SIAM J. IMAGING SCIENCES
Vol. 6, No. 4, pp. 1979–2009

Synchrosqueezed Wave Packet Transform for 2D Mode Decomposition*

Haizhao Yang† and Lexing Ying‡

Abstract. This paper introduces the synchrosqueezed wave packet transform as a method for analyzing two-dimensional images. This transform is a combination of wave packet transforms of a certain geometric scaling, a reallocation technique for sharpening phase space representations, and clustering algorithms for modal decomposition. For a function that is a superposition of several wave-like components with a highly oscillatory pattern satisfying certain separation conditions, we prove that the synchrosqueezed wave packet transform identifies these components and estimates their local wavevectors. A discrete version of this transform is discussed in detail, and numerical results are given to demonstrate the properties of the proposed transform.

Key words. wave packet transform, synchrosqueezing, clustering, local wavevector, phase space representation, empirical mode decomposition

AMS subject classifications. 42A99, 65T99

DOI. 10.1137/120891113

1. Introduction.

In many applications, a typical time signal can be viewed as a superposition of several simple components, each of which is localized in time-frequency (or phase space) representation and exhibits well-defined, often nonstationary, instantaneous frequency [18]. An important task in analyzing these signals is to identify these simple components and estimate their instantaneous frequencies. Time-frequency analysis provides a wide range of tools for this task. Most of these tools fall into two categories: linear and quadratic methods, each of which has its own strengths and weaknesses. The linear methods are typically efficient and easy to reconstruct but provide poor resolution. The quadratic methods, on the other hand, provide better resolution but suffer from higher computational cost, a more difficult reconstruction process, and nonphysical interference between multiple components. Among the approaches proposed to remedy this problem, the reallocation (or reassignment) methods [1, 7, 8, 11] can be viewed as standard linear methods, followed by reassigning values of the time-frequency representation based upon their local oscillation. One such method is the synchrosqueezed wavelet transform, which was proposed in [11] and given rigorous justification for an important class of signals (superpositions of approximate sinusoidal waves with well-separated frequencies at each location) in [10].

An obvious question, which is motivated by various applications [26, 28], is whether the synchrosqueezing idea can be extended to two-dimensional (2D) images. For example, in

*Received by the editors September 12, 2012; accepted for publication (in revised form) June 24, 2013; published electronically October 22, 2013.
http://www.siam.org/journals/siims/6-4/89111.html

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

‡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.

seismic imaging analysis, different local wavevectors correspond to different seismic events, which typically link to different geological features. A naïve attempt would simply combine the 2D wavelet transform with the synchrosqueezing approach. The resulting synchrosqueezed 2D wavelet transform would be capable of separating components that have different wavevectors at each location, just as the 1D transform does for 1D signals. However, in many situations this is not enough, since a typical 2D image can have components whose wavevectors have the same magnitude but point in different directions, as shown in Figure 1(left). In fact, images from many applications related to high-frequency wave propagation have this feature. In order to distinguish these modes, we propose the synchrosqueezed wave packet transform, which combines the synchrosqueezing idea with wave packets of an appropriate geometric scaling. The key feature is that these wave packets have finer and, more importantly, directional support in the 2D Fourier domain, which allows anisotropic angular separation in the Fourier domain, i.e., distinguishing components oscillating in different directions, as shown in Figure 1(right). As far as we know, the synchrosqueezed wave packet transform is the first method to date equipped with this ability.

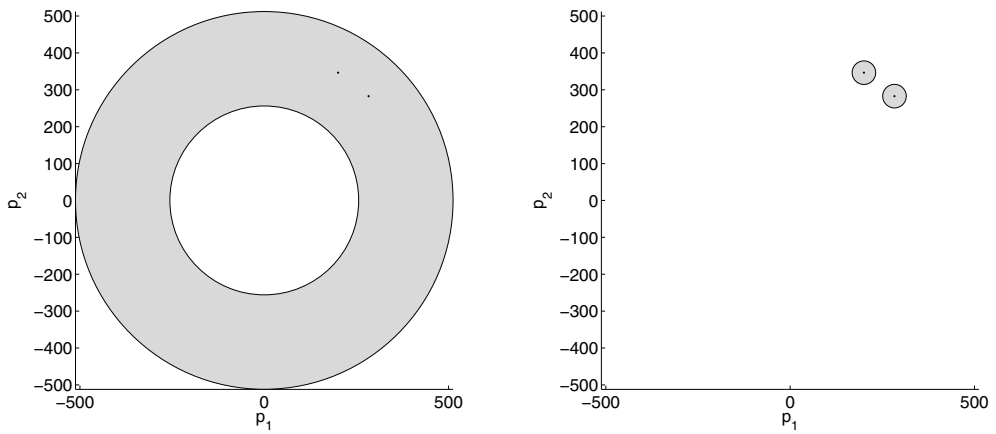

Figure 1. *Comparison of the resolutions of continuous wavelets (left) and continuous wave packets (right) in the Fourier domain. Consider the superposition of two plane waves $e^{2\pi ip\cdot x}$ and $e^{2\pi iq\cdot x}$ with the same frequency ($|p| = |q|$) but different wavevectors ($p \neq q$). The two dots in each plot show the support of the Fourier transforms of these two plane waves. Left: The gray region stands for the support of a continuous wavelet. Since the isotropic support of each wavelet either covers or misses both points p and q, the wavelet transform is not able to distinguish these two plane waves. Right: Each gray region represents the support of a wave packet. As long as p and q are well separated, they are in the support of two different wave packets. Hence these two plane waves can be distinguished from each other by the wave packet transform.*

1.1. Synchrosqueezed wave packet transform. In what follows, we introduce the synchrosqueezed wave packet transform, alongside several simple motivating examples. Let $w(x)$ for $x \in \mathbb{R}^2$ be the mother wave packet, which is used to define all wave packets through scaling, modulation, and translation. Suppose that $w(x)$ is in the Schwartz class, and that the Fourier transform $\hat{w}(\xi)$ is a radial, real-valued, nonnegative, smooth function with support equal to the unit ball $B_1(0)$ in the Fourier domain. Based on $w(x)$, we can define a family of wave packets through scaling, modulation, and translation as follows, controlled by a geometric parameter s.

Definition 1.1. *Given the mother wave packet $w(x)$ and the parameter $s \in (1/2, 1)$, the family of wave packets $\{w_{pb}(x) : p, b \in \mathbb{R}^2, |p| \geq 1\}$ is defined as*

$$w_{pb}(x) = |p|^s w(|p|^s(x-b))e^{2\pi i(x-b)\cdot p},$$

or, equivalently, in the Fourier domain, as

$$\widehat{w_{pb}}(\xi) = |p|^{-s} e^{-2\pi i b\cdot\xi}\hat{w}(|p|^{-s}(\xi - p)).$$

It is essential for our purposes that s is between $1/2$ and 1. As we shall see, the upper bound $s < 1$ enables the wave packets to detect oscillations in different directions, while the lower bound $s > 1/2$ makes the support of the wave packets sufficiently small for local wavevector estimation. It is clear from the definition that the Fourier transform $\widehat{w_{pb}}(\xi)$ is supported in $B_{|p|^s}(p)$, a ball centered at p with radius $|p|^s$. On the other hand, $w_{pb}(x)$ is centered in space at b with an essential support of width $O(|p|^{-s})$. Further, $\{w_{pb}(x) : p, b \in \mathbb{R}^2, |p| \geq 1\}$ are all appropriately scaled to have the same L^2 norm with the mother wave packet $w(x)$. Notice that if s were equal to 1, these functions would be qualitatively similar to the standard 2D wavelets. On the other hand, if s were equal to $1/2$, we would obtain the wave atoms defined in [12].

In this definition, we require $|p| \geq 1$. The reason is that, when $|p| < 1$, the above consideration regarding the shape of the wave packets is no longer valid. However, since we are mostly concerned with the high frequencies as the signals of interest here are oscillatory, the case $|p| < 1$ is essentially irrelevant.

Equipped with this family of wave packets, we can define the wave packet transform as follows.

Definition 1.2. *The wave packet transform of a function $f(x)$ is a function*

$$W_f(p,b) = \langle w_{pb}, f\rangle = \int \overline{w_{pb}(x)} f(x)dx \tag{1.1}$$
$$= \langle \widehat{w_{pb}}, \hat{f}\rangle = \int \overline{\widehat{w_{pb}}(\xi)}\hat{f}(\xi)d\xi$$

for $p, b \in \mathbb{R}^2, |p| \geq 1$.

If the Fourier transform $\hat{f}(\xi)$ vanishes for $|\xi| < 1$, it is easy to check that the L^2 norms of $W_f(p,b)$ and $f(x)$ are equivalent, up to a uniform constant factor; i.e.,

$$\int |W_f(p,b)|^2 dpdb \approx \int |f(x)|^2 dx. \tag{1.2}$$

As a simple example, let us consider the wave packet transform for a plane wave function

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

where α and β are nonzero constants of order $O(1)$ and N is a sufficiently large constant. The

local wavevector is $N\beta$, and applying the wave packet transform to $f(x)$ gives

$$\begin{aligned} W_f(p,b) &= \int_{\mathbb{R}^2} \alpha e^{2\pi i N\beta\cdot x}|p|^s w(|p|^s(x-b))e^{-2\pi i(x-b)\cdot p}dx \\ &= |p|^{-s}\alpha\int_{\mathbb{R}^2} e^{2\pi i N\beta\cdot(b+|p|^{-s}y)}w(y)e^{-2\pi i|p|^{-s}p\cdot y}dy \\ &= |p|^{-s}\alpha e^{2\pi i N\beta\cdot b}\overline{\hat{w}(|p|^{-s}(N\beta-p))}. \end{aligned}$$

Since $\hat{w}(\xi)$ is compactly supported in the unit ball, for each fixed b the coefficients $W_f(p,b)$ are nonzero if p satisfies

$$|p-N\beta| \le |p|^s.$$

This implies that, for each b, $W_f(p,b)$ has an essential support of width $O(|N\beta|^s)$ around the wavevector $N\beta$ in the p variable. The essential observation of synchrosqueezing is that the oscillation of $W_f(p,b)$ in the b variable in fact encodes the correct wavevector $N\beta$, independently of the amplitude α or the position b. More precisely, the derivative of $W_f(p,b)$ with respect to b and $W_f(p,b)$ satisfies the following equation:

$$\frac{\nabla_b W_f(p,b)}{2\pi i W_f(p,b)} = \frac{2\pi i N\beta|p|^{-s}\alpha e^{2\pi i N\beta\cdot b}\overline{\hat{w}(|p|^{-s}(N\beta-p))}}{2\pi i|p|^{-s}\alpha e^{2\pi i N\beta\cdot b}\overline{\hat{w}(|p|^{-s}(N\beta-p))}} = N\beta$$

for $W_f(p,b)\neq 0$.

Let us consider now a general function of the form

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

with smooth amplitude $\alpha(x)$, smooth phase $\phi(x)$, and sufficiently large N. As we shall see, for each b the wave packet transform $W_f(p,b)$ is essentially supported in the following set:

$$\{p : |p-N\nabla\phi(b)| \lesssim |p|^s\}.$$

This motivates us to define the local wavevector estimation for a general function $f(x)$ as follows.

Definition 1.3. *The local wavevector estimation of a function $f(x)$ at (p,b) is*

$$v_f(p,b) = \frac{\nabla_b W_f(p,b)}{2\pi i W_f(p,b)} \tag{1.3}$$

for $p,b\in\mathbb{R}^2$ with $W_f(p,b)\neq 0$.

Though $v_f(p,b)$ is defined here for any p,b with $W_f(p,b)\neq 0$, we want to emphasize that it is relevant only when $|W_f(p,b)|$ is above a certain threshold value. This will be made more precise in the analysis and implementation in later sections.

Given the wavevector estimation $v_f(p,b)$, the synchrosqueezing step reallocates the information in the phase space and provides a sharpened phase space representation of $f(x)$.

Definition 1.4. *Given $f(x)$, $W_f(p,b)$, and $v_f(p,b)$, the synchrosqueezed energy distribution $T_f(v,b)$ is defined by*

$$T_f(v,b) = \int |W_f(p,b)|^2 \delta(\Re v_f(p,b) - v) dp \tag{1.4}$$

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

As we shall see, for $f(x) = \alpha(x)e^{2\pi i N\phi(x)}$ with sufficiently smooth amplitude $\alpha(x)$ and sufficiently steep phase $N\phi(x)$, we can show that for each b the estimation $v_f(p,b)$ indeed approximates $N\nabla\phi(b)$ independently of p as long as $W_f(p,b)$ is nonnegligible. As a direct consequence, for each b, the essential support of $T_f(v,b)$ in the v variable concentrates near $N\nabla\phi(b)$ (see Figure 2 for an example). In addition, we have the following property:

$$\int T_f(v,b) dv db = \int |W_f(p,b)|^2 \delta(\Re v_f(p,b) - v) dv dp db = \int |W_f(p,b)|^2 dp db \approx \|f\|_2^2$$

from Fubini's theorem and the norm equivalence (1.2), for any $f(x)$ with its Fourier transform vanishing for $|\xi| < 1$.

