Robustness Analysis of Synchrosqueezed Transforms

Haizhao Yang† and Lexing Ying‡

† Department of Mathematics, Stanford University
‡ Department of Mathematics and ICME, Stanford University

October 2014

Abstract

Identifying and extracting principle wave-like components underlying a complex physical phenomenon are of great importance in modern data science. It is difficult to estimate all the wave-like components simultaneously from their superposition in order to reduce the influence of a sifting bias, which is crucial to many scientific problems. The newly developed synchrosqueezed transform has been proved a good option for this simultaneous analysis. Although its mathematical background is clear and is well-developed in a noiseless model, there is relatively little study on its robustness under noise.

This paper is concerned with several fundamental robustness properties of synchrosqueezed transforms. We prove that it is possible to develop compactly supported synchrosqueezed transforms for oscillatory component analysis and give the conditions for accurate and robust estimation. Considering a generalized Gaussian random noise, we address the multiscale robustness problem of a wide range of existing synchrosqueezed transforms in one and two dimensions. It is shown that their multiscale robustness can be improved by tuning their corresponding multiscale geometry in the frequency domain. This dependence is clarified by quantitative probability analysis. As a supplement, new insights and numerical implementations are introduced for estimates with better accuracy and robustness. A software package together with several heavily noisy examples is provided to demonstrate these proposed properties.

Keywords. Wave-like components, multiscale data analysis, instantaneous (local) properties, synchrosqueezed transforms, noise robustness, generalized Gaussian random noise.

AMS subject classifications: 42A99 and 65T99.

1 Introduction

1.1 Problem statement

Oscillatory signals with non-linear and non-stationary wave-like patterns are ubiquitous in science and engineering, e.g., clinical data, seismic data, climate data, astronomical data, and materials science. Analyzing instantaneous properties (e.g., instantaneous frequencies, instantaneous amplitudes and instantaneous phases) or local properties (concepts for 2D signals similar to “instantaneous” in 1D) of signals has been an important topic for over two decades. In many applications, these signals can be modeled as a non-parametric superposition of
several wave-like components with slowly varying amplitudes, frequencies or wave vectors, contaminated by noise. For example, a complex signal

\[ f(x) = \sum_{k=1}^{K} \alpha_k(x)e^{2\pi i N_k \phi_k(x)} + T(x) + e(x), \]

where \( \alpha_k(x) \) is the instantaneous (local) amplitude, \( 2\pi N_k \phi_k(x) \) is the instantaneous (local) phase, \( N_k \phi'_k(x) \) is the instantaneous frequency (or \( N_k \nabla \phi_k(x) \) as the local wave vector), \( T(x) \) is a smooth trend function, and \( e(x) \) is a noisy perturbation term. Although \( \alpha_k(x) \) is conventionally treated as a smooth positive function, it is more realistic to consider \( \alpha_k(x) \) with an unknown and bounded support. Identifying wave-like components can also be understood as searching for appropriate governing equations to describe complex physical problems [34].

In this paper, we will focus on signals without a smooth trend function \( T(x) \):

\[ f(x) = \sum_{k=1}^{K} \alpha_k(x)e^{2\pi i N_k \phi_k(x)} + e(x), \] (1)

since \( T(x) \) can be recovered by removing wave-like components \( \alpha_k(x)e^{2\pi i N_k \phi_k(x)} \) and denoising methods for a smooth function has been well established [9, 10, 26, 27]. It follows from the definition that \( \alpha_k(x)e^{2\pi i N_k \phi_k(x)} \) is a highly oscillatory component with a frequency content also rapidly changing with \( x \). An immediate challenge from this rapid change is that an instantaneous frequency or the magnitude of a local wave vector may quickly increase to the sampling rate, e.g., power-law chirps in gravitational waves [18]. Another challenge comes from a large number of different \( N_k \) in various scales caused by the phenomenon of general wave shapes [53, 58, 59] or equivalently intrawaves [34, 57]. When noise meets these multiscale oscillatory components, an efficient and accurate tool with multiscale robustness to identify and analyze these wave-like components is of great value.

1.2 Synchrosqueezed transforms

A powerful tool for analyzing signal (1) is the synchrosqueezed time-frequency analysis consisting of a linear time-frequency analysis tool and a synchrosqueezing technique initialized in [23]. At each time or space location, the synchrosqueezing process reassigns values of the time-frequency representation based upon their local oscillation to achieve a sharpened time-frequency representation. Different to other time-frequency reassignment methods [4, 16, 17, 23] and time-frequency distribution methods [18], the synchrosqueezing process enjoys a simple and efficient reconstruction formula. This is especially important to high dimensional applications. Synchrosqueezed transforms are local in nature due to the local essence of linear time-frequency transforms before synchrosqueezing. It is also flexible to choose different linear transforms according to different data characteristics. Thus, synchrosqueezed transforms are non-parametric and adaptive to different data, e.g., ECG signals with spikes [53, 58], waveforms even with discontinuity [58], wave propagation with defects and sharp boundaries [59, 61]. Finally, synchrosqueezed transforms are visually informative. Hence, they allow human interaction in spectral analysis and inspire new thoughts for better understanding of oscillatory signals.

A variety of synchrosqueezed transforms have been proposed to study signal (1), e.g., the synchrosqueezed wavelet transform (SSWT) in [21, 22], the synchrosqueezed short time Fourier transform (SSSTFT) in [18], the synchrosqueezed wave packet transform (SSWPT)
in [58, 60] and the synchrosqueezed curvelet transform (SSCT) in [61]. Rigorous analysis has proved that these transforms can accurately decompose a class of superpositions of wave-like components and estimate their instantaneous (local) properties if the given signal is noiseless. To improve the synchrosqueezing operator in the presence of strongly non-linear instantaneous frequencies, some further methods have been proposed in [5, 40] based on an extra investigation of the non-linearity of instantaneous frequencies. All these synchrosqueezed transforms are compactly supported in the frequency domain to ensure their estimation accuracy. To better analyze signals with a trend in real-time computation, a recent paper [20] proposes a new synchrosqueezing method based on carefully designed wavelets with sufficient vanishing moments and a minimum support in the time domain. However, mathematical analysis on the accuracy of this new synchrosqueezing method is still under development. The underlying bottleneck is: whether a compactly supported synchrosqueezed transform is accurate enough to analyze signal (1).

Although the literature on synchrosqueezed transforms for noiseless data is fairly developed and they have been successfully applied to various real problems with noisy data, rigorous robustness analysis of these transforms is still limited. In a recent paper [47] that addresses the robustness analysis, it is assumed that (1) contains only Gaussian white noise with a variance much smaller than $\epsilon^2$, where $\epsilon \ll 1$ is the error tolerance of the estimation accuracy in [22]. This requirement is too restricted in real applications. To deal with heavier noise, a recent paper [32] proves the robustness against a generalized stationary Gaussian noise and analyzes statistical properties of (1) when it has a trend with heteroscedasticity. However, this proof is valid only for the 1D SSWT in [22] in analyzing wave-like components with instantaneous frequencies of constant order. It also relies on wavelets compactly supported in the frequency domain, which is consequently not good for trend estimation as discussed in [20]. Many fundamental problems remain open:

1. In harmonic analysis, compactly supported wavelet transforms are useful in various applications. However, they cannot be simply “synchrosqueezed”. Synchrosqueezing an arbitrary time-frequency analysis tool might not give an accurate estimate. Hence, it is important to clarify the condition for good synchrosqueezed transforms compactly supported in the time/space domain.

2. In probability analysis, up to now, people only know the robustness of 1D SSWT in [22]. Since there are massive noisy oscillatory signals in multidimensional space, it is of great value to explore the robustness of various synchrosqueezed transforms in different dimensions and optimize their robustness. In a difficult and important case, the robustness analysis of compactly supported synchrosqueezed transforms is quite challenging.

3. From a multiscale point of view, the robustness of SSWT in [22] is valid when non-linear wave-like components are of the same scale, as we shall see later. It is important to propose synchrosqueezed transforms with multiscale robustness and show how to optimize this robustness.

4. In numerical analysis, it is of practical interest to identify which numerical implementation can realize the robustness properties of the current synchrosqueezing framework.

Before formally answering these questions, it is necessary to discuss the different languages of various synchrosqueezed transforms. In the analysis of existing synchrosqueezed
transforms [22, 32, 47, 48], the authors are assuming a class of well-separated superpositions of intrinsic mode type functions. If we rephrase the definition of the 1D SSWT in [22, 32, 47] using a statement convenient for multiscale analysis, then this class of well-separated superpositions can be defined through the following definitions.

**Definition 1.1.** (An intrinsic mode type function for the 1D SSWT). A continuous function \( f : \mathbb{R} \to \mathbb{C}, f \in L^\infty(\mathbb{R}) \) is said to be intrinsic-mode-type (IMT) with accuracy \( \epsilon > 0 \) if \( f(x) = \alpha(x)e^{2\pi i N\phi(x)} \) with \( \alpha(x) \) and \( \phi(x) \) having the following properties:

\[
\alpha \in C^1(\mathbb{R}) \cap L^\infty(\mathbb{R}), \quad \phi \in C^2(\mathbb{R})
\]

\[
\inf_{x \in \mathbb{R}} \phi'(x) > 0, \quad \sup_{x \in \mathbb{R}} \phi'(x) < \infty, \quad \sup_{x \in \mathbb{R}} |\phi''(x)| < \infty,
\]

\[
|\alpha'(x)| \leq \epsilon|N\phi'(x)|, |\phi''(x)| \leq \epsilon|\phi'(x)|, \quad \forall x \in \mathbb{R}.
\]

To guarantee accurate estimates of non-linear wave-like components provided by the SSWT, the approach in [22] needs to assume \( N \) to be sufficiently small. To make things concrete, consider the requirement of Equation (3.5) in [22], which reads \( \epsilon < N^{-3/2} \) in the language of this paper. For example, if \( \epsilon = 0.01 \), then \( N \) has to be less than 21.5. Since larger \( \epsilon \) allows stronger non-linearity in Definition 1.1, Equation (3.5) says that high frequency wave-like components have to be nearly linear, which is impractical for a superposition of multiscale non-linear wave-like components. Indeed, multiscale components are common in nature, which motivates the work in [58, 60, 61] and this paper.

**Definition 1.2.** (A superposition of well-separated intrinsic mode functions for the 1D SSWT). A function \( f : \mathbb{R} \to \mathbb{C} \) is said to be a superposition of well-separated intrinsic mode functions, up to accuracy \( \epsilon \), with separation \( \Delta \), if there exists a finite \( K \), such that

\[
f(x) = \sum_{k=1}^{K} f_k(x) = \sum_{k=1}^{K} \alpha_k(x)e^{2\pi i N_k \phi_k(x)},
\]

where each \( f_k \) is an IMT, and where moreover their respective phase functions \( \phi_k \) satisfy

\[
N_k \phi_k'(x) > N_{k-1} \phi_{k-1}'(x)
\]

and

\[
|N_k \phi_k'(x) - N_{k-1} \phi_{k-1}'(x)| \geq \Delta \left[ N_k \phi_k'(x) + N_{k-1} \phi_{k-1}'(x) \right], \quad \forall x \in \mathbb{R}.
\]

The well-separation condition can be essentially understood as the condition that the instantaneous frequencies \( N_k \phi_k'(x) \) do not cross over the same support of a wavelet in the time-frequency domain. This is a special case of the well-separation condition that we adopt later. To guarantee the above well-separation condition, the sequence \( \{N_k\} \) has to increase in an order like \( 2^k \), which might be impractical for multiscale wave-like components in real applications.

Using our new language convenient for multiscale analysis for synchrosqueezed transforms, we answer the above questions here shortly and leave detailed illustrations later. We shall prove that it is possible to construct compactly supported synchrosqueezed transforms in the time domain. Their mother wave packets (“wavelets” in a special case) should obey certain smoothness requirement determined by the corresponding geometric scaling of the wave packet transform before synchrosqueezing. Previous synchrosqueezed transforms in [22, 58, 60, 61] and their extensions with compact supports in the time/space domain are robust against a generalized stationary Gaussian random noise. Rigorous probability analysis for their multiscale robustness with different geometric scaling is provided. It will be shown that a trade-off between the multiscale robustness and the estimation accuracy has to be balanced.
1.3 Significance of this work

Analyzing signals in (1) is also called a mode decomposition problem. A famous empirical mode decomposition (EMD) method has initialized a very active research line in advanced and adaptive data analysis. This method was first proposed by Huang et al. in [36] and refined in [37]. It has good numerical performance in decomposing a class of superpositions of oscillatory components and has been widely used in various applications, even though the mathematics behind this method is still unknown. However, the good properties of the EMD method are fragile. It is well known that EMD methods are not robust against noise. Therefore, synchrosqueezed transforms with well-developed mathematical background and provable robustness are important alternatives. This is illustrated in a recent review [45] by comparing several advanced tools for spectral estimations, e.g., the EMD method, the short-time Fourier transform, the SSWT, some basics pursuit method and some matching pursuit method. We expect synchrosqueezed transforms can provide new insights for oscillatory component analysis to help us understand the nature, since in some cases the EMD method would give misleading results [58, 60].

Oscillatory patterns are not only ubiquitous in natural signals and images, but also in broader mathematical research. First, a recent paper [34] builds up the connection between the decomposition of oscillatory components and classical second order differential equations. Each oscillatory components can be associated with a solution of a second order ordinary differential equation of the form \( x'' + p(x, t)x' + q(x, t) = 0 \) using a data-driven data analysis method [33, 35]. The non-linearity in \( p \) and \( q \) is equivalent to the non-linearity of the oscillatory components. It would be of interest to see whether we can connect high dimensional non-linear components to the non-linearity of partial differential equations. Second, the atom decomposition method is a well-established method to study oscillatory operators and to compute high frequency wave equations. Wave propagation decomposition methods, e.g., a wave field is decomposed into a superposition of Gaussian beams for further evolution of the high frequency waves, have been investigated in [2, 3]. There are serious indications that these tasks could be realized with better efficiency using fast synchrosqueezed transforms.

Statistical literature on oscillatory estimation is fairly developed, but a multiscale oscillatory estimation with a possible trend is perhaps more recent. Some existing models, e.g., the seasonal auto-regressive integrated moving average [7] and the trend and seasonal components algorithm [24], focus on forecasting. They might not be suitable for time-varying historical components as discussed in [32]. Some methods are based on a global assumption with precise known properties of the signal and perform a generalized likelihood ratio test. Global assumptions could be too restrictive in analyzing local information hidden in a general time-varying component. The resulting statistics might be sensitive to the length of the given signal. Some models are fully non-parametric and local in nature. They can even detect an oscillatory component from totally unknown and fully noisy data [15]. However, they are focusing on detecting and analyzing only one oscillatory component and cannot be applied to more complex data. Hence, the non-parametric robust analysis tool for multiscale components discussed in this paper is new and adaptive to a general problem.

1.4 Organization

The rest of this paper is organized as follows. In Section 2 to 4, main theorems for the 1D synchrosqueezed wave packet transform, the 2D synchrosqueezed wave packet transform and the 2D synchrosqueezed curvelet transform are presented, respectively. In each
of these sections, after a brief introduction to these transforms, new theorems about compactly supported synchrosqueezed transforms and their robustness to bounded perturbation and generalized stationary Gaussian noise are demonstrated. In Section 5, a few algorithms and their implementations are introduced in detail for robust instantaneous frequency/local wave vector estimates. Several numerical examples with heavy noise are provided to demonstrate the proposed properties. In Section 6, we conclude this paper with a discussion of future work.

2 1D synchrosqueezed wave packet transform (SSWPT)

2.1 The definition of the 1D SSWPT

We briefly introduce the 1D SSWPT proposed in [58]. Previously, the SSWPT is built on mother wave packets compactly supported in the frequency domain. In order to include SSWPTs with smaller essential supports in the time/space domain, this paper discusses a wider class of mother wave packets defined below.

**Definition 2.1.** A mother wave packet \( w(x) \in C^m(\mathbb{R}) \) is of type \((\epsilon, m)\) for some \( \epsilon > 0 \), and some non-negative integer \( m \), if \( \hat{w}(\xi) \) is a real-valued function with an essential support in the ball \( B_1(0) \) centered at the origin with a radius 1 satisfying that:

\[
|\hat{w}(\xi)| \leq \frac{\epsilon}{(1 + |\xi|)^m},
\]

for \( |\xi| > 1 \).

Since \( w \in C^m(\mathbb{R}) \), the above decaying requirement is easy to satisfy. Actually, we can further assume \( \hat{w}(\xi) \) is essentially supported in a ball \( B_d(0) \) with \( d \in (0, 1] \) to adapt signals with close instantaneous frequencies. However, \( d \) is just a constant in later asymptotic analysis. Hence, we omit its discussion and consider it as 1 in the analysis but implement it in the numerical tool. We can use this mother wave packet \( w(x) \) to define a family of wave packets through scaling, modulation, and translation, controlled by a geometric parameter \( s \).

**Definition 2.2.** Given the mother wave packet \( w(x) \) of type \((\epsilon, m)\) and a parameter \( s \in (1/2, 1) \), the family of wave packets \( \{w_{ab}(x) : a \geq 1, b \in \mathbb{R}\} \) is defined as

\[
w_{ab}(x) = a^{s/2} w(a^s(x - b)) e^{2\pi i a(x - b)},
\]
or equivalently, in the Fourier domain as

\[
\hat{w}_{ab}(\xi) = a^{-s/2} e^{-2\pi ib\xi} \hat{w}(a^{-s}(\xi - a)).
\]

These definitions allow us to construct a family of compactly supported wave packets, which will be useful in practice. Notice that if \( s \) were equal to 1, these functions would be qualitatively similar to the standard wavelets. On the other hand, if \( s \) were equal to 1/2, we would obtain the wave atoms defined in [25]. But \( s \in (1/2, 1) \) is essential as we shall see in the main theorems later.

The instantaneous frequency of the low frequency part is not well defined as discussed in [42]. For this reason, it is enough to consider wave packets with \( a \geq 1 \). High frequency components can be identified and extracted independently of the low frequency part so that the low frequency part can be recovered by removing high frequency components.
Definition 2.3. The 1D wave packet transform of a function \( f(x) \) is a function

\[
W_f(a,b) = \langle f, w_{ab} \rangle = \int f(x) w_{ab}(x) dx
\]

for \( a \geq 1, b \in \mathbb{R} \).

Definition 2.4. Instantaneous frequency estimation:

Let \( f \in L^\infty(\mathbb{R}) \). The instantaneous frequency estimation function \( v_f(a,b) \) for \( a \in [1, \infty) \) and \( b \in \mathbb{R} \) of \( f \) is defined by

\[
v_f(a,b) = \begin{cases} 
\frac{\partial_b W_f(a,b)}{2\pi i W_f(a,b)}, & \text{for } |W_f(a,b)| > 0; \\
\infty, & \text{otherwise.}
\end{cases}
\]

It will be proved that, for a class of wave-like functions \( f(x) = \alpha(x)e^{2\pi i N \phi(x)} \), \( v_f(a,b) \) precisely approximates \( N \phi'(b) \) independently of \( a \) as long as \( |W_f(a,b)| \) is large enough. Hence, if we squeeze the coefficients \( W_f(a,b) \) together based upon the same instantaneous frequency information function \( v_f(a,b) \), then we would obtain a sharpened time-frequency representation of \( f(x) \). This motivates the definition of the synchrosqueezed energy distribution as follows.

Definition 2.5. Given \( f(x) \), the synchrosqueezed energy distribution \( T_f(v,b) \) is defined by

\[
T_f(v,b) = \int |W_f(a,b)|^2 \delta(\Re v_f(a,b) - v) da
\]

for \( v, b \in \mathbb{R} \).

Here \( \delta \) denotes the Dirac delta function and \( \Re v_f(a,b) \) means the real part of a complex number \( v_f(a,b) \). For a multi-component signal \( f(x) \), the synchrosqueezed energy of each component will concentrate around its corresponding instantaneous frequency. Hence, the SSWPT can provide information about their instantaneous frequencies.

2.2 Analysis of the 1D SSWPT

In this section, we provide rigorous analysis of the 1D SSWPT generated from mother wave packets of type \((\epsilon, m)\) to analyze a superposition of components in noiseless, perturbed and noisy cases. The proofs of these theorems are provided in Appendix A.

Definition 2.6. A function \( f(x) = \alpha(x)e^{2\pi i N \phi(x)} \) is an intrinsic mode type function (IMT) of type \((M,N)\), if \( \alpha(x) \) and \( \phi(x) \) satisfy the conditions below.

\[
\alpha(x) \in C^\infty, \quad |\alpha'(x)| \leq M, \quad 1/M \leq \alpha(x) \leq M
\]

\[
\phi(x) \in C^\infty, \quad 1/M \leq |\phi'(x)| \leq M, \quad |\phi''(x)| \leq M.
\]

Definition 2.7. A function \( f(x) \) is a well-separated superposition of type \((M,N,K,s)\), if

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

where each \( f_k(x) = \alpha_k(x)e^{2\pi i N_k \phi_k(x)} \) is an IMT of type \((M,N_k)\) such that \( N_k \geq N \) and the phase functions satisfy the separation condition: for any pair \((a,b)\), there exists at most one \( k \) such that

\[
a^{-s}|a - N_k \phi'_k(b)| < 1.
\]

We denote by \( F(M,N,K,s) \) the set of all such functions.
Theorem 2.8 below shows that the estimates of instantaneous frequencies \( \{ N_k \phi_k' (x) \}_{k=1}^K \) are accurate if a given superposition of IMTs is noiseless. In what follows, when we write \( O ( \cdot ) , \lesssim , \text{or} \gtrsim , \) the implicit constants may depend on \( M \) and \( K \).

