, b) \in R_{f,\varepsilon} \cap Z_{f,k}$ ,

$$\frac{|v_f(a, \theta, b) - N \nabla \phi_k(b)|}{|N \nabla \phi_k(b)|} \lesssim \sqrt{\varepsilon}.$$

For simplicity, the notation  $O(\cdot)$ ,  $\lesssim$ , and  $\gtrsim$  is used when the implicit constants may only depend on  $M$  and  $K$ . The proof of the theorem relies on several lemmas. The following one estimates  $W_f(a, \theta, b)$ .

**LEMMA 2.4.** Suppose

$$\Omega = \left\{ (a, \theta) : a \in \left( \frac{N}{2M}, 2MN \right), \exists k \text{ such that (s.t.) } |\theta_{\nabla \phi_k(b)} - \theta| < \theta_0 \right\},$$

where  $\theta_0 = \arcsin\left(\left(\frac{M}{N}\right)^{t-s}\right)$ . Under the assumption of the theorem, the following estimation of  $W_f(a, \theta, b)$  holds for any  $\varepsilon$ , when  $N$  is larger than a constant  $N_0(M, s, t, \eta, \varepsilon) > 0$ :

- 1. If  $(a, \theta) \in \Omega$ ,

$$\begin{aligned} &W_f(a, \theta, b) \\ &= a^{-\frac{s+t}{2}} \left( \sum_{k: |\theta_{\nabla \phi_k(b)} - \theta| < \theta_0} f_k(b) \widehat{w}(A_a^{-1} R_\theta^{-1} (a \cdot e_\theta - N \nabla \phi_k(b))) + O(\varepsilon) \right). \end{aligned}$$

- 2. Otherwise,

$$W_f(a, \theta, b) = a^{-\frac{s+t}{2}} O(\varepsilon).$$

*Proof.* We only need to discuss the case when  $K = 1$ . The result for general  $K$  is an easy extension by the linearity of general curvelet transform. Suppose  $f(x)$  contains a single banded intrinsic mode function of type  $(M, N, \eta)$

$$f(x) = e^{-(\phi(x)-c)^2/\sigma^2} \alpha(x) e^{2\pi i N \phi(x)}.$$

We claim that when  $N$  is large enough, the approximation of  $W_f(a, \theta, b)$  holds. By the definition of general curvelet transform, it holds that

$$\begin{aligned} W_f(a, \theta, b) &= \int_{R^2} f(x) a^{\frac{s+t}{2}} w(A_a R_\theta^{-1} (x - b)) e^{-2\pi i a (x-b) \cdot e_\theta} dx \\ &= a^{-\frac{s+t}{2}} \int_{R^2} f(b + R_\theta A_a^{-1} y) w(y) e^{-2\pi i a^{1-t} y_1} dy. \end{aligned}$$

*Step 1.* We start with the proof of (ii) first.

Let  $h(y) = w(y)e^{-(\phi(b+R_\theta A_a^{-1}y)-c)^2/\sigma^2} \alpha(b + R_\theta A_a^{-1}y)$  and  $g(y) = 2\pi(N\phi(b + R_\theta A_a^{-1}y) - a^{1-t}y_1)$ , and then we have

$$W_f(a, \theta, b) = a^{-\frac{s+t}{2}} \int_{\mathbb{R}^2} h(y)e^{ig(y)} dy$$

with real smooth functions  $h(y)$  and  $g(y)$ . Consider the differential operator

$$L = \frac{1}{i} \frac{\langle \nabla g, \nabla \rangle}{|\nabla g|^2}.$$

If  $|\nabla g|$  does not vanish, we have

$$Le^{ig} = \frac{\langle \nabla g, i\nabla g e^{ig} \rangle}{i|\nabla g|^2} = e^{ig}.$$

By the definition of  $w(y)$ , we know  $h(y)$  is decaying rapidly at infinity. Then we can apply integration by parts to get

$$\int_{\mathbb{R}^2} h e^{ig} dy = \int_{\mathbb{R}^2} h (L e^{ig}) dy = - \int_{\mathbb{R}^2} \nabla \cdot \left( \frac{h \nabla g}{i|\nabla g|^2} \right) e^{ig} dy.$$

Hence, we need to estimate  $|\nabla \cdot (\frac{h \nabla g}{i|\nabla g|^2})|$ . Because

$$\nabla \cdot \left( \frac{h \nabla g}{i|\nabla g|^2} \right) = \frac{1}{i} \left( \frac{\nabla h \cdot \nabla g}{|\nabla g|^2} + h \nabla \cdot \left( \frac{\nabla g}{|\nabla g|^2} \right) \right)$$

and  $|h(y)| \lesssim 1$ , we only need to estimate  $|\frac{\nabla h \cdot \nabla g}{|\nabla g|^2}|$  and  $|\frac{\partial^2 g}{\partial y_i \partial y_j} \frac{1}{|\nabla g|^2}|$  for  $i, j = 1, 2$ .

Let  $z = (z_1, z_2)^T = R_\theta^{-1} \nabla \phi(b + R_\theta A_a^{-1}y)$ ,  $v_1 = N A_a^{-1} R_\theta^{-1} \nabla \phi(b + R_\theta A_a^{-1}y)$ , and  $v_2 = (a^{1-t}, 0)^T$ , and then  $\nabla g(y) = 2\pi(v_1 - v_2) = 2\pi((Nz_1 - a)a^{-t}, Na^{-s}z_2)$ .

*Case 1.*  $a \notin (\frac{N}{2M}, 2MN)$ .

When  $a \geq 2MN$ ,

$$|\nabla g(y)| \geq a^{1-t} - MN a^{-t} = \frac{a^{1-t}}{2} + \left(\frac{a}{2} - MN\right) a^{-t} \geq \frac{a^{1-t}}{2} \gtrsim N^{1-t}.$$

When  $a \leq \frac{N}{2M}$ ,

$$|\nabla g(y)| \gtrsim \frac{Na^{-t}}{M} - a^{1-t} \geq \frac{Na^{-t}}{2M} \gtrsim N^{1-t}.$$

So

$$(6) \quad |\nabla g(y)| \gtrsim N^{1-t}$$

for  $a \notin (\frac{N}{2M}, 2MN)$ .

If  $a \geq 2MN$ , then  $|\frac{\partial^2 g}{\partial y_i \partial y_j}| \lesssim Na^{-2s} \lesssim N^{1-2s}$ , implying that

$$\left| \frac{\partial^2 g}{\partial y_i \partial y_j} \frac{1}{|\nabla g|^2} \right| \lesssim N^{1-2s} / N^{2-2t} = \frac{1}{N^{1-2(t-s)}}.$$

Since  $|z| \geq \frac{1}{M}$ , either  $|z_1| \geq \frac{1}{\sqrt{2}M}$  or  $|z_2| \geq \frac{1}{\sqrt{2}M}$  holds. If  $a \leq \frac{N}{2M}$ , then

$$\begin{aligned} \left| \frac{\partial^2 g}{\partial y_i \partial y_j} \frac{1}{|\nabla g|^2} \right| &\lesssim \frac{Na^{-2s}}{(Nz_1 - a)^2 a^{-2t} + N^2 a^{-2s} z_2^2} \\ &= \frac{1}{(z_1 - \frac{a}{N})^2 Na^{-2(t-s)} + Nz_2^2} \\ &\lesssim \max \left\{ \frac{1}{N^{1-2(t-s)}}, \frac{1}{N} \right\} \\ &= \frac{1}{N^{1-2(t-s)}}. \end{aligned}$$

In sum,

$$(7) \quad \left| \frac{\partial^2 g}{\partial y_i \partial y_j} \frac{1}{|\nabla g|^2} \right| \lesssim \frac{1}{N^{1-2(t-s)}}$$

for  $a \notin (\frac{N}{2M}, 2MN)$ .

Notice that the dominant term of  $\nabla h$  is

$$w(y)\alpha(b + R_\theta A_a^{-1}y)e^{-(\phi(b + R_\theta A_a^{-1}y) - c)^2/\sigma^2} \cdot \frac{-2(\phi(b + R_\theta A_a^{-1}y) - c)}{\sigma^2} A_a^{-1}z$$

and the other terms are of order 1. Because  $e^{-\frac{x^2}{\sigma^2}} \cdot \frac{|x|}{\sigma^2} \leq e^{-\frac{1}{2}} \cdot \frac{1}{\sigma\sqrt{2}}$ ,

$$\left| \frac{\nabla h \cdot \nabla g}{|\nabla g|^2} \right| \lesssim \frac{1}{\sigma} \left| \frac{(A_a^{-1}z) \cdot \nabla g}{|\nabla g|^2} \right| + \left| \frac{1}{|\nabla g|} \right| \lesssim N^\eta \left| \frac{(A_a^{-1}z) \cdot \nabla g}{|\nabla g|^2} \right| + \frac{1}{N^{1-t}}.$$

Recall that  $\nabla g = 2\pi(NA_a^{-1}z - (a^{1-t}, 0)^T)$ , and then

$$\frac{(A_a^{-1}z) \cdot \nabla g}{|\nabla g|^2} \approx \frac{(Nz_1 - a)a^{-2t}z_1 + Na^{-2s}z_2^2}{(Nz_1 - a)^2 a^{-2t} + N^2 a^{-2s}z_2^2}.$$

If  $z_1 z_2 \neq 0$ , then  $|\frac{Na^{-2s}z_2^2}{N^2 a^{-2s}z_2^2}| = \frac{1}{N}$  and  $|\frac{(Nz_1 - a)a^{-2t}z_1}{(Nz_1 - a)^2 a^{-2t}}| \approx \frac{1}{|Nz_1 - a|} \approx \frac{1}{N}$ , which implies that  $|\frac{(A_a^{-1}z) \cdot \nabla g}{|\nabla g|^2}| \lesssim \frac{1}{N}$ . If  $z_1 z_2 = 0$ , then it is easy to check that  $|\frac{(A_a^{-1}z) \cdot \nabla g}{|\nabla g|^2}| \approx \frac{1}{N}$ . Hence,

$$(8) \quad \left| \frac{\nabla h \cdot \nabla g}{|\nabla g|^2} \right| \lesssim N^\eta \left| \frac{(A_a^{-1}z) \cdot \nabla g}{|\nabla g|^2} \right| + \frac{1}{N^{1-t}} \lesssim \frac{1}{N^{1-\eta}} + \frac{1}{N^{1-t}} \lesssim \frac{1}{N^{1-t}}$$

for  $a \notin (\frac{N}{2M}, 2MN)$ .

By (7) and (8), we have

$$\begin{aligned} \left| \int_{\mathbb{R}^2} h e^{ig} dy \right| &= \left| \int_{\mathbb{R}^2} \nabla \cdot \left( \frac{h \nabla g}{i|\nabla g|^2} \right) e^{ig} dy \right| \\ &\lesssim \left| \nabla \cdot \left( \frac{h \nabla g}{i|\nabla g|^2} \right) \right| (\|w\|_{L^1} + \|\nabla w\|_{L^1}) \lesssim \frac{1}{N^{1-t}} \end{aligned}$$

for  $a \notin (\frac{N}{2M}, 2MN)$ . So,

$$W_f(a, \theta, b) = a^{-\frac{s+t}{2}} O(\varepsilon)$$

when  $N \gtrsim \varepsilon^{\frac{-1}{1-t}}$  and  $a \notin (\frac{N}{2M}, 2MN)$ .