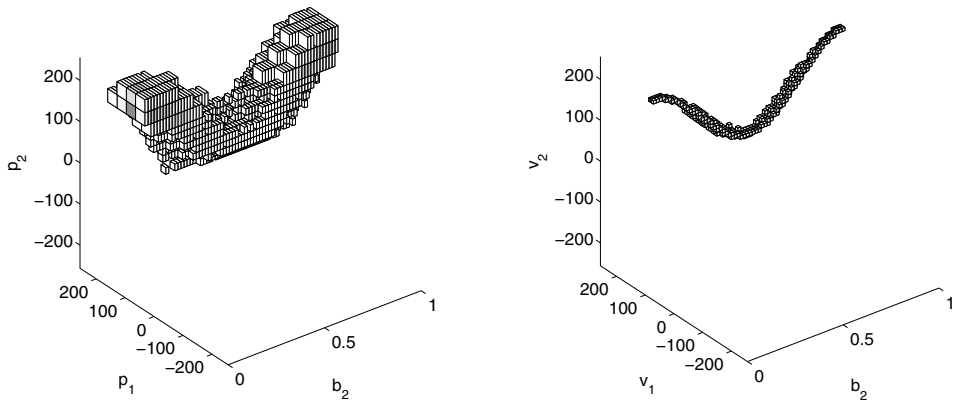

Figure 2. *Synchrosqueezed wave packet transform applied to a deformed plane wave $f(x) = \alpha(x)e^{2\pi i N\phi(x)}$, which is the first component of the signal in Example 2 in section 4. Left: The essential support of the wave packet transform $W_f(p,b)$ at $b_1 = 1$. Right: The essential support of the synchrosqueezed energy distribution $T_f(v,b)$ at the same b_1 value. $W_f(p,b)$ has been reallocated to form a sharp phase space representation $T_f(v,b)$.*

Let us now informally discuss why the synchrosqueezed wave packet transform allows one to identify individual pieces of a superposition of multiple components. For simplicity, let

$$f(x) = \alpha_1(x)e^{2\pi i N\phi_1(x)} + \alpha_2(x)e^{2\pi i N\phi_2(x)},$$

with smooth amplitudes $\alpha_1(x)$ and $\alpha_2(x)$ and smooth phases $N\phi_1(x)$ and $N\phi_2(x)$ for sufficiently large N (see Figure 3(top-left)). Let us assume that at each position the local wavevectors $N\nabla\phi_1(x)$ and $N\nabla\phi_2(x)$ are sufficiently large and well separated from each other

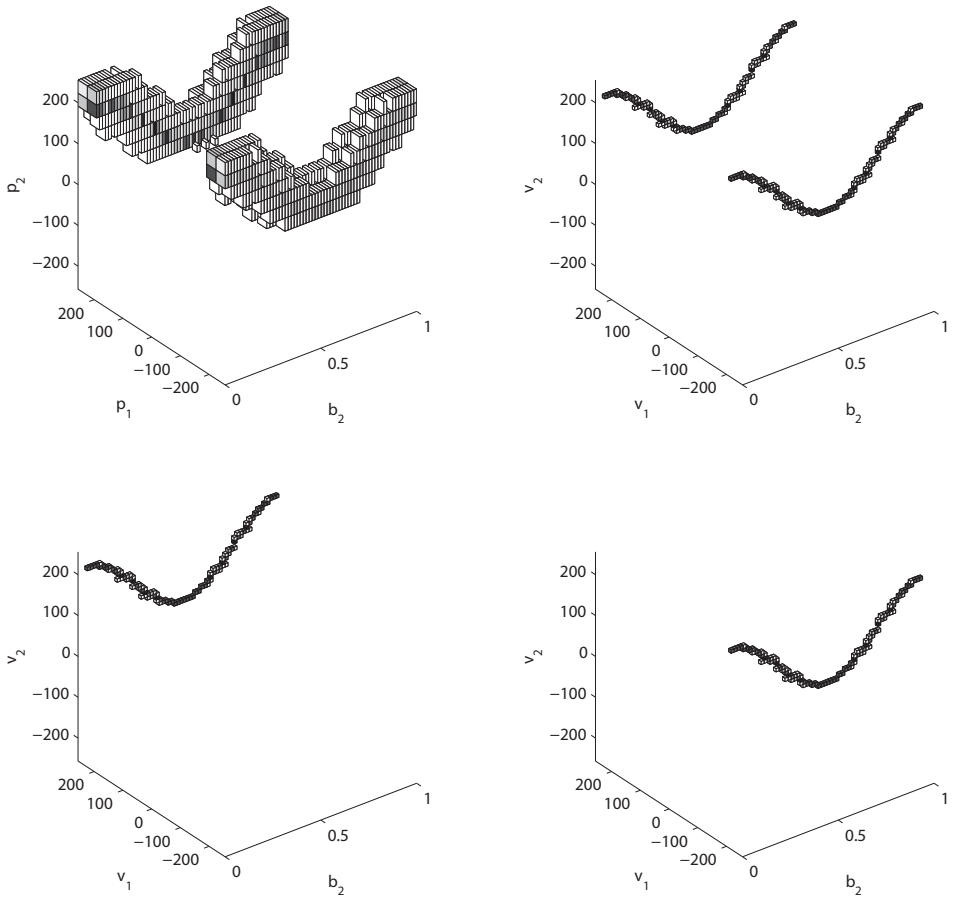

Figure 3. *Synchrosqueezed wave packet transform applied to a superposition of two deformed plane waves,* $f(x) = e^{2\pi i N\phi_1(x)} + e^{2\pi i N\phi_2(x)}$*, given in Example 2 in section 4. Top-left: The essential support of the wave packet transform* $W_f(p,b)$ *at* $b_1 = 1$*. Top-right: The essential support of the synchrosqueezed energy distribution* $T_f(v,b)$ *at the same* b_1 *value. At a fixed* b_1 *value,* $T_f(v,b)$ *is more concentrated near two curves. In general, in the full 4D phase space,* $T_f(v,b)$ *is concentrated near two separated 2D surfaces in the 4D phase space, which makes it easier to separate these components with clustering techniques. The second row shows the support of two sets* U_1 *and* U_2 *after the clustering algorithm is applied. Each set contains the synchrosqueezed energy distribution of one deformed plane wave.*

(this will be made precise later). From the above discussion, we know that for each p the wave packet transform $W_f(p,b)$ is essentially supported in two sets,

$$\{(p,b) : |p - N\nabla\phi_1(b)| \lesssim |p|^s\}, \qquad \{(p,b) : |p - N\nabla\phi_2(b)| \lesssim |p|^s\}.$$

Since both $|N\nabla\phi_1(b)|$ and $|N\nabla\phi_2(b)|$ are large and $1/2 \le s \le 1$, the first set is within distance $O(|N\nabla\phi_1(b)|^s)$ from $N\nabla\phi_1(b)$, and the second one is within distance $O(|N\nabla\phi_2(b)|^s)$ from $N\nabla\phi_2(b)$. Since $N\nabla\phi_1(x)$ and $N\nabla\phi_2(b)$ are sufficiently well separated, these two sets are essentially disjoint. Therefore,

$$v_f(p,b) \approx N\nabla\phi_1(b)$$

for p in the first set, and

$$v_f(p,b) \approx N\nabla\phi_2(b)$$

for p in the second one (see Figure 3(top-right) for an example). After synchrosqueezing, the energy distribution $T_f(v,b)$ then essentially concentrates near two two-dimensional surfaces, $S_1 = \{(N\nabla\phi_1(b), b) : b \in \mathbb{R}^2\}$ and $S_2 = \{(N\nabla\phi_2(b), b) : b \in \mathbb{R}^2\}$, in the four-dimensional phase space. Since S_1 and S_2 are well separated, one expects the *essential* support of $T_f(v,b)$ to separate into two disjoint regions U_1 and U_2, where $U_1 \supset S_1$ and $U_2 \supset S_2$ contain the synchrosqueezed coefficients from the first mode $\alpha_1(x)e^{2\pi iN\phi_1(x)}$ and the second mode $\alpha_2(x)e^{2\pi iN\phi_2(x)}$, respectively. Typically, these two sets U_1 and U_2 can be identified with standard clustering algorithms (see the second row of Figure 3).

Once U_1 and U_2 are identified, we can extract individual modes with

$$f_1(x) = \int_{p,b:\Re v_f(p,b)\in U_1} \tilde{w}_{pb}(x) W_f(p,b)\,dpdb, \tag{1.5}$$

$$f_2(x) = \int_{p,b:\Re v_f(p,b)\in U_2} \tilde{w}_{pb}(x) W_f(p,b)\,dpdb,$$

where the set of functions $\{\tilde{w}_{pb}(x)\}$ is a dual frame of $\{w_{pb}(x)\}$.

1.2. Related work. Extracting individual components and estimating local wavevectors is an essential problem in adaptive data analysis. There has been a long history behind applying linear and quadratic methods from applied harmonic analysis to this problem; this line of research is summarized in [21], for example. The synchrosqueezing method was originally proposed for auditory signal processing in [11] and has been shown to provide good results for 1D signals even under a substantial amount of noise. The recent work in [10] provides an important step towards understanding synchrosqueezing on 1D mode decomposition, and it is generalized to a 2D model in [9]. Since the 2D synchrosqueezed wavelet transform in [9] uses only a 1D spectral line of the wavenumber (the length of the anisotropic wavevector) for decomposition, it cannot distinguish two modes with the same wavenumber but different wavevectors. Even if two modes have different but similar wavenumbers, the 2D synchrosqueezed wavelet transform is not able to distinguish them. [29] has discussed this phenomenon in a simplified model.

A different but related approach for component extraction is the empirical mode decomposition (EMD) initiated and refined by Huang and coworkers [17, 18]. Given a superposition of simple components, called intrinsic mode functions, this approach recursively extracts these components starting from the most oscillatory. Typically, the most oscillatory component is estimated by computing local minima and maxima and applying spline interpolation to these extrema to estimate the envelope function. Though this method has been widely used for real data analysis, it is rather sensitive to noise, and its mathematical analysis is still under development. To address the problem of noise, several variants of EMD have been proposed more recently [15, 30] and shown to improve results for data from practical applications. However, mathematical analysis is still not a resolved problem and is currently under active research. Following the same methodology of extracting each component one-by-one from the most oscillatory one, Hou and Shi have developed an optimization framework for solving

the mode decomposition problem, based on total variation [13] or sparse representation in a data-driven dictionary of basis functions [14]. Each component extracted in the optimization scheme belongs to the class of intrinsic mode functions defined in EMD.

In the spirit of EMD, many methods in higher-dimensional cases have been proposed recently. They fall into two types: one is based on surface interpolation [23, 24, 19, 20], and the other one is based on decomposition of 1D slices [16, 31]. These methods cannot distinguish two modes with the same wavenumber but different wavevectors by lack of anisotropic angular separation. Let us consider a simple superposition of the form

$$f(x) = \cos(2\pi(x_1 + x_2)) + \cos(2\pi(x_1 - x_2)).$$

Surface interpolation methods consider $f(x)$ as one single 2D mode by definition. On the other hand, because each 1D slice of $f(x)$ contains two modes with the same wavenumber, the methods based on 1D decomposition would consider the superposition as one single mode as well. Even if the wavenumbers are different but close, the number of modes answered by the EMD methods would still be one, as explained in [27].

Comparing these two approaches, we note that the 1D models considered in the EMD approaches are typically more general. Nevertheless, the simpler models considered here and in [10, 11] allow for a more thorough analysis of synchrosqueezed transforms applied to these signals and serve as a first step for the analysis of more complicated models considered in the EMD approach.

The rest of the paper is organized as follows. Section 2 contains the main theoretical result of this paper. We prove that when the local wavevectors of multiple components are well separated at each position, the synchrosqueezed wave packet transform is able to estimate the local wavevectors. In section 3, a discrete version of the synchrosqueezed wave packet transform is introduced in detail. Section 4 provides several numerical examples on local wavevector estimation and mode decomposition. Finally, we conclude with some discussions in section 5.