**Theorem 2.8.** Suppose the mother wave packet is of type \(( \epsilon, m )\), for any fixed \( \epsilon \in (0, 1) \) and any fixed integer \( m \geq 0 \). For a function \( f(x) \), we define

\[
R_\epsilon = \{ (a, b) : |W_f(a, b)| \geq a^{-s/2}\sqrt{\epsilon} \},
\]

\[
S_\epsilon = \{ (a, b) : |W_f(a, b)| \geq \sqrt{\epsilon} \},
\]

and

\[
Z_k = \{ (a, b) : |a - N_k \phi_k' (b)| \leq a^s \}
\]

for \( 1 \leq k \leq K \). For fixed \( M \) and \( K \), there exists a constant \( N_0 (M, K, s, \epsilon) \simeq \max \left\{ \epsilon^{\frac{s}{s-1}}, \epsilon^{\frac{s}{1-s}} \right\} \) such that for any \( N > N_0 (M, K, s, \epsilon) \) and \( f(x) \in F(M, N, K, s) \) the following statements hold.

(i) \( \{ Z_k : 1 \leq k \leq K \} \) are disjoint and \( S_\epsilon \subset R_\epsilon \subset \bigcup_{1 \leq k \leq K} Z_k \);

(ii) For any \((a, b) \in R_\epsilon \cap Z_k\),

\[
\frac{|v_f(a, b) - N_k \phi_k' (b)|}{|N_k \phi_k' (b)|} \lesssim \sqrt{\epsilon};
\]

(iii) For any \((a, b) \in S_\epsilon \cap Z_k\),

\[
\frac{|v_f(a, b) - N_k \phi_k' (b)|}{|N_k \phi_k' (b)|} \lesssim \frac{\sqrt{\epsilon}}{N_k^{s/2}}.
\]

The proof of Theorem 2.8 relies on two lemmas as follows to estimate the asymptotic behavior of \( W_f(a, b) \) and \( \partial_b W_f(a, b) \) as \( N \) going to infinity.

**Lemma 2.9.** Suppose \( \Omega_a = \{ k : a \in [\frac{N_a}{2M}, 2MN_a] \} \). Under the assumption of Theorem 2.8 we have

\[
W_f(a, b) = a^{-s/2} \left( \sum_{k \in \Omega_a} \alpha_k (b) e^{2\pi i N_k \phi_k (b)} \hat{w} \left( (a - N_k \phi_k' (b)) a^{-s} \right) + O (\epsilon) \right),
\]

when \( N > N_0 (M, K, s, \epsilon) \simeq \max \left\{ \epsilon^{\frac{s}{s-1}}, \epsilon^{\frac{s}{1-s}} \right\} \).

The next lemma is to estimate \( \partial_b W_f(a, b) \) when \( \Omega_a = \{ k : a \in [\frac{N_a}{2M}, 2MN_a] \} \) is not empty, i.e., when \( W_f(a, b) \) is relevant.

**Lemma 2.10.** Suppose \( \Omega_a = \{ k : a \in [\frac{N_a}{2M}, 2MN_a] \} \) is not empty. Under the assumption of Theorem 2.8 we have

\[
\partial_b W_f(a, b)
= a^{-s/2} \left( \sum_{k \in \Omega_a} 2\pi i N_k \alpha_k (b) \phi_k' (b) e^{2\pi i N_k \phi_k (b)} \hat{w} \left( (a - N_k \phi_k' (b)) a^{-s} \right) + aO (\epsilon) \right),
\]

when \( N > N_0 (M, K, s, \epsilon) \simeq \max \left\{ \epsilon^{\frac{s}{s-1}}, \epsilon^{\frac{s}{1-s}} \right\} \).
Lemma \ref{lem:1} and \ref{lem:2} are proved in \cite{58} for mother wave packets \(w\) of type \((0, m)\) for any non-negative integer \(m\). The proof does not rely on the compact support of \(\hat{w}\) and consequently these lemmas remain true for \(w\) of type \((\epsilon, m)\). Theorem \ref{thm:2} is a direct result of Lemma \ref{lem:1} and Lemma \ref{lem:2} as proved in Appendix A. The results in Part \((ii)\) and \((iii)\) are similar, but Part \((iii)\) is more useful in practical implementation with a uniform threshold.

Theorem \ref{thm:2} shows that the instantaneous frequency estimation function \(v_f(a, b)\) can estimate \(N_k\phi_k'(x)\) accurately for a class of noiseless superpositions of IMTs if their phases are sufficiently steep. This guarantees the well concentration of the synchrosqueezed energy distribution \(T_f(v, b)\) around \(N_k\phi_k'(x)\). If the superposition is perturbed slightly by a contaminant, Theorem \ref{thm:2} below shows that these conclusions are still valid with a reasonable error determined by the magnitude of the perturbation.

**Theorem 2.11.** Suppose the mother wave packet is of type \((\epsilon, m)\), for any fixed \(\epsilon \in (0, 1)\) and any fixed integer \(m \geq 0\). Suppose \(g(x) = f(x) + e(x)\), where \(e(x) \in L^{\infty}\) is a small error term that satisfies \(\|e\|_{L^\infty} \leq \sqrt{\epsilon_1}\) for some \(\epsilon_1 > 0\). For any \(p \in (0, \frac{3}{2}]\), let \(\delta = \sqrt{\epsilon} + \epsilon_1^{1-p}\). Define

\[
R_\delta = \{(a, b) : |W_g(a, b)| \geq a^{-s/2} \delta\},
\]
\[
S_\delta = \{(a, b) : |W_g(a, b)| \geq \delta\},
\]

and

\[
Z_k = \{(a, b) : |a - N_k\phi_k'(b)| \leq a^s\}
\]

for \(1 \leq k \leq K\). For fixed \(M\) and \(K\), there exists a constant \(N_0 (M, K, s, \epsilon) \simeq \max\{\epsilon^{1/2}, \epsilon\}\) such that for any \(N > N_0 (M, K, s, \epsilon)\) and \(f(x) \in F(M, N, K, s)\) the following statements hold.

\(i)\) \(\{Z_k : 1 \leq k \leq K\}\) are disjoint and \(S_\delta \subset R_\delta \subset \bigcup_{1 \leq k \leq K} Z_k\);

\(ii)\) For any \((a, b) \in R_\delta \cap Z_k\),

\[
\frac{|v_g(a, b) - N_k\phi_k'(b)|}{|N_k\phi_k'(b)|} \lesssim \sqrt{\epsilon} + \epsilon_1^p;
\]

\(iii)\) For any \((a, b) \in S_\delta \cap Z_k\),

\[
\frac{|v_g(a, b) - N_k\phi_k'(b)|}{|N_k\phi_k'(b)|} \lesssim \frac{\sqrt{\epsilon} + \epsilon_1^p}{N_k^{s/2}}.
\]

We introduce the parameter \(p\) to clarify the relation among the perturbation level, the threshold and the accuracy for better understanding the influence of perturbation or noise. For the same purpose, a parameter \(q\) will be introduced in the coming theorems. Theorem \ref{thm:3} shows that the instantaneous frequency estimates provided by the SSWPT are still reasonable when the given signal is contaminated by a bounded perturbation. Actually, if the threshold \(\delta\) is larger, e.g., \(\delta \geq \sqrt{\frac{\epsilon}{\epsilon_1}}\), the relative estimate errors in \((ii)\) and \((iii)\) are bounded by \(\sqrt{\epsilon}\) and \(\frac{\sqrt{\epsilon}}{N_k^{s/2}}\), respectively. This also implies that the instantaneous frequency can be better estimated by selecting the wave packet coefficient with the largest magnitude. However, when the perturbation is overwhelming, e.g., the wave packet coefficients
of a component are below the threshold in \((ii)\), it is difficult to estimate instantaneous frequencies.

Next, we will illustrate the robustness properties of the SSWPT in the presence of random perturbation. \[3, 39, 41, 46, 49\] are referred to for basic facts about generalized random fields and complex Gaussian processes that are used throughout this paper. To warm up, we start with additive Gaussian white noise in Theorem 2.12 and extend it to a general zero mean stationary Gaussian noise in Theorem 2.13. Let \(n\) be the dimension of given data. \(n = 1\) in this section and \(n = 2\) in later sections. If we fix a probability space \((\mathbb{R}^n, \mu)\) and assume that \(L^2(\mathbb{R}^n, \mu)\) is separable, a stationary Gaussian process \(e\) on \(\mathbb{R}^n\) is an \(L^2(\mathbb{R}^n, \mu)\)-valued distribution \[46\], i.e., a continuous linear functional in \(\mathcal{D}'(\mathbb{R}^n, L^2(\mathbb{R}^n, \mu))\) such that

\[
e: C_0^\infty(\mathbb{R}^n) \to L^2(\mathbb{R}^n, \mu),
\]

which can be continuously extended to

\[
e: L^1 C^r(\mathbb{R}^n) \to L^2(\mathbb{R}^n, \mu),
\]

for some \(r \in \mathbb{N}\) or \(r = \infty\) depending on \(e\). We assume that \(r\) is small enough such that the family of wave packets we constructed and their derivatives belong to \(C^r(\mathbb{R}^n)\).

Suppose \(e\) has a mean functional \(\mathcal{T}: L^1 C^r(\mathbb{R}^n) \to L^1 C^r(\mathbb{R}^n)\) and a covariance functional \(\mathcal{R}: L^1 C^r(\mathbb{R}^n) \to L^1 C^r(\mathbb{R}^n)\), then we have

1. For any finite collection \(\{f_k\} \subset L^1 C^r(\mathbb{R}^n)\), \(\{e(f_k)\}\) are jointly Gaussian variables and their joint distribution is translation invariant for all translates of \(f_k\);

2. \(\mathbb{E}[e(f)] = \mathcal{T} f\) and \(\mathbb{E}[e(f_1) e(f_2)] = \langle f_1, \mathcal{R} f_2 \rangle\), where \(\langle \cdot, \cdot \rangle\) is the \(L^2\) inner product.

Gaussian white noise is a special case of stationary Gaussian processes with \(\mathcal{T} = 0\) and \(\mathcal{R}\) being the identical functional. For the convenience of notations, for any wave packet \(w_{ab}(x)\), \(e(w_{ab})\) and \(e(\partial_b w_{ab})\) are denoted as \(W_e(a, b)\) and \(\partial_b W_e(a, b)\), respectively. We assume that \(e\) has an explicit power spectral function denoted by \(\tilde{e}(\xi)\). \(|\cdot|\) will represent the \(L^2\) norm.

**Theorem 2.12.** Suppose the mother wave packet is of type \((\epsilon, m)\), for any fixed \(\epsilon \in (0, 1)\) and any fixed integer \(m \geq \frac{2}{\epsilon} + 4\). Suppose \(g(x) = f(x) + e\), where \(e\) is zero mean Gaussian white noise with a variance \(\epsilon_1^{1+q}\) for some \(q > 0\) and some \(\epsilon_1 > 0\). For any \(p \in (0, \frac{1}{2}]\), let \(\delta = \sqrt{\epsilon} + \epsilon_1^{1-p}\). Define

\[
R_\delta = \{(a, b) : |W_g(a, b)| \geq a^{-s/2} \delta\},
\]

\[
S_\delta = \{(a, b) : |W_g(a, b)| \geq \delta\},
\]

and

\[
Z_k = \{(a, b) : |a - N_k \phi_k^I(b)| \leq a^s\}
\]

for \(1 \leq k \leq K\). For fixed \(M\) and \(K\), there exists a constant \(N_0(M, K, s, \epsilon) \simeq \max \left\{ \epsilon^{-1/s}, \epsilon^{-1} \right\} \) such that for any \(N > N_0(M, K, s, \epsilon)\) and \(f(x) \in F(M, N, K, s)\) the following statements hold.

(i) \(\{Z_k : 1 \leq k \leq K\}\) are disjoint.

(ii) If \((a, b) \in R_\delta\), then \((a, b) \in \bigcup_{1 \leq k \leq K} Z_k\) with a probability at least

\[
1 - e^{-O(N_k^{-s} \epsilon_1^{-q})} + O \left( \frac{\epsilon}{N_k^{m(1-s)}} \right).
\]
(iii) If \((a, b) \in S_\delta\), then \((a, b) \in \bigcup_{1 \leq k \leq K} Z_k\) with a probability at least
\[
1 - e^{-\epsilon_1\|w\|^{-2}} + O \left( \frac{\epsilon}{N_k^{m(1-s)}} \right).
\]

(iv) If \((a, b) \in R_\delta \cap Z_k\) for some \(k\), then
\[
\left| v_g(a, b) - N_k \phi_k(a) \right| / \left| N_k \phi_k(a) \right| \lesssim \sqrt{\epsilon} + \epsilon_1^p
\]
is true with a probability at least
\[
\left( 1 - e^{-O(N_k^{2-3s\epsilon_1^{-q}})} \right) \left( 1 - e^{-O(N_k^{-s\epsilon_1^{-q}})} \right) + O \left( \frac{\epsilon}{N_k^{(m-4)(1-s)-2}} \right).
\]

(v) If \((a, b) \in S_\delta \cap Z_k\), then
\[
\left| v_g(a, b) - N_k \phi_k(a) \right| / \left| N_k \phi_k(a) \right| \lesssim \sqrt{\epsilon} + \epsilon_1^p
\]
is true with a probability at least
\[
\left( 1 - e^{-O(N_k^{2-2s\epsilon_1^{-q}})} \right) \left( 1 - e^{-O(N_k^{-2s\epsilon_1^{-q}})} \right) + O \left( \frac{\epsilon}{N_k^{(m-4)(1-s)-2}} \right).
\]

Thus far, we considered the robustness to small perturbation and Gaussian white noise. Next, we will show that Theorem 2.13 can be extended to a broader class of colored noise.

**Theorem 2.13.** Suppose the mother wave packet is of type \((\epsilon, m)\), for any fixed \(\epsilon \in (0, 1)\) and any fixed integer \(m \geq \frac{2}{s} + 4\). Suppose \(g(x) = f(x) + \epsilon\), where \(\epsilon\) is a zero mean stationary Gaussian process. Let \(\hat{e}(\xi)\) denote the spectrum of \(\epsilon\), max\(\xi |\hat{e}(\xi)| \leq \epsilon^{-1}\) and \(M_a = \max_{|\xi| \leq \epsilon} \hat{e}(a^s \xi + a)\). For any \(p \in (0, \frac{1}{2})\) and \(q > 0\), let \(\delta_a = M_a^{(\frac{1}{2}-p)/(1+q)} + \sqrt{\epsilon}\),
\[
R_{\delta_a} = \{(a, b) : |W_g(a, b)| \geq a^{-s/2} \delta_a \},
\]
\[
S_{\delta_a} = \{(a, b) : |W_g(a, b)| \geq \delta_a \},
\]
and
\[
Z_k = \{(a, b) : |a - N_k \phi_k(b)| \leq a^s \}
\]
for \(1 \leq k \leq K\). For fixed \(M\) and \(K\), there exists a constant \(N_0(M, K, s, \epsilon) \simeq \max \left\{ \epsilon^{\frac{1}{2s-1}}, \epsilon^{\frac{1}{1-s}} \right\}\) such that for any \(N > N_0(M, K, s, \epsilon)\) and \(f(x) \in F(M, N, K, s)\) the following statements hold.

(i) \(\{Z_k : 1 \leq k \leq K\}\) are disjoint.

(ii) If \((a, b) \in R_{\delta_a}\), then \((a, b) \in \bigcup_{1 \leq k \leq K} Z_k\) with a probability at least
\[
1 - e^{-O(N_k^{-s} M_a^{-q/(1+q)})} + O \left( \frac{\epsilon}{N_k^{m(1-s)}} \right).
\]
(iii) If \((a, b) \in S_\delta a\), then \((a, b) \in \bigcup_{1 \leq k \leq K} Z_k\) with a probability at least

\[
1 - e^{-O(M_a^{-q/(1+q)})} + O\left(\frac{\epsilon}{N_k^m(1-s)}\right),
\]

(iv) If \((a, b) \in R_\delta a \cap Z_k\) for some \(k\), then

\[
\frac{|v_g(a, b) - N_k\phi_k(b)|}{|N_k\phi_k'(b)|} \lesssim \sqrt{\epsilon + M_a^{p/(1+q)}}
\]

is true with a probability at least

\[
\left(1 - e^{-O(N_k^{2-3s}M_a^{-q/(1+q)})}\right)\left(1 - e^{-O(N_k^{-2s}M_a^{-q/(1+q)})}\right) + O\left(\frac{\epsilon}{N_k^{m(4)(1-s)-2}}\right).
\]

(v) If \((a, b) \in S_\delta a \cap Z_k\), then

\[
\frac{|v_g(a, b) - N_k\phi_k(b)|}{|N_k\phi_k'(b)|} \lesssim N_k^{-s/2} \left(\sqrt{\epsilon + M_a^{p/(1+q)}}\right)
\]

is true with a probability at least

\[
\left(1 - e^{-O(N_k^{2-3s}M_a^{-q/(1+q)})}\right)\left(1 - e^{-O(N_k^{-2s}M_a^{-q/(1+q)})}\right) + O\left(\frac{\epsilon}{N_k^{m(4)(1-s)-2}}\right).
\]

Theorem 2.12 and 2.13 illustrate that when the sampling rate of a given signal is high enough such that the wave-like components are relatively smooth in terms of the noise, the SSWPT can estimate the instantaneous frequencies of these components accurately with a high probability. In particular, Theorem 2.13 says that if the spectrum of noise is not overwhelming the wave packet coefficients of IMTs, the SSWPT can provide accurate estimates with a high probability. Part (ii) and (iii) in the last two theorems demonstrate that the influence of noise can be significantly reduced with a proper threshold after the wave packet transform and we could obtain useful information with a high probability. Part (iv) and (v) show that the synchrosqueezing process is able to concentrate the wave packet representation to the instantaneous frequencies with a reasonable probability after a properly thresholding. Hence, the essential support of the synchrosqueezed energy distribution helps to estimate the instantaneous frequencies statistically.

In the above discussion, we have not optimized the dependence of \(N\) on \(\epsilon\). There are two extra steps to minimize the lower bound for \(N\). Comparing Definition 1.1 and Definition 2.6, it is clear that we have allowed fully non-linearity to IMTs in the previous theorems. The requirement \(\epsilon^{1-s}\) can be reduced to a constant order if we restrict to a slightly smaller class of IMTs with weaker non-linearity. For example, if \(N_k \lesssim \epsilon^{-1/s}\) or \(N_k \lesssim \epsilon^{-1/(2s-1)}\), then we impose extra condition \(|a'_k(x)| \leq \epsilon N_k^s|\phi_k'(x)|\) or \(|\phi''_k(x)| \leq \epsilon N_k^{2s-1}|\phi_k'(x)|\), respectively. A careful inspection of the proof of Lemma 2.9 and 2.10 in the Taylor expansion approximation shows that these lemmas are still true. Hence, the synchrosqueezed transforms remain accurate.

Another step is to look at \(\epsilon^{1-s}\), which comes from the decaying estimate of wave packet coefficients \(W_f(a, b)\) when their scales \(a\) do not match the oscillation \(N_k\) of IMTs. If we further take advantage of the decay speed of the mother wave packet, we will see \(|W_f(a, b)|\)
would decay much faster when this mismatch occurs. For example, a mother wave packet in $C^m$ satisfies that
\[
\hat{w}(\xi) \leq C_m (1 + |\xi|)^{-m}.
\]
Since the mother wave packet is decaying rapidly at infinity, we can simply assume that the smooth amplitude function of the IMT has a compact support large enough and only need to bound $|W_f(a, b)|$ for $b$ at the support center. Since $\phi \in C^\infty$, by the diffeomorphism equivalence in Lemma 2.2 in [25], which is also valid for the wave packet transform by careful inspection, it is sufficient to assume $\phi(x) = x$. It follows from the discussion in Lemma 2.3 in [25] that we only require the following bound for previous theorems:
\[
|W_f(a, b)| \leq a^{-s/2}C_m (1 + |a - N|)^{-m/2} \leq \epsilon.
\]
Thus, we see $|W_f(a, b)|$ decays rapidly when $a \notin \left[\frac{N}{2M}, 2MN\right]$ for a reasonable large $N$. In practice, $\epsilon$ cannot be too small for numerical purposes and the number of periods of the input data is large enough so that a non-linear wave-like component is well defined. Hence, the above requirement is not a main issue.