Case 2.  $a \in (\frac{N}{2M}, 2MN)$  and  $|\theta_{\nabla\phi(b)} - \theta| \geq \theta_0$ .

Observing that  $\nabla g(y) = 2\pi A_a^{-1} R_\theta^{-1} (N \nabla \phi(b + R_\theta A_a^{-1} y) - a \cdot e_\theta)$ , we can expect  $|\nabla g|$  is large when  $\theta_{\nabla \phi(b)}$  is far away from  $\theta$ . Notice that  $w(y)$  is in the Schwartz class,  $\exists C_m > 0$  such that  $|w(y)| \leq \frac{C_m}{y^m}$  for  $|y| \geq 1$  and any  $m$  large enough. So

$$W_f(a, \theta, b) = a^{-\frac{s+t}{2}} \left( \int_{|y| \lesssim \varepsilon^{-1/m}} f(b + R_\theta A_a^{-1} y) w(y) e^{-2\pi i a^{1-t} y_1} dy + O(\varepsilon) \right).$$

Define  $D = \{y : |y| \lesssim \varepsilon^{-1/m}\}$  and  $D_+ = \{y : |y| \lesssim \varepsilon^{-1/m} + 1\}$ . Suppose  $X_D(y)$  is a positive and smooth function compactly supported in  $D_+$  such that  $X_D(y) = 1$  if  $y \in D$ ,  $\|X_D\|_{L^\infty} \leq 1$ ; then

$$W_f(a, \theta, b) = a^{-\frac{s+t}{2}} \left( O(\varepsilon) + \int_{D_+} X_D(y) h(y) e^{ig(y)} dy \right).$$

If  $|\nabla g(y)|$  is not vanishing in  $D_+$ , then apply the integral by parts to get

$$\int_{D_+} X_D h e^{ig} dy = \int_{D_+} X_D h (L e^{ig}) dy = - \int_{D_+} \nabla \cdot \left( \frac{X_D h \nabla g}{i |\nabla g|^2} \right) e^{ig} dy.$$

We are going to estimate  $|\nabla g(y)|$  when  $a \in (\frac{N}{2M}, 2MN)$  and  $|\theta_{\nabla \phi(b)} - \theta| \geq \theta_0$ . By Taylor expansion,

$$\nabla \phi(b + R_\theta A_a^{-1} y) = \nabla \phi(b) + \nabla^2 \phi(b^*) R_\theta A_a^{-1} y,$$

where  $b^*$  is between  $b$  and  $b + R_\theta A_a^{-1} y$ . Notice that

$$|\nabla^2 \phi(b^*) R_\theta A_a^{-1} y| \leq a^{-s} |\nabla^2 \phi(b^*)| |y| \lesssim M a^{-s} (\varepsilon^{-1/m} + 1) \leq \frac{\sin(\theta_0)}{2M}$$

when  $|y| \lesssim \varepsilon^{-1/m} + 1$  and  $(\frac{2M^2}{\sin(\theta_0)})^{1/s} (\varepsilon^{-1/m} + 1)^{1/s} \leq a$ . The latter one holds when  $N \gtrsim (\varepsilon^{-1/m} + 1)^{1/(2s-t)}$  for  $a \in (\frac{N}{2M}, 2MN)$ . So, when these conditions are satisfied, we have

$$\nabla \phi(b + R_\theta A_a^{-1} y) = \nabla \phi(b) + v$$

with  $|v| \leq \frac{\sin(\theta_0)}{2M}$ . Recall the fact  $|\theta_{\nabla \phi(b)} - \theta| \geq \theta_0$ ; then it holds that

$$\begin{aligned} & |A_a^{-1} R_\theta^{-1} (N \nabla \phi(b + R_\theta A_a^{-1} y) - a \cdot e_\theta)| \\ & \geq |N A_a^{-1} R_\theta^{-1} \nabla \phi(b) - (a^{1-t}, 0)^T| - N |A_a^{-1} R_\theta^{-1} v| \\ & \geq \sqrt{(r \cos \alpha - a)^2 a^{-2t} + r^2 a^{-2s} \sin^2 \alpha} - \frac{N}{2M} \sin \theta_0 a^{-s} \\ & \geq r a^{-s} \sin \theta_0 - \frac{N}{2M} \sin \theta_0 a^{-s} \\ & \geq \frac{N}{2M} \sin \theta_0 a^{-s} \\ & \gtrsim N^{1-t}, \end{aligned}$$

where  $\alpha = \theta_{\nabla \phi(b)} - \theta$  and  $r = |N \nabla \phi(b)| \geq \frac{N}{M}$ . Hence, we have

$$(9) \quad |\nabla g(y)| \gtrsim N^{1-t}$$

when  $a \in (\frac{N}{2M}, 2MN)$ ,  $|\theta_{\nabla \phi(b)} - \theta| \geq \theta_0$ ,  $N \gtrsim (\varepsilon^{-1/m} + 1)^{1/(2s-t)}$ , and  $y \in D_+$ .

Next, we move on to estimate  $|\frac{\nabla(X_D h) \cdot \nabla g}{|\nabla g|^2}|$  and  $|\frac{\partial^2 g}{\partial y_i \partial y_j} \frac{1}{|\nabla g|^2}|$  for  $i, j = 1, 2$ , under the conditions that  $a \in (\frac{N}{2M}, 2MN)$ ,  $|\theta_{\nabla \phi(b)} - \theta| \geq \theta_0$ ,  $N \gtrsim (\varepsilon^{-1/m} + 1)^{1/(2s-t)}$ , and  $y \in D_+$ . First,

$$(10) \quad \left| \frac{\partial^2 g}{\partial y_i \partial y_j} \frac{1}{|\nabla g|^2} \right| \leq \frac{Na^{-2s}}{|\nabla g|^2} \leq \frac{N^{1-2s}}{N^{2-2t}} = \frac{1}{N^{1-2(t-s)}}.$$

Second, as for  $|\frac{\nabla(X_D h) \cdot \nabla g}{|\nabla g|^2}|$ , we only need to estimate  $|\frac{(A_a^{-1}z) \cdot \nabla g}{|\nabla g|^2}|$  for a similar reason as in the last case. As we have shown,

$$\frac{(A_a^{-1}z) \cdot \nabla g}{|\nabla g|^2} \approx \frac{(Nz_1 - a)a^{-2t}z_1 + Na^{-2s}z_2^2}{(Nz_1 - a)^2a^{-2t} + N^2a^{-2s}z_2^2}.$$

If  $z_1 = 0$ , then  $|\frac{(A_a^{-1}z) \cdot \nabla g}{|\nabla g|^2}| \approx \frac{1}{N}$ . If  $z_1 \neq 0$  and  $|\frac{z_2}{z_1}| \gtrsim \frac{a^s}{a^t}$ , then  $|z_2| \gtrsim \frac{a^s}{Ma^t}$ , since  $|z| \geq \frac{1}{M}$ . Hence,

$$\begin{aligned} \left| \frac{(A_a^{-1}z) \cdot \nabla g}{|\nabla g|^2} \right| &\lesssim \frac{|(Nz_1 - a)a^{-2t}z_1| + |Na^{-2s}z_2^2|}{N^2a^{-2s}z_2^2} \\ &\lesssim \frac{|Nz_1 - a| \cdot |z_1|}{N^2a^{2(t-s)}z_2^2} + \frac{1}{N} \\ &\lesssim \frac{1}{Na^{t-s}|z_2|} + \frac{1}{N} \\ &\lesssim \frac{1}{N}. \end{aligned}$$

If  $z_1 \neq 0$  and  $|\frac{z_2}{z_1}| \lesssim \frac{a^s}{a^t}$ , then

$$\begin{aligned} \left| \frac{(A_a^{-1}z) \cdot \nabla g}{|\nabla g|^2} \right| &\leq \frac{|(Nz_1 - a)a^{-2t}z_1| + |Na^{-2s}z_2^2|}{|\nabla g|^2} \\ &\lesssim \frac{(|Nz_1| + a)a^{-2t}|z_1| + Na^{-2t}z_1^2}{N^{2-2t}} \\ &\lesssim \frac{1}{N}. \end{aligned}$$

In sum,

$$\left| \frac{(A_a^{-1}z) \cdot \nabla g}{|\nabla g|^2} \right| \lesssim \frac{1}{N},$$

which implies that

$$(11) \quad \left| \frac{\nabla(X_D h) \cdot \nabla g}{|\nabla g|^2} \right| \lesssim \frac{1}{N^{1-t}}.$$

By (10) and (11), we have

$$\begin{aligned} &\left| \int_{D_+} \nabla \cdot \left( \frac{X_D h \nabla g}{i|\nabla g|^2} \right) e^{ig} dy \right| \\ &\lesssim \left| \nabla \cdot \left( \frac{X_D h \nabla g}{i|\nabla g|^2} \right) \right| (\|X_D w\|_{L^1} + \|\nabla(X_D w)\|_{L^1}) \lesssim \frac{1}{N^{1-t}} \end{aligned}$$

for  $a \in (\frac{N}{2M}, 2MN)$ ,  $|\theta_{\nabla\phi(b)} - \theta| \geq \theta_0$ , and  $N \gtrsim (\varepsilon^{-1/m} + 1)^{1/(2s-t)}$ . So,

$$W_f(a, \theta, b) = a^{-\frac{s+t}{2}} O(\varepsilon)$$

when  $a \in (\frac{N}{2M}, 2MN)$ ,  $|\theta_{\nabla\phi(b)} - \theta| \geq \theta_0$ , and

$$N \gtrsim \max \left\{ \left( \varepsilon^{\frac{-1}{m}} + 1 \right)^{\frac{1}{2s-t}}, \varepsilon^{\frac{-1}{1-t}} \right\}.$$

From the discussion in the two cases above, we see that

$$W_f(a, \theta, b) = a^{-\frac{s+t}{2}} O(\varepsilon)$$

if  $a \notin (\frac{N}{2M}, 2MN)$  or  $|\theta_{\nabla\phi(b)} - \theta| \geq \theta_0$  when

$$(12) \quad N \gtrsim \max \left\{ \left( \varepsilon^{\frac{-1}{m}} + 1 \right)^{\frac{1}{2s-t}}, \varepsilon^{\frac{-1}{1-t}} \right\},$$

where  $m$  is any fixed positive integer. Hence, the proof of (ii) when  $K = 1$  is done.

*Step 2.* Henceforth, we move on to prove (i), i.e., to discuss the approximation of  $W_f(a, \theta, b)$ , when  $a \in (\frac{N}{2M}, 2MN)$  and  $|\theta_{\nabla\phi(b)} - \theta| < \theta_0$ . Recall that

$$W_f(a, \theta, b) = a^{-\frac{s+t}{2}} \left( \int_{y \in D} f(b + R_\theta A_a^{-1} y) w(y) e^{-2\pi i a^{1-t} y_1} dy + O(\varepsilon) \right).$$

Our goal is to get the estimate

$$(13) \quad W_f(a, \theta, b) = a^{-\frac{s+t}{2}} \left( \int_{y \in D} f(b) w(y) e^{2\pi i (N \nabla\phi(b) \cdot (R_\theta A_a^{-1} y) - a^{1-t} y_1)} dy + O(\varepsilon) \right)$$

for  $N$  large enough.