2. Analysis of the transform.

In this section, we show that the synchrosqueezed wave packet transform can distinguish well-separated local wavevectors for a superposition of multiple components. We start by providing precise definitions for these components and the superposition in two dimensions, following the model used in Daubechies, Lu, and Wu [10].

Definition 2.1. *A function $f(x) = \alpha(x)e^{2\pi i N\phi(x)}$ is an intrinsic mode function of type (M, N) if $\alpha(x)$ and $\phi(x)$ satisfy*

$$\alpha(x) \in C^\infty, \quad |\nabla\alpha| \le M, \quad 1/M \le \alpha \le M$$
$$\phi(x) \in C^\infty, \quad 1/M \le |\nabla\phi| \le M, \quad |\nabla^2\phi| \le M.$$

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

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

where each $f_k(x) = \alpha_k(x)e^{2\pi i N\phi_k(x)}$ *is an intrinsic mode function of type* (M,N) *and the phase functions satisfy the separation condition*

$$|N\nabla\phi_k(b) - N\nabla\phi_l(b)| \geq 2^{1+s}(|N\nabla\phi_k(b)|^s + |N\nabla\phi_l(b)|^s)$$

for any $1 \leq k,l \leq K$. *We denote by* $F(M,N,K)$ *the set of all such functions.*

Let us recall that $W_f(p,b)$ is the wave packet transform with geometric scaling parameter $s \in (1/2,1)$ of a function $f(x)$ and that $v_f(p,b)$ is the local wavevector estimation. The following theorem is our main theoretical result.

Theorem 2.3. *For a function* $f(x)$ *and* $\varepsilon > 0$ *we define*

$$R_{f,\varepsilon} = \{(p,b) : |W_f(p,b)| \geq |p|^{-s}\sqrt{\varepsilon}\}$$

and

$$Z_{f,k} = \{(p,b) : |p - N\nabla\phi_k(b)| \leq |p|^s\}$$

for $1 \leq k \leq K$. *For fixed* M *and* K *there exists a constant* $\varepsilon_0(M,K) > 0$ *such that for any* $\varepsilon \in (0,\varepsilon_0)$ *there exists a constant* $N_0(M,K,s,\varepsilon) > 0$ *such that for any* $N > N_0(M,K,s,\varepsilon)$ *and* $f(x) \in F(M,N,K)$ *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* $(p,b) \in R_{f,\varepsilon} \cap Z_{f,k}$,

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

In what follows, when we write $O(\cdot)$, $\lesssim$, or $\gtrsim$, the implicit constants may depend on M and K. The proof of the theorem relies on several lemmas. The following one estimates $W_f(p,b)$.

Lemma 2.4. *Under the assumption of the theorem, we have*

(2.1)
$$W_f(p,b) = \begin{cases} |p|^{-s}O(\varepsilon), & |p| \notin [\frac{N}{2M}, 2MN], \\ |p|^{-s}\left(\sum_{k=1}^K \alpha_k(b)e^{2\pi i N\phi_k(b)}\hat{w}\left(|p|^{-s}(p - N\nabla\phi_k(b))\right) + O(\varepsilon)\right), & |p| \in [\frac{N}{2M}, 2MN]. \end{cases}$$

Proof. Let us first estimate $W_f(p,b)$ by assuming that $f(x)$ contains a single intrinsic mode function of type (M,N):

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

Using the definition of the wave packet transform, we have the following expression for $W_f(p,b)$:

$$\begin{aligned} W_f(p,b) &= \int \alpha(x)e^{2\pi i N\phi(x)}|p|^s w(|p|^s(x-b))e^{-2\pi i(x-b)\cdot p}dx \\ &= \int \alpha(b+|p|^{-s}y)e^{2\pi i N\phi(b+|p|^{-s}y)}|p|^s w(y)e^{-2\pi i|p|^{-s}y\cdot p}d(|p|^{-s}y) \\ &= |p|^{-s}\int \alpha(b+|p|^{-s}y)w(y)e^{2\pi i(N\phi(b+|p|^{-s}y)-|p|^{-s}y\cdot p)}dy. \end{aligned}$$

We claim that when N is sufficiently large,

$$W_f(p,b) = \begin{cases} |p|^{-s}O(\varepsilon), & |p| \notin [\frac{N}{2M}, 2MN], \\ |p|^{-s}\left(\alpha(b)e^{2\pi i N\phi(b)}\hat{w}\left(|p|^{-s}(p - N\nabla\phi(b))\right) + O(\varepsilon)\right), & |p| \in [\frac{N}{2M}, 2MN]. \end{cases} \tag{2.2}$$

First, let us consider the case $|p| \notin [\frac{N}{2M}, 2MN]$. Consider the integral

$$\int h(y)e^{ig(y)}dy$$

for smooth real functions $h(y)$ and $g(y)$, along with the differential operator

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

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

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

Assuming that $h(y)$ decays sufficiently quickly at infinity, we perform integration by parts r times to get

$$\int he^{ig}dy = \int h(L^r e^{ig})dy = \int ((L^*)^r h)e^{ig}dy,$$

where L^* is the adjoint of L. In the current setting, $W_f(p,b) = |p|^{-s}\int h(y)e^{ig(y)}dy$ with

$$h(y) = \alpha(b + |p|^{-s}y)w(y), \quad g(y) = 2\pi(N\phi(b + |p|^{-s}y) - |p|^{-s}y \cdot p),$$

where $h(y)$ clearly decays rapidly at infinity since $w(y)$ is in the Schwartz class. In order to understand the impact of L and L^*, we need to bound the norm of

$$\nabla g(y) = 2\pi\left(N\nabla\phi(b + |p|^{-s}y) - p\right)|p|^{-s}$$

from below when $|p| \notin [\frac{N}{2M}, 2MN]$. If $|p| < \frac{N}{2M}$, then

$$|\nabla g| \gtrsim (|N\nabla\phi| - |p|)|p|^{-s} \gtrsim |N\nabla\phi||p|^{-s}/2 \gtrsim N^{1-s}.$$

If $|p| > 2MN$, then

$$|\nabla g| \gtrsim (|p| - |N\nabla\phi|)|p|^{-s} \gtrsim |p| \cdot |p|^{-s}/2 \gtrsim (|p|)^{1-s} \gtrsim N^{1-s}.$$

Hence $|\nabla g| \gtrsim N^{1-s}$ if $|p| \notin [\frac{N}{2M}, 2MN]$. Since $|\nabla g| \neq 0$ and each L^* contributes a factor of order $1/|\nabla g|$,

$$\left|\int e^{ig(y)}((L^*)^r h)(y)dy\right| \lesssim N^{-(1-s)r}.$$

When $N \gtrsim \varepsilon^{-1/((1-s)r)}$, we obtain

$$\left|\int e^{ig(y)}((L^*)^r h)(y)dy\right| \lesssim \varepsilon.$$

Using the fact that $W_f(p,b) = |p|^{-s}\int h(y)e^{ig(y)}dy$, we have $|W_f(p,b)| \lesssim |p|^{-s}\varepsilon$.

Second, let us address the case $|p| \in [\frac{N}{2M}, 2MN]$. We want to approximate $W_f(p,b)$ with

$$|p|^{-s}\alpha(b)e^{2\pi i N\phi(x)}\hat{w}\left(|p|^{-s}(p - N\nabla\phi(b))\right).$$

Since $w(y)$ is in the Schwartz class, we can assume that $|w(y)| \leq \frac{C_m}{|y|^m}$ for some sufficient large m with C_m for $|y| \geq 1$. Therefore, the integration over $|y| \gtrsim \varepsilon^{-1/m}$ yields a contribution of at most order $O(\varepsilon)$. We can then estimate

$$|W_f(p,b)| = |p|^{-s}\left(\int_{|y|\lesssim\varepsilon^{-1/m}} \alpha(b+|p|^{-s}y)w(y)e^{2\pi i(N\phi(b+|p|^{-s}y)-|p|^{-s}y\cdot p)}dy + O(\varepsilon)\right).$$

A Taylor expansion of $\alpha(x)$ and $\phi(x)$ shows that

$$\alpha(b+|p|^{-s}y) = \alpha(b) + \nabla\alpha(b^*)\cdot|p|^{-s}y$$

and

$$\phi(b+|p|^{-s}y) = \phi(b) + \nabla\phi(b)\cdot(|p|^{-s}y) + \frac{1}{2}(|p|^{-s}y)^t\nabla^2\phi(b^*)(|p|^{-s}y),$$

where in each case b^* is a point between b and $b+|p|^{-s}y$. We want to drop the last term from the above formulas without introducing a relative error larger than $O(\varepsilon)$. We begin with the estimate

$$\int_{|y|\lesssim\varepsilon^{-1/m}} |\nabla\alpha\cdot|p|^{-s}yw(y)|dy \lesssim \varepsilon,$$

which holds if $\varepsilon^{-2/m}|\nabla\alpha\cdot|p|^{-s}y| \lesssim \varepsilon$, which is true when $|p|^{-s} \lesssim \varepsilon^{1+3/m}$. Since $|p| \in [\frac{N}{2M}, 2MN]$, the above holds if

$$N \gtrsim \varepsilon^{-(1+3/m)/s}. \tag{2.3}$$

We also need

$$\int_{|y|\lesssim\varepsilon^{-1/m}} |\alpha(b)w(y)e^{2\pi i(N\phi(b)+N\nabla\phi(b)\cdot|p|^{-s}y-|p|^{-s}y\cdot p)}|\cdot|e^{2\pi iN/2(|p|^{-s}y)^t\nabla^2\phi(|p|^{-s}y)} - 1|dy \lesssim \varepsilon.$$

Since $|e^{ix}-1| \leq |x|$, the above inequality is equivalent to

$$\int_{|y|\lesssim\varepsilon^{-1/m}} \alpha(b)w(y)e^{2\pi i(N\phi(b)+N\nabla\phi(b)\cdot|p|^{-s}y-|p|^{-s}y\cdot p)}|2\pi N/2(|p|^{-s}y)^t\nabla^2\phi(|p|^{-s}y)|dy \lesssim \varepsilon,$$

which is true if $\varepsilon^{-2/m}N(|p|^{-s}y)^t\nabla^2\phi(|p|^{-s}y) \lesssim \varepsilon$, which in turn holds if $N|p|^{-2s}|y|^2 \lesssim \varepsilon^{1+2/m}$. Because $|y| \lesssim \varepsilon^{-\frac{1}{m}}$ and $|p| \in [\frac{N}{2M}, 2MN]$, the above inequality is valid when

$$N \gtrsim \varepsilon^{-(1+4/m)/(2s-1)}. \tag{2.4}$$

In summary, for N larger than the maximum of the right-hand sides of (2.3) and (2.4), if $|p| \in [\frac{N}{2M}, 2MN]$, then we have

$$\begin{aligned} W_f(p,b) &= |p|^{-s}\left(\int_{|y|\lesssim\varepsilon^{-1/m}} \alpha(b)w(y)e^{2\pi i(N\phi(b)+N\nabla\phi(b)\cdot|p|^{-s}y-|p|^{-s}y\cdot p)}dy + O(\varepsilon)\right) \\ &= |p|^{-s}\left(\int_{|y|\lesssim\varepsilon^{-1/m}} \left(\alpha(b)e^{2\pi iN\phi(b)}\right) w(y)e^{2\pi i(N\nabla\phi(b)-p)\cdot|p|^{-s}y}dy + O(\varepsilon)\right) \\ &= |p|^{-s}\left(\int_{\mathbb{R}^2} \left(\alpha(b)e^{2\pi iN\phi(b)}\right) w(y)e^{2\pi i(N\nabla\phi(b)-p)\cdot|p|^{-s}y}dy + O(\varepsilon)\right) \\ &= |p|^{-s}\left(\alpha(b)e^{2\pi iN\phi(b)}\hat{w}\left(|p|^{-s}(p-N\nabla\phi(b))\right) + O(\varepsilon)\right), \end{aligned}$$

where the third line uses the fact that the integration of $w(y)$ outside the set $\{y : |y| \lesssim \varepsilon^{-1/m}\}$ is again of order $O(\varepsilon)$.