We close this section with a few extra remarks. First, $s \in (1/2, 1)$ is essential in those theorems if we do not impose extra condition on the non-linearity of IMTs. Second, as pointed out in [58], another advantage of allowing $s \in (1/2, 1)$ is that a smaller $s$ leads to a better scale resolution to distinguish two IMTs with close instantaneous frequencies or a sequence of IMTs with instantaneous frequencies spreading out in the time-frequency domain. We refer to [58] for a detailed discussion. Finally, the theorems above provide a new insight that a smaller $s$ yields a synchrosqueezed transform with better robustness. This new insight is especially important when designing synchrosqueezed transforms compactly supported or decaying fast in the time domain. Theorem 2.12 and 2.13 show that the parameter $m$ in the mother wave packet has to be large enough, satisfying $m \geq 1 - s + 4$. In a special case, if a compactly supported synchrosqueezed wavelet transform (corresponding to $s = 1$) is preferable in some application, its mother wavelet is better to be $C^\infty$.

3 2D synchrosqueezed wave packet transform (SSWPT)

3.1 The definition of the 2D SSWPT

In a similar way, we will briefly introduce the 2D synchrosqueezed wave packet transform proposed in [60]. We can also introduce a 2D mother wave packet $w(x) \in C^m(\mathbb{R}^2)$ of type $(\epsilon, m)$ such that $\hat{w}(\xi)$ has an essential support in the ball $B_1(0)$ centered at the frequency origin with a radius 1, i.e.,
\[
|\hat{w}(\xi)| \leq \frac{\epsilon}{(1 + |\xi|)^m},
\]
for $|\xi| > 1$ and some non-negative integer $m$. A family of 2D wave packets is obtained by isotropic dilations, rotations and translations of the mother wave packet as follows, controlled by a geometric parameter $s$.

**Definition 3.1.** Given the mother wave packet $w(x)$ of type $(\epsilon, m)$ and the parameter $s \in (1/2, 1)$, the family of wave packets $\{w_{ab}(x) : a, b \in \mathbb{R}^2, |a| \geq 1\}$ are defined as
\[
w_{ab}(x) = |a|^s w(|a|^s (x - b)) e^{2\pi i (x - b) \cdot a},
\]
or equivalently in the Fourier domain
\[
\hat{w}_{ab}(\xi) = |a|^{-s} e^{-2\pi ib \cdot \xi} \hat{w}(|a|^{-s} (\xi - a)).
\]
Some properties can be seen immediately from the definition: the Fourier transform \( \hat{w}_{ab}(\xi) \) is essentially supported in \( B_{|a|^s}(a) \), a ball centered at \( a \) with a radius \( |a|^s \); \( w_{ab}(x) \) is centered in space at \( b \) with an essential support of width \( O(|a|^s) \); \( \{w_{ab}(x) : a, b \in \mathbb{R}^2, |a| \geq 1\} \) 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 \( 1/2 \), we would obtain the wave atoms defined in [25]. If \( s \) were equal to 1, these functions would be qualitatively similar to the standard 2D wavelets. In general, a 2D SSWPT with a smaller \( s \) value is better distinguishing two IMTs with close propagating directions. This is the motivation to propose 2D SSWPT rather than directly generalizing the 1D SSWT in [21, 22, 23, 55].

With this family of wave packets, we define the wave packet transform as follows.

**Definition 3.2.** The wave packet transform of a function \( f(x) \) is a function

\[
W_f(a, b) = \langle f, w_{ab} \rangle = \int_{\mathbb{R}^2} f(x) \overline{w_{ab}(x)} \, dx
\]

for \( a, b \in \mathbb{R}^2, |a| \geq 1 \).

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

\[
\int_{\mathbb{R}^4} |W_f(a, b)|^2 \, da \, db \approx \int_{\mathbb{R}^2} |f(x)|^2 \, dx.
\]

**Definition 3.3.** The local wave vector estimation of a function \( f(x) \) at \((a, b) \in \mathbb{R}^4\) is

\[
v_f(a, b) = \begin{cases} \nabla_b W_f(a, b) & \text{for } W_f(a, b) \neq 0; \\ (\infty, \infty) & \text{otherwise.} \end{cases}
\]

Given the wave vector estimation \( v_f(a, b) \), the synchrosqueezing step reallocates the information in the phase space and provides a sharpened phase space representation of \( f(x) \) in the following way.

**Definition 3.4.** Given \( f(x) \), the synchrosqueezed energy distribution \( T_f(v, b) \) is defined by

\[
T_f(v, b) = \int_{\mathbb{R}^2} |W_f(a, b)|^2 \delta(\Re v_f(a, b) - v) \, da
\]

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

In addition, we have the following property

\[
\int T_f(v, b) \, dv \, db = \int |W_f(a, b)|^2 \delta(\Re v_f(a, b) - v) \, dv \, db = \int |W_f(a, b)|^2 \, da \, db \approx \|f\|_2^2
\]

from Fubini’s theorem and the norm equivalence [2], for any \( f(x) \) with its Fourier transform vanishing for \( |\xi| < 1 \).

### 3.2 Analysis of the 2D SSWPT

In this section, we first revisit the results in [60]. The synchrosqueezed wave packet transform can estimate well-separated local wave vectors of a superposition of 2D IMTs defined below. The robustness properties of these estimates will be analyzed afterward.
Definition 3.5. A function \( f(x) = \alpha(x)e^{2\pi i N \phi(x)} \) is an intrinsic mode type function (IMT) of type \((M, N)\) if \( \alpha(x) \) and \( \phi(x) \) satisfy

\[
\alpha(x) \in C^\infty, \quad |\nabla \alpha(x)| \leq M, \quad 1/M \leq \alpha(x) \leq M \\
\phi(x) \in C^\infty, \quad 1/M \leq |\nabla \phi(x)| \leq M, \quad |\nabla^2 \phi(x)| \leq M.
\]

Definition 3.6. A function \( f(x) \) is a well-separated superposition of type \((M, N, K, s)\) if

\[
f(x) = \sum_{k=1}^{K} f_k(x)
\]

where each \( f_k(x) = \alpha_k(x)e^{2\pi i N_k \phi_k(x)} \) is an IMT of type \((M, N_k)\) with \( N_k \geq N \) and the phase functions satisfy the separation condition: for any \((a, b) \in \mathbb{R}^2\), there exists at most one \( f_k \) satisfying that

\[
|a|^{-s} |a - N_k \nabla \phi_k(b)| \leq 1.
\]

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

The following theorem illustrates the main results of 2D SSWPT for a superposition of IMTs without noise or perturbation.

Theorem 3.7. Suppose the 2D mother wave packet is of type \((\epsilon, m)\), for any fixed \( \epsilon \in (0, 1) \) and any fixed integer \( m \geq 0 \). For a function \( f(x) \), we define

\[
R_\epsilon = \{(a, b) : |W_f(a, b)| \geq |a|^{-s}\sqrt{\epsilon}\}, \\
S_\epsilon = \{(a, b) : |W_f(a, b)| \geq \sqrt{\epsilon}\},
\]

and

\[
Z_k = \{(a, b) : |a - N_k \nabla \phi_k(b)| \leq |a|^s\}
\]

for \( 1 \leq k \leq K \). For fixed \( M \) and \( K \) there exists a constant \( N_0(M, K, s, \epsilon) \approx \max \{\epsilon^{\frac{2}{s^2 - 1}}, \epsilon^{\frac{1}{s^2}}\} \) such that for any \( N > N_0 \) and \( f(x) \in F(M, N, K, s) \) the following statements hold.

(i) \( \{Z_k : 1 \leq k \leq K\} \) are disjoint and \( S_\epsilon \subset R_\epsilon \subset \bigcup_{1 \leq k \leq K} Z_k \);

(ii) For any \((a, b) \in R_\epsilon \cap Z_k\),

\[
\frac{|v_f(a, b) - N_k \nabla \phi_k(b)|}{|N_k \nabla \phi_k(b)|} \lesssim \sqrt{\epsilon};
\]

(iii) For any \((a, b) \in S_\epsilon \cap Z_k\),

\[
\frac{|v_f(a, b) - N_k \nabla \phi_k(b)|}{|N_k \nabla \phi_k(b)|} \lesssim N_k^{-s} \sqrt{\epsilon}.
\]

The proof of this theorem relies on two lemmas below.

Lemma 3.8. Suppose \( \Omega_a = \{k : |a| \in \left[\frac{N}{2M}, 2MN\right]\} \). Under the assumption of Theorem 3.7 we have

\[
W_f(a, b) = |a|^{-s} \left( \sum_{k \in \Omega_a} \alpha_k(b)e^{2\pi i N_k \phi_k(b)} \tilde{w} \left( |a|^{-s} (a - N_k \nabla \phi_k(b)) \right) + O(\epsilon) \right),
\]

when \( N > N_0(M, K, s, \epsilon) \approx \max \{\epsilon^{\frac{2}{s^2 - 1}}, \epsilon^{\frac{1}{s^2}}\} \).
The next lemma estimates $\nabla_b W_f(a, b)$ when $\Omega_a$ is not empty, i.e., the case where $W_f(a, b)$ is non-negligible.

**Lemma 3.9.** Suppose $\Omega_a = \{k : |a| \in [\frac{N_k}{2M}, 2MN_k]\}$ is not empty. Under the assumption of Theorem 3.7, we have

$$\nabla_b W_f(a, b) = 2\pi i |a|^{-s} \left( \sum_{k \in \Omega_a} N_k \nabla \phi_k(b) \alpha_k(b) e^{2\pi i N_k \phi_k(b)} \hat{w} \left( |a|^{-s} (a - N_k \nabla \phi_k(b)) + |a|O(\epsilon) \right) \right),$$

when $N > N_0 (M, K, s, \epsilon)$ such that for any $\epsilon$, we have

$$\nabla \phi_k(b) \simeq \left\{ \epsilon^{\frac{2}{2s-1}}, \epsilon^{\frac{1}{1-s}} \right\}.$$

Similar to Theorem 2.8 in Section 2, Theorem 3.7 is an easy generalization of Theorem 2.3 in [60] for the purpose of a wider class of mother wave packets and a uniform threshold in the discrete SSWPT. We will focus on the robustness of the 2D SSWPT illustrated in the next two theorems. Appendix B is referred to for the proofs of these theorems.

**Theorem 3.10.** Suppose the 2D mother wave packet is of type $(\epsilon, m)$, for any fixed $\epsilon \in (0, 1)$ and any fixed integer $m \geq 0$. Suppose $g(x) = f(x) + e(x)$, where $e(x) \in L^\infty$ is a small error term that satisfies $\|e\|_{L^\infty} \leq \epsilon_1$ for some $\epsilon_1 > 0$. For $p \in (0, \frac{1}{2}]$, let $\delta = \sqrt{\epsilon} + \epsilon_1^{\frac{1}{2} - p}$. Define

$$R_\delta = \{(a, b) : |W_f(a, b)| \geq |a|^{-s}\delta\},$$

$$S_\delta = \{(a, b) : |W_f(a, b)| \geq \delta\},$$

and

$$Z_k = \{(a, b) : |a - N_k \nabla \phi_k(b)| \leq |a|^s\}$$

for $1 \leq k \leq K$. For fixed $M$ and $K$, there exists a constant $N_0 (M, K, s, \epsilon) \simeq \max \left\{ \epsilon^{\frac{2}{2s-1}}, \epsilon^{\frac{1}{1-s}} \right\}$ such that for any $N > N_0$ and $f(x) \in F(M, N, K, s)$ the following statements hold.

(i) $\{Z_k : 1 \leq k \leq K\}$ are disjoint and $S_\delta \subset R_\delta \subset \bigcup_{1 \leq k \leq K} Z_k$;

(ii) For any $(a, b) \in R_\delta \cap Z_k$,

$$\frac{|v_\delta(a, b) - N_k \nabla \phi_k(b)|}{|N_k \nabla \phi_k(b)|} \lesssim \sqrt{\epsilon} + \epsilon_1^p;$$

(iii) For any $(a, b) \in S_\delta \cap Z_k$,

$$\frac{|v_\delta(a, b) - N_k \nabla \phi_k(b)|}{|N_k \nabla \phi_k(b)|} \lesssim N_k^{-s} \left( \sqrt{\epsilon} + \epsilon_1^p \right).$$

Theorem 3.10 shows that the 2D SSWPT is robust to a bounded perturbation. Actually, if the threshold $\delta$ is larger, e.g., $\delta \geq \sqrt{\frac{\epsilon}{\epsilon_0}}$, the relative estimate errors in (ii) and (iii) are bounded by $\sqrt{\epsilon}$ and $\frac{\sqrt{\epsilon}}{N_k}$, respectively. Hence, the local wave vector estimates are better if the wave packet coefficient with the largest magnitude is selected. However, when the perturbation is overwhelming, e.g., the wave packet coefficients of a 2D IMT are below the threshold in (ii), it is difficult to estimate its local wave vector. Next, Theorem 3.11 will illustrate the robustness properties of the 2D SSWPT to a zero mean stationary Gaussian noise.
Theorem 3.11. Suppose the 2D mother wave packet is of type \((\epsilon, m)\), for any fixed \(\epsilon \in (0, 1)\) and any fixed integer \(m \geq \max \left\{ \frac{2(1+s)}{1-s}, \frac{2}{1-s} + 4 \right\}\). Suppose \(g(x) = f(x) + e\), where \(e\) is a zero mean stationary Gaussian process with a spectrum denoted by \(\tilde{c}(\xi)\) and \(\max_{\xi} |\tilde{c}(\xi)| \leq \epsilon^{-1}\). Define \(M_a = \max_{|\xi|<d} \hat{c}(|a|^{s} \xi + a)\). For any \(p \in (0, \frac{1}{2}]\) and \(q > 0\), let \(\delta_a = M_a^{(\frac{1}{2}-p)/(1+q)} + \sqrt{\epsilon}\),

\[
R_{\delta_a} = \{(a, b) : |W_g(a, b)| \geq |a|^{-s} \delta_a\},
\]

and

\[
S_{\delta_a} = \{(a, b) : |W_g(a, b)| \geq \delta_a\},
\]

and

\[
Z_k = \{(a, b) : |a - N_k \nabla_b \phi_k(b)| \leq |a|^s\}
\]

for \(1 \leq k \leq K\). For fixed \(M\) and \(K\), there exists a constant \(N_0 (M, K, s, \epsilon) \simeq \max \left\{ \epsilon^{\frac{2}{2s-1}}, \epsilon^{\frac{s+1}{s+2}} \right\}\)

such that for any \(N > N_0\) and \(f(x) \in F(M, K, N, s)\) the following statements hold.

(i) \(\{Z_k : 1 \leq k \leq K\}\) are disjoint.

(ii) If \((a, b) \in R_{\delta_a}\), then \((a, b) \in \bigcup_{1 \leq k \leq K} Z_k\) with a probability at least

\[
1 - e^{-O(N_k^{-2s}M_a^{-q/(1+q)})} + O \left( \frac{\epsilon}{N_k^m(1-s)} \right).
\]

(iii) If \((a, b) \in S_{\delta_a}\), then \((a, b) \in \bigcup_{1 \leq k \leq K} Z_k\) with a probability at least

\[
1 - e^{-O(M_a^{-q/(1+q)})} + O \left( \frac{\epsilon}{N_k^m(1-s)} \right).
\]

(iv) If \((a, b) \in R_{\delta_a} \cap Z_k\) for some \(k\), then

\[
\frac{|v_g(a, b) - N_k \nabla_b \phi_k(b)|}{|N_k \nabla_b \phi_k(b)|} \lesssim \sqrt{\epsilon} + M_a^{p/(1+q)}
\]

is true with a probability at least

\[
\left( 1 - e^{-O(N_k^{-2s}M_a^{-q/(1+q)})} \right) \left( 1 - e^{-O(N_k^{-2s}M_a^{-q/(1+q)})} \right) \left( 1 - e^{-O(N_k^{-2s}M_a^{-q/(1+q)})} \right)
\]

\[
+ O \left( \frac{\epsilon}{N_k^{-m-4(1-s)-2}} \right) + O \left( \frac{\epsilon}{N_k^{-m-2-(m+2)s}} \right).
\]

(v) If \((a, b) \in S_{\delta_a} \cap Z_k\) for some \(k\), then

\[
\frac{|v_g(a, b) - N_k \nabla_b \phi_k(b)|}{|N_k \nabla_b \phi_k(b)|} \lesssim N_k^{-s} \left( \sqrt{\epsilon} + M_a^{p/(1+q)} \right)
\]

is true with a probability at least

\[
\left( 1 - e^{-O(N_k^{-2s}M_a^{-q/(1+q)})} \right) \left( 1 - e^{-O(N_k^{-2s}M_a^{-q/(1+q)})} \right) \left( 1 - e^{-O(N_k^{-2s}M_a^{-q/(1+q)})} \right)
\]

\[
+ O \left( \frac{\epsilon}{N_k^{-m-4(1-s)-2}} \right) + O \left( \frac{\epsilon}{N_k^{-m-2-(m+2)s}} \right).
\]
We would like to remark that the requirement of \( N_0(M,K,s,\epsilon) \simeq \max \{ \epsilon^{\frac{s-2}{2s-1}}, \epsilon^{\frac{1}{1-s}} \} \)
can be relieved if we impose a weak assumption on the non-linearity of IMTs, as discussed
in the end of Section 2. For example,

\[ |\nabla \alpha_k(x)| \leq \epsilon N_k^s |\nabla \phi_k(x)| \quad \text{and} \quad |\nabla^2 \phi_k(x)| \leq \epsilon N_k^{2s-1} |\nabla \phi_k(x)|. \]

Hence, the theorems introduced in this section is multiscale indeed.

Theorem 3.10 and 3.11 show that the local wave vector estimates by the 2D SSWPT are
robust against bounded perturbation and additive Gaussian random noise, if a threshold
is properly chosen after the wave packet transform. First, the robustness becomes stronger
as \( s \) gets smaller. Second, similar to the 1D case, as we increase the sampling rate of the
signal to make IMTs relatively smoother compared to Gaussian random noise, the SSWPT
can estimate local wave vectors accurately with a high probability.

4 2D synchrosqueezed curvelet transform (SSCT)

In some applications such as wave field separation problems \cite{43, 51} and ground roll re-
moval problems \cite{8, 28, 62} in geophysics, IMTs to be analyzed and decomposed would have
bounded supports in space, sometimes even banded supports. This motivates the design
of the SSCT as a better tool to estimate local wave vectors of banded IMTs with close
propagating directions in \cite{61}. As we shall see in the following theorems, the geometric
scaling of the SSCT is crucial to obtaining an accurate estimate of local wave vectors.

4.1 The definition of the 2D SSCT

The SSCT consists of a generalized curvelet transform with a radial scaling parameter
\( t < 1 \) and an angular scaling parameter \( s \in (1/2, t) \), followed by a synchrosqueezing step
for sharpening the phase space representation. The assumption \( \frac{1}{2} < s < t < 1 \) is essential
for accurate estimates of local wave vectors. Most importantly, \( s < t \) guarantees precise
estimates in the case of banded IMTs. Before discussing the SSCT, we briefly introduce
some notations and definitions of 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 in the frequency
domain.

2. The rotation angle \( \theta \) and the rotation matrix

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

3. The unit vector \( u_\theta = (\cos \theta, \sin \theta)^T \) of the rotation angle \( \theta \) and \( T \) denotes a transpose.

4. \( \theta_\alpha \) represents the argument of a given vector \( \alpha \).
5. \( w(x) \) of \( x \in \mathbb{R}^2 \) denotes the mother curvelet, which belongs to the class of mother wave packets of some type \((\epsilon, m)\) in Section 3 and obeys the admissibility condition: \( \exists 0 < c_1 < c_2 < \infty \) such that
\[
c_1 \leq \int_0^{2\pi} \int_1^{\infty} a^{-(t+s)} |\mathcal{W}(A_a^{-1}R_\theta^{-1}(\xi - a \cdot u_\theta))|^2 ad\theta \leq c_2
\]
for any \(|\xi| \geq 1\).

With the notations above, it is ready to define a family of curvelets through scaling, modulation, and translation as follows, controlled by the geometric parameter \( s \) and \( t \).

**Definition 4.1.** Given geometric scaling parameters \( \frac{1}{2} < s < t < 1 \) and a mother curvelet of type \((\epsilon, m)\), the family of curvelets \( \{w_{a\theta b}(x) \mid a \in [1, \infty), \theta \in [0, 2\pi), b \in \mathbb{R}^2\} \) is constructed as
\[
w_{a\theta b}(x) = a^{\frac{t-s}{2}} e^{2\pi i a(x-b) \cdot u_\theta} w(A_a R_\theta^{-1}(x-b)),
\]
or equivalently, in the frequency domain
\[
\hat{w}_{a\theta b}(\xi) = \hat{w}(A_a^{-1}R_\theta^{-1}(\xi - a \cdot u_\theta)) e^{-2\pi i b \cdot \xi} a^{\frac{t-s}{2}}.
\]

It is clear from the definition that the Fourier transform \( \hat{w}_{a\theta b}(\xi) \) has an ellipse-like essential support \( \{\xi : |A_a^{-1}R_\theta^{-1}(\xi - a \cdot u_\theta)| \leq 1\} \) centered at \( a \cdot u_\theta \) with a major radius \( a^t \) and a minor radius \( a^s \). Meanwhile, \( w_{a\theta b}(x) \) is centered in space at \( b \) with an essential support of length \( O(a^{-s}) \) and width \( O(a^{-s}) \). By this appropriate construction, each curvelet is scaled to have the same \( L^2 \) norm with the mother curvelet \( w(x) \). The general curvelet transform can also be considered as a generalization of the wave packet transform in Section 3 with two different scaling parameters \( s \) and \( t \). This family of functions is quantitatively similar to wavelets when \( s = t = 1 \), wave atoms \cite{25} when \( s = t = \frac{1}{2} \), and curvelets \cite{11, 12, 13} when \( s = \frac{1}{2} \) and \( t = 1 \). In real applications, it is beneficial to adaptively tune \( s \) and \( t \) for better estimates of local wave vectors in complex data structures.