First, we are going to show

$$(14) \quad \begin{aligned} & W_f(a, \theta, b) \\ &= a^{-\frac{s+t}{2}} \left( \int_{y \in D} e^{-\frac{(\phi(b)-c)^2}{\sigma^2}} \alpha(b + R_\theta A_a^{-1} y) w(y) e^{2\pi i (N \phi(b + R_\theta A_a^{-1} y) - a^{1-t} y_1)} dy + O(\varepsilon) \right) \end{aligned}$$

for sufficiently large  $N$ . Taylor expansion is applied again to obtain the following three expansions:

$$\phi(b + R_\theta A_a^{-1} y) = \phi(b) + \nabla\phi(b) \cdot (R_\theta A_a^{-1} y) + \frac{1}{2} (R_\theta A_a^{-1} y)^T \nabla^2 \phi(b^*) (R_\theta A_a^{-1} y),$$

where  $b^*$  is between  $b$  and  $b + R_\theta A_a^{-1} y$ ;

$$\begin{aligned} & e^{-\frac{(\phi(b + R_\theta A_a^{-1} y) - c)^2}{\sigma^2}} \\ &= e^{-\frac{(\phi(b) + \nabla\phi(b) \cdot (R_\theta A_a^{-1} y) + \frac{1}{2} (R_\theta A_a^{-1} y)^T \nabla^2 \phi(b^*) (R_\theta A_a^{-1} y) - c)^2}{\sigma^2}} \\ &= e^{-\frac{(\phi(b) - c)^2}{\sigma^2}} + e^{-\frac{(\lambda - c)^2}{\sigma^2}} \cdot \frac{-2(\lambda - c)}{\sigma^2} \left( \nabla\phi(b) \cdot (R_\theta A_a^{-1} y) \right. \\ & \quad \left. + \frac{1}{2} (R_\theta A_a^{-1} y)^T \nabla^2 \phi(b^*) (R_\theta A_a^{-1} y) \right), \end{aligned}$$

where  $\lambda \in [\phi(b), \phi(b) + \nabla\phi(b) \cdot (R_\theta A_a^{-1}y) + \frac{1}{2}(R_\theta A_a^{-1}y)^T \nabla^2\phi(b^*)(R_\theta A_a^{-1}y)]$ ;

$$\alpha(b + R_\theta A_a^{-1}y) = \alpha(b) + \nabla\alpha(b^{**}) \cdot (R_\theta A_a^{-1}y),$$

where  $b^{**}$  is between  $b$  and  $b + R_\theta A_a^{-1}y$ .

The above Taylor expansions help us to estimate the effect of phase function  $\phi(x)$  in the Gaussian term. We claim two estimates as follows:

$$I_1 = \int_{y \in D} \left| e^{-\frac{(\lambda-c)^2}{\sigma^2}} \cdot \frac{-2(\lambda-c)}{\sigma^2} \nabla\phi(b) \cdot (R_\theta A_a^{-1}y) \alpha(b + R_\theta A_a^{-1}y) w(y) \right| dy \leq O(\varepsilon)$$

and

$$I_2 = \int_{y \in D} \left| e^{-\frac{(\lambda-c)^2}{\sigma^2}} \cdot \frac{-2(\lambda-c)}{\sigma^2} \frac{1}{2} (R_\theta A_a^{-1}y)^T \nabla^2\phi(b^*)(R_\theta A_a^{-1}y) \alpha(b + R_\theta A_a^{-1}y) w(y) \right| dy \leq O(\varepsilon).$$

Because  $e^{-\frac{x^2}{\sigma^2}} \cdot \frac{|x|}{\sigma^2} \leq e^{-\frac{1}{2}} \cdot \frac{1}{\sigma\sqrt{2}}$ , we know

$$I_2 \lesssim \frac{1}{\sigma} \int_{y \in D} |y|^2 a^{-2s} dy \lesssim \frac{1}{\sigma} a^{-2s} \varepsilon^{-\frac{4}{m}} < \varepsilon$$

if  $a \gtrsim \sigma^{-\frac{1}{2s}} \varepsilon^{-\frac{1+\frac{4}{m}}{2s}}$ , which is true when

$$(15) \quad N \gtrsim \varepsilon^{-\frac{1+\frac{4}{m}}{2s-\eta}}.$$

As for  $I_1$ , notice that  $|\theta_{\nabla\phi(b)} - \theta| < \theta_0$ , and then  $|\theta_{R_\theta^{-1}\nabla\phi(b)}| < \theta_0$ . Let  $\tilde{\theta} = \theta_{R_\theta^{-1}\nabla\phi(b)}$  and  $y = (y_1, y_2)^T$ , and then for  $a \in (\frac{N}{2M}, 2MN)$

$$\begin{aligned} I_1 &\lesssim \frac{1}{\sigma} \int_{y \in D} |\nabla\phi(b) \cdot (R_\theta A_a^{-1}y)| dy \\ &\lesssim \frac{M}{\sigma} \int_{y \in D} \left| \frac{y_1}{a^t} \cos \tilde{\theta} + \frac{y_2}{a^s} \sin \tilde{\theta} \right| dy \\ &\lesssim \frac{Md}{\sigma} \int_{y \in D} \max_{\gamma \in [0, 2\pi]} \left| \frac{\cos \gamma \cos \tilde{\theta}}{a^t} + \frac{\sin \gamma \sin \tilde{\theta}}{a^s} \right| dy \\ &\lesssim \frac{Md^3 L}{\sigma}, \end{aligned}$$

where  $d \approx \varepsilon^{-\frac{1}{m}}$  is the radius of  $D$  and

$$L = \sqrt{\frac{\cos^2 \tilde{\theta}}{a^{2t}} + \frac{\sin^2 \tilde{\theta}}{a^{2s}}} \leq \sqrt{\frac{1}{a^{2t}} + \frac{\sin^2 \theta_0}{a^{2s}}} \lesssim \max \left\{ \frac{1}{a^t}, \frac{|\sin \theta_0|}{a^s} \right\} \lesssim N^{-t}.$$

So

$$I_1 \lesssim \frac{Md^3 L}{\sigma} \lesssim \frac{Md^3 N^{-t}}{\sigma} \lesssim O(\varepsilon)$$

if

$$(16) \quad N \gtrsim \varepsilon^{-\frac{1+\frac{3}{m}}{t-\eta}}.$$



A direct result of the estimate of  $I_1$  and  $I_2$  is (14) for

$$(17) \quad N \gtrsim \max \left\{ \varepsilon^{-\frac{1+\frac{3}{m}}{t-\eta}}, \varepsilon^{-\frac{1+\frac{4}{m}}{2s-\eta}} \right\}.$$

Second, we need to show

$$(18) \quad W_f(a, \theta, b) = a^{-\frac{s+t}{2}} \left( \int_{y \in D} e^{-\frac{(\phi(b)-c)^2}{\sigma^2}} \alpha(b) w(y) e^{2\pi i(N\phi(b+R_\theta A_a^{-1}y)-a^{1-t}y_1)} dy + O(\varepsilon) \right),$$

which relies on the analysis of the effect of  $\phi(x)$  on  $\alpha(x)$  as follows. Since  $a \in (\frac{N}{2M}, 2MN)$ ,

$$\begin{aligned} I_3 &= \int_{y \in D} e^{-\frac{(\phi(b)-c)^2}{\sigma^2}} |\nabla \alpha \cdot (R_\theta A_a^{-1}y) w(y)| dy \\ &\lesssim \int_{y \in D} |R_\theta A_a^{-1}y| dy \\ &\lesssim a^{-s} \varepsilon^{-\frac{3}{m}} \\ &\lesssim O(\varepsilon) \end{aligned}$$

holds when

$$N \gtrsim \varepsilon^{-\frac{1+\frac{3}{m}}{s}}.$$

Then we derive (18) by the estimate of  $I_3$  and (14) for  $N \gtrsim \varepsilon^{-\frac{1+\frac{3}{m}}{s}}$ .

Finally, we should estimate the nonlinear effect of  $\phi(x)$  on the oscillatory pattern and show (13) for sufficiently large  $N$ . If

$$N \gtrsim \varepsilon^{-\frac{1+\frac{4}{m}}{2s-\eta}},$$

then

$$\begin{aligned} I_4 &= \int_{y \in D} \left| e^{2\pi i(N\phi(b)+N\nabla\phi(b)\cdot(R_\theta A_a^{-1}y)-a^{1-t}y_1)} \right. \\ &\quad \left. \cdot \left| e^{2\pi i\frac{N}{2}(R_\theta A_a^{-1}y)^T \nabla^2 \phi(R_\theta A_a^{-1}y)} - 1 \right| dy \right| \\ &\lesssim \int_{y \in D} |N(R_\theta A_a^{-1}y)^T \nabla^2 \phi(R_\theta A_a^{-1}y)| dy \\ &\lesssim \int_{y \in D} Na^{-2s}|y|^2 dy \\ &\lesssim Na^{-2s} \varepsilon^{-\frac{4}{m}} \\ &\lesssim O(\varepsilon) \end{aligned}$$

holds by the fact that  $|e^{ix} - 1| \leq |x|$  and  $a \in (\frac{N}{2M}, 2MN)$ . Then, by (18) and  $I_4$ , we have

$$\begin{aligned} W_f(a, \theta, b) &= a^{-\frac{s+t}{2}} \left( f(b) \int_{y \in D} w(y) e^{2\pi i(N\nabla\phi(b)\cdot(R_\theta A_a^{-1}y)-a^{1-t}y_1)} dy + O(\varepsilon) \right) \\ &= a^{-\frac{s+t}{2}} \left( f(b) \int_{R^2} w(y) e^{2\pi i(N A_a^{-1} R_\theta^{-1} \nabla\phi(b) - (a^{1-t}, 0)^T) \cdot y} dy + O(\varepsilon) \right) \\ &= a^{-\frac{s+t}{2}} \left( f(b) \widehat{w}(A_a^{-1} R_\theta^{-1} (a \cdot e_\theta - N\nabla\phi(b))) + O(\varepsilon) \right) \end{aligned}$$

for  $a \in (\frac{N}{2M}, 2MN)$  and  $|\theta_{\nabla\phi(b)} - \theta| < \theta_0$  if

$$(19) \quad N \gtrsim \max \left\{ \varepsilon^{-\frac{1+\frac{3}{m}}{t-\eta}}, \varepsilon^{-\frac{1+\frac{4}{m}}{2s-\eta}}, \varepsilon^{-\frac{1+\frac{3}{m}}{s}}, \varepsilon^{-\frac{1+\frac{4}{m}}{2s-1}} \right\},$$

where  $m$  is any fixed positive integer.

By (12) and (19), the requirement for  $N$  is

$$N \gtrsim N_0 = \max \left\{ (\varepsilon^{\frac{-1}{m}} + 1)^{\frac{1}{2s-t}}, \varepsilon^{\frac{-1}{1-t}}, \varepsilon^{-\frac{1+\frac{3}{m}}{t-\eta}}, \varepsilon^{-\frac{1+\frac{4}{m}}{2s-\eta}}, \varepsilon^{-\frac{1+\frac{3}{m}}{s}}, \varepsilon^{-\frac{1+\frac{4}{m}}{2s-1}} \right\},$$

where  $m$  is any fixed positive integer. Hence, this completes the proof of (i) when  $K = 1$ .

In sum, we have proved this lemma when  $K = 1$ . The conclusion is also true for general  $K$  by the linearity of general curvelet transform.  $\square$

To prove Theorem 2.3, we need one more lemma which estimates  $\nabla_b W_f(a, \theta, b)$ .