Now let us return to the general case, where $f(x)$ is a superposition of K well-separated intrinsic mode components:

$$f(x) = \sum_{k=1}^{K} f_k(x) = \sum_{k=1}^{K} \alpha_k(x)e^{2\pi iN\phi_k(x)},$$

as in Definition 2.2. By linearity of the wave packet transform and (2.2), we find

(2.5)
$$W_f(p,b) = \begin{cases} |p|^{-s}O(\varepsilon), & |p| \notin [\frac{N}{2M}, 2MN], \\ |p|^{-s}\left(\sum_{k=1}^{K} \alpha_k(b)e^{2\pi iN\phi_k(b)}\hat{w}\left(|p|^{-s}(p-N\nabla\phi_k(b))\right) + O(\varepsilon)\right), & |p| \in [\frac{N}{2M}, 2MN]. \end{cases}$$

The next lemma estimates $\nabla_b W_f(p,b)$ when $|p| \in [\frac{N}{2M}, 2MN]$, i.e., the case where $W_f(p,b)$ is nonnegligible.

Lemma 2.5. *Under the assumption of the theorem, we have*

$$\nabla_b W_f(p,b) = 2\pi iN|p|^{-s}\left(\sum_{k=1}^{K} \nabla\phi_k(b)\alpha_k(b)e^{2\pi iN\phi_k(b)}\hat{w}(|p|^{-s}(p-N\nabla\phi_k(b))) + O(\varepsilon)\right) \tag{2.6}$$

when $|p| \in [\frac{N}{2M}, 2MN]$.

Proof. The proof is similar to that of Lemma 2.4. Assume that $f(x)$ contains a single intrinsic mode function, i.e.,

$$f(x) = \alpha(x)e^{2\pi iN\phi(x)};$$

then

$$\begin{aligned}\nabla_b W_f(p,b) &= \int_{\mathbb{R}^2} \alpha(x) e^{2\pi i N\phi(x)} |p|^s \left(\nabla w(|p|^s(x-b))(-|p|^s) + 2\pi i p w(|p|^s(x-b))\right) e^{-2\pi i(x-b)\cdot p} dx \\ &= \int_{\mathbb{R}^2} \alpha(b+|p|^{-s}y) e^{2\pi i N\phi(b+|p|^{-s}y)} |p|^{-s} \left(\nabla w(y)(-|p|^s) + 2\pi i p w(y)\right) e^{-2\pi i |p|^{-s} y\cdot p} dy \\ &= \int_{\mathbb{R}^2} \alpha(b+|p|^{-s}y) e^{2\pi i N\phi(b+|p|^{-s}y)} |p|^{-s} \nabla w(y)(-|p|^s) e^{-2\pi i |p|^{-s} y\cdot p} dy \\ &\quad + \int_{\mathbb{R}^2} \alpha(b+|p|^{-s}y) e^{2\pi i N\phi(b+|p|^{-s}y)} |p|^{-s} 2\pi i p w(y) e^{-2\pi i |p|^{-s} y\cdot p} dy.\end{aligned}$$

Forming a Taylor expansion and following the same argument as in the proof of Lemma 2.4 gives the following approximation for $|p| \in [\frac{N}{2M}, 2MN]$:

$$\begin{aligned}\nabla_b W_f(p,b) &= \left(-2\pi i |p|^{-s}(p - N\nabla\phi(b))\alpha(b) e^{2\pi i N\phi(b)} \hat{w}(|p|^{-s}(p - N\nabla\phi(b))) + O(\varepsilon)\right) \\ &\quad + 2\pi i |p|^{-s} p \left(\alpha(b) e^{2\pi i N\phi(b)} \hat{w}\left(|p|^{-s}(p - N\nabla\phi(b))\right) + O(\varepsilon)\right) \\ &= 2\pi i N |p|^{-s} \left(\nabla\phi(b)\alpha(b) e^{2\pi i N\phi(b)} \hat{w}(|p|^{-s}(p - N\nabla\phi(b))) + O(\varepsilon)\right).\end{aligned}$$

For $f(x) = \sum_{k=1}^K f_k(x) = \sum_{k=1}^K \alpha_k(x) e^{2\pi i N\phi_k(x)}$, taking the sum over K terms gives

$$\nabla_b W_f(p,b) = 2\pi i N |p|^{-s} \left(\sum_{k=1}^K \left(\nabla\phi_k(b)\alpha_k(x) e^{2\pi i N\phi_k(b)} \hat{w}(|p|^{-s}(p - N\nabla\phi_k(b)))\right) + O(\varepsilon)\right)$$

for $|p| \in [\frac{N}{2M}, 2MN]$. ∎

We are now ready to prove the theorem.

Proof. Let us first consider (i). Suppose there exists $(p,b) \in Z_{f,k} \cap Z_{f,l}$ with $k \neq l$. Then

$$|p - N\nabla\phi_k(b)| \leq |p|^s, \quad |p - N\nabla\phi_l(b)| \leq |p|^s,$$

which implies

$$|p| \leq |p|^s + |N\nabla\phi_k(b)|, \quad |p| \leq |p|^s + |N\nabla\phi_l(b)|.$$

Since $s < 1$, we have

$$|p| \leq 2|N\nabla\phi_k(b)| \quad \text{and} \quad |p| \leq 2|N\nabla\phi_l(b)|$$

if N satisfies

$$N \geq M 2^{1/(1-s)}. \tag{2.7}$$

Therefore,

$$\begin{aligned}|N\nabla\phi_k(b) - N\nabla\phi_l(b)| &\leq |p - N\nabla\phi_k(b)| + |p - N\nabla\phi_l(b)| \\ &\leq 2|p|^s \leq 2 \cdot 2^s (|N\nabla\phi_k(b)|^s + |N\nabla\phi_l(b)|^s),\end{aligned}$$

which contradicts the separation assumption. Thus, all $Z_{f,k}$ are disjoint.

Let (p,b) be a point in $R_{f,\varepsilon} = \{(p,b) : |W_f(p,b)| \geq |p|^{-s}\sqrt{\varepsilon}\}$. Now let us define $N_0(M,K,s,\varepsilon)$ be the maximum of the right-hand sides of (2.3), (2.4), and (2.7). From the above lemma, we see that, for ε sufficiently small, if $N \geq N_0(M,K,s,\varepsilon)$ and $|p| \in [\frac{N}{2M}, 2MN]$, we have

$$W_f(p,b) = |p|^{-s}\left(\sum_{k=1}^{K}\alpha_k(b)e^{2\pi iN\phi_k(b)}\hat{w}\left(|p|^{-s}(p-N\nabla\phi_k(b))\right)+O(\varepsilon)\right).$$

Therefore, there exists k between 1 and K such that $\hat{w}\left(|p|^{-s}(p-N\nabla\phi_k(b))\right)$ is nonzero. From the definition of $\hat{w}(\xi)$, we see that this implies $(p,b) \in Z_{f,k}$. Hence $R_{f,\varepsilon} \subset \bigcup_{k=1}^{K} Z_{f,k}$.

To show (ii), let us recall that $v_f(p,b)$ is defined as

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

for $W_f(p,b) \neq 0$. If $(p,b) \in R_{f,\varepsilon} \cap Z_{f,k}$, then

$$W_f(p,b) = |p|^{-s}\left(\alpha_k(b)e^{2\pi iN\phi_k(b)}\hat{w}\left(|p|^{-s}(p-N\nabla\phi(b))\right)+O(\varepsilon)\right)$$

and

$$\nabla_b W_f(p,b) = 2\pi iN|p|^{-s}\left(\nabla\phi_k(b)\alpha_k(b)e^{2\pi iN\phi_k(b)}\hat{w}(|p|^{-s}(p-N\nabla\phi_k(b)))+O(\varepsilon)\right),$$

as the other terms drop out since $\{Z_{f,k}\}$ are disjoint. Hence

$$v_f(p,b) = \frac{N\nabla\phi_k(b)\left(\alpha_k(b)e^{2\pi iN\phi_k(b)}\hat{w}\left(|p|^{-s}(p-N\nabla\phi_k(b))\right)+O(\varepsilon)\right)}{\left(\alpha_k(b)e^{2\pi iN\phi_k(b)}\hat{w}\left(|p|^{-s}(p-N\nabla\phi_k(b))\right)+O(\varepsilon)\right)}.$$

Let us denote the term $\alpha_k(b)e^{2\pi iN\phi_k(b)}\hat{w}\left(|p|^{-s}(p-N\nabla\phi_k(b))\right)$ by g. Then

$$v_f(p,b) = \frac{N\nabla\phi_k(b)\left(g+O(\varepsilon)\right)}{g+O(\varepsilon)}.$$

Since $|W_f(p,b)| \geq |p|^{-s}\sqrt{\varepsilon}$ for $(p,b) \in R_{f,\varepsilon}$, $|g| \gtrsim \sqrt{\varepsilon}$, and therefore

$$\frac{|v_f(p,b)-N\nabla\phi_k(b)|}{|N\nabla\phi_k(b)|} \lesssim \left|\frac{O(\varepsilon)}{g+O(\varepsilon)}\right| \lesssim \sqrt{\varepsilon}. \qquad \blacksquare$$

The assumption $s \in (1/2,1)$ is essential to the proof. The upper bound $s<1$ enables the wave packets to detect oscillations in different directions. The lower bound $s>1/2$ ensures that the support of the wave packets is sufficiently small in space so that the second order properties of the phase function (such as the curvature of the wave front) do not affect the synchrosqueezing estimate of the local wavevectors. As shown in (2.3), (2.4), and (2.7), the constant $N_0(M,K,s,\varepsilon)$ in Theorem 2.3 goes to infinity when s approaches either 1 or $\frac{1}{2}$. Therefore, in practice, s should be chosen to be well separated from both 1/2 and 1.

In [10], the authors show that, for synchrosqueezed wavelet transform, each intrinsic mode function or component can be reconstructed from the synchrosqueezed coefficients by making

use of a reconstruction formula that integrates the continuous wavelet coefficient over the scale parameter with an appropriate weight. They also prove an error bound on the reconstructed intrinsic mode functions. In the current setting, however, we are not aware of a similar reconstruction formula for the wave packet. Therefore, our reconstruction step is based on a Calderon-type reconstruction formula for the wave packets, as illustrated in (1.5). A similar approach based on the Calderon reconstruction formula for the wavelets is in fact used in the numerical examples of [10], as it is more robust in the noisy case. However, we have not been able to derive a rigorous error bound for this Calderon-type reconstruction formula for the wave packets at this point.

Since we require N to be sufficiently large in Theorem 2.3, a function defined in Definition 2.2 is a superposition of highly oscillatory components. It is easy to extend Theorem 2.3 to the case where we have different levels of oscillation, i.e., we have different N_k large enough. In practical applications, a function might also contain a low-frequency component. For such a low-frequency component, the local wavevector is not well-defined, as it is impossible to perform a phase-amplitude decomposition as is given in Definition 2.1 for a low-frequency signal. Thus Theorem 2.3 does not apply to such a superposition. However, in practice, we observe that the synchrosqueezing step can still separate the support of different components quite well: typically the support of high-frequency components is squeezed into regions $Z_{f,k}$, while the support of the low frequency component remains at the low-frequency part of the Fourier domain. Therefore, by applying the reconstruction formula to the coefficients of the low-frequency component, one is still able to identify the low-frequency component quite accurately, even though one cannot estimate its local wavevector.

3. Implementation of the transform.

In this section, we describe in detail the discrete synchrosqueezed wave packet transform. Let us first recall the continuous setting. For a given superposition $f(x)$ of several well-separated components, the synchrosqueezed wave packet transform consists of the following steps:

(i) Apply the wave packet transform to obtain $W_f(p,b)$ and the gradient $\nabla_b W_f(p,b)$.
(ii) Compute the approximate local wavevector $v_f(p,b)$ and perform synchrosqueezing to get $T_f(v,b)$.
(iii) Use a clustering algorithm to identify the support of the new phase space representation $T_f(v,b)$ of different intrinsic mode functions.
(iv) Reconstruct each intrinsic mode function using the dual frame.

In order to realize these steps in the discrete setting, we first introduce a discrete implementation of the wave packet transform in section 3.1. The full discrete algorithm will then be discussed in section 3.2.