Similar to the classical curvelet transform, the general curvelet transform is defined as follows.

**Definition 4.2.** The general curvelet transform of a function \( f(x) \) is a function
\[
W_f(a, \theta, b) = \langle f, w_{a\theta b} \rangle = \int_{\mathbb{R}^2} f(x) \overline{w_{a\theta b}(x)} dx
\]
for \( a \in [1, \infty), \theta \in [0, 2\pi), b \in \mathbb{R}^2 \).

If the Fourier transform \( \hat{f}(\xi) \) 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 \cite{14}, i.e.,
\[
c_1 \int |f(x)|^2 dx \leq \int |W_f(a, \theta, b)|^2 ad\theta db \leq c_2 \int |f(x)|^2 dx.
\]

**Definition 4.3.** The local wave vector estimation of a function \( f(x) \) at \((a, \theta, b)\) for \( a \in [1, \infty), \theta \in [0, 2\pi), b \in \mathbb{R}^2 \) is
\[
v_f(a, \theta, b) = \begin{cases} \frac{\nabla_b W_f(a, \theta, b)}{2\pi W_f(a, \theta, b)}, & \text{for } W_f(a, \theta, b) \neq 0; \\ (\infty, \infty), & \text{otherwise.} \end{cases}
\]
Since \( v_f(a, \theta, b) \) estimates the local wave vectors accurately, as we shall see, reallocating the coefficients with the same \( v_f \) together would generate a sharpened phase space representation of \( f(x) \). This motivates the design of the synchrosqueezed energy distribution as follows.

**Definition 4.4.** Given \( f(x) \), the synchrosqueezed energy distribution \( T_f(v, b) \) is

\[
T_f(v, b) = \int |W_f(a, \theta, b)|^2 \delta(\Re v_f(a, \theta, b) - v) \, ad\theta
\]

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

For \( f(x) \) with Fourier transform vanishing for \( |\xi| < 1 \), the following norm equivalence holds

\[
\int T_f(v, b) \, dv \, db = \int |W_f(a, \theta, b)|^2 \, ad\theta \, db \approx \|f\|_2
\]

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

### 4.2 Analysis of the 2D SSCT

To model a wave-like component with a band-shape support, we are going to analyze components of the form

\[
f(x) = e^{-\left(\phi(x) - c\right)^2/\sigma^2} \alpha(x) e^{2\pi i N \phi(x)},
\]

where \( \alpha(x) \) is a smooth amplitude function, \( \phi(x) \) a smooth phase function, and \( \sigma \) is a band parameter that controls the width of the signal.

To understand how large the bandwidth should be so as to obtain accurate local wave vector estimates by the SSCT, we assume \( \sigma = \Theta(N^{-\eta}) \) and show that the SSCT gives good estimates when \( \eta < t \) and \( N \) is sufficiently large. 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 \( e^{-\left(\phi(x) - c\right)^2/\sigma^2} \alpha(x) e^{2\pi i N \phi(x)} \) can be almost as narrow as the width of a general curvelet that sharing the same wave number \( O(N) \), when \( N \) is sufficiently large.

**Definition 4.5.** For any \( c \in \mathbb{R}, N > 0 \) and \( M > 0 \), a function \( f(x) = e^{-\left(\phi(x) - c\right)^2/\sigma^2} \alpha(x) e^{2\pi i N \phi(x)} \) is a banded intrinsic mode function of type \((M, N, \eta)\), if \( \alpha(x) \) and \( \phi(x) \) satisfy

\[
\alpha(x) \in C^\infty, \quad |\nabla \alpha(x)| \leq M, \quad 1/M \leq \alpha(x) \leq M, \\
\phi(x) \in C^\infty, \quad 1/M \leq |\nabla \phi(x)| \leq M, \quad |\nabla^2 \phi(x)| \leq M, \\
\text{and} \quad \sigma \geq N^{-\eta}.
\]

**Definition 4.6.** A function \( f(x) \) is a well-separated superposition of type \((M, N, \eta, s, t, K)\) if

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

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

\[
|A_\alpha^{-1} R_\theta^{-1} (a \cdot u_\theta - N_k \nabla \phi_k(b))| \leq 1.
\]

We denote by \( F(M, N, \eta, s, t, K) \) the set of all such functions.
The first theorem in this section demonstrates the accuracy of the local wave vector estimation of the banded IMTs when the given data does not contain noise.

**Theorem 4.7.** Suppose the 2D mother curvelet is of type $(\epsilon, m)$, for any fixed $\epsilon \in (0, 1)$ and any fixed integer $m \geq 0$. For a function $f(x)$, we define

$$R_\epsilon = \left\{ (a, \theta, b) : |W_f(a, \theta, b)| \geq a^{-\frac{\pi + \epsilon}{2}} \sqrt{\epsilon} \right\},$$

$$S_\epsilon = \left\{ (a, \theta, b) : |W_f(a, \theta, b)| \geq \sqrt{\epsilon} \right\},$$

and

$$Z_k = \left\{ (a, \theta, b) : |A_n^1 R_\theta^{-1} (a \cdot u_\theta - N_k \nabla \phi_k(b))| \leq 1 \right\}$$

for $1 \leq k \leq K$. For fixed $M$, $s$, $t$, $\eta$, and $\epsilon$, there exists $N_0 (M, s, t, \eta, \epsilon) \simeq \max \left\{ \epsilon^{\frac{1}{2}}, \epsilon^{\frac{1}{2}}, \epsilon^{\frac{2}{3}} \right\}$ such that for any $N > N_0 (M, s, t, \eta, \epsilon)$ and $f(x) \in F (M, N, \eta, s, t, K)$ the following statements hold.

(i) $\{ Z_k : 1 \leq k \leq K \}$ are disjoint and $S_\epsilon \subset R_\epsilon \subset \bigcup_{1 \leq k \leq K} Z_k$.

(ii) For any $(a, \theta, b) \in R_\epsilon \cap Z_k$,

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

(iii) For any $(a, \theta, b) \in S_\epsilon \cap Z_k$,

$$\frac{|v_f(a, \theta, b) - N_k \nabla \phi_k(b)|}{|N_k \nabla \phi_k(b)|} \lesssim N_k^{-\frac{\pi + \epsilon}{2}} \sqrt{\epsilon}.$$

The proof of Theorem 4.7 relies on two lemmas below to estimate $W_f(a, \theta, b)$ and $\nabla_b W_f(a, \theta, b)$.

**Lemma 4.8.** Suppose

$$\Omega_{\alpha \theta b} = \left\{ k : a \in \left( \frac{N_k}{2M}, 2MN_k \right), |\theta \nabla \phi_k - \theta| < \arcsin \left( \left( \frac{M}{N_k} \right)^{t-s} \right) \right\}.$$

Under the assumption of Theorem 4.7, the following estimation of $W_f(a, \theta, b)$ holds when $N > N_0 (M, s, t, \eta, \epsilon) \simeq \max \left\{ \epsilon^{\frac{1}{2}}, \epsilon^{\frac{1}{2}}, \epsilon^{\frac{2}{3}} \right\}$:

$$W_f(a, \theta, b) = a^{-\frac{\pi + \epsilon}{2}} \left( \sum_{k \in \Omega_{\alpha \theta b}} f_k(b) \tilde{w} \left( A_n^1 R_\theta^{-1} (a \cdot u_\theta - N_k \nabla \phi_k(b)) \right) + O(\epsilon) \right).$$

**Lemma 4.9.** Suppose

$$\Omega_{\alpha \theta b} = \left\{ k : a \in \left( \frac{N_k}{2M}, 2MN_k \right), |\theta \nabla \phi_k(b) - \theta| < \arcsin \left( \left( \frac{M}{N_k} \right)^{t-s} \right) \right\}$$

is not empty. Under the assumption of Theorem 4.7 there exists a constant $N_0 (M, s, t, \eta, \epsilon) \simeq \max \left\{ \epsilon^{\frac{1}{2}}, \epsilon^{\frac{1}{2}}, \epsilon^{\frac{2}{3}} \right\}$ such that if $N > N_0 (M, s, t, \eta, \epsilon)$, then we have

$$\nabla_b W_f(a, \theta, b) = a^{-\frac{\pi + \epsilon}{2}} \left( 2\pi i \sum_{k \in \Omega_{\alpha \theta b}} \nabla \phi_k(b) f_k(b) \tilde{w} \left( A_n^1 R_\theta^{-1} (a \cdot u_\theta - N_k \nabla \phi(b)) \right) + aO(\epsilon) \right).$$
Similar to Theorem 2.8 in Section 2 and Theorem 3.7 in Section 3, Theorem 4.7 is an easy generalization of Theorem 2.3 in [61], for the purpose of a wider class of mother wave packets and a uniform threshold in the discrete SSCT. We skip its proof and focus on the robustness of the 2D SSCT illustrated in the next two theorems. We refer to their proofs in Appendix C.

**Theorem 4.10.** Suppose the 2D mother wave packet is of type \((\epsilon, m)\), for any fixed \(\epsilon \in (0, 1)\) and any fixed integer \(m \geq 0\). Suppose \(g(x) = f(x) + e(x)\), where \(e(x) \in L^\infty\) is a small error term that satisfies \(\|e\|_{L^\infty} \leq \epsilon_1\) for some \(\epsilon_1 \geq 0\). For any \(p \in (0, \frac{1}{2}]\), let \(\delta = \sqrt{\epsilon + \epsilon_1^p}\). Define

\[
R_\delta = \left\{ (a, \theta, b) : |W_f(a, \theta, b)| \geq a^{-\frac{m+1}{2}} \delta \right\},
\]

and

\[
S_\delta = \left\{ (a, \theta, b) : |W_f(a, \theta, b)| \geq \delta \right\},
\]

and

\[
Z_k = \left\{ (a, \theta, b) : |A_a^{-1} R_\theta^{-1} (a \cdot u_\theta - N_k \nabla \phi_k(b))| \leq 1 \right\}
\]

for \(1 \leq k \leq K\). For fixed \(M, s, t, \eta, \) and \(\epsilon\), there exists \(N_0(M, s, t, \eta, \epsilon) \simeq \max \left\{ \epsilon_1^{\frac{1}{2}}, \epsilon_1^{\frac{1}{4}}, \epsilon_1^{\frac{3}{2}} \right\}\) such that for any \(N > N_0(M, s, t, \eta, \epsilon)\) and \(f(x) \in F(M, N, s, t, K)\) the following statements hold.

(i) \(\{Z_k : 1 \leq k \leq K\}\) are disjoint and \(S_\delta \subset R_\delta \subset \bigcup_{1 \leq k \leq K} Z_k\).

(ii) For any \((a, \theta, b) \in R_\delta \cap Z_k\),

\[
\frac{|v_g(a, \theta, b) - N_k \nabla \phi_k(b)|}{|N_k \nabla \phi_k(b)|} \lesssim \sqrt{\epsilon + \epsilon_1^p}.
\]

(iii) For any \((a, \theta, b) \in S_\delta \cap Z_k\),

\[
\frac{|v_g(a, \theta, b) - N_k \nabla \phi_k(b)|}{|N_k \nabla \phi_k(b)|} \lesssim N_k^{-\frac{m+1}{2}}(\sqrt{\epsilon + \epsilon_1^p}).
\]

Theorem 4.10 justifies the robustness of the 2D SSCT to a bounded perturbation. Next theorem below will illustrate its robustness against additive zero mean stationary Gaussian noise.

**Theorem 4.11.** Suppose the 2D mother wave packet is of type \((\epsilon, m)\), for any fixed \(\epsilon \in (0, 1)\) and any fixed integer \(m \geq \max \left\{ \frac{2(1+q)}{1+q}, \frac{q}{2} + 4 \right\}\). Suppose \(g(x) = f(x) + e,\) where \(e\) is a zero mean stationary Gaussian process with a spectrum denoted by \(\hat{e}(\xi)\) and \(\max_{\xi} |\hat{e}(\xi)| \leq \epsilon^{-1}\). Define \(M_a = \max_{|\xi| < d} \hat{e}(R_a A_\xi - a \cdot u_0)\). For any \(p \in (0, \frac{1}{2}]\) and \(q > 0\), let \(\delta_a = M_a^{(1-p)/(1+q)} + \sqrt{\epsilon},\)

\[
R_{\delta_a} = \left\{ (a, \theta, b) : |W_f(a, \theta, b)| \geq a^{-\frac{m+1}{2}} \delta_a \right\},
\]

and

\[
S_{\delta_a} = \left\{ (a, \theta, b) : |W_f(a, \theta, b)| \geq \delta_a \right\},
\]

and

\[
Z_k = \left\{ (a, \theta, b) : |A_a^{-1} R_\theta^{-1} (a \cdot u_\theta - N_k \nabla \phi_k(b))| \leq 1 \right\}
\]

for \(1 \leq k \leq K\). For fixed \(M, s, t, \eta, \) and \(\epsilon\), there exists \(N_0(M, s, t, \eta, \epsilon) \simeq \max \left\{ \epsilon_1^{\frac{1}{2}}, \epsilon_1^{\frac{1}{4}}, \epsilon_1^{\frac{3}{2}} \right\}\) such that for any \(N > N_0(M, s, t, \eta, \epsilon)\) and \(f(x) \in F(M, N, s, t, K)\) the following statements hold.

22
(i) \( \{Z_k : 1 \leq k \leq K\} \) are disjoint.

(ii) If \((a, \theta, b) \in R_{\delta_a}\), then \((a, \theta, b) \in \bigcup_{1 \leq k \leq K} Z_k\) with a probability at least

\[
1 - e^{-O(N_k^{-(s+\ell)}M_a^{-q/(1+q)})} + O\left(\frac{\epsilon}{N_k^{m(1-\ell)}}\right).
\]

(iii) If \((a, \theta, b) \in S_{\delta_a}\), then \((a, \theta, b) \in \bigcup_{1 \leq k \leq K} Z_k\) with a probability at least

\[
1 - e^{-M_a^{-q/(1+q)}} + O\left(\frac{\epsilon}{N_k^{m(1-\ell)}}\right).
\]

(iv) If \((a, \theta, b) \in R_{\delta_a} \cap Z_k\) for some \(k\), then

\[
\frac{|v_g(a, \theta, b) - N_k \nabla \phi_k(b)|}{|N_k \nabla \phi_k(b)|} \lesssim \sqrt{\epsilon} + M_a^{p/(1+q)}
\]

is true with a probability at least

\[
\left(1 - e^{-O(N_k^{2s-3q}M_a^{-q/(1+q)})}\right)\left(1 - e^{-O(N_k^{(3s+1)}M_a^{-q/(1+q)})}\right)\left(1 - e^{-O(N_k^{2s-1}M_a^{-q/(1+q)})}\right)
\]

\[+O\left(\frac{\epsilon}{N_k^{(m-1)(1-\ell)-2}}\right) + O\left(\frac{\epsilon}{N_k^{m-2-2m-2\ell}}\right).
\]

(v) If \((a, \theta, b) \in S_{\delta_a} \cap Z_k\) for some \(k\), then

\[
\frac{|v_g(a, \theta, b) - N_k \nabla \phi_k(b)|}{|N_k \nabla \phi_k(b)|} \lesssim N_k^{-(s+\ell)/2}\left(\sqrt{\epsilon} + M_a^{p/(1+q)}\right)
\]

is true with a probability at least

\[
\left(1 - e^{-O(N_k^{2s-2q}M_a^{-q/(1+q)})}\right)\left(1 - e^{-O(N_k^{2s}M_a^{-q/(1+q)})}\right)\left(1 - e^{-O(N_k^{2s-2}M_a^{-q/(1+q)})}\right)
\]

\[+O\left(\frac{\epsilon}{N_k^{(m-1)(1-\ell)-2}}\right) + O\left(\frac{\epsilon}{N_k^{m-2-2m-2\ell}}\right).
\]

Similar to the discussion in previous sections, the requirement for \(N_0 = \max\left\{\frac{1}{1-\ell}, \frac{2}{1-\eta}, \frac{2}{2s-1}\right\}\) can be further optimized if we impose extra conditions on the non-linearity of IMTs and consider the polynomial decaying of the mother curvelet in the frequency domain.

Up to now, we have proved that the SSCT is able to accurately and robustly estimate the local wave vectors of banded IMTs, if their essential supports can be modeled by a Gaussian function with an essential support larger than the width of a curvelet sharing the same order of oscillation. Before closing this section, we would like to emphasize a new understanding of the results obtained in those theorems in this section: if the amplitude function of an IMT has a vanishing boundary, then the vanishing rate can be almost as fast as the oscillation. If an IMT has a sharp boundary, the estimates provided by synchrosqueezed transforms are reliable at the locations almost \(O(1)\) wave lengths away from the boundary (see Figure [right]). As a corollary in 1D cases, a similar conclusion is true as illustrated in Figure [left].
5 Implementation and numerical results

In this section, we provide numerical examples to demonstrate the conclusions of those theorems in this paper and explain several ideas to obtain reliable instantaneous frequency or local wave vector information from extremely noisy data. We have developed SynLab, a collection of MATLAB implementation for the SSWPT and the SSCT that we have made publicly available at: [www.stanford.edu/~haizhao/codes](http://www.stanford.edu/~haizhao/codes). All numerical examples presented in this paper can be found in this toolbox.

In all examples, we assume the given data \( g(x) = f(x) + e(x) \) is defined in \([0, 1]^n\), where \( f(x) \) is the target signal, \( e(x) \) is Gaussian white noise with a distribution \( \sigma^2 \mathcal{N}(0, 1) \), and \( n \) is the number of dimensions. We would only focus on testing the robust performance of the SSWPT, since the SSCT has similar properties. Detailed implementations of these transforms have been discussed in [58, 59, 60, 61]. As we have seen in the theorems, a proper thresholding adaptive to the noise level after the wave packet/general curvelet transform is important to obtain an accurate instantaneous frequency/local wave vector estimate. We refer to [26, 27] for estimating noise level and [47] for designing thresholds for the SSWT. The generalization of these techniques for the SSWPT and the SSCT is straightforward.

Our main purpose in this section is to show the robustness of synchrosqueezed transforms with various scaling parameters. We compare the performance of the SSWPT with \( s = 1/2 + k/8 \), where \( k = 1, 2 \) and \( 3 \), in both noiseless cases and highly noisy cases. For the purpose of showing how the synchrosqueezing process is affected by heavy noise, we are using a small uniform threshold \( \delta = 10^{-2} \) (rather than a threshold adaptive to noise level) and setting \( \sigma^2 \) such that the noise is overwhelming the original signal in all of our synthetic toy models. The accuracy tolerance \( \epsilon = 10^{-4} \).

5.1 Robustness tests for 1D SST

We start from testing the 1D SSWPT. In some real applications, IMTs are only supported in a bounded domain or they have sharp changes in instantaneous frequencies. Hence, we would like to test a benchmark signal in which there is a component with a bounded support and an oscillatory instantaneous frequency, and a component with an exponential instantaneous frequency (see Figure 2). Of a special interest to test the performance of synchrosqueezed transforms for impulsive waves, a wavelet component is added in this signal.
Figure 2: Left: A 1D synthetic benchmark signal. It is normalized using $L^\infty$ norm. Right: A noisy signal generated with Gaussian white noise $0.75\mathcal{N}(0, 1)$.

at $x = 0.2$. The synthetic benchmark signal is generated using the example functions:

\[
\begin{align*}
    f_1(x) &= 0.6 \cos(700\pi x); \\
    f_2(x) &= 0.8 \cos(300\pi x); \\
    f_3(x) &= 0.7 \cos(1300\pi x + 5 \sin(20\pi x)); \\
    f_4(x) &= \sin \left( \frac{80\pi 100^{5/4}}{\ln(100)} \right); \\
    f_5(x) &= 3e^{-50t^2} \cos(50t).
\end{align*}
\]

The sampling rate of this signal is 8192 Hz and the instantaneous frequency range is $150 - 1600$ Hz. The Gaussian white noise in this example is $0.75\mathcal{N}(0, 1)$. To make a fair comparison, we have tuned the size of the essential support of mother wave packets to obtain a good result for each kind of synchrosqueezed transforms.