LEMMA 2.5. *Under the assumption of the theorem, there exists a constant  $N_0(M, s, t, \eta, \varepsilon) > 0$  such that if  $N > N_0(M, s, t, \eta, \varepsilon)$ , then we have*

$$\begin{aligned} &\nabla_b W_f(a, \theta, b) \\ &= a^{-\frac{s+t}{2}} \left( 2\pi i N \sum_{k: |\theta_{\nabla\phi_k(b)} - \theta| < \theta_0} \nabla\phi_k(b) f_k(b) \widehat{w} (A_a^{-1} R_\theta^{-1} (a \cdot e_\theta - N \nabla\phi(b))) + O(\varepsilon) \right) \end{aligned}$$

when

$$(a, \theta) \in \Omega = \left\{ (a, \theta) : a \in \left( \frac{N}{2M}, 2MN \right), \exists k \text{ s.t. } |\theta_{\nabla\phi_k(b)} - \theta| < \theta_0 \right\}.$$

*Proof.* The proof is similar to the one of Lemma 2.4. We only need to discuss the case  $K = 1$ , and the case  $K > 1$  holds by the linearity of general curvelet transform. Suppose

$$f(x) = e^{-\frac{(\phi(x)-c)^2}{\sigma^2}} \alpha(x) e^{2\pi i N \phi(x)},$$

and we have

$$\begin{aligned} &\nabla_b W_f(a, \theta, b) \\ &= \int_{\mathbb{R}^2} f(x) a^{\frac{s+t}{2}} \left( (-R_\theta A_a) \nabla w (A_a R_\theta^{-1} (x - b)) \right. \\ &\quad \left. + 2\pi i a e_\theta w (A_a R_\theta^{-1} (x - b)) \right) e^{-2\pi i a (x-b) \cdot e_\theta} dx \\ &= \int_{\mathbb{R}^2} f(b + R_\theta A_a^{-1} y) a^{-\frac{s+t}{2}} \left( (-R_\theta A_a) \nabla w (y) + 2\pi i a e_\theta w (y) \right) e^{-2\pi i a^{1-t} y_1} dy \\ &= a^{-\frac{s+t}{2}} \left( f(b) \int_{\mathbb{R}^2} ((-R_\theta A_a) \nabla w (y) + 2\pi i a e_\theta w (y)) \right. \\ &\quad \left. e^{-2\pi i ((a^{1-t}, 0)^T - N A_a^{-1} R_\theta^{-1} \nabla\phi(b)) \cdot y} dy + O(\varepsilon) \right) \\ &= a^{-\frac{s+t}{2}} \left( 2\pi i N \nabla\phi(b) f(b) \widehat{w} (A_a^{-1} R_\theta^{-1} (a \cdot e_\theta - N \nabla\phi(b))) + O(\varepsilon) \right) \end{aligned}$$

for  $a \in (\frac{N}{2M}, 2MN)$  and  $|\theta_{\nabla\phi(b)} - \theta| < \theta_0$  if  $N$  satisfies the condition in Lemma 2.4. Therefore, if  $f$  has  $K$  components, we know

$$\begin{aligned} & \nabla_b W_f(a, \theta, b) \\ &= a^{-\frac{s+t}{2}} \left( \sum_{k: |\theta_{\nabla\phi_k(b)} - \theta| < \theta_0} 2\pi i N \nabla\phi_k(b) f_k(b) \widehat{w}(A_a^{-1} R_\theta^{-1}(a \cdot e_\theta - N \nabla\phi_k(b))) + O(\varepsilon) \right) \end{aligned}$$

for  $(a, \theta) \in \Omega$  and  $N$  is larger than the same constant  $N_0$  in Lemma 2.4.  $\square$

With the above two lemmas proved, it is enough to prove Theorem 2.3.

*Proof.* We shall start from (i).  $\{Z_{f,k} : 1 \leq k \leq K\}$  are disjoint as soon as  $f(x)$  is a superposition of well-separated components. Let  $(a, \theta, b) \in R_{f,\varepsilon}$ . By Lemma 2.4,  $(a, \theta) \in \Omega$ . So, we have

$$\begin{aligned} & W_f(a, \theta, b) \\ &= a^{-\frac{s+t}{2}} \left( \sum_{k: |\theta_{\nabla\phi_k(b)} - \theta| < \theta_0} f_k(b) \widehat{w}(A_a^{-1} R_\theta^{-1}(a \cdot e_\theta - N \nabla\phi_k(b))) + O(\varepsilon) \right). \end{aligned}$$

Therefore,  $\exists k$  such that  $\widehat{w}(A_a^{-1} R_\theta^{-1}(a \cdot e_\theta - N \nabla\phi_k(b))) \neq 0$ . By the definition of  $Z_{f,k}$ , we see that  $(a, \theta, b) \in Z_{f,k}$ . Hence,  $R_{f,\varepsilon} \subset \cup_{k=1}^K Z_{f,k}$ .

To show (ii), notice that  $(a, \theta, b) \in R_{f,\varepsilon} \cup Z_{f,k}$ , and then

$$W_f(a, \theta, b) = a^{-\frac{s+t}{2}} \left( f_k(b) \widehat{w}(A_a^{-1} R_\theta^{-1}(a \cdot e_\theta - N \nabla\phi_k(b))) + O(\varepsilon) \right)$$

and

$$\begin{aligned} & \nabla_b W_f(a, \theta, b) \\ &= a^{-\frac{s+t}{2}} \left( 2\pi i N \nabla\phi_k(b) f_k(b) \widehat{w}(A_a^{-1} R_\theta^{-1}(a \cdot e_\theta - N \nabla\phi_k(b))) + O(\varepsilon) \right). \end{aligned}$$

Let  $g = f_k(b) \widehat{w}(A_a^{-1} R_\theta^{-1}(a \cdot e_\theta - N \nabla\phi_k(b)))$ , and then

$$v_f(a, \theta, b) = \frac{N \nabla\phi_k(b) g + O(\varepsilon)}{g + O(\varepsilon)}.$$

Since  $|W_f(a, \theta, b)| \geq a^{-\frac{s+t}{2}} \sqrt{\varepsilon}$  for  $(a, \theta, b) \in R_{f,\varepsilon}$ ,  $|g| \gtrsim \sqrt{\varepsilon}$ . So

$$\frac{|v_f(a, \theta, b) - N \nabla\phi_k(b)|}{|N \nabla\phi_k(b)|} \lesssim \left| \frac{O(\varepsilon)}{g + O(\varepsilon)} \right| \lesssim \sqrt{\varepsilon}. \quad \square$$

The assumptions  $\frac{1}{2} < s < t < 1$  and  $\eta < t$  are essential to the proof. However, we have not arrived at a clear opinion on the optimal values of these parameters. The difference  $t - s$  allows us to construct directional needle-like curvelets in order to approximate banded wave-like components or wavefronts and capture the oscillatory behavior better. When  $t$  and  $\eta$  approach 1, and  $s$  gets close to  $\frac{1}{2}$ , we can expect that the SSCT can separate banded components of width approximately  $O(N^{-1})$  if  $N$  is large enough. On the other hand, the lower bound  $s > 1/2$  ensures that the support

of each curvelet is sufficiently small in space so that the second order properties of the phase function (such as the curvature of wavefronts) do not affect the estimate of local wave-vectors. The upper bound  $t < 1$  guarantees sufficient resolution to detect different components with large wavenumbers.

In Theorem 2.3, although the lower bound of  $N$  could be optimized,  $N$  is required to be sufficiently large so that the local wave-vector can be precisely captured by synchrosqueezing. In other words, for a fixed  $f(x)$  of a type  $F(M, N, \eta, K)$ , there is a lower bound on the accuracy parameter  $\epsilon$ . On the other hand, the local wave-vector is not well defined for a low frequency component. In fact, in the presence of such a component, each high oscillatory component is still squeezed into a well-separated sharpened representation in the high frequency part of Fourier domain. Therefore, the low frequency component would be identified precisely by subtracting high frequency components.

**3. Implementation of the transform.** In this section, we describe the discrete SSCT and the mode decomposition in detail. Subsection 1.2 has discussed the key ideas of mode decomposition by SSCT. Let us describe the whole framework now. Suppose  $f(x)$  is a superposition of several well-separated components; the mode decomposition by SSCT consists of the following steps:

- (i) Apply the general curvelet transform to obtain  $W_f(a, \theta, b)$  and the gradient  $\nabla_b W_f(a, \theta, b)$ .
- (ii) Compute the local wave-vector estimate  $v_f(a, \theta, b)$  and concentrate the energy around it to get  $T_f(v, b)$ .
- (iii) Separate the essential supports of the concentrated phase space energy distribution  $T_f(v, b)$  into several components by clustering techniques.
- (iv) Restrict  $W_f(a, \theta, b)$  to each resulting component and reconstruct corresponding intrinsic mode functions using the dual frame.

We first introduce a discrete implementation of the general curvelet transform in section 3.1 for steps (i) and (iv). Clustering methods will be discussed later in section 3.2. The full discrete algorithm will then be summarized in section 3.3.

**3.1. Discrete general curvelet transforms.** For simplicity, we consider functions that are periodic over the unit square  $[0, 1]^2$  in two dimensions. If it is not the case, the functions will be periodized by multiplying a smooth decaying function near the boundary of  $[0, 1]^2$ . Let

$$X = \{(n_1/L, n_2/L) : 0 \leq n_1, n_2, < L, n_1, n_2 \in \mathbb{Z}\}$$

be the  $L \times L$  spatial grid at which these functions are sampled. The corresponding  $L \times L$  Fourier grid is

$$\Xi = \{(\xi_1, \xi_2) : -L/2 \leq \xi_1, \xi_2 < L/2, \xi_1, \xi_2 \in \mathbb{Z}\}.$$

For a function  $f(x) \in \ell^2(X)$ , the discrete forward Fourier transform is defined by

$$\hat{f}(\xi) = \frac{1}{L} \sum_{x \in X} e^{-2\pi i x \cdot \xi} f(x).$$

For a function  $g(\xi) \in \ell^2(\Xi)$ , the discrete inverse Fourier transform is

$$\check{g}(x) = \frac{1}{L} \sum_{\xi \in \Xi} e^{2\pi i x \cdot \xi} g(\xi).$$

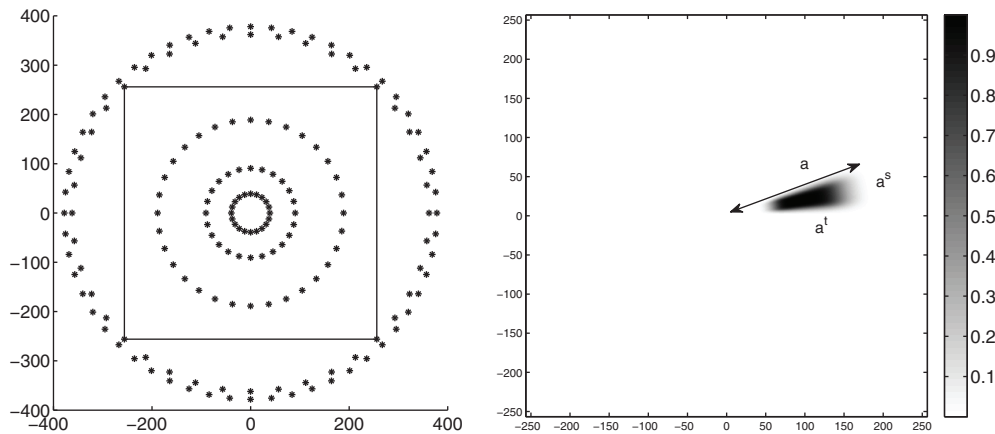


FIG. 4. Left: Sampled point set  $P$  in Fourier domain for an image of size  $512 \times 512$ . Each point represents the center of the support of a window function. The window function centered at the origin is supported on a disk and is not indicated in this picture. The size of finest scale is set to be small (e.g., 16) in order to save memory. Right: An example of fan-shaped window function  $g_{a,\theta}(\xi)$ .