3.1. Discrete wave packet transform.

For simplicity, we consider functions that are periodic over the unit square $[0,1)^2$ in two dimensions. Let

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

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

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

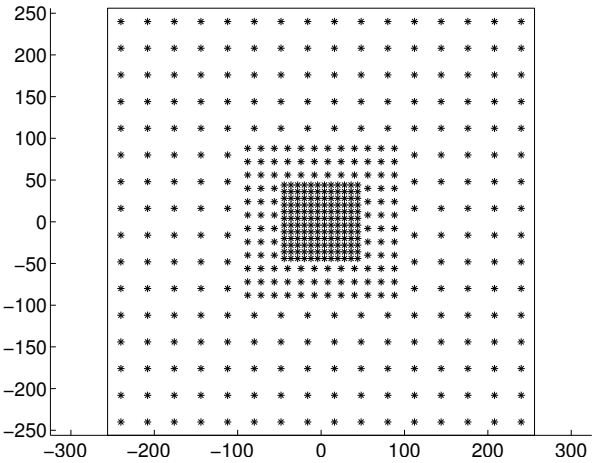

Figure 4. *The sample set P. Each star represents a point p in P. Each p is associated with a window function $g_p(\xi)$ in the frequency domain.*

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

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

while the discrete inverse Fourier transform of $g(\xi) \in \ell^2(\Xi)$ is

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

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

In order to design a discrete wave packet transform, we need to specify how to decimate the momentum space and the position space. Let us first consider the momentum space. In the continuous setting, the Fourier transform $\widehat{w_{pb}}(\xi)$ of the wave packets for a fixed p value has the profile

$$|p|^{-s} \hat{w}(|p|^{-s}(\xi - p)), \tag{3.1}$$

modulo complex modulation. In the discrete setting, we sample the Fourier domain $[-L/2, L/2]^2$ with a set of points P (as shown in Figure 4) and associate with each $p \in P$ a window function $g_p(\xi)$ that behaves qualitatively as $\hat{w}(|p|^{-s}(\xi - p))$. More precisely, $g_p(\xi)$ is required to satisfy the following conditions:

- $g_p(\xi)$ is nonnegative and centered at p with a compact support of width $L_p = O(|p|^s)$;
- $g_p(|p|^s \tau + p)$ is a sufficiently smooth function of τ, so that the discrete wave packets decay rapidly in the spatial domain;
- $C_1 \le \int |g_p(|p|^s \tau + p)|^2 d\tau \le C_2$ for constants $C_1, C_2 > 0$ which are independent of p;
- in addition, for any $\xi \in [-L/2, L/2)^2$, $\sum_{p \in P} |g_p(\xi)|^2 = 1$.

One possible way to specify the set P and the functions $\{g_p(\xi), p \in P\}$ is to follow the constructions of the the wave atom frame in [12] or the Gaussian wave packets of [25]. In both constructions, the parabolic scaling $s = 1/2$ is used in order to represent the oscillatory patterns efficiently. However, in the current setting, the proposed wave packet transform requires $s \in (1/2, 1)$, and hence one needs to increase the support of $g_p(\xi)$ accordingly. We refer to [12, 25] for more detailed discussions. The above conditions for $g_p(\xi)$, $p \in P$, also impose a constraint on the sampling density of the set P. In the frequency plane, the set P becomes dense near the origin and sparser for large ξ. A straightforward calculation shows that the total number of samples in P is of order $O(L^{2-2s})$.

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

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

As we shall see, the only requirement is that $L_B \ge \max_{p\in P} L_p$ so that the discrete wave packets can form a frame.

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

$$\widehat{w_{pb}}(\xi) = \frac{1}{L_p} e^{-2\pi i b\cdot\xi} g_p(\xi)$$

for $\xi \in \Xi$. Since $g_p(\xi)$ is centered at p and has a support of width $L_p = O(|p|^s)$, this function fits into the scaling of wave packets. Applying the discrete inverse Fourier transform provides its spatial description,

$$w_{pb}(x) = \frac{1}{L \cdot L_p} \sum_{\xi\in\Xi} e^{2\pi i (x-b)\cdot\xi} g_p(\xi).$$

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

$$W_f(p, b) = \langle w_{pb}, f\rangle = \langle \widehat{w_{pb}}, \hat{f}\rangle = \int \overline{\widehat{w_{pb}(\xi)}} \hat{f}(\xi) d\xi = \frac{1}{L_p} \sum_{\xi\in\Xi} e^{2\pi i b\cdot\xi} g_p(\xi) \hat{f}(\xi). \tag{3.2}$$

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

$$\langle g, h\rangle = \sum_{p\in P, b\in B} \overline{g(p, b)} h(p, b)(L_p/L_B)^2.$$

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

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

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

Proof. From the definition of the wave packet transform, we have

$$\begin{aligned}\sum_{p\in P,b\in B}|W_f(p,b)|^2(L_p/L_B)^2 &= \sum_{p\in P,b\in B}\left|\sum_{\xi\in\Xi}\frac{1}{L_p}e^{2\pi ib\cdot\xi}g_p(\xi)\hat{f}(\xi)\right|^2\left(\frac{L_p}{L_B}\right)^2\\ &= \sum_{p\in P,b\in B}\left|\sum_{\xi\in\Xi}\frac{1}{L_B}e^{2\pi ib\cdot\xi}g_p(\xi)\hat{f}(\xi)\right|^2\\ &= \sum_{p\in P}\sum_{\xi\in\Xi}\left|g_p(\xi)\hat{f}(\xi)\right|^2\\ &= \sum_{\xi\in\Xi}|\hat{f}(\xi)|^2. \qquad \blacksquare\end{aligned}$$

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

$$W_h^t(x) := \sum_{p\in P,b\in B}h(p,b)w_{pb}(x)(L_p/L_B)^2. \tag{3.3}$$

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

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

$$f(x) = \sum_{p\in P,b\in B}W_f(p,b)w_{pb}(x)(L_p/L_B)^2.$$

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

$$\begin{aligned}&\sum_{p\in P,b\in B}\left(\sum_{\eta\in\Xi}\frac{1}{L_p}e^{2\pi ib\cdot\eta}g_p(\eta)\hat{f}(\eta)\right)\cdot\frac{1}{L_p}e^{-2\pi ib\cdot\xi}g_p(\xi)\left(\frac{L_p}{L_B}\right)^2\\ =&\sum_{p\in P}\left(\sum_{\eta\in\Xi}\frac{1}{L_B^2}\left(\sum_{b\in B}e^{2\pi ib\cdot(\eta-\xi)}g_p(\eta)\hat{f}(\eta)\right)\right)g_p(\xi)\\ =&\sum_{p\in P}(g_p(\xi))^2\hat{f}(\xi) = \hat{f}(\xi),\end{aligned}$$

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

Let us now turn to the discrete approximation of $\nabla_b W_f(p,b)$. From the continuous definition (1.1), we have

$$\nabla_b W_f(p,b) = \nabla_b\langle\widehat{w_{pb}},\hat{f}\rangle = \langle-2\pi i\xi\widehat{w_{pb}}(\xi),\hat{f}(\xi)\rangle.$$

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

$$\nabla_b W_f(p,b) = \sum_{\xi\in\Xi}\frac{1}{L_p}2\pi i\xi e^{2\pi ib\cdot\xi}g_p(\xi)\hat{f}(\xi). \tag{3.4}$$

The above definitions give rise to fast algorithms for computing the forward wave packet transform, its transpose, and the discrete gradient operator. All three algorithms heavily rely on the fast Fourier transform (FFT). For the forward transform, writing (3.2) as

$$W_f(p,b) = \frac{L_B}{L_p} \cdot \left(\frac{1}{L_B} \sum_{\xi \in \Xi} e^{2\pi i b \cdot \xi} g_p(\xi) \hat{f}(\xi) \right)$$

suggests the following algorithm.

Algorithm 3.3. *Forward transform from $f(x)$ to $W_f(p,b)$:*

Compute $\hat{f}(\xi)$ with $\xi \in \Xi$ from $f(x)$ with $x \in X$ using an $L \times L$ forward FFT.
for** each $p \in P$ **do
  Form $g_p(\xi)\hat{f}(\xi)$ on the support of $g_p(\xi)$.
  Wrap the result modulo L_B onto the domain $[-L_B/2, L_B/2)^2$.
  Apply an $L_B \times L_B$ inverse FFT to the wrapped result.
  Multiply the result by L_B/L_p to get $W_f(p,b)$ for all $b \in B$.
end for

The transpose operator (3.3) can be written equivalently in the Fourier domain as

$$\hat{W}_h^t(\xi) = \sum_{p \in P, b \in B} h(p,b) \frac{1}{L_p} e^{-2\pi i b \cdot \xi} g_p(\xi) \left(\frac{L_p}{L_B} \right)^2 = \sum_{p \in P} \left(\sum_{b \in B} \frac{1}{L_B} h(p,b) \frac{L_p}{L_B} e^{-2\pi i b \cdot \xi} \right) g_p(\xi),$$

which suggests the following algorithm for the transpose operator.

Algorithm 3.4. *Transpose operator from $h(p,b)$ to $W_h^t(x)$:*

for** each $p \in P$ **do
  Multiply $h(p,b)$ for each $b \in B$ by L_p/L_B.
  Apply an $L_B \times L_B$ forward FFT to the product.
  Unwrap the result modulo L_B onto the support of $g_p(\xi)$.
  Multiply the unwrapped data with $g_p(\xi)$ and add the product to get $\hat{f}(\xi)$.
end for
Compute $f(x)$ with $x \in X$ from $\hat{f}(\xi)$ with $\xi \in \Xi$ using an $L \times L$ inverse FFT.

To implement the discrete gradient operator, we rewrite (3.4) as

$$\nabla_b W_f(p,b) = \frac{L_B}{L_p} \left(\sum_{\xi \in \Xi} \frac{1}{L_B} e^{2\pi i b \cdot \xi} 2\pi i \xi g_p(\xi) \hat{f}(\xi) \right).$$

This suggests the following algorithm.

Algorithm 3.5. *Discrete gradient operator from $f(x)$ to $\operatorname{grad}_b W_f(p,b)$:*

Compute $\hat{f}(\xi)$ with $\xi \in \Xi$ from $f(x)$ with $x \in X$ using an $L \times L$ forward FFT.
for** each $p \in P$ **do
  Form $2\pi i \xi g_p(\xi)\hat{f}(\xi)$ on the support of $g_p(\xi)$.
  Wrap the result modulo L_B onto the domain $[-L_B/2, L_B/2)^2$.
  Apply an $L_B \times L_B$ inverse FFT to each component of the wrapped result.

6: *Multiply the result by* L_B/L_p *to get* $\nabla_b W_f(p,b)$ *for all* $b \in B$.
7: ***end for***

As we mentioned earlier, the conditions on $\{g_p(\xi), p \in P\}$ imply that there are $O(L^{2(1-s)})$ samples in set P. A straightforward calculation shows that the computational cost of all three algorithms is $O(L^2 \log L + L^{2(1-s)} L_B^2 \log L_B)$ with $L_B \geq \max_{p\in P} L_p = O(L^s)$. If we choose L_B to be of the same order as L^s, the complexity of these algorithms is $O(L^2 \log L)$, which is the cost of an FFT on an $L \times L$ Cartesian grid.

3.2. Description of the full algorithm. With the discrete transforms and their fast algorithms available, we now go through the steps of the synchrosqueezed wave packet transform.

For a given function $f(x)$ defined on $x \in X$, we apply Algorithm 3.3 to compute $W_f(p,b)$ and Algorithm 3.5 to compute $\nabla_b W_f(p,b)$. The approximate local wavevector $v_f(p,b)$ is then estimated by

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

for $p \in P, b \in B$ with $W_f(p,b) \neq 0$. In view of Theorem 2.3, a threshold $|W_f(p,b)| \geq |p|^{-s}\sqrt{\varepsilon}$ $(p \geq 1)$ is necessary. Following Theorem 2.3, we define a discrete set $R_{f,\varepsilon}$ with

$$R_{f,\varepsilon} = \{(p,b) : p \in P, b \in B, |W_f(p,b)| \geq |p|^{-s}\sqrt{\varepsilon}\},$$

and $v_f(p,b)$ provides an approximate estimate for the local wavevector only for $(p,b) \in R_{f,\varepsilon}$.

To specify the synchrosqueezed energy distribution $T_f(v,b)$, we first place in the Fourier domain a 2D Cartesian grid of stepsize Δ:

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

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 the discrete synchrosqueezed energy distribution is defined as

$$T_f(v,b) = \sum_{(p,b)\in R_{f,\varepsilon}: \Re v_f(p,b)\in D_v} |W_f(p,b)|^2 \cdot (L_p/L_B)^2.$$

It is straightforward to check that

$$\sum_{v\in V, b\in B} T_f(v,b) = \sum_{(p,b)\in R_{f,\varepsilon}} |W_f(p,b)|^2 (L_p/L_B)^2 \leq \|f\|_2^2,$$

where the last inequality comes from Proposition 3.1 and the fact that $R_{f,\varepsilon}$ is a subset of $P \times B$.