Although we are not aware of the optimal value of the scaling parameter $s$, it is clear from Theorem 2.12 and 2.13 that the synchrosqueezed transform with a smaller $s$ is more robust. As shown in the second and the third rows in Figure 3, in the noisy cases, the synchrosqueezed energy distribution with $s = 0.625$ (in the first column) is better than the one with $s = 0.75$ (in the second column), which is better than the one with $s = 0.875$ (in the last column). This agrees with the conclusion in Theorem 2.12 and 2.13 that a smaller $s$ results in a higher probability to obtain a better instantaneous frequency estimate.

Another key point of Theorem 2.12 and 2.13 is that a wave packet coefficient with a larger magnitude gives a better instantaneous frequency estimate with a higher probability. A highly redundant wave packet transform with a denser translation grid in space and scaling grid in frequency would have wave packets better fitting local oscillation of IMTs. In another word, there would be more coefficients with larger magnitudes. The resulting synchrosqueezed energy distribution has higher non-zero energy concentrating around instantaneous frequencies. This is also validated in Figure 3. The synchrosqueezed energy distributions in the third row are obtained by a SSWPT with a 16 times denser grid in frequency than the grid used in the second row. Hence, instantaneous frequencies are much more visible if a SSWPT with a higher redundancy is applied.

\footnote{Prepared by Professor Mirko van der Baan and available at [45, 50].}
It is also interesting to observe that the synchrosqueezed transform with a smaller $s$ is better to capture the component boundaries, e.g. at $x = 0.39, 0.59$ and $0.77$ and is more robust to an impulsive perturbation (see Figure 2 and 3 at $x = 0.2$ and an example of $\alpha$ stable noise in Figure 4 and 5). Boundaries and impulse perturbation would produce frequency aliasing. The SSWPT with a smaller $s$ has wave packets with a smaller support in frequency and a larger support in space. Hence, it is more robust to frequency aliasing in the sense that the influence of impulsive perturbation is smoothed out and the synchrosqueezed energy of the target components might not get dispersed when it meets the frequency aliasing, as shown in Figure 5.

However, if $s$ is small, the instantaneous frequency estimate might be smoothed out and it is difficult to observe detailed information of instantaneous frequencies. As shown in the first row of Figure 3, when the input signal is noiseless, the synchrosqueezed transforms with $s = 0.75$ and $0.875$ have better accuracy than the one with $s = 0.625$. In short, it is important to have tunable scaling parameters to design problem dependent synchrosqueezed transforms, which has been implemented in the SynLab toolbox.

5.2 Robustness tests for 2D SSTs

We now explore the performance of the 2D SSWPT using a single IMT in Figure 6. The function

$$f(x) = e^{2\pi i(60(x_1 + 0.05 \sin(2\pi x_1))+60(x_2 + 0.05 \sin(2\pi x_2)))}$$

(3)

is uniformly sampled in $[0, 1]^2$ with a sampling rate 512 Hz and is disturbed by additive Gaussian white noise $5\mathcal{N}(0, 1)$. The 2D SSWPTs with $s = 0.625, 0.75$ and $0.875$ are applied to this noisy example and their results are shown in Figure 7. Since the synchrosqueezed energy distribution $T_f(x_1, x_2, k_1, k_2)$ of an image is a function in $\mathbb{R}^4$, we fix $x_2 = 0$, stack the results in $k_2$, and visualize $\int_{\mathbb{R}} T_f(x_1, 0, k_1, k_2)dk_2$.

The results in Figure 7 again validate the theoretical conclusion in Theorem 3.11 that a smaller scaling parameter $s$ and a higher redundancy yield to a better robustness. A new idea here to achieve a better robustness is to design adaptive synchrosqueezed transforms tracing the possible frequency band of IMTs. A band-limited synchrosqueezed transform is designed in [59] for an efficient tool to analyze atomic crystal images. Numerical experiments will show that this method is strongly robust to noise. This inspires the idea of adaptive synchrosqueezed transforms above. We will do a simple experiment to justify this idea. In this experiment, we apply the band-limited SSWPT to the same 2D noisy image and present the results in the last row of Figure 7. Comparing to the results in the second row of Figure 7, the band-limited SSWPT clearly outperforms the original SSWPT.

5.3 Component test

We will present the last example to validate the results of previous theorems. Suppose we look at a region in the time-frequency or phase space domain and we know there might be only one IMT in this region. This assumption is reasonable because after synchrosqueezed transforms people might be interested in the synchrosqueezed energy in a particular region: is this corresponding to a component or just heavy noise? A straightforward solution is that, at each time or space grid point, we only reassign those coefficients with the largest magnitude. By Theorem 3.11, if there is an IMT, we can obtain a sketch of its instantaneous frequency or local wave vector with a high probability. If there is only noise, we would obtain random reassigned energy with a high probability. Using this idea, we apply the
Figure 3: Synchrosqueezed energy distributions with $s = 0.625$ (left column), $s = 0.75$ (middle column) and $s = 0.875$ (right column). In the first row, we apply the SSWPT to clean data. In the second row, the SSWPT with a smaller redundancy is applied to the noisy data with $0.75N(0, 1)$ noise in Figure 2. In the last row, a highly redundant SSWPT is applied to the same noisy data.

band-limited SSWPT with $s = 0.625$ and 10 times redundancy to a noisy version of the image in Figure 6 left. From left to right, Figure 8 shows the results of a noisy image (3) with $5N(0, 1)$ noise, a noisy image (3) with $10N(0, 1)$ noise, and an image with only noise, respectively. A reliable sketch of the local wave vector is still visible even if the input image is highly noisy.

5.4 Real examples

In the last part of this section, we introduce a newly developed atomic crystal image analysis method based on synchrosqueezed transforms [59] to demonstrate the robustness of synchrosqueezed transforms in real applications. In materials science, information hidden in an atomic crystal image, e.g., grain boundaries, isolated defects, deformation field of each grain, is important for better understanding the properties of materials. As seen in Figure 9 (a), an atomic crystal image can be considered as an assemble of several general IMTs [59].
Figure 4: Left: A 1D synthetic benchmark signal. Right: A signal contaminated by an α stable random noise with parameters α = 1, dispersion= 0.9, δ = 1, N = 8192. The noise is rescaled to have a $15 L^\infty$-norm by dividing a constant factor.

Figure 5: Synchrosqueezed energy distributions with $s = 0.625$ (left), $s = 0.75$ (middle) and $s = 0.875$ (right) using highly redundant SSWPTs. The synchrosqueezed energy with a smaller $s$ is smoother and the influence of impulsive noise is weaker.

Figure 6: Left: A 2D noiseless IMT. Right: A noisy IMT generated with Gaussian white noise $5N(0,1)$. 
Figure 7: Stacked synchrosqueezed energy distribution \( \int_{\mathbb{R}} T_f(x_1, 0, k_1, k_2) dk_2 \) of the noisy 2D signal in Figure 6. From left to right, \( s = 0.625, 0.75 \) and 0.875. From top to bottom: standard redundancy, 10 times redundancy and 10 times redundancy with a SSWPT restricted to a frequency band from 20 to 120 Hz.

where a general IMT is a superposition of a few IMTs with similar local wave vectors, e.g., with local wave vectors \( \{(n_k \partial_{b_1} \phi(b), m_k \partial_{b_2} \phi(b))\} \) for some \( \phi(b) \) and a few pairs \( (n_k, m_k) \). The method in [59] automatically determines a frequency band of the input image and applies a band-limited SSWPT to estimate the synchrosqueezed energy of each IMT. The location of the essential synchrosqueezed energy reveals grain boundaries, isolated defects and deformation fields (denoted by \( \nabla F(x_1, x_2) \in \mathbb{R}^{2 \times 2} \)). Integrating \( \nabla F(x_1, x_2) \) around a defect region can estimate the Burgers vector corresponding to the defect region. The distortion volume of \( \nabla F(x_1, x_2) \), i.e., \( \det(\nabla F(x_1, x_2)) - 1 \) can reflect the strain stress on the grains (e.g. see Figure 9 (c)).

We apply the method in [59] to a phase field crystal image (Figure 9 (a)) and show the detected grain boundaries and isolated defects in Figure 9 (b), and the distortion volume in Figure 9 (c). To demonstrate the robustness, we generate additive Gaussian white noise with a distribution 0.5\( \mathcal{N}(0, 1) \) and 1.4\( \mathcal{N}(0, 1) \), respectively and present the noisy results in the second and the third rows of Figure 9. In the results of extremely noisy cases, even
Figure 8: Top row: The synchrosqueezed energy distribution of the highly redundant band-limited SSWPT with a frequency band 20 to 120 Hz. Bottom row: reassigned wave packet coefficients with the largest magnitude at a space location. Left column: $5\mathcal{N}(0,1)$ noise. Middle column: $10\mathcal{N}(0,1)$ noise. Right column: noise only.

If no crystal structure visible by human eyes, the SSWPT method is still able to reveal grain boundaries and isolated defects with a reasonable accuracy. The distortion volume in Figure 9 (f) and (i) still roughly reflects the strain stress encoded by color.

6 Conclusion

In theory, this paper has proved the multiscale robustness of a wide range of multidimensional multiscale synchrosqueezed transforms, considering zero mean stationary Gaussian random noise and small perturbation. These transforms can have compact supports in the time/space domain if they meet certain smoothness condition. The multiscale robustness becomes stronger as the corresponding scaling parameter becomes smaller. In the presence of wave-like components with bounded supports, it is also pointed out that instantaneous frequency/local wave vector estimates stay accurate until only a few wave lengths away from the boundary. Numerically, this paper has presented several approaches to improve the performance of these synchrosqueezed transforms under heavy noise. A software package SynLab implements the algorithms proposed in this paper in MATLAB and is available at [www.stanford.edu/~haizhao/codes](http://www.stanford.edu/~haizhao/codes).

This work is an initial step for theoretical analysis of compactly supported synchrosqueezed transforms. The probability results can be further optimized if we take good advantage of the geometry of high dimensional Gaussian distributions. Since some synchrosqueezing methods based on extra estimates of the non-linearity of instantaneous frequencies have obtained better results than the original synchrosqueezing operator, it would be interesting to analyze their robustness theoretically.
Figure 9: Atomic crystal image analysis using 2D synchrosqueezed transforms. First row: Results of noiseless data. Second row: Results of noisy data with Gaussian white noise $0.5\mathcal{N}(0, 1)$. Third row: Results of noisy data with Gaussian white noise $1.4\mathcal{N}(0, 1)$. First column: input images. Second column: detected grain boundaries and isolated defects. Third column: distortion volume. Zoomed-in images show that our method can still identify isolated defects even if noise is heavy.

Finally, it would be significant to develop a statistical theory of multi wave-like component detections. In previous literature, existing methods can only detect a single component with known starting and ending points [15][19]. In practice, real data may contain unknown number of components with unknown boundaries, and even with crossover instantaneous frequencies or local wave vectors. Hence, an automatically statistical detecting tool is more realistic and valuable.

Acknowledgments. H.Y. and L.Y. were partially supported by the National Science Foundation under award DMS-0846501 and the KAUST Global Collaborative Research.
Grant, 2011-2014. We are grateful to Jean-Baptiste Tary and Mirko van der Baan for the discussion of benchmark signals. We also thank Charles K. Chui, Jianfeng Lu and Segey Fomel for the discussion of the implementation and the application of synchrosqueezed transforms.

A Proofs of the theorems in Section 2

A.1 Proof of Theorem 2.8

Proof. Part (i) follows from the well-separation condition.

Let \((a, b)\) be a point in \(R_\epsilon\). From Lemma 2.9, we see that for \(\epsilon\) sufficiently small, if \(N \geq N_0(M, K, s, \epsilon)\), then there exists \(k\) between 1 and \(K\) such that \(\hat{w}(a^{-s}(a - N_k\phi_k'(b)))\) is non-zero. From the definition of \(\hat{w}(\xi)\), we see that this implies \((a, b) \in Z_k\). Hence \(R_\epsilon \subset \bigcup_{k=1}^{K} Z_k\). It’s obvious \(S_\epsilon \subset R_\epsilon\). Hence, (i) is true.

To show (ii), let us recall that \(v_f(a, b)\) is defined as

\[
v_f(a, b) = \frac{\partial_b W_f(a, b)}{2\pi i W_f(a, b)}
\]

for \(W_f(a, b) \neq 0\). If \((a, b) \in R_\epsilon \cap Z_k\), then

\[
W_f(a, b) = a^{-s/2} \left( \alpha_k(b) e^{2\pi i N_k\phi_k(b)} \hat{w}(a^{-s}(a - N_k\phi_k(b))) + O(\epsilon) \right)
\]

and

\[
\partial_b W_f(a, b) = a^{-s/2} \left( 2\pi i N_k\phi_k'(b) \alpha_k(b) e^{2\pi i N_k\phi_k(b)} \hat{w}(a^{-s}(a - N_k\phi_k(b))) + O(\epsilon) \right)
\]

as the other terms are of \(O(\epsilon)\) since \(\{Z_k\}\) are disjoint. Hence

\[
v_f(a, b) = \frac{N_k\phi_k'(b) \left( \alpha_k(b) e^{2\pi i N_k\phi_k(b)} \hat{w}(a^{-s}(a - N_k\phi_k(b))) + O(\epsilon) \right)}{\left( \alpha_k(b) e^{2\pi i N_k\phi_k(b)} \hat{w}(a^{-s}(a - N_k\phi_k(b))) + O(\epsilon) \right)}.
\]

Let us denote the term \(\alpha_k(b) e^{2\pi i N_k\phi_k(b)} \hat{w}(a^{-s}(a - N_k\phi_k(b)))\) by \(g\). Then

\[
v_f(a, b) = \frac{N_k\phi_k'(b) (g + O(\epsilon))}{g + O(\epsilon)}.
\]

Since \(|W_f(a, b)| \geq a^{-s/2}\sqrt{\epsilon}\) for \((a, b) \in R_\epsilon\), \(|g| \gtrsim \sqrt{\epsilon}\), and therefore

\[
\frac{|v_f(a, b) - N_k\phi_k'(b)|}{|N_k\phi_k'(b)|} \lesssim \frac{O(\epsilon)}{|g + O(\epsilon)|} \lesssim \sqrt{\epsilon}.
\]

Similarly, if \((a, b) \in S_\epsilon \cap Z_k\), then

\[
\frac{|v_f(a, b) - N_k\phi_k'(b)|}{|N_k\phi_k'(b)|} \lesssim \frac{O(\epsilon)}{|g + O(\epsilon)|} \lesssim \frac{\sqrt{\epsilon}}{N_k^{s/2}},
\]

since the new \(|g| \gtrsim N_k^{s/2}\sqrt{\epsilon}\) in this case. \(\square\)
A.2 Proof of Theorem 2.11

Proof. We estimate several inequalities first. By the definition of the wave packet transform of \( e(x) \), we have

\[
W_e(a, b) = \int_{\mathbb{R}} e(x) a^{s/2} w(s(x - b)) e^{-2\pi i (x-b)a} dx \\
= a^{-s/2} \int_{\mathbb{R}} e(a^{-s} y + b) w(y) e^{-2\pi i a^{1-s} y} dy.
\]

Hence,

\[
|W_e(a, b)| \lesssim \|e\|_{L^\infty} a^{-s/2} \leq \sqrt{\epsilon_1} a^{-s/2}.
\] (4)

Applying the same strategy, we have

\[
\partial_b W_e(a, b) = 2\pi i a^{1+s/2} \int_{\mathbb{R}} e(x) w'(s(x-b)) e^{-2\pi i (x-b)a} dx - a^{3s/2} \int_{\mathbb{R}} e(x) w'(s(x-b)) e^{-2\pi i (x-b)a} dx
\]

\[
= 2\pi i a^{-s/2} \int_{\mathbb{R}} e(a^{-s} y + b) w(y) e^{-2\pi ia^{1-s} y} dy - a^{s/2} \int_{\mathbb{R}} e(a^{-s} y + b) w'(y) e^{-2\pi ia^{1-s} y} dy.
\]

Hence,

\[
|\partial_b W_e(a, b)| \lesssim \sqrt{\epsilon_1} \left( a^{1-s/2} + a^{s/2} \right).
\] (5)

if \((a, b) \in R_\delta\), then \(|W_f(a, b)| \geq a^{-s/2} \delta\). Together with Equation (4), it holds that

\[
|W_f(a, b)| \geq |W_f(a, b)| - |W_e(a, b)| \geq a^{-s/2} (\delta - \sqrt{\epsilon_1}) \geq a^{-s/2} \sqrt{\epsilon}.
\] (6)

Hence, \(S_\delta \subset R_\delta \subset R_\epsilon\), where \(R_{\epsilon}\) is defined in Theorem 2.8 and is a subset of \( \bigcup_{1 \leq k \leq K} Z_k \). So, (i) is true by Theorem 2.8.

Now, let us prove (ii). Since \(R_\delta \subset R_\epsilon\), \((a, b) \in R_\delta \cap Z_k\) implies \((a, b) \in R_\epsilon \cap Z_k\). Hence, by Theorem 2.8, it holds that

\[
\frac{|v_f(a, b) - N_k \phi_k'(b)|}{|N_k \phi_k'(b)|} \lesssim \sqrt{\epsilon},
\] (7)

when \(N\) is larger than a constant \(N_0\) \((M, K, s, \epsilon) \simeq \max \left\{ \epsilon^{\frac{1}{s-1}}, \epsilon^{\frac{1}{s+1}} \right\}\). Notice that \((a, b) \in Z_k\) implies \(a \simeq N_k\). Hence, by Equation (4) to (7),

\[
\frac{|v_g(a, b) - N_k \phi_k'(b)|}{|N_k \phi_k'(b)|} \lesssim \sqrt{\epsilon} + \sqrt{\epsilon} + \frac{\sqrt{\epsilon_1}}{\delta} + \frac{\sqrt{\epsilon_1} (a^{1-s/2} + a^{s/2})}{\sqrt{\epsilon}}
\]

\[
\lesssim \sqrt{\epsilon} + \frac{\sqrt{\epsilon_1}}{\delta} = \sqrt{\epsilon} + \epsilon_1^p,
\]
when $N > N_0$. Hence, (ii) is proved. The proof of (iii) is similar. If $(a, b) \in S_\delta \cap Z_k$, then
\[
\frac{|v_f(a, b) - N_k\phi_k'(b)|}{|N_k\phi_k'(b)|} \leq \frac{|v_f(a, b) - N_k\phi_k'(b)|}{|N_k\phi_k'(b)|} + \frac{|\partial_b W_f(a, b) - \partial_b W_e(a, b)|}{|2\pi i W_f(a, b)|}.
\]
\[
\lesssim \sqrt{e} \frac{\sqrt{\epsilon_1} + \sqrt{\epsilon_1} + \sqrt{\epsilon_1} + \sqrt{\epsilon_1}}{N_k\delta}
\]
when $N > N_0$.

\[\]
If we define

\[ V = \left( \begin{array}{c} \| \tilde{w} \|^2 \\ \langle 2\pi i \xi \tilde{w}_{ab}, \tilde{w}_{ab} \rangle \\ \langle 2\pi i \xi \tilde{w}_{ab}, 2\pi i \xi \tilde{w}_{ab} \rangle \end{array} \right), \]

then \( \epsilon_1^{1+q} V \) is the covariance matrix of \((W_e(a, b), \partial_b W_e(a, b))\) and its distribution is described by the joint probability density

\[ e^{-\epsilon_1^{1+q} z^* V^{-1} z} \]

\[ \pi^{-2} \epsilon_1^{2(q+1)} \det V, \]

where \( z = (z_1, z_2)^T \), \( T \) and * denote the transpose operator and conjugate transpose operator. \( V \) is an invertible and self-adjoint matrix, since \( W_e(a, b) \) and \( \partial_b W_e(a, b) \) are linearly independent. Hence, there exist a diagonal matrix \( D \) and a unitary matrix \( U \) such that \( V^{-1} = U^* D U \).