In both transforms, the factor  $1/L$  ensures that these discrete transforms are isometric between  $\ell_2(X)$  and  $\ell_2(\Xi)$ .

In order to design a discrete curvelet transform, we need to specify how to decimate the Fourier domain  $(a, \theta)$  and the position space  $b$ . Let us first consider the Fourier domain  $(a, \theta)$ . In the continuous setting, the Fourier transform  $\widehat{w_{a\theta b}}(\xi)$  for a fixed  $(a, \theta)$  value have the profile

$$(20) \quad a^{-\frac{s+t}{2}} \widehat{w}(A_a^{-1} R_\theta^{-1}(\xi - a \cdot e_\theta)),$$

modulo complex modulation. In the discrete setting, we sample the Fourier domain  $[-L/2, L/2]^2$  with a set of points  $P$  (Figure 4 left) and associate with each  $(a, \theta) \in P$  a window function  $g_{a,\theta}(\xi)$  (Figure 4 right) that behaves qualitatively as  $\widehat{w}(A_a^{-1} R_\theta^{-1}(\xi - a \cdot e_\theta))$ . More precisely,  $g_{a,\theta}(\xi)$  is required to satisfy the following conditions:

- $g_{a,\theta}(\xi)$  is nonnegative and centered at  $a \cdot e_\theta$  with a compact fan-shaped support of length  $O(a^t)$  and width  $O(a^s)$ , which is approximately a directional elliptical support  $\{\xi : |A_a^{-1} R_\theta^{-1}(\xi - a \cdot e_\theta)| \leq 1\}$ .
- $g_{a,\theta}(R_\theta A_a \tau + a \cdot e_\theta)$  is a sufficiently smooth function of  $\tau$ , thus making the discrete curvelets decay rapidly in the spatial domain.
- $C_1 \leq \int |g_{a,\theta}(R_\theta A_a \tau + a \cdot e_\theta)|^2 d\tau \leq C_2$  for positive constants  $C_1$  and  $C_2$ , independent of  $(a, \theta)$ .
- In addition, for any  $\xi \in [-L/2, L/2]^2$ ,  $\sum_{(a,\theta) \in P} |g_{a,\theta}(\xi)|^2 = 1$ .

We follow the discretization and construction of frames in [4] to specify the set  $P$  and window functions, and refer to [3] for detail implementation. The difference here is that we do not restrict angular scaling parameter to  $s = \frac{1}{2}$  and radial scaling parameter to  $t = 1$ . This allows us to adaptively adjust the size of tiles according to data structure. In the construction of the tiling in this article, the scaling parameters  $s$  and  $t$  remain constant as the scale changes.

The decimation of the position space  $b$  is much easier; we simply discretize it with an  $L_B \times L_B$  uniform grid as follows:

$$B = \{(n_1/L_B, n_2/L_B) : 0 \leq n_1, n_2 < L_B, n_1, n_2 \in \mathbb{Z}\}.$$

The only requirement is that  $L_B$  is large enough so that a sampling grid of size  $L_B \times L_B$  can cover the supports of all window functions.

For each fixed  $(a, \theta) \in P$  and  $b \in B$ , the discrete curvelet, still denoted by  $w_{a\theta b}(x)$  without causing much confusion, is defined through its Fourier transform as

$$\widehat{w_{a\theta b}}(\xi) = \frac{1}{L_a} e^{-2\pi i b \cdot \xi} g_{a,\theta}(\xi)$$

for  $\xi \in \Xi$  with  $L_a = a^{\frac{s+t}{2}}$ . Applying the discrete inverse Fourier transform provides its spatial description

$$w_{a\theta b}(x) = \frac{1}{L \cdot L_a} \sum_{\xi \in \Xi} e^{2\pi i(x-b) \cdot \xi} g_{a,\theta}(\xi).$$

For a function  $f(x)$  defined on  $x \in X$ , the discrete curvelet transform is a map from  $\ell_2(X)$  to  $\ell_2(P \times B)$ , defined by

$$(21) \quad W_f(a, \theta, b) = \langle w_{a\theta b}, f \rangle = \langle \widehat{w_{a\theta b}}, \widehat{f} \rangle = \frac{1}{L_a} \sum_{\xi \in \Xi} e^{2\pi i b \cdot \xi} g_{a,\theta}(\xi) \widehat{f}(\xi).$$

We can introduce an inner product on the space  $\ell_2(P \times B)$  as follows: for any two functions  $g(a, \theta, b)$  and  $h(a, \theta, b)$ ,

$$\langle g, h \rangle = \sum_{(a,\theta) \in P, b \in B} \overline{g(a, \theta, b)} h(a, \theta, b).$$

The following result shows that  $\{w_{a\theta b} : (a, \theta, b) \in P \times B\}$  forms a tight frame when equipped with this inner product.

PROPOSITION 3.1. *For any function  $f(x)$  for  $x \in X$ , we have*

$$\sum_{(a,\theta) \in P, b \in B} |W_f(a, \theta, b)|^2 (L_a/L_B)^2 = \|f\|_2^2.$$

*Proof.* From the definition of the curvelet transform, we have

$$\begin{aligned} \sum_{(a,\theta) \in P, b \in B} |W_f(a, \theta, b)|^2 (L_a/L_B)^2 &= \sum_{(a,\theta) \in P, b \in B} \left| \sum_{\xi \in \Xi} \frac{1}{L_a} e^{2\pi i b \cdot \xi} g_{a,\theta}(\xi) \widehat{f}(\xi) \right|^2 (L_a/L_B)^2 \\ &= \sum_{(a,\theta) \in P} \sum_{\xi \in \Xi} |g_{a,\theta}(\xi) \widehat{f}(\xi)|^2 \\ &= \sum_{\xi \in \Xi} |\widehat{f}(\xi)|^2. \quad \square \end{aligned}$$

For a function  $h(a, \theta, b)$  in  $\ell_2(P \times B)$ , the transpose of the curvelet transform is given by

$$(22) \quad W_h^t(x) := \sum_{(a,\theta) \in P, b \in B} h(a, \theta, b) w_{a\theta b}(x) (L_a/L_B)^2.$$

The next result shows that this transpose operator allows us to reconstruct  $f(x), x \in X$  from its curvelet transform  $W_f(a, \theta, b), (a, \theta, b) \in P \times B$ .

PROPOSITION 3.2. *For any function  $f(x)$  with  $x \in X$ ,*

$$f(x) = \sum_{(a,\theta) \in P, b \in B} W_f(a, \theta, b) w_{a\theta b}(x) (L_a/L_B)^2.$$

*Proof.* Let us consider the Fourier transform of the right-hand side. It is equal to

$$\begin{aligned} & \sum_{(a,\theta) \in P, b \in B} \left( \sum_{\eta \in \Xi} \frac{1}{L_a} e^{2\pi i b \cdot \eta} g_{a,\theta}(\eta) \widehat{f}(\eta) \right) \cdot \frac{1}{L_a} e^{-2\pi i b \cdot \xi} g_{a,\theta}(\xi) (L_a/L_B)^2 \\ &= \sum_{(a,\theta) \in P} \left( \sum_{\eta \in \Xi} \frac{1}{L_B^2} \left( \sum_{b \in B} e^{2\pi i b \cdot (\eta - \xi)} g_{a,\theta}(\eta) \widehat{f}(\eta) \right) \right) g_{a,\theta}(\xi) \\ &= \sum_{(a,\theta) \in P} (g_{a,\theta}(\xi))^2 \widehat{f}(\xi) = \widehat{f}(\xi), \end{aligned}$$

where the second step uses the fact that in the  $\eta$  sum only the term with  $\eta = \xi$  yields a nonzero contribution.  $\square$

Let us now turn to the discrete approximation of  $\nabla_b W_f(a, \theta, b)$ . From the continuous definition 1.2, we have

$$\nabla_b W_f(a, \theta, b) = \nabla_b \langle \widehat{w}_{a\theta b}, \widehat{f} \rangle = \langle -2\pi i \xi \widehat{w}_{a\theta b}(\xi), \widehat{f}(\xi) \rangle.$$

Therefore, we define the discrete gradient  $\nabla_b W_f(a, \theta, b)$  in a similar way,

$$(23) \quad \nabla_b W_f(a, \theta, b) = \sum_{\xi \in \Xi} \frac{1}{L_a} 2\pi i \xi e^{2\pi i b \cdot \xi} g_{a,\theta}(\xi) \widehat{f}(\xi).$$

The above definitions give rise to fast algorithms for computing the forward general curvelet transform, its transpose, and the discrete gradient operator. All three algorithms heavily rely on the fast Fourier transform. The detailed implementation of these fast algorithms has been discussed in [33]. The computational cost of all three algorithms is  $O(L^2 \log L + L^{2-s-t} L_B^2 \log L_B)$  with  $L_B$  large enough so that a grid of size  $L_B \times L_B$  can cover the supports of all window functions. If we choose  $L_B$  to be of the same order as  $L^t$ , the complexity of these algorithms is  $O(L^{2+t-s} \log L)$ .

**3.2. Clustering in the phase space.** In the proof of Theorem 2.3, the radial separation and angular separation conditions play an important role in describing the well-separated condition. Therefore, the polar coordinate is used to quantify distance in the Fourier domain, which motivates the following clustering method used in the numerical examples of this article. Before introducing the algorithm, some notation is defined below.

1. We associate any point  $p$  in the four-dimensional (4D) phase space with  $(x_p, a_p, \theta_p)$ , where  $x_p$  is the projection of  $p$  in the 2D spatial domain and  $(a_p \cos \theta_p, a_p \sin \theta_p)$  is the projection of  $p$  in the 2D Fourier domain.
2. We say that  $(p, q)$  is a pair of adjacent points with parameters  $(d_0, \theta_0, R_0)$  if
  - $|x_p - x_q| \leq d_0$ ,
  - $|a_p - a_q| \leq R_0$ ,
  - $\min\{|\theta_p - \theta_q|, 2\pi - |\theta_p - \theta_q|\} \leq \theta_0$ .

By the definitions above, after fixing parameters  $d_0$ ,  $\theta_0$ , and  $R_0$ , we can construct an undirected graph  $G$ , in which the nodes are the points to be clustered and two nodes are connected if their corresponding points are a pair of adjacent points with parameters  $(d_0, \theta_0, R_0)$ . Identifying clusters in the 4D phase space is equivalent to identifying all connected components in the graph  $G$ . This is a traditional research topic in graph theory and computer science. It is straightforward to compute these components in linear time (in terms of the numbers of the nodes and edges of  $G$ )

using either the breadth-first search, which is applied in this paper, or the depth-first search (refer to Chapter IV in [8]).