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

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

From the previous discussion we know that, for each $b \in B$, $v_f(p,b)$ points approximately to one of $N\nabla\phi_k(b)$, depending on p. Therefore, after synchrosqueezing, $T_f(v,b)$ is essentially supported in the phase space near the K "discrete" surfaces $\{(N\phi_k(b),b), b \in B\}$. The next step is to decompose the essential support of $T_f(v,b)$ into K clusters, one for each intrinsic mode function, through a spectral clustering method. We make use of the algorithm proposed in [22]. For a fixed set S of n points $\{s_1,\dots,s_n\}$ in $\mathbb{R}^l$ and an integer K, this method partitions these points into K clusters as follows.

Algorithm 3.6. *General spectral clustering on set* $S = s_1,\dots,s_n$*:*

1: *Construct the matrix* $A = (\alpha_{ij})_{ij} \in R^{n\times n}$ *with distance function* $\alpha_{ij} = \exp(-|s_i - s_j|^2/\sigma^2)$ *if* $i \neq j$*, and* $\alpha_{ii} = 0\ \forall i$*. Here* σ *is an input parameter.*

2: *Let* D *be a diagonal matrix such that* $D_{ii} = \sum_{j=1}^{n} \alpha_{ij}$*, and define the Laplacian-type matrix* $L = D^{-\frac{1}{2}} A D^{-\frac{1}{2}}$.

3: *Choose the* K *largest orthogonal eigenvectors of* L*, say* $v_1,\dots,v_K$*, and stack them horizontally to get the matrix* $V = [v_1, v_2, \dots, v_K] \in R^{n\times K}$*. The entries of* V *are denoted by* v_{ij}.

4: *Define the matrix* $M = (m_{ij})$ *with* $m_{ij} = v_{ij}/(\sum_j v_{ij}^2)^{1/2}$*, which means normalizing the rows of* V.

5: *Consider each row of* M *as a point in* $\mathbb{R}^K$*, and then partition these* n *points into* K *clusters with the* K*-means algorithm.*

6: *If row* i *of* M *is assigned to cluster* j*, then assign the original point* s_i *to cluster* j.

In the current setting, we choose a threshold parameter $\eta > 0$, define the set S to be

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

and apply the above algorithm to S. The resulting clusters are defined to be $U_1,\dots,U_K$. In many cases, the number of components K is not known a priori and needs to be discovered from the function $T_f(v,b)$. To do that, we use the set of nonnegligible entries for each b, i.e., $\{v \in V, T_f(v,b) \geq \eta\}$, to decide the local number of clusters and take the maximum number of all $b \in B$ to be the cluster number K.

In the final step, we recover each intrinsic mode function by computing

$$f_k(x) = \sum_{(p,b):\Re v_f(p,b)\in U_k} W_f(p,b) w_{pb}(x) (L_p/L_B)^2.$$

This step can be carried out efficiently by restricting $W_f(p,b)$ to the set $\{(p,b) : \Re v_f(p,b) \in U_k\}$ and applying Algorithm 3.4 to the restriction for each k.

4. Numerical results.

This section presents several numerical examples to illustrate the proposed synchrosqueezed wave packet transforms. Throughout all examples, the threshold value ε is 10^{-4}, and the size L of the Cartesian grid X of the discrete algorithm is 512. In the implementation of the discrete wave packet transforms, the scaling parameter s is equal to 2/3, which is a good balance as discussed previously. All these examples are implemented using a PC with Intel Xeon(R), CPU E5-1620, 3.60 GHz.

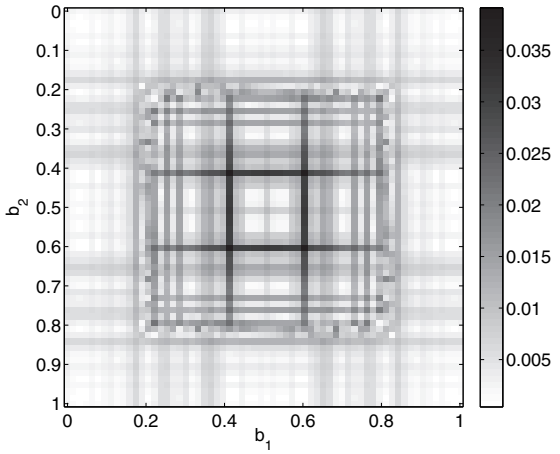

Figure 5. *Example* 1. *Relative error* $R(b)$ *of local wavevector estimation.*

4.1. Local wavevector extraction. We first test the accuracy of the estimated local wavevector $v_f(p,b)$. Let $f(x)$ be a deformed plane wave:

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

Theorem 2.3 shows that, for each fixed point b in space, the estimate $v_f(p,b)$ approximates the local wavevector at b for any p that satisfies the condition $|W_f(p,b)| \geq |p|^{-s}\sqrt{\varepsilon}$. In the discrete setting, since $L_p = O(|p|^s) \geq 1$, the corresponding threshold criterion becomes $|W_f(p,b)L_p| \geq \sqrt{\varepsilon}$. Though $v_f(p,b)$ for any such p provides an estimate of the local wavevector at b, it is more useful to combine them together to obtain a unique local wavevector estimate for each fixed b. More precisely, we define the *mean* local wavevector estimate at b to be

$$v_f^m(b) = \frac{\sum_p |W_f(p,b)|^2 v_f(p,b)}{\sum_p |W_f(p,b)|^2},$$

where the sum in p is taken over all p that satisfy $|W_f(p,b)L_p| \geq \sqrt{\varepsilon}$. Using this estimate, we can define the (discrete) relative error $R(b)$ between $v_f^m(b)$ and the exact local frequency $N\nabla\phi(b)$ as

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

Example 1. We perform the above test on a deformed plane wave $f(x)$ with $\alpha(x) = 1$, $\phi(x) = \phi(x_1,x_2) = x_1 + x_2 + \beta\sin(2\pi x_1) + \beta\sin(2\pi x_2)$ with $\beta = 0.1$, and $N = 135$. The relative error $R(b)$ shown in Figure 5 is of order 10^{-2}, which agrees with Theorem 2.3 in that the relative approximation error is $O(\sqrt{\varepsilon})$.

4.2. Intrinsic mode decomposition.

Example 2. Here $f(x)$ is a sum of two deformed plane waves:

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

with $N = 135$ and $\beta = 0.1$. The algorithm described in section 3.2 is applied to $f(x)$ to extract these two components. Figure 6 summarizes the results of this test. The first row shows the superposition $f(x)$ (left) and the synchrosqueezed energy distribution $T_f(v, b)$ with b_1 fixed at 1 (right). For a fixed b_1 value, $T_f(v, b)$ concentrates near two curves. More generally, in phase space, $T_f(v, b)$ concentrates near two 2D surfaces. The second row shows the two sets U_1 and U_2 after the clustering steps. Finally, the third row plots the two reconstructed components.

The proposed synchrosqueezed wave packet transform is also robust to noise. To demonstrate this, let $f(x)$ be the superposition of two deformed plane waves and a noise term,

$$f(x) = e^{2\pi i N\phi_1(x)} + e^{2\pi i N\phi_2(x)} + n(x), \tag{4.1}$$

where $n(x)$ is an isotropic complex Gaussian random noise with zero mean and variance σ^2. In order to reduce the influence of noise, we set up a threshold parameter $\delta \approx C\sigma^2$, where C is an application-dependent constant, and keep only the values of $T_f(v, b)$ that are greater than δ to get the essential support of synchrosqueezed energy distribution. We use the signal-to-noise ratio (SNR) and peak signal-to-noise ratio (PSNR) to quantitatively demonstrate the robustness against noise. Here the SNR is defined by

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

Let $\tilde{f}_i$ stand for the ith reconstructed mode. The corresponding PSNR_i is defined by

$$\mathrm{PSNR}_i[dB] = 10 \log_{10}\left(\frac{\mathrm{MAX}_{f_i}^2}{\mathrm{MSE}_i}\right)$$

where

$$\mathrm{MAX}_{f_i} = \max |f_i(x)|$$

and

$$\mathrm{MSE}_i = \frac{||f_i - \tilde{f}||_2^2}{L^2}.$$

In this test, we set SNR to be 3, 0, and -3. Table 1 summarizes the PSNR values for the two reconstructed modes in each case. Figure 7 demonstrates the result of the synchrosqueezed wave packet transform in more detail. The first column of images shows the essential support of $T_f(v, b)$ for $b_1 = 1$ in each case. We observe that the supports of $T_f(v, b)$ are almost identical to those of the noiseless case, except for a little energy loss caused by noise. This demonstrates that the proposed method is robust to noise in estimating local wavevectors. The second column of images shows the second reconstructed mode of (4.1) in each case. These images are almost very close to the original one, with only some energy loss caused by noise. This shows that the proposed method is robust to noise in reconstructing individual mode functions.

In the analysis, we have assumed so far that each component of the superposition covers the whole domain. In many real applications, one or more components of the superposition might be *incomplete* and cover only part of the domain. The analysis of this general case is

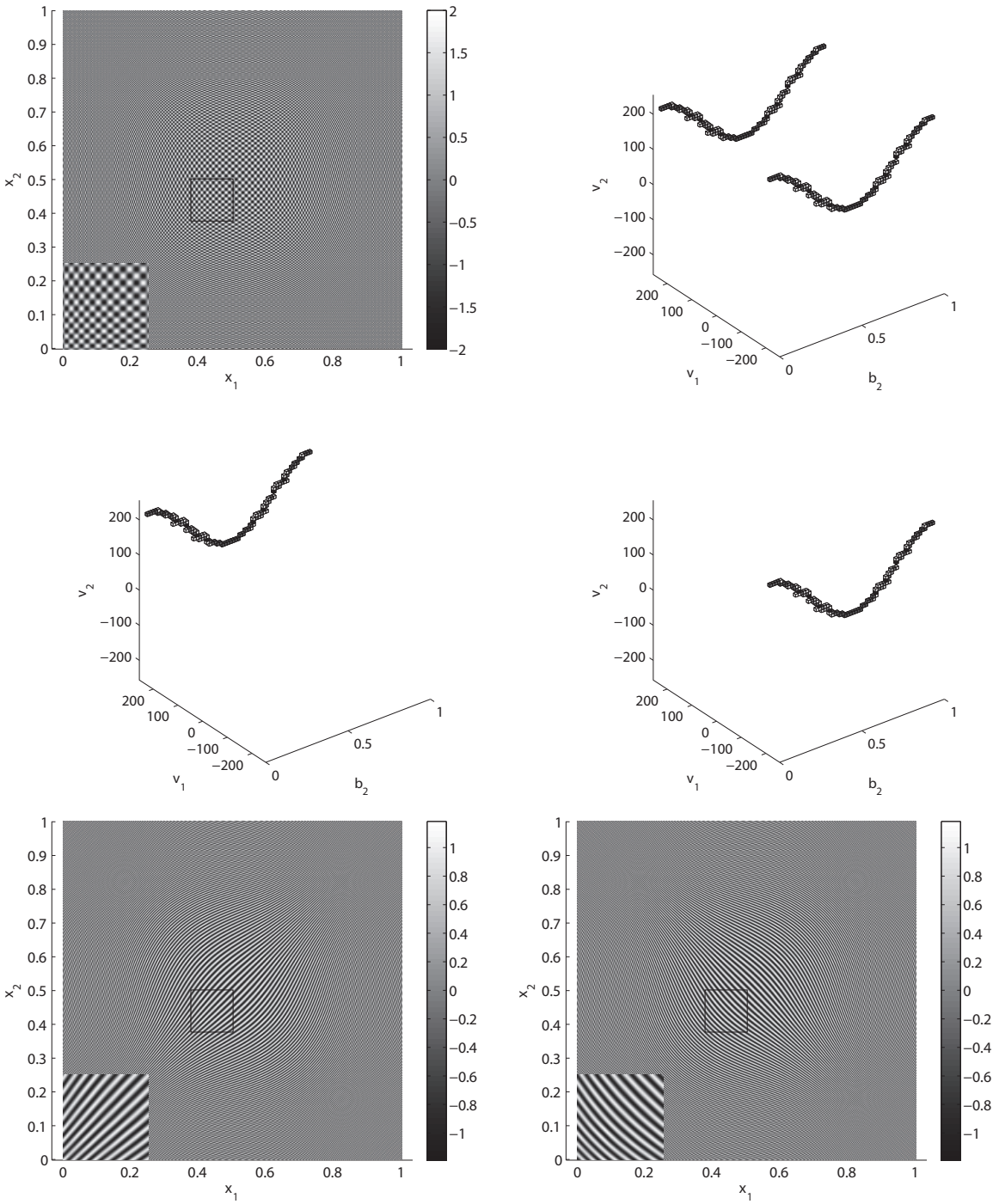

Figure 6. *Example 2. Mode decomposition without noise. Top-left: A superposition of two deformed plane waves with the inset showing a zoomed-in view of the highlighted rectangle. Top-right: Synchrosqueezed energy distribution $T_f(v, b)$ at $b_1 = 1$. Second row: The support of $T_f(v, b)$ is clustered into two subsets. Third row: The two reconstructed components. The run time of mode decomposition by synchrosqueezed wave packet transform is 263 seconds.*