Part (i) is true by previous theorems. Define the following events

\[ G_1 = \left\{ |W_e(a, b)| < a^{-s/2} \sqrt{\epsilon_1} \right\}, \]

\[ G_2 = \left\{ |W_e(a, b)| < \sqrt{\epsilon_1} \right\}, \]

\[ G_3 = \left\{ |\partial_b W_e(a, b)| < \sqrt{\epsilon_1} \left( a^{1-s/2} + a^{s/2} \right) \right\}, \]

\[ H_k = \left\{ \left| \frac{\epsilon g(a, b) - N_k \phi_k(b)}{|N_k \phi_k(b)} \right| \leq \sqrt{\epsilon} + \epsilon_1^p \right\}, \]

and

\[ J_k = \left\{ \left| \frac{\epsilon g(a, b) - N_k \phi_k(b)}{|N_k \phi_k(b)} \right| \leq \sqrt{\epsilon} + \epsilon_1^p \left\{ \frac{1}{N_k^{s/2}} \right\} \right\}, \]

for \( 1 \leq k \leq K \). To conclude Part (ii) to (v), we need to estimate the probability \( P(G_1) \), \( P(G_2) \), \( P(G_1 \cap G_3) \), \( P(G_2 \cap G_3) \), \( P(H_k) \) and \( P(J_k) \). By the calculations above, we have

\[ P(G_1) = \int_{|z_1| < a^{-s/2} \sqrt{\epsilon_1}} e^{-\epsilon_1^{1+q} |z_1|^2 \|w\|^2} \, dz_1 \]

\[ = \frac{2}{\epsilon_1^{1+q} \|w\|^2} \int_0^{a^{-s/2} \sqrt{\epsilon_1}} r e^{-\epsilon_1^{1+q} r^2 \|w\|^2} \, dr \]

\[ = \int_0^{a^{-s/2} \epsilon_1^{-q/2} \|w\|^{-1}} 2 r e^{-r^2} \, dr \]

\[ = 1 - e^{-a^{-s} \epsilon_1^{-q} \|w\|^{-2}}, \]

and similarly

\[ P(G_2) = \int_{|z_1| < \sqrt{\epsilon_1}} e^{-\epsilon_1^{1+q} |z_1|^2 \|w\|^2} \, dz_1 = 1 - e^{-\epsilon_1^{-q} \|w\|^{-2}}. \]

We are ready to summarize and conclude (ii) and (iii). If \((a, b) \in R_\delta\), then

\[ |W_e(a, b) + W_f(a, b)| \geq a^{-s/2} \left( \epsilon_1^{1/2-p} + \sqrt{\epsilon} \right). \]

(8)

If \((a, b) \notin \bigcup_{1 \leq k \leq K} Z_k\), then by Lemma 2.10,

\[ |W_f(a, b)| \leq a^{-s/2} \epsilon. \]

(9)
Equation (8) and (9) lead to $|W_e(a,b)| \geq a^{-s/2}/\sqrt{c_1}$. Hence,

$$P \left( (a,b) \notin \bigcup_{1 \leq k \leq K} Z_k \right) \leq P \left( |W_e(a,b)| \geq a^{-s/2}/\sqrt{c_1} \right) = 1 - P(G_1).$$

This means that if $(a,b) \in R_\delta$, then $(a,b) \in \bigcup_{1 \leq k \leq K} Z_k$ with a probability at least $P(G_1) = 1 - e^{-a^{-s/2}c_1^{-q}/w^{-2}} = 1 - e^{-O(N_k^{-s/2}c_1^{-q})}$, since $a \simeq N_k$ if $(a,b) \in Z_k$. So, (ii) is true. A similar argument applied to $(a,b) \in S_\delta$ shows that $(a,b) \in \bigcup_{1 \leq k \leq K} Z_k$ with a probability at least $P(G_2) = 1 - e^{-c_1^{-q}/w^{-2}}$. Hence, (iii) is proved.

Note that any rotated polydisk of radius $r$ in $(z_1, z_2) \in \mathbb{C}^2$ contains a smaller polydisk of radius $2^{-1/2}r$ that is aligned with the $z_1$ and $z_2$ planes. If we define a transform $z' = Uz$ and introduce notations $\delta_1 = a^{-s/2}/\sqrt{c_1}, \delta_2 = \sqrt{c_1}, \delta_3 = (a^{-s/2} + a^{s/2})/\sqrt{c_1}, d_1 = \min\{\delta_2, \delta_3\}$, and $d_2 = \min\{\delta_2/\sqrt{2}, \delta_3/\sqrt{2}\}$, then

$$P(G_1 \cap G_3) = \int_{|z_1| < \delta_1, |z_2| < \delta_3} \frac{e^{-\epsilon_1^{-1}(1+q)z_1V^{-1}z_2}}{\pi^2 \epsilon_1^{2(1+q)}} \det V \, dz_1 dz_2$$

$$= \int_{|z_1| < \delta_1, |z_2| < \delta_3} \frac{e^{-\epsilon_1^{-1}(1+q)(D_{11}|z_1|^2 + D_{22}|z_2|^2)}}{\pi^2 \epsilon_1^{2(1+q)}} \det V \, dz_1 dz_2$$

$$= \int_{|z_1|^2 + |z_2|^2 < 2d_1^2} \frac{e^{-\epsilon_1^{-1}(1+q)(D_{11}|z_1|^2 + D_{22}|z_2|^2)}}{\pi^2 \epsilon_1^{2(1+q)}} \det V \, dz_1 dz_2$$

$$= \int_{|z_1|^2 < d_1^2, |z_2|^2 < d_1^2} \frac{4}{\pi^2 \epsilon_1^{2(1+q)}} \det V \int_0^{d_1} r_1 e^{-\frac{D_{11}r_1^2}{1+q}} dr_1 \int_0^{d_1} r_2 e^{-\frac{D_{22}r_2^2}{1+q}} dr_2$$

$$= \left(1 - e^{-\frac{D_{11}d_1^2}{1+q}}\right) \left(1 - e^{-\frac{D_{22}d_1^2}{1+q}}\right),$$

and similarly

$$P(G_2 \cap G_3) = \int_{|z_1| < \delta_2, |z_2| < \delta_3} \frac{e^{-\epsilon_1^{-1}(1+q)z_1V^{-1}z_2}}{\pi^2 \epsilon_1^{2(1+q)}} \det V \, dz_1 dz_2 \geq \left(1 - e^{-\frac{D_{11}d_1^2}{1+q}}\right) \left(1 - e^{-\frac{D_{22}d_1^2}{1+q}}\right).$$

Suppose that $\Xi$ is a real random variable with a probability density function $h(\xi) = \frac{\hat{\omega}(\xi)^2}{\|\hat{\omega}\|^2}$, then

$$D_{11}^{-1}D_{22}^{-1} = \det (V)$$

$$= 4\pi^2\|\hat{\omega}\|^4 \left( \int_{\mathbb{R}} (a^s\xi + a)^2 \frac{|\hat{\omega}(\xi)|^2}{\|\hat{\omega}\|^2} d\xi - \left( \int_{\mathbb{R}} (a^s\xi + a) \frac{|\hat{\omega}(\xi)|^2}{\|\hat{\omega}\|^2} d\xi \right)^2 \right)$$

$$= 4\pi^2\|\hat{\omega}\|^4 \text{Var} (a^s\Xi + a)$$

$$\simeq a^{2s}. $$

36
and
\[
D_{11} + D_{22} = \det \left( V^{-1} \left( \|\bar{w}\|^2 + \langle 2\pi i \xi \bar{w}_{ab}, 2\pi i \xi \bar{w}_{ab} \rangle \right) \right) \\
\approx 1 + 4\pi^2 \mathbb{E} \left[ (a^s \Xi + a)^2 \right] \\
\approx \mathbb{E} \left[ (\Xi + a^{1-s})^2 \right] \\
\approx a^{2(1-s)}.
\]
This implies \( D_{11} \approx a^{2(1-s)} \) and \( D_{22} \approx a^{-2} \). Therefore,
\[
P(G_1 \cap G_3) \geq \left( 1 - e^{-\frac{D_{11} d_1^2}{c_1 + q}} \right) \left( 1 - e^{-\frac{D_{22} d_2^2}{c_1 + q}} \right) = \left( 1 - e^{-O(a^{2-3s} \epsilon_1^{-q})} \right) \left( 1 - e^{-O(a^{-s-2} \epsilon_1^{-q})} \right),
\]
and
\[
P(G_2 \cap G_3) \geq \left( 1 - e^{-\frac{D_1 d_3^2}{c_1 + q}} \right) \left( 1 - e^{-\frac{D_2 d_2^2}{c_1 + q}} \right) = \left( 1 - e^{-O(a^{2-2s} \epsilon_1^{-q})} \right) \left( 1 - e^{-O(a^{-2} \epsilon_1^{-q})} \right).
\]
By Theorem 2.11 if \((a, b) \in R_\delta \cap Z_k\) for some \(k\), then
\[
P(H_k) \geq P(H_k \mid G_1 \cap G_3) P(G_1 \cap G_3) = P(G_1 \cap G_3) \geq \left( 1 - e^{-O(a^{2-3s} \epsilon_1^{-q})} \right) \left( 1 - e^{-O(a^{-s-2} \epsilon_1^{-q})} \right).
\]
Note that \(a \approx N_k\) when \(a \in Z_k\), then
\[
P(H_k) \geq \left( 1 - e^{-O(N_k^{2-3s} \epsilon_1^{-q})} \right) \left( 1 - e^{-O(N_k^{-s-2} \epsilon_1^{-q})} \right).
\]
Similarly, if \((a, b) \in S_\delta \cap Z_k\) for some \(k\), then
\[
P(J_k) \geq P(J_k \mid G_2 \cap G_3) P(G_2 \cap G_3) = P(G_2 \cap G_3) \geq \left( 1 - e^{-O(N_k^{2-2s} \epsilon_1^{-q})} \right) \left( 1 - e^{-O(N_k^{-2} \epsilon_1^{-q})} \right).
\]
These arguments prove (iv) and (v).

Step 2: we go on to prove this theorem when the mother wave packet is of type \((\epsilon, m)\) with \(m \geq \frac{2}{1-s} + 4\). We would like to emphasize that the requirement is crucial to the following asymptotic analysis and it keeps the error caused by the non-compact support of \(\bar{w}\) reasonably small.

The sketch of the proof is similar to the first step, but \(W_\epsilon(a, b)\) and \((W_\epsilon(a, b), \partial_b W_\epsilon(a, b))\) are Gaussian variables not circularly symmetric. Suppose they have covariance matrices \(C_1\) and \(C_2\), pseudo-covariance matrices \(P_1\) and \(P_2\), respectively. We can still check that they have zero mean, \(C_1 = \epsilon_1^{1+q} \|w\|^2\) and \(C_2 = \epsilon_1^{1+q} V\), where \(V\) is defined in the first step. By
the definition of the mother wave packet of type \((\epsilon, m)\), we have

\[
|\mathbb{E}[W_\epsilon(a, b)W_\epsilon(a, b)]| \\
\leq \epsilon_1^{1+q} \int_{\mathbb{R}} |\hat{w}_{ab}(\xi)\hat{w}_{ab}(-\xi)| d\xi \\
\leq \epsilon_1^{1+q} \int_{\mathbb{R}} |\hat{w}(\xi - a^{1-s})\hat{w}(-\xi - a^{1-s})| d\xi \\
\leq \epsilon_1^{1+q} \left( \int_{\xi>0} |\hat{w}(\xi - a^{1-s})\hat{w}(-\xi - a^{1-s})| d\xi + \int_{\xi<0} |\hat{w}(\xi - a^{1-s})\hat{w}(-\xi - a^{1-s})| d\xi \right) \\
\leq \frac{\epsilon_1^{1+q}\epsilon}{(a^{1-s} - 1)m} \left( \int_{\xi>0} |\hat{w}(\xi - a^{1-s})| d\xi + \int_{\xi<0} |\hat{w}(-\xi - a^{1-s})| d\xi \right) \\
\approx \frac{2\epsilon_1^{1+q}\epsilon}{a^{m(1-s)}} \int_{\mathbb{R}} |\hat{w}(\xi)| d\xi.
\]

Similarly, we know

\[
\mathbb{E} \left[ (\partial_b W_\epsilon(a, b))^2 \right] \leq \frac{8\pi^2\epsilon_1^{1+q}\epsilon}{a^{m(1-s)}} \int_{\mathbb{R}} |\xi^2 \hat{w}(\xi)| d\xi,
\]

\[
\mathbb{E} \left[ \partial_b W_\epsilon(a, b)w_\epsilon(a, b) \right] \leq \frac{4\pi\epsilon_1^{1+q}\epsilon}{a^{m(1-s)}} \int_{\mathbb{R}} |\xi\hat{w}(\xi)| d\xi,
\]

\[
\mathbb{E} \left[ \partial_b W_\epsilon(a, b)\partial_b W_\epsilon(a, b) \right] = \epsilon_1^{1+q}\langle \partial_b w_{ab}, \partial_b w_{ab} \rangle = \epsilon_1^{1+q}/2\pi i\xi \hat{w}_{ab}, 2\pi i\xi \hat{w}_{ab}
\]

and

\[
\mathbb{E} \left[ W_\epsilon(a, b)\partial_b W_\epsilon(a, b) \right] = \epsilon_1^{1+p}\langle w_{ab}, \partial_b w_{ab} \rangle = \epsilon_1^{1+q}/\hat{w}_{ab}, 2\pi i\xi \hat{w}_{ab}.
\]

Hence, the magnitude of every entry in \(P_1\) and \(P_2\) is bounded by \(O\left(\frac{\epsilon_1^{1+q}\epsilon}{a^{m(1-s)}}\right)\). Since the covariance matrix of \((W_\epsilon(a, b), W_\epsilon^*(a, b))\) is

\[
V_1 = \begin{pmatrix} C_1 & P_1 \\ P_1^* & C_1^* \end{pmatrix},
\]

according to Equation (27) in \([11]\), the distribution of \(W_\epsilon(a, b)\) is described by the following distribution

\[
e^{-\frac{1}{2}(z_1, z_1^*)V_1^{-1}(z_1, z_1^*)^T} \frac{1}{\sqrt{\pi\sqrt{\det V_1}}},
\]

which is

\[
e^{-\frac{C_1 |z_1|^2 - \Re(P_1^* z_1^*)}{C_1^2 - P_1^* P_1}} \frac{1}{\pi\sqrt{C_1^2 - P_1^* P_1}}.
\]

Notice that

\[
\frac{C_1}{\sqrt{C_1^2 - P_1^* P_1}} = 1 + O\left(\frac{P_1 P_1^*}{C_1^2}\right) = 1 + O\left(\frac{\epsilon^2}{a^{2m(1-s)}}\right),
\]

\[
\frac{C_1 |z_1|^2 - \Re(P_1^* z_1^2)}{C_1 |z_1|^2} = 1 + O\left(\frac{\epsilon}{a^{m(1-s)}}\right),
\]

38
and
\[ \frac{C_1^2}{C_1^2 - P_1^*} = 1 + O \left( \frac{\epsilon^2}{a^{2m(1-s)}} \right). \]

Hence,
\[ e^{-\frac{1}{2}V_1^{-1}(z_1,z_1^*)} \pi \sqrt{\det V_1} = e^{-\epsilon_1^{1+q} |z_1||w|^2} \left( 1 + O \left( \frac{\epsilon |z_1|^2}{\epsilon_1^{1+q} a^{m(1-s)}} \right) \right). \]

By the same argument, the covariance matrix of \((W_e(a,b), \partial_b W_e(a,b), W_e^*(a,b), \partial_b W_e^*(a,b))\) is
\[ V_2 = \begin{pmatrix} C_2 & P_2 \\ P_2^* & C_2^* \end{pmatrix}. \]

Let \(z = (z_1, z_2)^T\), where \(T\) and \(*\) denote the transpose operator and conjugate transpose operator, respectively. Then the distribution of \((W_e(a,b), \partial_b W_e(a,b))\) is described by the joint probability density
\[ \frac{e^{-\frac{1}{2}V_2^{-1}(z,z)}}{\pi^2 \sqrt{\det V_2}} \]
\[ = e^{-\epsilon_1^{1+q} |z_1||w|^2} \left( 1 + O \left( \frac{\epsilon |z_1|^2}{\epsilon_1^{1+q} a^{m(1-s)}} \right) \right). \]

Notice that \(C_2 = \epsilon_1^{1+q} V \) and \(V\) has eigenvalues of order \(a^2\) and \(a^{2(s-1)}\). Hence, \(C_2\) has eigenvalues of order \(\epsilon_1^{1+q} a^2\) and \(\epsilon_1^{1+q} a^{2(s-1)}\). Recall that the magnitude of every entry in \(P_2\) is bounded by \(O \left( \frac{\epsilon_1^{1+q}}{a^{m(1-s)}} \right)\). This means that \(V_2\) is nearly dominated by diagonal blocks \(C_2\) and \(C_2^*\). Basic spectral theory for linear transforms shows that
\[ V_2^{-1} = \begin{pmatrix} C_2^{-1} \\ (C_2^*)^{-1} \end{pmatrix} + P_e, \]
where \(P_e\) is a matrix with 2-norm bounded by
\[ O \left( \frac{\epsilon_1^{1+q} \epsilon}{a^{m(1-s)}} \right) O(\epsilon_1^{1+q} a^{2(s-1)})^{-2} = O \left( \epsilon_1^{-1} \epsilon a^{m-4}(s-1) \right). \]

\(\frac{m-6}{m-4} \geq s\) ensures the above spectral analysis. Since every entry of \(P_2\) is bounded by \(O \left( \frac{\epsilon_1^{1+q} \epsilon}{a^{m(1-s)}} \right)\),
\[ \det V_2 = (\det C_2)^2 + O \left( \frac{\epsilon_1^{4(1+q)} \epsilon}{a^{m-2-(m+2)s}} \right), \]
where the residual comes from the entry bound and the eigenvalues of \(C_2\). Hence \(\text{(10)}\) is actually
\[ e^{-\epsilon_1^{1+q} z^* V^{-1} z} e^{-\frac{1}{2}V_2^{-1}(z_1,z_1^*)} \pi^2 \epsilon_1^{2(1+q)} \sqrt{(\det V_2)^2 + O \left( \frac{\epsilon}{a^{m-2-(m+2)s}} \right)}. \]

By the same argument in the first step, we can show that there exist a diagonal matrix \(D = \text{diag} \{ a^{2(1-s)}, a^{-2} \}\) and a unitary matrix \(U\) such that \(V^{-1} = U^* DU\). Part \((i)\) is still true by previous theorems. To conclude Part \((ii)\) to \((v)\), we still need to estimate
the probability of those events defined in the first step, i.e., \( P(G_1), P(G_2), P(G_1 \cap G_3), \)
\( P(G_2 \cap G_3), P(H_k) \) and \( P(J_k) \). By the calculations above, we have

\[
P(G_1) = \int_{|z_1| < a^{-s/2} \sqrt{e_1}} \frac{e^{-\frac{1}{2}(z_1^* z_1)V_1^{-1}(z_1, z_1)^T}}{\pi \sqrt{\text{det} V_1}} dz_1
\]

\[
= \int_{|z_1| < a^{-s/2} \sqrt{e_1}} \frac{e^{-\epsilon_1^{-1}(1+q)|z_1|^2||w||^{-2}}}{\pi \epsilon_1^{1+q} ||w||^2} \left( 1 + O \left( \frac{\epsilon |z_1|^2}{\epsilon_1^{1+q} a^{m(1-s)}} \right) \right) dz_1
\]

\[
= \frac{2}{\epsilon_1^{1+q} ||w||^2} \int_{0}^{a^{-s/2} \epsilon_1^{-q/2} ||w||^{-1}} \left( r + O \left( \frac{\epsilon}{\epsilon_1^{1+q} a^{m(1-s)}} \right) \right) e^{-\epsilon_1^{-1}(1+q) r^2 ||w||^{-2}} dr
\]

\[
= \int_{0}^{\frac{2}{\epsilon_1^{1+q} ||w||^2}} e^{-\epsilon_1^{-1} r^2} dr + O \left( \frac{\epsilon}{a^{m(1-s)}} \right) \int_{0}^{a^{-s/2} \epsilon_1^{-q/2} ||w||^{-1}} e^{-\epsilon_1^{-1} r^2} dr
\]

\[
= 1 - e^{-\epsilon_1^{-1} ||w||^{-2}} + O \left( \frac{\epsilon}{a^{m(1-s)}} \right).
\]

Hence, we can conclude \((ii)\) and \((iii)\) follows the same proof in the first step. Next, we look at the last two part of this theorem.