The cost of computation and memory of these two searching algorithms is extremely high. Suppose the size of given data  $f(x)$  is  $L \times L$  and there are  $K$  components with wavenumbers of  $O(L)$ . By Theorem 2.3, each synchrosqueezed energy distribution  $T_{f_k}(v, b)$  is surrounding its 2D wave-vector surface within a distance of  $O(L\sqrt{\epsilon})$ . Hence, the total number of nonzero grid points in the 4D phase space s.t.  $T_f(v, b) \geq \delta$  is of order  $KL^4\epsilon$ , which is an impractical number for clustering. To reduce the cost, we should apply similar clustering methods first in the 2D Fourier domain at each location, which results in  $O(K)$  clusters at each location. Afterward, a clustering method is applied to the point set of reduced size of  $O(KL^2)$  in 4D phase space. In the case of a very large  $L$ , a space efficient algorithm, the iterative deepening depth-first search algorithm [22], may be a good alternative.

**3.3. Description of the full algorithm.** With the fast discrete synchrosqueezed transforms and clustering algorithms available, we now go through the steps of the SSCT.

For a given function  $f(x)$  defined on  $x \in X$ , we apply fast algorithms to compute  $W_f(a, \theta, b)$  and  $\nabla_b W_f(a, \theta, b)$ . Then the local wave-vector estimate  $v_f(a, \theta, b)$  is computed by

$$v_f(a, \theta, b) = \frac{\nabla_b W_f(a, \theta, b)}{2\pi i W_f(a, \theta, b)}$$

for  $(a, \theta) \in P, b \in B$  with  $W_f(a, \theta, b) \neq 0$  (indeed,  $|W_f(a, \theta, b)| \geq \sqrt{\epsilon}$  in the numerical implementation).

The energy resulting in  $\Re v_f(a, \theta, b)$  should be stacked up to obtain  $T_f(\Re v_f(a, \theta, b), b)$ . To realize this step, a two-dimensional Cartesian grid of step size  $\Delta$  is generated to discretize the Fourier domain of  $T_f(v, b)$  in variable  $v$  as follows:

$$V = \{(n_1\Delta, n_2\Delta) : n_1, n_2 \in \mathbb{Z}\}.$$

At each  $v = (n_1\Delta, n_2\Delta) \in V$ , we associate a cell  $D_v$  centered at  $v$

$$D_v = \left[ \left( n_1 - \frac{1}{2} \right) \Delta, \left( n_1 + \frac{1}{2} \right) \Delta \right) \times \left[ \left( n_2 - \frac{1}{2} \right) \Delta, \left( n_2 + \frac{1}{2} \right) \Delta \right).$$

Then  $T_f(v, b)$  is estimated by

$$T_f(v, b) = \sum_{(a, \theta, b): \Re v_f(a, \theta, b) \in D_v} |W_f(a, \theta, b)|^2 (L_a/L_B)^2.$$

Suppose that  $f(x)$  is a superposition of  $K$  well-separated banded intrinsic mode functions:

$$f(x) = \sum_{k=1}^K f_k(x) = \sum_{k=1}^K e^{-(\phi_k(x)-c_k)^2/\sigma_k^2} \alpha_k(x) e^{2\pi i N \phi_k(x)}.$$

In the discrete implementation, we choose a threshold parameter  $\delta > 0$  and define the set  $S$  to be

$$\{(v, b) : v \in V, b \in B, T_f(v, b) \geq \delta\}.$$

After synchrosqueezing,  $T_f(v, b)$  is essentially supported in the phase space near  $K$  “discrete” surfaces  $\{(N\phi_k(b), b), b \in B\}$ . Hence, under the separation condition given



by Theorem 2.3,  $S$  will have  $K$  well-separated clusters  $U_1, \dots, U_K$ , and they would be identified by the clustering method in the last subsection.

Once we discover  $U_1, \dots, U_K$ , we can define  $W_{f_k}(a, \theta, b)$  by restricting  $W_f(a, \theta, b)$  to the set  $\{(a, \theta, b) : \Re v_f(a, \theta, b) \in U_k\}$ . Then, we can recover each intrinsic mode function efficiently using the fast algorithm discussed to compute

$$f_k(x) = \sum_{(a, \theta) \in P, b \in B} W_{f_k}(a, \theta, b) w_{a\theta b}(x) (L_a/L_B)^2.$$

**4. Numerical results.** In this section, we start with error analysis of local wave-vector estimation using SSCT and compare it with SSWPT. Afterward, some mode decomposition examples of synthetic and real data will be presented to illustrate the efficiency of proposed SSCT. For all the synthetic examples in this section, the size  $L$  of the Cartesian grid  $X$  of the discrete algorithm is 512, the threshold value  $\epsilon = 10^{-4}$  for  $W_f(a, \theta, b)$ . The scaling parameters of SSCT are  $t = 1 - \frac{1}{8}$  and  $s = \frac{1}{2} + \frac{1}{8}$ , as an appropriate balance as discussed previously. In the meantime, we chose  $t = s = \frac{1}{2} + \frac{1}{8}$  to construct discrete synchrosqueezed wave packet transform for a reasonable comparison. In all the decomposition problems, section 3.2 with application dependent parameters is applied, and it provides desired solutions. We will only present relevant recovered components to save space.

**4.1. Instantaneous wave-vector estimation.** In Theorem 2.3, we have seen that the estimate  $v_f(a, \theta, b)$  approximates the local wave-vector at  $b$  if  $|W_f(a, \theta, b)| \geq a^{-\frac{s+t}{2}} \sqrt{\epsilon}$ . Since  $a \geq 1$ , as we discussed after the Definition 1.1, it is useful to consider a simple and universal threshold criteria  $|W_f(a, \theta, b)| \geq \sqrt{\epsilon}$ , which amounts to a smaller region of the essential support of  $W_f(a, \theta, b)$ . In such a region, though  $v_f(a, \theta, b)$  provides an accurate estimate of the local wave-vector at each  $b$ , it is more rational to average them up to obtain a unique local wave-vector estimate for each fixed  $b$ . By the definition of synchrosqueezed energy distribution,  $T_f(\Re v_f(a, \theta, b), b)$  truly reflects a natural weight of  $v_f(a, \theta, b)$  in variables  $a$  and  $\theta$ . Hence, we define the *mean* local wave-vector estimate at  $b$  to be

$$v_f^m(b) = \frac{\sum_{(a, \theta)} |T_f(\Re v_f(a, \theta, b), b)| v_f(a, \theta, b)}{\sum_{(a, \theta)} |T_f(\Re v_f(a, \theta, b), b)|}.$$

In the presence of noise, a threshold  $\delta$  proportional to noise level is set up for  $T_f(\Re v_f(a, \theta, b), b)$  to uncover the dominant estimate. Correspondingly, we define the thresholded *mean* local waveform estimate as

$$v_f^{m, \delta}(b) = \frac{\sum_{(a, \theta) \in \Omega_\delta(b)} |T_f(\Re v_f(a, \theta, b), b)| v_f(a, \theta, b)}{\sum_{(a, \theta) \in \Omega_\delta(b)} |T_f(\Re v_f(a, \theta, b), b)|},$$

where  $\Omega_\delta(b) = \{(a, \theta) : |T_f(\Re v_f(a, \theta, b), b)| \geq \delta\}$ . In a noiseless case,  $v_f^m(b) = v_f^{m, 0}(b)$ . Using this estimate, we can define the relative error  $R_\delta(b)$  between  $v_f^{m, \delta}(b)$  and the exact local wave-vector  $N\nabla\phi(b)$  as

$$R_\delta(b) = \frac{|v_f^{m, \delta}(b) - N\nabla\phi(b)|}{|N\nabla\phi(b)|}.$$

*Example 1.* We test the accuracy for a noise free deformed plane wave  $f(x) = \alpha(x)e^{2\pi i N\phi(x)}$  with  $\alpha(x) = 1$ ,  $\phi(x) = \phi(x_1, x_2) = x_1 + (1 - x_2) + 0.1 \sin(2\pi x_1) +$

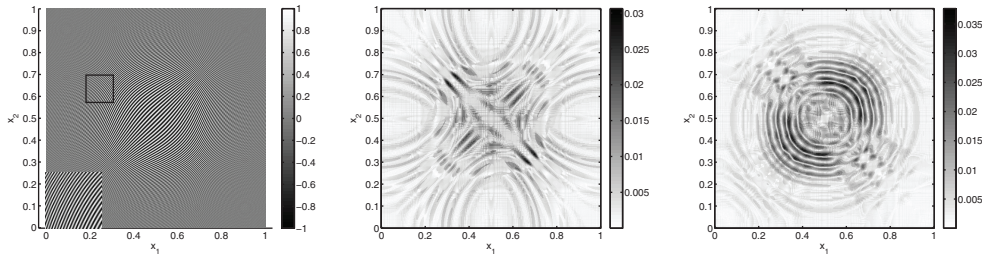


FIG. 5. Left: A deformed plane wave propagating in the full space with zoomed-in data indicated by a rectangle. Middle: Relative error  $R_0(b)$  of local wave-vector estimation using SSCT. Right: Relative error  $R_0(b)$  of local wave-vector estimation given by SSWPT.

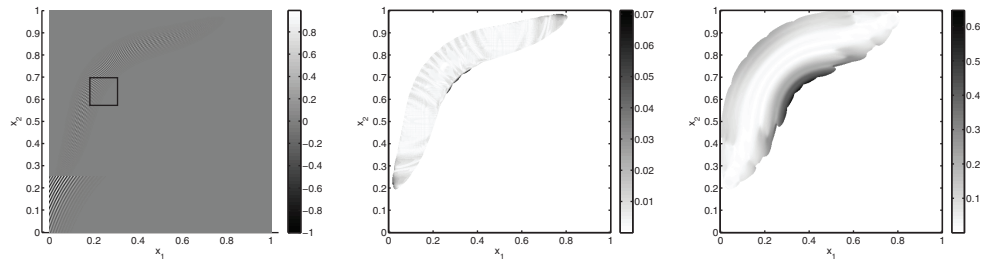


FIG. 6. Left: A banded deformed plane wave. The zoomed-in data comes from the small rectangle. Middle: Relative error  $R_0(b)$  of local wave-vector estimation using SSCT. Right: Relative error  $R_0(b)$  of local wave-vector estimation given by SSWPT.

$0.1 \sin(2\pi(1 - x_2))$ , and  $N = 135$  (see Figure 5 left). It is a special case in Definition 2.1 with banded parameter  $\sigma = \infty$ . The relative error  $R_0(b)$  of SSCT shown in Figure 5 (middle) is of order  $10^{-2}$ , which agrees with Theorem 2.3 that the relative approximation error is of order  $O(\sqrt{\varepsilon})$ . The SSWPT and the SSCT share the same accuracy in this case, shown by Figure 5 middle and right.

We compare the efficiency of SSCT and SSWPT in a noiseless case of a banded deformed plane wave  $f(x) = e^{-(\phi(x)-c)^2/\sigma^2} \alpha(x) e^{2\pi i N \phi(x)}$  with the same parameters in the last example and two more parameters  $c = 0.7$  and  $\sigma = \frac{4}{135}$ . As we discussed at the beginning of this subsection,  $v_f(a, \theta, b)$  is only computed in the relevant region  $|W_f(a, \theta, b)| \geq \sqrt{\varepsilon}$ . So, the relative error will be set to be zero at the position  $b$  such that  $|W_f(a, \theta, b)| < \sqrt{\varepsilon} \forall(a, \theta)$ . The numerical result in Figure 6 matches well with our theoretical prediction, showing that SSCT estimates local wave-vectors of this banded wave-like component within a relative error of order  $O(\sqrt{\varepsilon})$ . However, SSWPT fails the truth, as we discussed in the section of introduction.