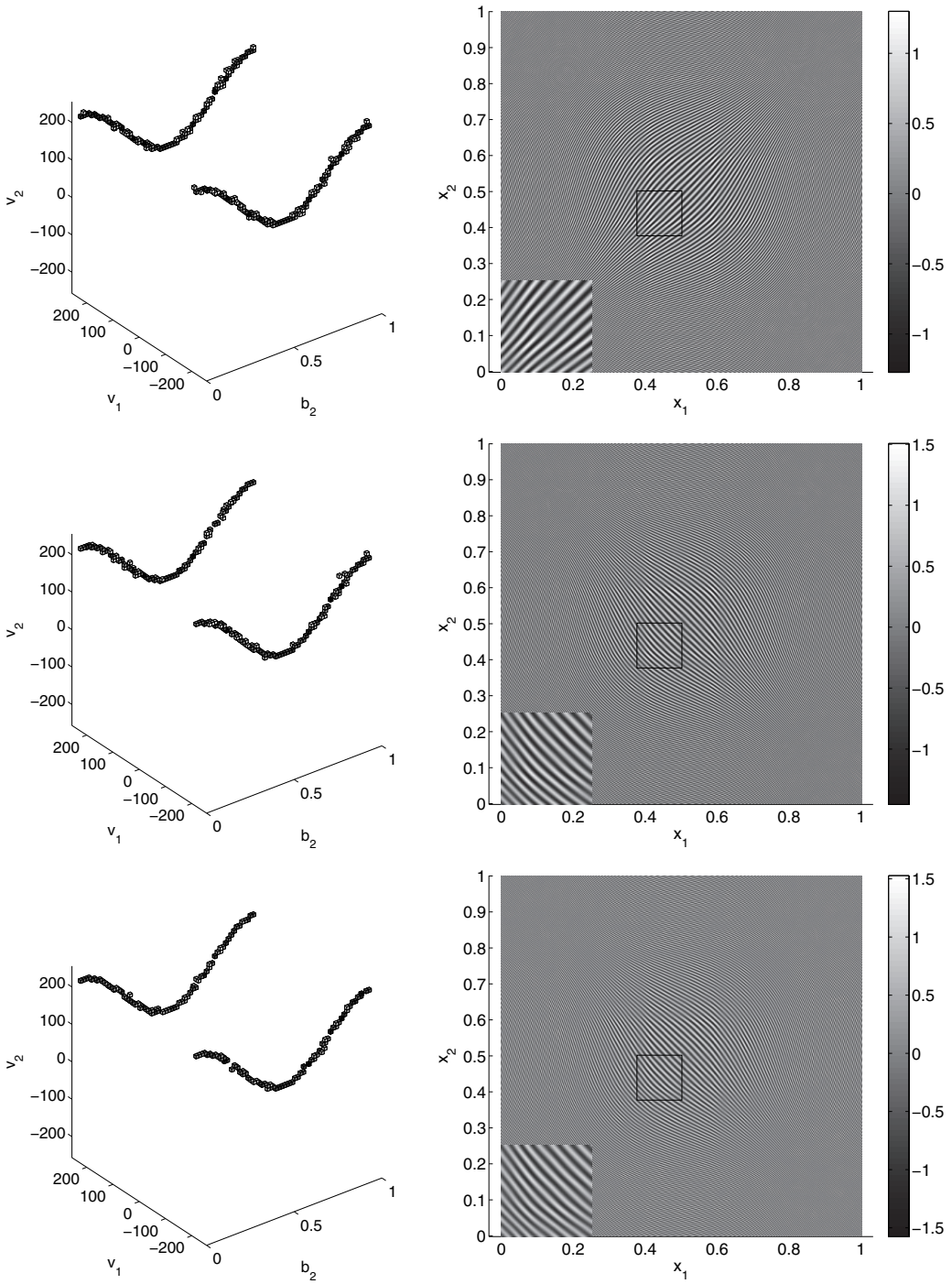

Figure 7. *Mode decomposition with noise. Top row: SNR = 3. Middle row: SNR = 0. Bottom row: SNR = −3. First column: The essential support of $T_f(v, b)$ for $b_1 = 1$. Second column: The second reconstructed mode with the insets showing a zoomed-in view of the highlighted rectangles. The run time of the example with SNR = 0 is 265 seconds, which is almost the same as in noiseless case.*

Table 1
SNR and PSNR for each mode.

SNR	∞	3	0	-3
PSNR_1	54.19	24.86	22.87	18.78
PSNR_2	54.13	24.79	23.02	18.67

more complicated due to the boundary of these incomplete components. On the other hand, one expects that the synchrosqueezed energy distribution $T_f(v,b)$ should still be supported near several 2D surfaces in the phase space, though some of them might be incomplete. In this example, we show that the synchrosqueezed wave packet transform still works quite well under this more general setting. Here we choose $f(x)$ to be the superposition of two components, one of which is incomplete:

$$\begin{aligned} f(x) &= \chi(x)\cdot e^{2\pi i N\phi_1(x)} + e^{2\pi i N\phi_2(x)},\\ \phi_1(x) &= \phi_1(x_1,x_2) = -(x_1+\beta\sin(2\pi x_1)) + (x_2+\beta\sin(2\pi x_2)),\\ \phi_2(x) &= \phi_2(x_1,x_2) = (x_1+\beta\sin(2\pi x_1)) - (x_2+\beta\sin(2\pi x_2)), \end{aligned}$$

where $N = 135$, $\beta = 0.1$, and $\chi(x)$ is an indicator function of an ellipse in $[0,1)^2$. Figure 8 summarizes the results of this example. The first row shows the superposition $f(x)$ without noise. The second row plots the two reconstructed components. We note that the boundary of the second incomplete component is accurately captured by the proposed method.

Example 3. In the analysis, we have assumed that different intrinsic mode functions have well-separated local wavevectors at each point. Real data, however, may have multiple components that exhibit nearby or even identical local wavevectors at certain points. Though Theorem 2.3 does not apply to this setting any more, this example shows that often the synchrosqueezed wave packet can still identify the local wavevectors robustly. Let $f(x)$ be a superposition of two deformed plane waves given by

$$\begin{aligned} f(x) &= e^{2\pi i N\phi_1(x)} + e^{2\pi i N\phi_2(x)},\\ \phi_1(x) &= \phi_1(x_1,x_2) = x_1+x_2+\beta\sin(2\pi x_1)+\beta\sin(2\pi x_2),\\ \phi_2(x) &= \phi_2(x_1,x_2) = -x_1+x_2-\beta\cos(2\pi x_1)+\beta\cos(2\pi x_2), \end{aligned}$$

with $N = 135$ and $\beta = 0.1$. In this example, the wavevectors of two modes are the same when $x_1 = x_2 = \frac{3\pi}{8}$. We apply the synchrosqueezed wave packet transform to noiseless and noisy data and summarize the results in Figure 9. We observe that the synchrosqueezed wave packet transform continues to estimate the local wavevector accurately without coarsening the supports.

Example 4. We provide an example coming from a solution of the Helmholtz equation in a layered medium to show the performance of the synchrosqueezed wave packet transform on seismic wavefield decomposition. This example comes from a seismic inversion simulation provided by Sergey B. Fomel and Siwei Li at the Bureau of Economic Geology in the University of Texas at Austin.

In this example, the seismic source is near the top-left corner of the image, and the wave propagates downward. When the wave reaches the interface of the layered media, part of the

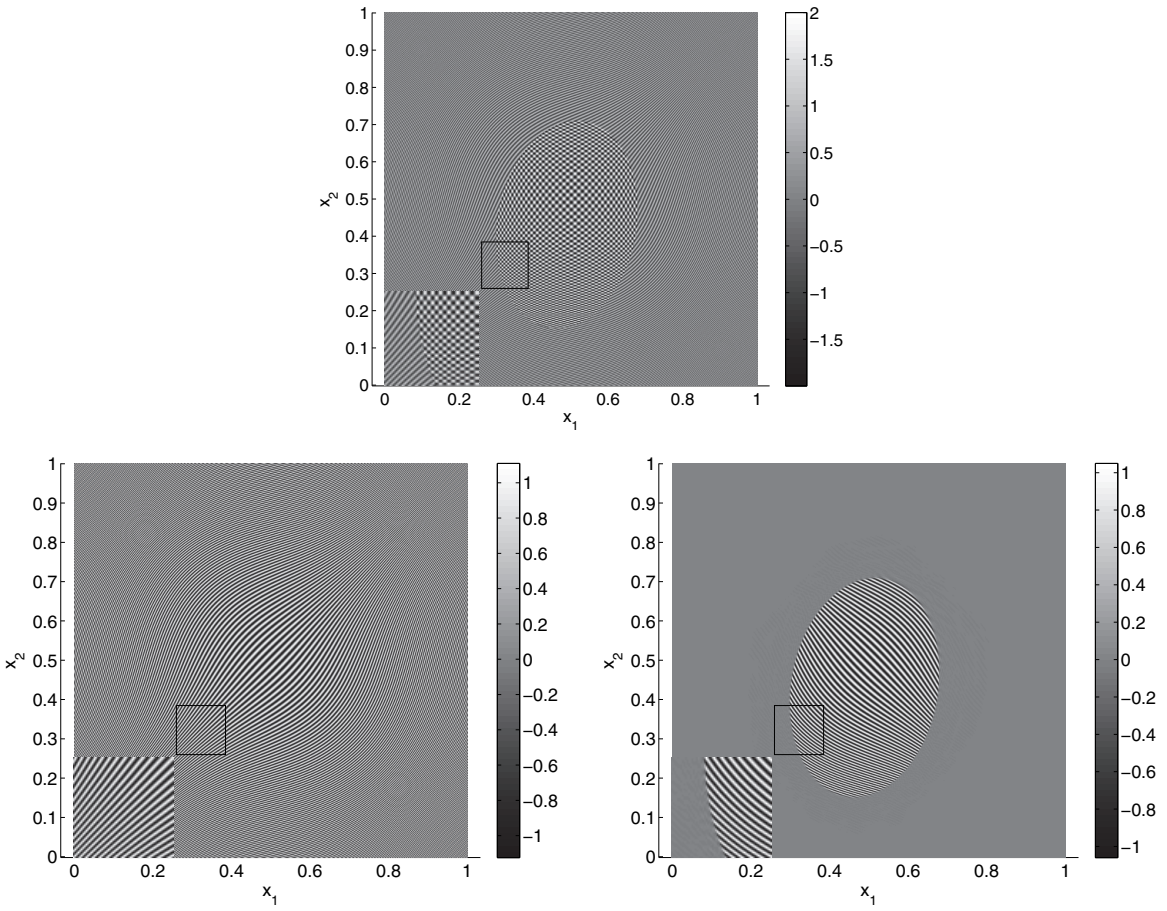

Figure 8. *Example 3. Mode decomposition with incomplete component. Top: A superposition of a complete deformed plane wave and an incomplete one. Second row: the two reconstructed components. The sharp cutoff boundary of the second incomplete component is clearly shown. In each plot, the inset is a zoomed-in view of the highlighted rectangle.*

wave continues propagating downward while the rest gets reflected (Figure 10(top)). Figure 10(bottom) shows the two components obtained from applying the synchrosqueezed wave packet transform to this image. The first component contains the downward propagating mode both above and below the interface of the layered media, while the second component contains the upward reflected mode only above the interface.

5. Discussion. The method proposed here is an initial step in mode decomposition for higher-dimensional signals. This is the first method to date that is designed to distinguish two modes with the same frequency but different directions. Several possible directions for future research are listed below.