Recall that we have defined a transform \( z' = U z \) and introduced notations \( \delta_1 = a^{-s/2} \sqrt{e_1}, \delta_2 = \sqrt{e_1}, \delta_3 = (a^{1-s/2} + a^{s/2}) \sqrt{e_1}, \delta_1 = \min \{ \frac{\delta_1}{\sqrt{2}}, \frac{\delta_2}{\sqrt{2}} \}, \) and \( \delta_2 = \min \{ \frac{\delta_3}{\sqrt{2}}, \frac{\delta_3}{\sqrt{2}} \} \) in the first step. Using the same notations and a similar argument, we have

\[
P(G_1 \cap G_3) = \int_{\{|z_1| < \delta_1, |z_2| < \delta_3\}} \frac{e^{-\frac{1}{2}(z_1^* z_1, z_2^* z_2)V_2^{-1}(z_1, z_2, z_1^* z_2)^T}}{\pi^2 \sqrt{\text{det} V_2}} dz_1 dz_2
\]

\[
= \int_{\{|z_1| < \delta_1, |z_2| < \delta_3\}} \frac{e^{-\epsilon_1^{-1}(1+q) z_1^* z_2^* V^{-1} z_1^* z_2^*} \epsilon^{-\frac{1}{2}(z_1^* z_1, z_2^* z_2) P_4(z_1, z_2, z_1^* z_2)^T}}{\pi^2 \epsilon_1^{2(1+q)} \sqrt{(\text{det} V)^2 + O \left( \frac{\epsilon}{a^{m-2(1-s)}} \right)}} dz_1 dz_2
\] (11)

Since

\[
\frac{\text{det} V}{\sqrt{(\text{det} V)^2 + O \left( \frac{\epsilon}{a^{m-2(1-s)}} \right)}} = 1 + O \left( \frac{\epsilon}{a^{m-2(1-s)}} \right),
\] (12)

we can drop out the term \( O \left( \frac{\epsilon}{a^{m-2(1-s)}} \right) \) in (11), which would generate an absolute error no more than \( O \left( \frac{\epsilon}{a^{m-2(1-s)}} \right) \) in the estimate of \( P(G_1 \cup G_3). \) Let

\[
g(z) = -\frac{1}{2} (z_1^*, z_2^*, z_1, z_2) P_4(z_1, z_2, z_1^* z_2)^T;
\]
\[ \begin{align*}
\int_{\{|z_1|<\delta_1,|z_2|<\delta_3\}} & \frac{e^{-(1+\eta)z^*V^{-1}z_{\theta}}}{\pi^2 \epsilon_1^{2(1+\eta)}} \, d\bar{z}_1 \, dz_2 \\
= & \int_{\{|z_1|<\delta_1,|z_2|<\delta_3\}} \frac{e^{-(1+\eta)(D_{11}|z_1|^2+D_{22}|z_2|^2)}e^{\theta(U^*z')}}{\pi^2 \epsilon_1^{2(1+\eta)}} \, d\bar{z}_1' \, dz_2' \\
\geq & \int_{\{|z_1'|<d_1,|z_2'|<d_1\}} \frac{e^{-(1+\eta)(D_{11}|z_1'|^2+D_{22}|z_2'|^2)}e^{\theta(U^*z')}}{\pi^2 \epsilon_1^{2(1+\eta)}} \, d\bar{z}_1' \, dz_2' \\
= & \frac{1}{\pi^2 \epsilon_1^{2(1+\eta)}} \det V \int_0^{d_1} \int_0^{d_1} \int_0^{2\pi} \int_0^{2\pi} r_1 r_2 e^{\frac{D_{11}r_1^2}{\epsilon_1^{1+\eta}}} e^{\frac{D_{22}r_2^2}{\epsilon_1^{1+\eta}}} e^{\tilde{g}(r_1,\theta_1, r_2, \theta_2)} d\theta_1 d\theta_2 dr_1 dr_2 \\
= & \frac{1}{\pi^2 \epsilon_1^{2(1+\eta)}} \det V \int_0^{d_1} \int_0^{d_1} \int_0^{2\pi} \int_0^{2\pi} r_1 r_2 e^{\frac{D_{11}r_1^2}{\epsilon_1^{1+\eta}}} e^{\frac{D_{22}r_2^2}{\epsilon_1^{1+\eta}}} \left( e^{\tilde{g}(r_1,\theta_1, r_2, \theta_2)} - 1 \right) d\theta_1 d\theta_2 dr_1 dr_2 \\
& + \left( 1 - e^{\frac{D_{11}d_1^2}{\epsilon_1^{1+\eta}}} \right) \left( 1 - e^{\frac{D_{22}d_1^2}{\epsilon_1^{1+\eta}}} \right), 
\end{align*} \]

where \( \tilde{g}(r_1, \theta_1, r_2, \theta_2) = g(U^*z') \). Recall that the 2-norm of \( P_\epsilon \) is bounded by \( O \left( \epsilon_1^{-(1+\eta)} \epsilon a^{(m-4)(s-1)} \right) \).

Hence,

\[ |\tilde{g}(r_1, \theta_1, r_2, \theta_2)| \leq O \left( \epsilon_1^{-(1+\eta)} \epsilon a^{(m-4)(s-1)} \right) (|z_1|^2 + |z_2|^2) = O \left( \epsilon_1^{-(1+\eta)} \epsilon a^{(m-4)(s-1)} \right) (r_1^2 + r_2^2). \]

Therefore, the first term in (13) is bounded by

\[ \begin{align*}
O \left( \epsilon a^{(m-4)(s-1)} \right) & \int_0^{d_1} \int_0^{d_1} r_1 r_2 e^{\frac{D_{11}r_1^2}{\epsilon_1^{1+\eta}}} e^{\frac{D_{22}r_2^2}{\epsilon_1^{1+\eta}}} (r_1^2 + r_2^2) \, dr_1 \, dr_2 \\
= & \int_0^{d_1} \int_0^{d_1} \int_0^{2\pi} \int_0^{2\pi} \left( \frac{D_{11}d_1^2}{\epsilon_1^{1+\eta}} \right) \left( \frac{D_{22}d_1^2}{\epsilon_1^{1+\eta}} \right) r_1 r_2 \left( \frac{r_1^2}{D_{11}} + \frac{r_2^2}{D_{22}} \right) e^{-r_1^2} e^{-r_2^2} \, dr_1 \, dr_2 \\
\leq & \int_0^{\infty} \int_0^{\infty} r_1 r_2 (r_1^2 + r_2^2) e^{-r_1^2} e^{-r_2^2} \, dr_1 \, dr_2 \\
= & \int_0^{\infty} \int_0^{\infty} \left( \frac{\epsilon}{a^{(m-4)(s-1)}} \right) \left( \frac{\epsilon}{a^{(m-4)(s-1)}} \right) e^{-r_1^2} e^{-r_2^2} \, dr_1 \, dr_2 \\
& = O \left( \frac{\epsilon}{a^{(m-4)(s-1)}} \right) \left( \frac{\epsilon}{a^{(m-4)(s-1)}} \right) \left( \frac{\epsilon}{a^{(m-4)(s-1)}} \right). 
\end{align*} \]

The analysis in (12) and (14) implies that

\[ P(G_1 \cup G_3) \geq \left( 1 - e^{\frac{D_{11}d_1^2}{\epsilon_1^{1+\eta}}} \right) \left( 1 - e^{\frac{D_{22}d_1^2}{\epsilon_1^{1+\eta}}} \right) + O \left( \frac{\epsilon}{a^{(m-4)(s-1)}} \right). \]
and similarly

\[
P(G_2 \cap G_3) = \int_{\{|z_1|<\delta_2,|z_2|<\delta_1\}} \frac{e^{-\frac{1}{2}(z_1^* z_2, z_1^* z_2)}V_2^{-1}(z_1, z_2, z_1^* z_2)^T}{\pi^2 \det V_2} dz_1dz_2 
\geq \left( 1 - e^{-\frac{D_1+D_2^2}{1+q}} \right) \left( 1 - e^{-\frac{D_2+D_2^2}{1+q}} \right) + O\left( \frac{e}{a(m-4)(1-s)-2} \right).
\]

The rest of the proof is exactly the same as the one in the first step and consequently we know this theorem is also true for a mother wave packets of type \((\epsilon, m)\) with \(m\) satisfying \(m \geq \frac{2}{1-s} + 4\). 

\[\square\]

A.4 Proof of Theorem 2.13

**Proof.** The proof of this theorem is nearly identical to Theorem 2.12 but for the covariance functional of the noise term which is now a general functional \(\mathcal{R} : L^1 \cap C^{m-1} \rightarrow L^1 \cap C^{m-1}\).

Step 1: In a similar structure, we prove the case when the mother wave packet is of type \((0,m)\).

We can still check that \(W_g(a,b) = W_f(a,b) + W_e(a,b)\) and \(\partial_b W_g(a,b) = \partial_b W_f(a,b) + \partial_b W_e(a,b)\) are Gaussian variables. Furthermore, \(W_e(a,b)\) and \((W_e(a,b), \partial_b W_e(a,b))\) are still circularly symmetric Gaussian variables. Since the Gaussian process \(\epsilon\) is zero mean, we have \(\mathbb{E}[W_e(a,b)] = 0\) and \(\mathbb{E}[\partial_b W_e(a,b)] = 0\). Note that \(\mathcal{R}\) can be “diagonalized” to a functional \(\mathcal{D}\) by the Fourier transform denoted as \(\mathcal{F}\) in the sense that

\[
\langle f_1, \mathcal{R} f_2 \rangle = \langle f_1, \mathcal{F}^* \mathcal{D} \mathcal{F} f_2 \rangle = \langle \hat{f}_1, \hat{\mathcal{D}} \hat{f}_2 \rangle
\]

for any \(f_1\) and \(f_2\) in \(L^1 \cap C^{m-1}\). Hence,

\[
\mathbb{E}\left[ W_e(a,b) W_e^*(a,b) \right] = \langle w_{ab}, \mathcal{R} w_{ab} \rangle = \langle \tilde{w}_{ab}, \tilde{\epsilon}^{(a^* \xi + a)} \rangle = \langle \tilde{w}, \tilde{\epsilon}^{(a^* \xi + a)} \rangle
\]

and

\[
\mathbb{E}\left[ W_e(a,b) W_e^*(a,b) \right] = \langle w_{ab}, \mathcal{R} w_{ab} \rangle = \langle \tilde{w}_{ab}, \tilde{\epsilon}(a^* \xi + a) \rangle = \int_{\mathbb{R}} \tilde{w}_{ab}(\xi) \tilde{w}_{ab}(-\xi) \tilde{\epsilon}(a^* \xi + a) d\xi = 0.
\]

If we introduce \(\sigma^2 = \langle \tilde{w}, \tilde{\epsilon}^{(a^* \xi + a)} \rangle\) for simplicity and a random variable \(\Xi\) with a probability density function \(\sigma^{-2} |\tilde{\epsilon}(a^* \xi + a)|^2\), then by a similar argument, we know

\[
\mathbb{E}\left[ \langle \partial_b W_e(a,b) \rangle^2 \right] = \mathbb{E}[\partial_b W_e(a,b) w_{ab}(a,b)] = 0,
\]

\[
\mathbb{E}\left[ \partial_b W_e(a,b) \partial_b W_e^*(a,b) \right] = \langle \partial_b w_{ab}, \mathcal{R} \partial_b w_{ab} \rangle = 4\pi \sigma^2 \mathbb{E}\left[ (a^* \Xi + a)^2 \right],
\]

and

\[
\mathbb{E}\left[ W_e(a,b) \partial_b W_e^*(a,b) \right] = \langle w_{ab}, \mathcal{R} \partial_b w_{ab} \rangle = 2\pi i \sigma^2 \mathbb{E}\left[ a^* \Xi + a \right].
\]

Hence, \(W_e(a,b)\) and \((W_e(a,b), \partial_b W_e(a,b))\) have zero pseudo-covariance matrices and they are circularly symmetric. Therefore, the distribution of \(W_e(a,b)\) is determined by its variance as follows

\[
e^{-\frac{\sigma^{-2}|z_1|^2}{\pi \sigma^2}}.
\]
If we define

\[
V = \begin{pmatrix}
1 & 2\pi i E [a^s \Xi + a] \\
-2\pi i E [a^s \Xi + a] & 4\pi^2 E [(a^s \Xi + a)]
\end{pmatrix},
\]

then \(\sigma^2 V\) is the covariance matrix of \((W_e(a, b), \partial_b W_e(a, b))\) and its distribution is described by the joint probability density

\[
e^{-z^T V^{-1} z} / (\pi^2 \sigma^4 \det V),
\]

where \(z = (z_1, z_2)^T\). \(V\) is an invertible and self-adjoint matrix, since \(W_e(a, b)\) and \(\partial_b W_e(a, b)\) are linearly independent. Hence, there exist a diagonal matrix \(D\) and a unitary matrix \(U\) such that \(V^{-1} = U^* DU\).

Part (i) is true by previous theorems. Define the following events

\[
G_1 = \{|W_e(a, b)| < a^{-s/2} M_a^{1/(2+2q)}\},
\]

\[
G_2 = \{|W_e(a, b)| < M_a^{1/(2+2q)}\},
\]

\[
G_3 = \{ |\partial_b W_e(a, b)| < M_a^{1/(2+2q)} \left( a^{1-s/2} + a^{s/2} \right) \},
\]

\[
H_k = \left\{ \frac{|v_g(a, b) - N_k \phi'_k(b)|}{|N_k \phi'_k(b)|} \lesssim \sqrt{\epsilon} + M_p^{1/(1+q)} \right\},
\]

and

\[
J_k = \left\{ \frac{|v_g(a, b) - N_k \phi'_k(b)|}{|N_k \phi'_k(b)|} \lesssim N_k^{-s/2} \left( \sqrt{\epsilon} + M_p^{1/(1+q)} \right) \right\},
\]

for \(1 \leq k \leq K\). Now we estimate the probability \(P(G_1), P(G_2), P(G_1 \cap G_3), P(G_2 \cap G_3), P(H_k)\) and \(P(J_k)\). By the calculations above, we have

\[
P(G_1) = \int_{|z_1| < a^{-s/2} M_a^{1/(2+2q)}} e^{-z^T V^{-1} z} / \pi \sigma^2 \, dz_1 = 1 - e^{-a^{-s} M_a^{1/(1+q)} \sigma^{-2}} \geq 1 - e^{-O(a^{-s} M_a^{-q/(1+q)})},
\]

and similarly

\[
P(G_2) \geq 1 - e^{-O(M_a^{-q/(1+q)})}.
\]

We are ready to summarize and conclude (ii) and (iii). If \((a, b) \in R_{\delta_1}\), then

\[
|W_e(a, b) + W_f(a, b)| \geq a^{-s/2} \left( M_a^{1/(2-p)} / (1+q) + \sqrt{\epsilon} \right).
\]

(15)

If \((a, b) \notin \bigcup_{1 \leq k \leq K} Z_k\), then by Lemma 2.10

\[
|W_f(a, b)| \leq a^{-s/2} \epsilon.
\]

(16)

Equation (15) and (16) lead to

\[
|W_e(a, b)| \geq a^{-s/2} M_a^{(1-p)/(1+q)}.\]

Hence,

\[
P \left( (a, b) \notin \bigcup_{1 \leq k \leq K} Z_k \right) \leq P \left( |W_e(a, b)| \geq a^{-s/2} M_a^{(1-p)/(1+q)} \right) = 1 - P(G_1).
\]

This means that if \((a, b) \in R_{\delta_1}\), then \((a, b) \in \bigcup_{1 \leq k \leq K} Z_k\) with a probability at least \(P(G_1) \geq 1 - e^{-O(a^{-s} M_a^{-q/(1+q)})} = 1 - e^{-O(N_k^{-s} M_a^{-q/(1+q)})}\), since \(a \simeq N_k\) if \((a, b) \in Z_k\). So, (ii) is true.
A similar argument applied to \((a, b) \in S_{\delta_n}\) shows that \((a, b) \in \bigcup_{1 \leq k \leq K} Z_k\) with a probability at least \(P(G_2) = 1 - e^{-O\left(M_{\alpha}^{-(1+q)}\right)}\). Hence, \((iii)\) is proved.

If we introduce notations \(\delta_1 = a^{-s/2}M_{\alpha}^{1/(2 + 2q)}, \delta_2 = M_{\alpha}^{1/(2 + 2q)}, \delta_3 = (a^{1-s/2} + a^{s/2}) M_{\alpha}^{1/(2 + 2q)}, \)
\[
d_1 = \min\{\frac{\delta_1}{\sqrt{2}}, \frac{\delta_3}{\sqrt{2}}\}, \text{ and } d_2 = \min\{\frac{\delta_2}{\sqrt{2}}, \frac{\delta_3}{\sqrt{2}}\},
\]
then it follows from the same proof in Theorem 2.12 that
\[
P(G_1 \cap G_3) = \int_{\{|z_1| < \delta_1, |z_2| < \delta_3\}} \frac{e^{-\sigma z^2 V^{-1}z}}{\pi^{2} \sigma^4 \det V} dz_1 dz_2 \geq \left(1 - e^{-\frac{d_1^2}{\sigma^2}}\right) \left(1 - e^{-\frac{d_2^2}{\sigma^2}}\right),
\]
and similarly
\[
P(G_2 \cap G_3) = \int_{\{|z_1| < \delta_2, |z_2| < \delta_3\}} \frac{e^{-\sigma z^2 V^{-1}z}}{\pi^{2} \sigma^4 \det V} dz_1 dz_2 \geq \left(1 - e^{-\frac{d_1^2}{\sigma^2}}\right) \left(1 - e^{-\frac{d_2^2}{\sigma^2}}\right).
\]

Note that
\[
D_{11}^{-1}D_{22}^{-1} = \det(V) = 4\pi^2 a^{2s} \text{Var}[\Xi]
\]
and
\[
D_{11} + D_{22} = \frac{1 + 4\pi^2 \text{E}[(a^s \Xi + a)^2]}{4\pi^2 a^{2s} \text{Var}[\Xi]}
\]
We assume \(D_{22} \leq D_{11}\). Since \(|\text{E}[\Xi]| \lesssim 1\) and \(\text{E}[\Xi^2] \lesssim 1\), then
\[
D_{22} = \frac{\det(V^{-1})}{D_{11}} \simeq \frac{1}{\det V(D_{11} + D_{22})} = \frac{1}{1 + 4\pi^2 \text{E}[(a^s \Xi + a)^2]} \simeq a^{-2},
\]
and
\[
D_{11} \simeq \frac{1 + 4\pi^2 \text{E}[(a^s \Xi + a)^2]}{4\pi^2 a^{2s} \text{Var}[\Xi]} \gtrsim a^{-2s}.
\]
This implies
\[
P(G_1 \cap G_3) \geq \left(1 - e^{-\frac{d_1^2}{\sigma^2}}\right) \left(1 - e^{-\frac{d_2^2}{\sigma^2}}\right)
\]
\[
\gtrsim \left(1 - e^{-O(a^{2-3s} M_{\alpha}^{-(1+q)})}\right) \left(1 - e^{-O(a^{-2s} M_{\alpha}^{-(1+q)})}\right),
\]
and
\[
P(G_2 \cap G_3) \geq \left(1 - e^{-\frac{d_1^2}{\sigma^2}}\right) \left(1 - e^{-\frac{d_2^2}{\sigma^2}}\right)
\]
\[
\gtrsim \left(1 - e^{-O(a^{2-3s} M_{\alpha}^{-(1+q)})}\right) \left(1 - e^{-O(a^{-2s} M_{\alpha}^{-(1+q)})}\right).
\]
By Theorem 2.11 if \((a, b) \in R_{\delta_n} \cap Z_k\) for some \(k\), then
\[
P(H_k) \geq P(H_k | G_1 \cap G_3) P(G_1 \cap G_3) = P(G_1 \cap G_3).
\]
Note that $a \simeq N_k$ when $a \in Z_k$, then

$$P(H_k) \geq \left(1 - e^{-O\left(N_k^{2-3s}M^{-q/(1+q)}\right)}\right) \left(1 - e^{-O\left(N_k^{-s}M^{-q/(1+q)}\right)}\right).$$

Similarly, if $(a, b) \in S_{\delta_a} \cap Z_k$ for some $k$, then

$$P(J_k) \geq \left(1 - e^{-O\left(N_k^{2-2s}M^{-q/(1+q)}\right)}\right) \left(1 - e^{-O\left(N_k^{-s}M^{-q/(1+q)}\right)}\right).$$

These arguments prove (iv) and (v).

Step 2: We discuss the case for a mother wave packet of type $(\epsilon, m)$ for $m \geq 2^s + 4$. Similar to what we have already seen in the second step of the proof of Theorem 2.12, $W_g(a, b) = W_f(a, b) + W_e(a, b)$ and $\partial_b W_g(a, b) = \partial_b W_f(a, b) + \partial_b W_e(a, b)$ are Gaussian, $W_e(a, b)$ and $(W_e(a, b), \partial_b W_e(a, b))$ are nearly circularly symmetric Gaussian variables. Using the same strategy in Step 2 in the proof of Theorem 2.12 and the notations in Step 1 in this theorem, we can still check that:

1. The distribution of $W_e(a, b)$ is well approximated by

$$e^{-\sigma^2|z_1|^2} \pi \sigma^2,$$

where $\sigma^2 = \langle \hat{w}, \hat{e}(a^s \xi + a) \hat{w} \rangle$.

2. The distribution of $(W_e(a, b), \partial_b W_e(a, b))$ is well approximated by

$$e^{-\sigma^2 z^* V^{-1} z} \pi^2 \sigma^4 \det V,$$

where

$$V = \begin{pmatrix} 1 & 2\pi i E \left[ a^s \Xi + a \right] \\ -2\pi i E \left[ a^s \Xi + a \right] & 4\pi^2 E \left[ (a^s \Xi + a)^2 \right] \end{pmatrix},$$

and $V$ has eigenvalues $D_{22}^{-1} \simeq a^2$ and $D_{11}^{-1} \lesssim a^2(s-1)$.