To quantitatively demonstrate the robustness against noise, we provide a series of tests of the above banded deformed plane wave with increasing noise levels. As usual, the noise level is described by the signal-to-noise ratio (SNR) defined by

$$SNR[dB] = 10 \log_{10} \left( \frac{Var f}{\sigma^2} \right).$$

Suppose  $n(x)$  is an isotropic complex Gaussian random noise with zero mean. We consider the noisy data

$$(24) \quad f(x) = e^{-(\phi(x)-c)^2/\sigma^2} \alpha(x) e^{2\pi i N \phi(x)} + n(x)$$

TABLE 1  
Maximum relative error of  $R_\delta(b)$  with different SNR.

SNR	$\infty$	3	0	-3	-6
$\delta$	3	3.5	4	4.5	5
$ R_\delta(b) _{\ell^\infty}$	0.03	0.03	0.04	0.05	0.065

with the same parameters as in previous noiseless banded example. Table 1 summarizes the results. The first row shows different noise levels, and the second row records the threshold  $\delta$  for  $T_f(a, \theta, b)$ . We observe that a proper threshold  $\delta$  can successfully reduce the influence of noise and keeps the local wave-vector estimate accurate and stable. A further study of the behavior of the synchrosqueezing operator on noise is currently ongoing, and we leave the discussion of thresholding to our future work.

#### 4.2. Intrinsic mode decomposition for synthetic data.

*Example 2.* In many applications, it is desired to extract each component from a superposition. To show that our algorithm may provide a solution, we present some numerical examples of mode decomposition for highly oscillatory synthetic seismic data in noiseless and noisy cases (see Figure 7 top). Figure 7 shows the results of the application of our algorithm described in section 3.3. On the left is a noiseless example, and the example on the right has some noise (SNR is  $-3.07$  dB). Each mode of given data is accurately recovered in the noiseless case. In the noisy case, different modes with different propagation characters are completely separated. Each recovered mode practically reflects the curvature of corresponding mode in the original data, though there is some energy loss due to threshold  $\delta$  to remove noise.

*Example 3.* In some other applications, one component might be disrupted (e.g., randomly shifted in this example), and it is required to remove such component and recover others. Here we randomly shift the first mode in Example 2 in the vertical direction and apply our algorithm to recover the second mode. The numerical results summarized in Figure 8 show the capability of our algorithm to solve such a problem with or without noise. In this problem, the disrupted component can be considered as noise with high energy, i.e., this is a problem with very small SNR. It is even more problematic that random shifting may create some texture similar to the mode to be recovered in some region. Fortunately, the synchrosqueezed representation is so concentrated that the resolution is still good enough to separate the mode from such similar texture by appropriately thresholding  $T_f(a, \theta, b)$ .

The left example in Figure 8 shows the result of noiseless data. The recovered mode looks almost the same as the one recovered in noiseless Example 2 (Figure 7 bottom left), except there is some energy loss due to thresholding. It is of interest to add some background noise to see how well our algorithm is performing. Figure 8, right, shows the result of noisy case. SNR is  $-0.90$  if we consider the energy of disrupted component as part of data energy. The result (see Figure 8 bottom right) is almost identical to the recovered mode in Figure 7, bottom left.

**4.3. Intrinsic mode decomposition for real data.** So far, the experiments shown are idealized, e.g., the boundary of each component is clear and smooth, and the amplitudes of each component are of the same level. In this subsection, we apply the SSCT to real seismic data and illustrate its good performance in a complicated circumstance.

*Example 4.* This is real seismic data with four main components and a band of energy loss near the bottom. The centered component is overlapping with others. Components in the bottom left and bottom right corners have irregular boundaries

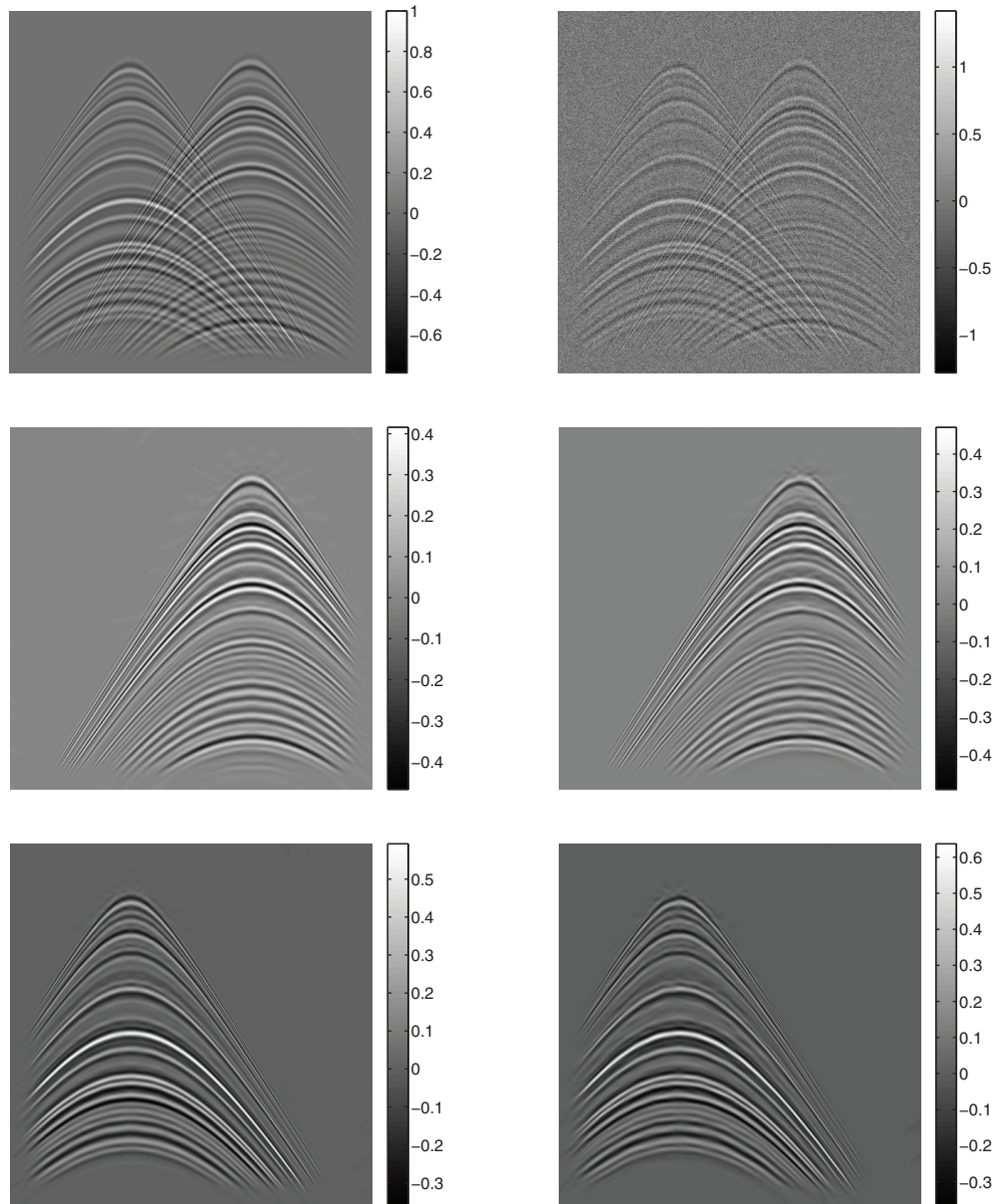


FIG. 7. *Example 2. Left: Mode decomposition without noise. Right: Mode decomposition with noise (SNR = -3.07). Top: A superposition of two components. Second row: The first recovered relevant mode. Third row: The second recovered relevant mode.*

and not well aligned textures. The component on top has obviously weaker energy than others. These characters cause large difficulty in identifying all these components accurately. As shown in Figure 9, the main textures and oscillatory patterns are recognized and recovered by our algorithm, though there is some loss of energy on the boundary of each component caused by thresholding.

**5. Discussion.** This paper has proposed the SSCT as an optimal tool to analyze a superposition of high dimensional banded wave-like components. It serves as

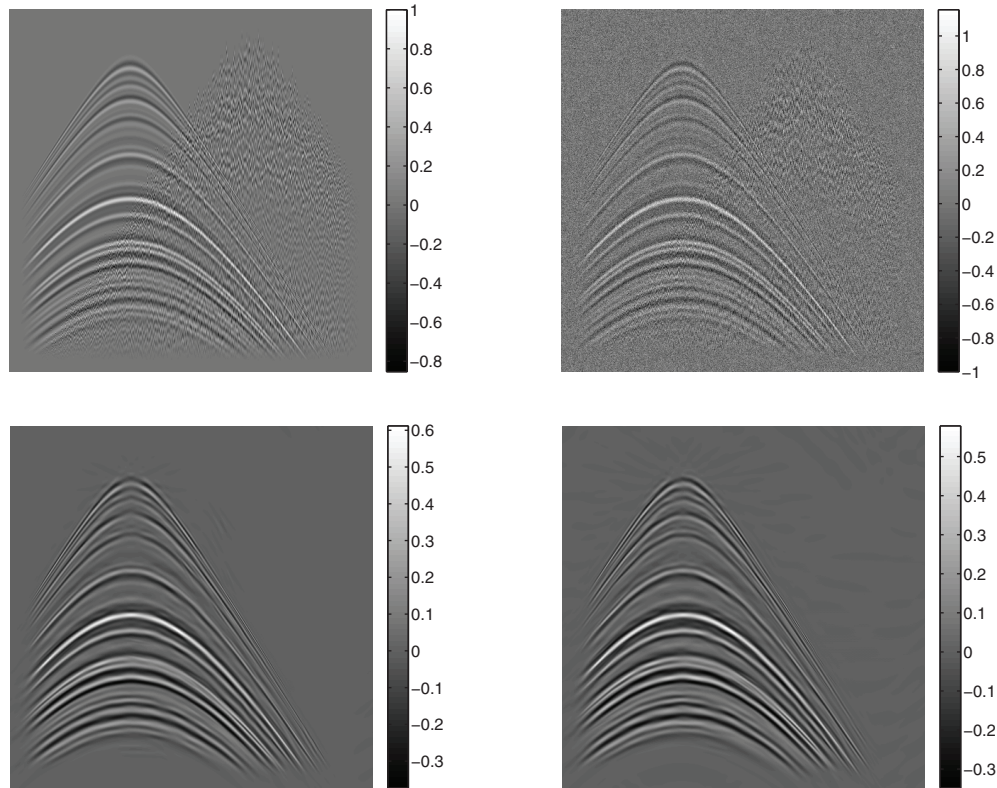


FIG. 8. *Example 3. Left: Mode identification without noise. Right: Mode identification with noise (SNR =  $-0.90$ ). Top: A superposition of two components, one of which is disrupted by random shifting and needs to be removed. Second row: The recovered relevant mode.*

the first example of adapting the synchrosqueezed transforms, by changing their geometric scaling parameters, to a superposition of components with specific structures. This framework has significant advantages over existing methods for 2D mode decompositions: it comes with a clean and solid theoretical analysis, and it can analyze and decompose a wide class of superpositions in many applications where related methods may not be well suited.

An appealing research direction is to study other types of data structures and other types of superpositions. In [33] and this article, the data is assumed to be a superposition of wave-like components. In more general circumstances, the oscillatory pattern should not be restricted to wave functions.