The synchrosqueezed wave packet transform has a geometric scaling parameter s, which is between 1/2 and 1. Theorem 2.3 shows that the constant $N_0(M, K, s, \varepsilon)$ goes to infinity as s approaches either 1/2 or 1. Therefore, in practice, s should be well separated from both

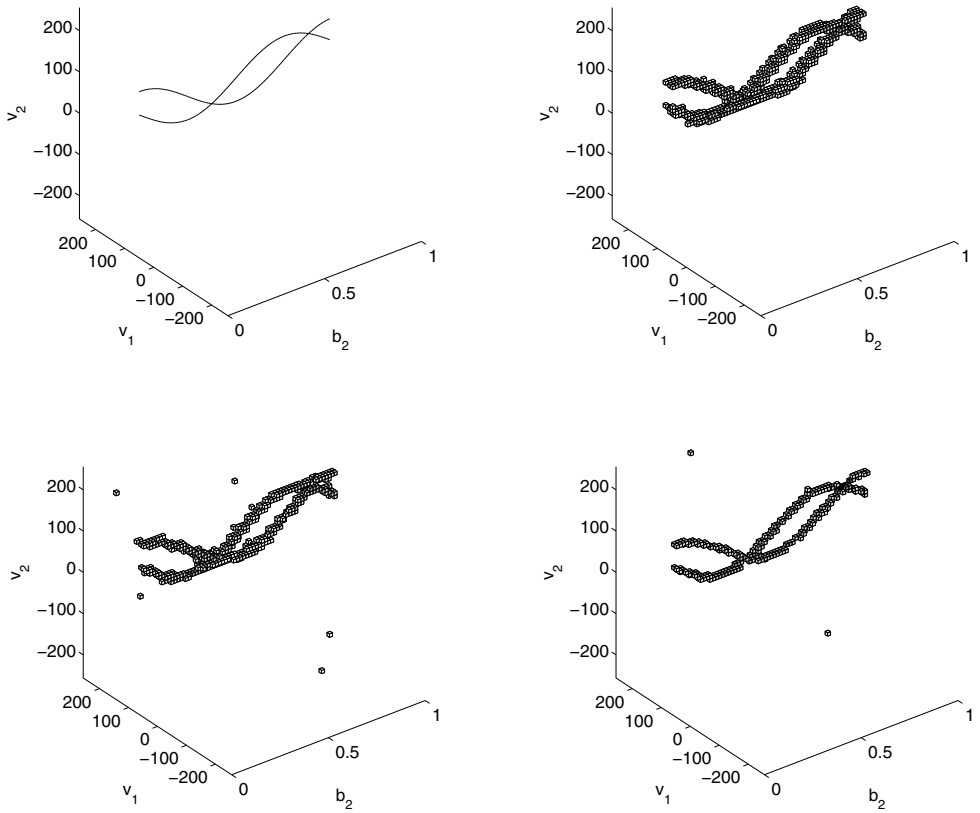

Figure 9. *Example 3. Synchrosqueezed wave packet transform on noiseless or noisy data with crossover wavevectors. Top-left: The exact local wavevectors at* $b_1 = \frac{3\pi}{8}$. *Top-right: Essential support of* $T_f(p, b)$ *for the noiseless data at* $b_1 = \frac{3\pi}{8}$. *Bottom: The essential support of* $T_f(p, b)$ *for the noisy data with SNR* $= 3$ *(left) and* -3 *(right) at* $b_1 = \frac{3\pi}{8}$.

values. A natural question is what the optimal choice for s is, i.e., what value of s minimizes the value $N_0(M, K, s, \varepsilon)$. Clearly, the answer depends on the data through the parameters M and K. The analysis used in the proof of Theorem 2.3 is far from providing an answer to this question. In practice, it is probably more relevant to develop an adaptive algorithm that adjusts the parameter s for a given image.

As we mentioned earlier, the case $s = 1/2$ is the wave atom construction proposed in [12]. Wave atoms provide better angular resolution as the support of wave atoms in the Fourier domain is more refined. However, as we pointed out, the synchrosqueezing step is no longer valid for local wavevector estimation, as the wave atoms are large enough in space to see the second order effects of the phase function. One natural question is whether it is possible to generalize or modify the synchrosqueezing idea so that it will work for the wave atom case.

Another closely related set of analyzing functions is the curvelet frame [2, 4]. Curvelets have been shown to be the optimal tool for representing images that are smooth except for

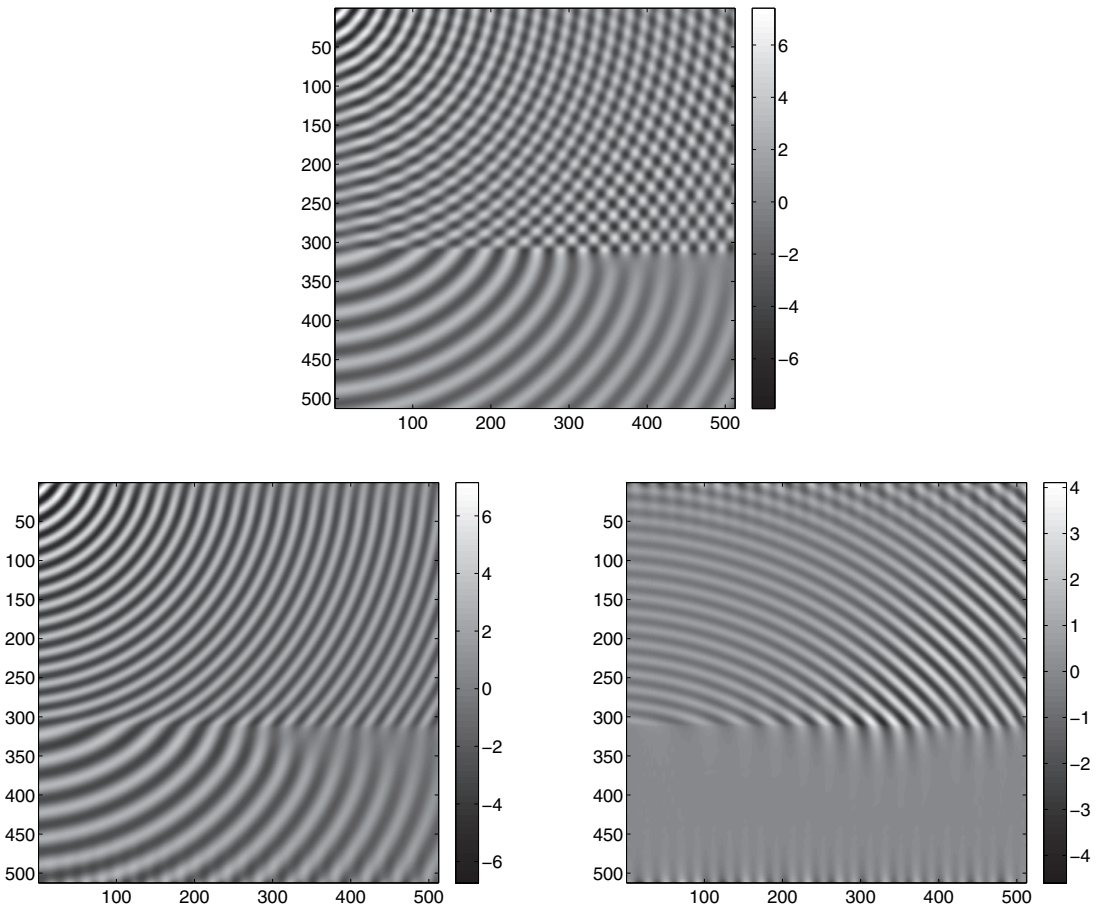

Figure 10. *Top: Synthetic seismic wavefield propagating through a layered media equation. Bottom: The two extracted components naturally correspond to the downward propagating wave (left) and the upward reflected wave (right).*

at isolated curvilinear discontinuities [3, 4, 5, 6]. The current approach can be extended to curvelets as long as the parabolic scaling case $s = 1/2$ can be addressed. A potential advantage of a synchrosqueezed curvelet transform is that it should be able to optimally identify isolated wavefronts, as opposed to the extended wave fields used in the current paper.

So far we have assumed that the local wavevectors of the different intrinsic mode functions are well separated at each point. Clearly this assumption may not hold in practice, as many images might have multiple nearby wavevectors at isolated points or curves. It would be useful to find more robust clustering algorithms to address such situations.

The current approach can be easily extended to three or higher numbers of dimensions. This direction should be relevant for applications, such as seismic imaging.

Acknowledgments. The authors thank Jianfeng Lu and Hau-Tieng Wu for discussion, Sergey Fomel and Siwei Li for providing the seismic example, and Jack Poulson for comments on the manuscript.

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] E. Candès, L. Demanet, D. Donoho, and L. Ying, *Fast discrete curvelet transforms*, Multiscale Model. Simul., 5 (2006), pp. 861–899.

[3] E. J. Candès and L. Demanet, *The curvelet representation of wave propagators is optimally sparse*, Comm. Pure Appl. Math., 58 (2005), pp. 1472–1528.

[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.* I. *Resolution of the wavefront set*, Appl. Comput. Harmon. Anal., 19 (2005), pp. 162–197.

[6] E. J. Candès and D. L. Donoho, *Continuous curvelet transform.* II. *Discretization and frames*, Appl. Comput. Harmon. Anal., 19 (2005), pp. 198–222.

[7] E. Chassande-Mottin, F. Auger, and P. Flandrin, *Time-frequency/time-scale reassignment*, in Wavelets and Signal Processing, Appl. Numer. Harmon. Anal., Birkhäuser Boston, Cambridge, MA, 2003, pp. 233–267.

[8] E. Chassande-Mottin, I. Daubechies, F. Auger, and P. Flandrin, *Differential reassignment*, IEEE Signal Process. Lett., 4 (1997), pp. 293–294.

[9] M. Clausel, T. Oberlin, and V. Perrier, *The Monogenic Synchrosqueezed Wavelet Transform: A Tool for the Decomposition/Demodulation of AM-FM Images*, arXiv:1211.5082, 2012.

[10] 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.

[11] 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.

[12] L. Demanet and L. Ying, *Wave atoms and sparsity of oscillatory patterns*, Appl. Comput. Harmon. Anal., 23 (2007), pp. 368–387.

[13] T. Y. Hou and Z. Shi, *Adaptive data analysis via sparse time-frequency representation*, Adv. Adapt. Data Anal., 3 (2011), pp. 1–28.

[14] T. Y. Hou and Z. Shi, *Data-driven time-frequency analysis*, Appl. Comput. Harmon. Anal., 35 (2013), pp. 284–308.

[15] 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.

[16] N. E. Huang, *Computer Implemented Empirical Mode Decomposition Apparatus, Method and Article of Manufacture for Two-Dimensional Signals*, US Patent 6,311,130 B1, granted Oct. 30, 2001.

[17] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu, *The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis*, R. Soc. Lond. Proc. Ser. A Math. Phys. Eng. Sci., 454 (1998), pp. 903–995.

[18] 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.

[19] A. Linderhed, *Variable sampling of the empirical mode decomposition of two-dimensional signals*, Int. J. Wavelets Multiresolut. Inf. Process, 3 (2005), pp. 435–452.

[20] A. Linderhed, *Image empirical mode decomposition: A new tool for image processing*, Adv. Adapt. Data Anal, 1 (2009), pp. 265–294.

[21] S. Mallat, *A Wavelet Tour of Signal Processing*, 3rd ed., Elsevier/Academic Press, Amsterdam, 2009.

[22] A. Y. Ng, M. I. Jordan, and Y. Weiss, *On spectral clustering: Analysis and an algorithm*, in Advances in Neural Information Processing Systems, Neural Inform. Process. Syst. 14, MIT Press, Cambridge, MA, 2001, pp. 849–856.

[23] J. C. Nunes, Y. Bouaoune, E. Delechelle, O. Niang, and P. Bunel, *Image analysis by bidimensional empirical mode decomposition*, Image Vision Comput., 21 (2003), pp. 1019–1026.

[24] J. C. Nunes, O. Niang, Y. Bouaoune, E. Delechelle, and Ph. Bunel, *Bidimensional empirical mode decomposition modified for texture analysis*, in Proceedings of the 13th Scandinavian Conference on Image Analysis (SCIA'03), Springer-Verlag, Berlin, 2003, pp. 171–177.

[25] J. Qian and L. Ying, *Fast multiscale Gaussian wavepacket transforms and multiscale Gaussian beams for the wave equation*, Multiscale Model. Simul., 8 (2010), pp. 1803–1837.
[26] 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.
[27] G. Rilling and P. Flandrin, *One or two frequencies? The empirical mode decomposition answers*, IEEE Trans. Signal Process., 56 (2008), pp. 85–95.
[28] M. van der Baan, *PP/PS wavefield separation by independent component analysis*, Geophys. J. Int., 166 (2006), pp. 339–348.
[29] H.-T. Wu, P. Flandrin, and I. Daubechies, *One or two frequencies? The synchrosqueezing answers*, Adv. Adapt. Data Anal., 3 (2011), pp. 29–39.
[30] 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.
[31] 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.