Another promising direction would be the robustness study and the optimization scheme for 2D mode decomposition. For one thing, hard thresholding can cause some energy loss while reducing the noise. It is important to study the behavior of the synchrosqueezing operator on noise and its thresholding strategies. For another thing, in other cases, some parts of the data are missing or have extremely weak energy. It is expected that an optimization scheme could estimate a clear structure of each component, even if there is missing data or severe noise.

Like the SSWPT, the current approach can be easily extended to 3D or higher dimensions. This direction should be relevant for applications.

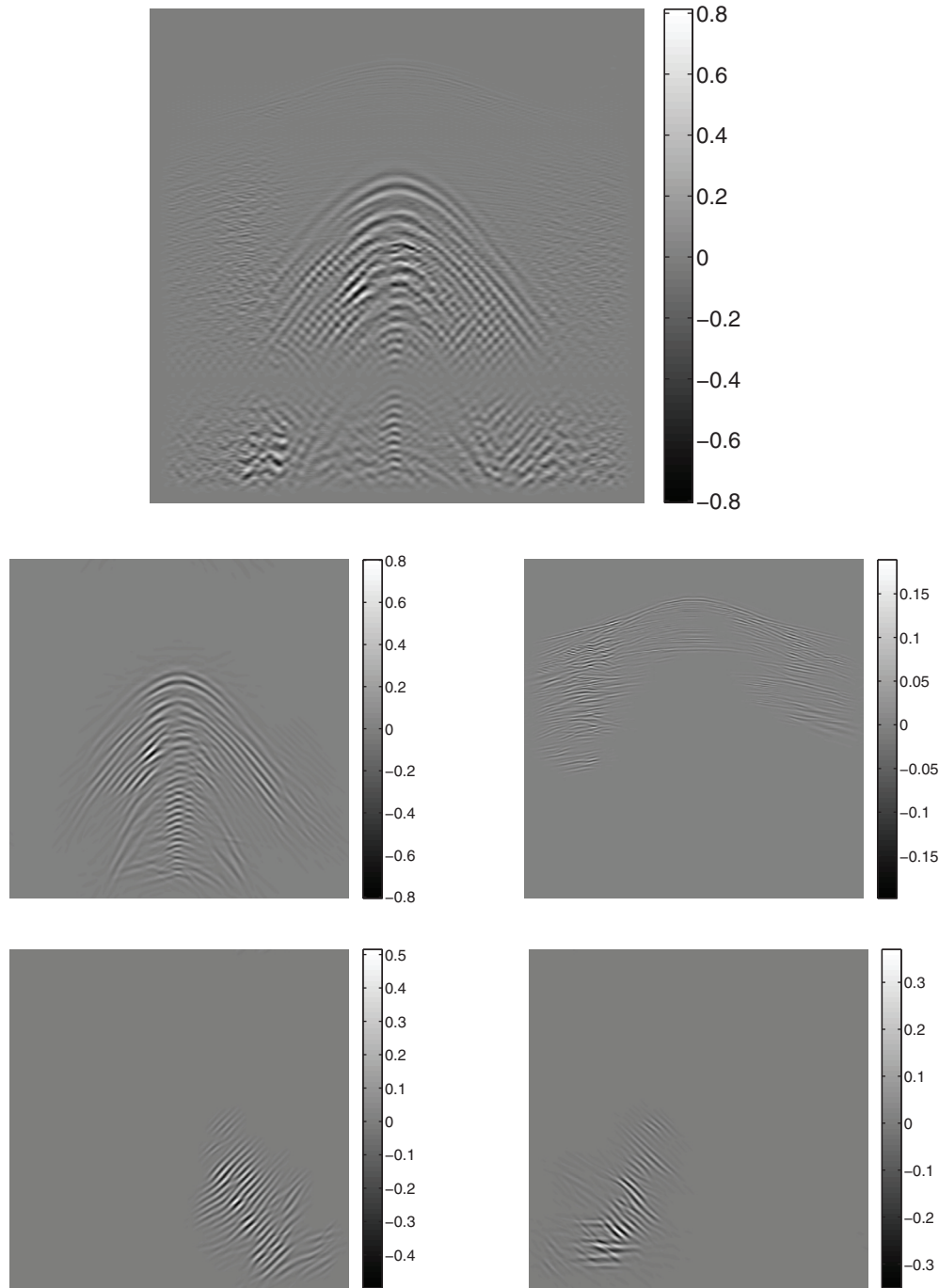


FIG. 9. *Example 4. Top: Real seismic data. Middle and bottom: Relevant recovered modes.*

**Acknowledgments.** H. Y. and L. Y. thank Jianfeng Lu, Jerome Gilles, and Giang Tran for discussions, and Sergey Fomel and Jingwei Hu for providing seismic application.

## REFERENCES

- [1] 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.
- [2] M. BROWN AND R. CLAPP, *(t, x) domain, pattern-based ground roll removal*, in Proceedings of the SEG Annual Meeting, Calgary, Alberta, Canada, 2000.
- [3] E. CANDÈS, L. DEMANET, D. DONOHO, AND L. YING, *Fast discrete curvelet transforms*, Multiscale Model. Simul., 5 (2006), pp. 861–899.
- [4] E. J. CANDÈS AND D. L. DONOHO, *New tight frames of curvelets and optimal representations of objects with piecewise  $C^2$  singularities*, Comm. Pure Appl. Math., 57 (2004), pp. 219–266.
- [5] E. J. CANDÈS AND D. L. DONOHO, *Continuous curvelet transform. II. Discretization and frames*, Appl. Comput. Harmon. Anal., 19 (2005), pp. 198–222.
- [6] E. CHASSANDE-MOTTIN, F. AUGER, AND P. FLANDRIN, *Time-frequency/time-scale reassignment*, in Wavelets and Signal Processing, Appl. Numer. Harmon. Anal., Birkhäuser Boston, Boston, MA, 2003, pp. 233–267.
- [7] E. CHASSANDE-MOTTIN, I. DAUBECHIES, F. AUGER, AND P. FLANDRIN, *Differential reassignment*, IEEE Signal Process. Lett., 4 (1997), pp. 293–294.
- [8] T. H. CORMEN, C. STEIN, R. L. RIVEST, AND C. E. LEISERSON, *Introduction to Algorithms*, 2nd ed., McGraw-Hill Higher Education, New York, 2001.
- [9] 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.
- [10] I. DAUBECHIES AND S. MAES, *A nonlinear squeezing of the continuous wavelet transform based on auditory nerve models*, in Wavelets in Medicine and Biology, CRC Press, Boca Raton, FL, 1996, pp. 527–546.
- [11] L. DEMANET AND L. YING, *Wave atoms and sparsity of oscillatory patterns*, Appl. Comput. Harmon. Anal., 23 (2007), pp. 368–387.
- [12] S. FOMEL, *Applications of plane-wave destruction filters*, Geophysics, 67 (2002), pp. 1946–1960.
- [13] J. GILLES, *Empirical wavelet transform*, IEEE Trans. Signal Process., 61 (2013), pp. 3999–4010.
- [14] J. GILLES, G. TRAN, AND S. OSHER, *2D empirical transforms. Wavelets, ridgelets and curvelets revisited*, SIAM J. Imaging Sci., 7 (2014), pp. 157–186.
- [15] T. HOU, Z. SHI, AND P. TAVALLALI, *Convergence of a data-driven time-frequency analysis method*, preprint, arXiv:1303.7048, 2013.
- [16] T. Y. HOU AND Z. SHI, *Adaptive data analysis via sparse time-frequency representation*, Adv. Adapt. Data Anal., 3 (2011), pp. 1–28.
- [17] T. Y. HOU AND Z. SHI, *Data-driven time-frequency analysis*, Appl. Comput. Harmon. Anal., 35 (2013), pp. 284–308.
- [18] T. Y. HOU, M. P. YAN, AND Z. WU, *A variant of the EMD method for multi-scale data*, Adv. Adapt. Data Anal., 1 (2009), pp. 483–516.
- [19] N. E. HUANG, *Computer Implemented Empirical Mode Decomposition Apparatus, Method and Article of Manufacture for Two-Dimensional Signals*, US Patent 6,311,130 B1, 2001.
- [20] N. E. HUANG, Z. SHEN, S. R. LONG, M. C. WU, H. H. SHIH, Q. ZHENG, N.-C. YEN, C. C. TUNG, AND H. H. LIU, *The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis*, R. Soc. Lond. Proc. Ser. A Math. Phys. Eng. Sci., 454 (1998), pp. 903–995.
- [21] N. E. HUANG, Z. WU, S. R. LONG, K. C. ARNOLD, X. CHEN, AND K. BLANK, *On instantaneous frequency*, Adv. Adapt. Data Anal., 1 (2009), pp. 177–229.
- [22] R. E. KORF, *Depth-first iterative deepening: An optimal admissible tree search*, Artificial Intelligence, 27 (1985), pp. 97–109.
- [23] A. LINDERHED, *Variable sampling of the empirical mode decomposition of two-dimensional signals*, Int. J. Wavelets Multiresolut. Inf. Process, 3 (2005), pp. 435–452.
- [24] A. LINDERHED, *Image empirical mode decomposition: A new tool for image processing*, Adv. Adapt. Data Anal., 1 (2009), pp. 265–294.
- [25] S. R. LONG, *Applications of HHT in image analysis*, Hilbert–Huang Transform and Its Applications, N. E. Huang and S. S. P. Shen, eds., World Scientific, River Edge, NJ, 2005, pp. 289–306.
- [26] J. C. NUNES, Y. BOUAOUNE, E. DELECELLE, O. NIANG, AND P. BUNEL, *Image analysis by bidimensional empirical mode decomposition*, Image Vision Comput., 21 (2003), pp. 1019–1026.
- [27] J. C. NUNES, O. NIANG, Y. BOUAOUNE, E. DELECELLE, AND P. BUNEL, *Bidimensional empirical mode decomposition modified for texture analysis*, Image Anal. Proc., 2749 (2003), pp. 171–177.

- [28] S. RICHWALSKI, K. ROY-CHOWDHURY, AND J. C. MONDT, *Multi-component wavefield separation applied to high-resolution surface seismic data*, J. Appl. Geophys., 46 (2001), pp. 101–114.
- [29] J.-L. STARCK, E. J. CANDÈS, AND D. L. DONOHO, *The curvelet transform for image denoising*, IEEE Trans. Image Process., 11 (2002), pp. 670–684.
- [30] M. VAN DER BAAN, *PP/PS wavefield separation by independent component analysis*, Geophys. J. Int., 166 (2006), pp. 339–348.
- [31] Z. WU AND N. E. HUANG, *Ensemble empirical mode decomposition: A noise-assisted data analysis method*, Adv. Adapt. Data Anal., 1 (2009), pp. 1–41.
- [32] Z. WU, N. E. HUANG, AND X. CHEN, *The multi-dimensional ensemble empirical mode decomposition method*, Adv. Adapt. Data Anal., 1 (2009), pp. 339–372.
- [33] H. YANG AND L. YING, *Synchrosqueezed wave packet transform for 2D mode decomposition*, SIAM J. Imaging Sci., 6 (2013), pp. 1979–2009.
- [34] C. YARHAM, U. BOENIGER, AND F. HERRMANN, *Curvelet based ground roll removal*, in Proceedings of the 2006 SEG Annual Meeting, New Orleans, LA, 2006.