Suppose $G_1, G_2, G_3$ are the events defined in the first step, then the well approximation here means that:

1. $P(G_1) = \int_{|z_1| < a^{-s/2} M_1^{1/(2+s)}} \frac{e^{-\sigma^2|z_1|^2}}{\pi \sigma^2} dz_1 + O\left(\frac{\epsilon}{a^{m(1-s)}}\right)$

$$= 1 - e^{-a^{-s} M_1^{1/(1+q)} + O\left(\frac{\epsilon}{a^{m(1-s)}}\right)} \sigma^2$$

$$\geq 1 - e^{-O\left(a^{-s} M_1^{-q/(1+q)}\right)} + O\left(\frac{\epsilon}{a^{m(1-s)}}\right),$$

similarly

$$P(G_2) \geq 1 - e^{-O\left(M_1^{-q/(1+q)}\right)} + O\left(\frac{\epsilon}{a^{m(1-s)}}\right).$$
So, 

\[
P(G_1 \cap G_3) = \int_{|z_1| < \delta_1, |z_2| < \delta_3} \frac{e^{-\sigma^{-2}z^* V^{-1}z}}{\pi^2 \sigma^4 \det V} dz_1 dz_2 + O \left( \frac{\epsilon}{a^{(m-4)(1-s)-2}} \right),
\]

and similarly

\[
P(G_2 \cap G_3) = \int_{|z_1| < \delta_2, |z_2| < \delta_3} \frac{e^{-\sigma^{-2}z^* V^{-1}z}}{\pi^2 \sigma^4 \det V} dz_1 dz_2 + O \left( \frac{\epsilon}{a^{(m-4)(1-s)-2}} \right),
\]

Following the proof in the first step, it is straightforward to see this theorem is true for a mother wave packet of type \((\epsilon, m)\) with \(m \geq \frac{2}{1-s} + 4\).

B Proofs for the theorems in Section 3

B.1 Proof of Theorem 3.7

Proof. This theorem is a straightforward generalization of Theorem 2.8 and and its first two parts have been proved for the case of mother wave packet of type \((0, m)\) for any \(m\) in 60. We would leave the proof for reader.

B.2 Proof of Theorem 3.10

Proof. We only sketch out the proof of this theorem, because its proof is similar to the proof in Theorem 2.11. By the definition of 2D wave packet transform and Lemma 3.8 and 3.9 we obtain the following two estimates:

\[
|W_e(a, b)| \lesssim \sqrt{\epsilon_1} |a|^{-s}, \tag{17}
\]

and

\[
|\nabla_b W_e(a, b)| \lesssim \sqrt{\epsilon_1} (1 + |a|^{1-s}). \tag{18}
\]

If \((a, b) \in R_\delta\), then \(W_g(a, b) \geq |a|^{-s} \delta\) and Equation (17) imply

\[
|W_f(a, b)| \geq |a|^{-s} \sqrt{\epsilon}. \tag{19}
\]

Hence, \(S_\delta \subset R_\delta \subset R_\epsilon\), where \(R_\epsilon\) is defined in Theorem 3.7 and is a subset of \(\bigcup_{1 \leq k \leq K} Z_k\). So, (i) is true by Theorem 3.7. As for (ii), since \(R_\delta \subset R_\epsilon\), then \((a, b) \in R_\delta \cap Z_k\) implies \((a, b) \in R_\epsilon \cap Z_k\) and \(|a| \approx N_k\). By Theorem 3.7 we have

\[
\frac{|v_f(a, b) - N_k \nabla \phi_k(b)|}{|N_k \nabla \phi_k(b)|} \lesssim \sqrt{\epsilon},
\]
when $N > N_0$. Hence,

$$\frac{|v_g(a, b) - N_k \nabla \phi_k(b)|}{|N_k \nabla \phi_k(b)|} \leq \frac{|v_f(a, b) - N_k \nabla \phi_k(b)|}{|N_k \nabla \phi_k(b)|} + \frac{|v_f(a, b) - v_g(a, b)|}{|N_k \nabla \phi_k(b)|}$$

$$\leq \sqrt{\epsilon} + \frac{|W_e(a, b)|}{|W_g(a, b)|} + \frac{\left|\nabla_b W_e(a, b)\right|}{N_k |W_g(a, b)|}$$

$$\leq \sqrt{\epsilon} + \sqrt{\epsilon_1}$$

$$\leq \sqrt{\epsilon} + \epsilon_1^p,$$

when $N > N_0$. With a similar argument, when $(a, b) \in S_\delta \cap Z_k$, we can show that

$$\frac{|v_g(a, b) - N_k \nabla \phi_k(b)|}{|N_k \nabla \phi_k(b)|} \leq \frac{\sqrt{\epsilon}}{N_k^s} + \frac{|W_e(a, b)|}{|W_g(a, b)|} + \frac{\left|\nabla_b W_e(a, b)\right|}{N_k |W_g(a, b)|}$$

$$\leq \frac{\sqrt{\epsilon}}{N_k^s} + \frac{\sqrt{\epsilon_1}}{a^s \delta}$$

$$\leq \frac{\sqrt{\epsilon}}{N_k^s} + \epsilon_1^p,$$

when $N > N_0$. \qed

### B.3 Proof of Theorem 3.11

*Proof.* The sketch of the proof of this theorem is similar to the proof of Theorem 2.12 and 2.13 but much trickier. Since $w_{ab} \in L^1 \cap C^m$ and $\nabla_b w_{ab} \in L^1 \cap C^{m-1} \times L^1 \cap C^{m-1}$, we know $W_g(a, b) = W_f(a, b) + W_e(a, b)$ and $\nabla_b W_g(a, b) = \nabla_b W_f(a, b) + \nabla_b W_e(a, b)$ are Gaussian random variables. By the properties of zero mean stationary Gaussian processes and the geometric supports of wave packets in the frequency domain, we can still check that $W_e(a, b)$ and $(W_e(a, b), \partial_b W_e(a, b), \partial_{bk} W_e(a, b))$ have nearly zero pseudo-covariance matrices. Hence, they are nearly circularly symmetric. We also divide the proof into two steps.

Step 1: We prove the case when the mother wave packet is of type $(0, m)$.

In this case, $W_e(a, b)$ and $(W_e(a, b), \partial_b W_e(a, b), \partial_{bk} W_e(a, b))$ are circularly symmetric. The variance of $W_e(a, b)$ is $\int_{\mathbb{R}^2} \hat{w}(\xi)^2 \varphi(\xi^* + a) d\xi$, which is denoted by $\sigma^2$. Suppose that $\Xi = (\Xi_1, \Xi_2)^T$ is a real random vector with a joint probability density function $h(\xi) = \sigma^{-2} |\hat{w}(\xi)|^2 \varphi(\xi^* + a)$, then the covariance matrix of $(W_e(a, b), \partial_b W_e(a, b), \partial_{bk} W_e(a, b))$ is $\varphi^2 V$, where $V$ is the matrix below:

$$\begin{pmatrix}
1 & -2\pi i \mathbb{E} [|a|^* \Xi_1 + a_1] & -2\pi i \mathbb{E} [|a|^* \Xi_2 + a_2] \\
2\pi i \mathbb{E} [|a|^* \Xi_1 + a_1] & 4\pi^2 \mathbb{E} \left[ |a|^2 \left( |a|^* \Xi_1 + a_1 \right) \right] & 4\pi^2 \mathbb{E} \left[ |a|^2 \left( |a|^* \Xi_2 + a_2 \right) \right] \\
2\pi i \mathbb{E} [|a|^* \Xi_2 + a_2] & 4\pi^2 \mathbb{E} \left[ |a|^2 \left( |a|^* \Xi_1 + a_1 \right) \right] & 4\pi^2 \mathbb{E} \left[ |a|^2 \left( |a|^* \Xi_2 + a_2 \right) \right]
\end{pmatrix}.$$

The distributions of $W_e(a, b)$ and $(W_e(a, b), \partial_b W_e(a, b), \partial_{bk} W_e(a, b))$ are described by the probability density functions

$$\frac{e^{-\sigma^{-2} |z|^2}}{\pi \sigma^{-2}}$$

47
and
\[ e^{-\sigma^{-2}z^*V^{-1}z} \]
\[ \frac{1}{\pi^3 \sigma^6 \det V}, \]
where \( z = (z_1, z_2, z_3)^T \). Part (i) is true by previous theorems. To prove Part (ii) to (v), we need to define the following events

\[
G_1 = \left\{ |W_e(a, b)| < |a|^{-s} M_1^{1/(2+2q)} \right\},
\]
\[
G_2 = \left\{ |W_e(a, b)| < M_1^{1/(2+2q)} \right\},
\]
\[
G_3 = \left\{ \nabla_b W_e(a, b) < M_1^{1/(2+2q)} (1 + |a|^{1-s}) \right\},
\]
\[
H_k = \left\{ \frac{v_g(a, b) - N_k \nabla_b \phi_k(b)}{|N_k \phi_k(b)|} \lesssim \sqrt{\epsilon} + M_2^{p/(1+q)} \right\},
\]
and
\[
J_k = \left\{ \frac{|v_g(a, b) - N_k \nabla_b \phi_k(b)|}{|N_k \phi_k(b)|} \lesssim N_k^{-s} \left( \sqrt{\epsilon} + M_2^{p/(1+q)} \right) \right\},
\]
for \( 1 \leq k \leq K \). Next, we are going to estimate the probability \( P(G_1), P(G_2), P(G_1 \cap G_3), P(G_2 \cap G_3), P(H_k) \) and \( P(J_k) \). By the calculations above, we have

\[
P(G_1) = \int_{|z_1| \leq |a|^{-s} M_1^{1/(2+2q)}} \frac{e^{-\sigma^{-2}|z_1|^2}}{\pi \sigma^{-2}} \, dz_1 = 1 - e^{-|a|^{-2s} \sigma^{-2} M_1^{1/(1+q)}} \geq 1 - e^{-|a|^{-2s} M_{a^{-q/(1+q)}}},
\]
and similarly

\[
P(G_2) = \int_{|z_1| \leq M_1^{1/(2+2q)}} \frac{e^{-\sigma^{-2}|z_1|^2}}{\pi \sigma^{-2}} \, dz_1 \geq 1 - e^{-M_a^{-q/(1+q)}}.
\]

We are now ready to conclude (ii) and (iii). If \((a, b) \in R_{b,a}\), then

\[
|W_e(a, b) + W_f(a, b)| \geq |a|^{-s} \left( M_1^{1/(2-p)/(1+q)} + \sqrt{\epsilon} \right). \tag{20}
\]

If \((a, b) \notin \bigcup_{1 \leq k \leq K} Z_k\), then by Lemma 3.9,

\[
|W_f(a, b)| \leq |a|^{-s} \epsilon. \tag{21}
\]

\(|W_e(a, b)| \geq |a|^{-s} M_1^{1/(2+2q)}\) follows from Equation (20) and (21). Hence,

\[
P \left( (a, b) \notin \bigcup_{1 \leq k \leq K} Z_k \right) \leq P \left( |W_e(a, b)| \geq |a|^{-s} M_1^{1/(2+2q)} \right) = 1 - P(G_1).
\]

This means that if \((a, b) \in R_{b,a}\), then \((a, b) \in \bigcup_{1 \leq k \leq K} Z_k\) with a probability at least \( P(G_1) \geq 1 - e^{-|a|^{-2s} M_{a^{-q/(1+q)}}} = 1 - e^{-O(N_k^{-2s} M_{a^{-q/(1+q)}})}\), since \(|a| \simeq N_k\) if \((a, b) \in Z_k\). So, (ii) is true. A similar argument applied to \((a, b) \in S_{b,a}\) shows that \((a, b) \in \bigcup_{1 \leq k \leq K} Z_k\) with a probability at least \( P(G_2) \geq 1 - e^{-O(M_{a^{-q/(1+q)}})}\). Hence, (iii) is proved.

Because \( V \) is invertible and self-adjoint, there exist a unitary matrix \( U \) and a diagonal matrix \( D \) such that \( V^{-1} = U^*DU \). For \( z \in \mathbb{C}^3 \), let \( z' = Uz \). Introduce notations \( \delta_1 = \)
\[ |a|^{-s} M^{1/(2+3\theta)}, \, \delta_2 = M^{1/(2+3\theta)}, \, \delta_3 = (1 + |a|^{1-s}) M^{1/(2+3\theta)}, \, d_1 = \min\{ \frac{\delta_1}{\sqrt{2}}, \frac{\delta_3}{2} \}, \] and \( d_2 = \min\{ \frac{\delta_2}{\sqrt{2}}, \frac{\delta_3}{2} \} \). Similar to the proof in Theorems 2.12 and 2.13 by a simple property of high dimensional polydisk, we have

\[
P(G_1 \cap G_3) = \int_{\{|z_1|<\delta_1, |z_2|<|z_3|<\delta_2\}} e^{-\sigma z_1 D_{11}^{-1} + D_{22}^{-1} + D_{33}^{-1}|z_1|^2} \pi^3 \sigma^6 \det V \, dz_1 dz_2 dz_3
\]

\[
\geq \int_{\{|z_1|<\delta_1, |z_2|<\delta_2, |z_3|<\delta_2\}} e^{-\sigma z_1 D_{11}^{-1} + D_{22}^{-1} + D_{33}^{-1}|z_1|^2} \pi^3 \sigma^6 \det V \, dz_1 dz_2 dz_3
\]

\[
= \int_{\{|z_1|<\delta_1, |z_2|<\delta_2, |z_3|<\delta_2\}} e^{-\sigma z_1 D_{11}^{-1} + D_{22}^{-1} + D_{33}^{-1}|z_1|^2} \pi^3 \sigma^6 \det V \, dz_1 dz_2 dz_3
\]

\[
\geq \left(1 - e^{-\frac{D_{11} d_1^2}{\sigma^4}}\right) \left(1 - e^{-\frac{D_{22} d_1^2}{\sigma^4}}\right) \left(1 - e^{-\frac{D_{33} d_1^2}{\sigma^4}}\right),
\]

and similarly

\[
P(G_2 \cap G_3) = \int_{\{|z_1|<\delta_2, |z_2|<|z_3|<\delta_2\}} e^{-\sigma z_1 D_{11}^{-1} + D_{22}^{-1} + D_{33}^{-1}|z_1|^2} \pi^3 \sigma^6 \det V \, dz_1 dz_2 dz_3
\]

\[
\geq \left(1 - e^{-\frac{D_{11} d_2^2}{\sigma^4}}\right) \left(1 - e^{-\frac{D_{22} d_2^2}{\sigma^4}}\right) \left(1 - e^{-\frac{D_{33} d_2^2}{\sigma^4}}\right).
\]

Next, we are going to estimate the asymptotic behavior of \( D_{11}, D_{22} \) and \( D_{33} \) as \( |a| \) increases. This relies on the estimates of \( D_{11} D_{22} D_{33}, D_{11} + D_{22} + D_{33} \) and \( D_{11}^{-1} + D_{22}^{-1} + D_{33}^{-1} \) as follows. Careful algebraic calculation shows that

\[
\det V = E \left[ (|a|^s \Xi_1 + a_1)^2 \right] E \left[ (|a|^s \Xi_2 + a_2)^2 \right] - E^2 \left[ (|a|^s \Xi_1 + a_1) (|a|^s \Xi_2 + a_2) \right]
\]

\[
+ 2E \left[ (|a|^s \Xi_1 + a_1) E \left[ (|a|^s \Xi_2 + a_2) E \left[ (|a|^s \Xi_1 + a_1) \right] (|a|^s \Xi_2 + a_2) \right] \right]
\]

\[
- E^2 \left[ (|a|^s \Xi_2 + a_2) \right] E \left[ (|a|^s \Xi_1 + a_1)^2 \right] \right] - E^2 \left[ (|a|^s \Xi_1 + a_1 \right] E \left[ (|a|^s \Xi_2 + a_2)^2 \right]
\]

\[
= E \left[ (|a|^s \Xi_1)^2 \right] E \left[ (|a|^s \Xi_2)^2 \right] - E^2 \left[ (|a|^s \Xi_1) \right] (|a|^s \Xi_2)\right]
\]

\[
+ 2E \left[ (|a|^s \Xi_1) \right] E \left[ (|a|^s \Xi_2) \right] E \left[ (|a|^s \Xi_1) \right] \left[ (|a|^s \Xi_2) \right]
\]

\[
- E^2 \left[ (|a|^s \Xi_2) \right] E \left[ (|a|^s \Xi_1)^2 \right] \right] - E^2 \left[ (|a|^s \Xi_1) \right] E \left[ (|a|^s \Xi_2)^2 \right]
\]

\[
= |a|^{4s} \left[ E [\Xi_1^2] E [\Xi_2^2] - E^2 [\Xi_1 \Xi_2] + 2E [\Xi_1] E [\Xi_2] E [\Xi_1 \Xi_2] - E^2 [\Xi_2] E [\Xi_1^2] - E^2 [\Xi_1] E [\Xi_2^2] \right]
\]

\[
\leq |a|^{4s}.
\]

Hence,

\[
D_{11} D_{22} D_{33} = \det \left( V^{-1} \right) \gtrsim |a|^{-4s}.
\] (22)
Similarly, we know
\[
\text{trace}(V^{-1}) = \frac{1}{\det V} \left( 16\pi^4 E \left( |a|^2 \Xi_1 + a_1 \right)^2 \right) \left[ (|a|^2 + a_2)^2 \right] - 16\pi^4 E^2 \left( |a|^2 \Xi_1 + a_1 \right) \left( |a|^2 + a_2 \right) + 4\pi^2 E \left( |a|^2 \Xi_1 + a_2 \right)^2 \\
- 4\pi^2 E^2 \left( |a|^2 \Xi_1 + a_1 \right) + 4\pi^2 E \left( |a|^2 \Xi_1 + a_1 \right)^2 - 4\pi^2 E^2 \left( |a|^2 \Xi_1 + a_1 \right)
\]
\[
\simeq \frac{1}{\det V} \left( E \left( |a|^2 \Xi_1 + a_1 \right)^2 \right) \left[ (|a|^2 + a_2)^2 \right] - E^2 \left( |a|^2 \Xi_1 + a_1 \right) \left( |a|^2 + a_2 \right) \left( |a|^2 \Xi_1 + a_1 \right)^2 \\
\simeq \frac{|a|^{2+2s}}{\det V}.
\]

Therefore,
\[
D_{11} + D_{22} + D_{33} = \text{trace}(V^{-1}) \simeq \frac{|a|^{2+2s}}{\det V}.
\]

Note that
\[
\text{trace}(V) = 1 + 4\pi^2 E \left( |a|^2 \Xi_1 + a_1 \right)^2 + 4\pi^2 E \left( |a|^2 + a_2 \right)^2 \simeq |a|^2,
\]
then
\[
D_{11}^{-1} + D_{22}^{-1} + D_{33}^{-1} \simeq |a|^2.
\]

Equation (22), (23) and (24) imply $D_{11} \geq |a|^{-2s}$, $D_{22} \simeq |a|^{-2s}$ and $D_{33} \simeq |a|^{-2}$. Therefore,
\[
P(G_1 \cap G_3) \geq \left( 1 - e^{-D_{11}^{2s+2}} \right) \left( 1 - e^{-D_{22}^{2s+2}} \right) \left( 1 - e^{-D_{33}^{2s+2}} \right) \\
\simeq \left( 1 - e^{-O(\epsilon^{2-4s} M_a^{-q/(1+q)})} \right) \left( 1 - e^{-O(\epsilon^{-4s} M_a^{-q/(1+q)})} \right) \left( 1 - e^{-O(\epsilon^{-2s} M_a^{-q/(1+q)})} \right).
\]

A similar argument leads to
\[
P(G_2 \cap G_3) \geq \left( 1 - e^{-D_{11}^{2s+2}} \right) \left( 1 - e^{-D_{22}^{2s+2}} \right) \left( 1 - e^{-D_{33}^{2s+2}} \right) \\
\simeq \left( 1 - e^{-O(\epsilon^{2-4s} M_a^{-q/(1+q)})} \right) \left( 1 - e^{-O(\epsilon^{-2s} M_a^{-q/(1+q)})} \right) \left( 1 - e^{-O(\epsilon^{-2s} M_a^{-q/(1+q)})} \right).
\]

By Theorem 3.10 if $(a, b) \in R_{\delta_k} \cap Z_k$ for some $k$, then
\[
P(H_k) \geq P(H_k | G_1 \cap G_3) P(G_1 \cap G_3) = P(G_1 \cap G_3).
\]

Note that $|a| \simeq N_k$ when $(a, b) \in Z_k$, then
\[
P(H_k) \geq \left( 1 - e^{-O(\epsilon^{2-4s} M_a^{-q/(1+q)})} \right) \left( 1 - e^{-O(\epsilon^{-4s} M_a^{-q/(1+q)})} \right) \left( 1 - e^{-O(\epsilon^{-2s} M_a^{-q/(1+q)})} \right).
\]

Similarly, if $(a, b) \in S_{\delta} \cap Z_k$ for some $k$, then
\[
P(J_k) \geq \left( 1 - e^{-O(\epsilon^{2-4s} M_a^{-q/(1+q)})} \right) \left( 1 - e^{-O(\epsilon^{-2s} M_a^{-q/(1+q)})} \right) \left( 1 - e^{-O(\epsilon^{-2s} M_a^{-q/(1+q)})} \right).
\]

These arguments prove $(iv)$ and $(v)$. 

50
Step 2: we now prove the case for a mother wave packet of type $(\epsilon, m)$ such that $m \geq \max \left\{ \frac{2(1+s)}{1-s}, \frac{2}{1-s} + 4 \right\}$. Larger $m$ keeps our approximation errors sufficiently small.

Now $W_e(a, b)$ and $(W_e(a, b), \partial_{b_i} W_e(a, b), \partial_{b_i} W_e(a, b))$ have nearly zero Pseudo-covariance matrices and they are nearly circularly symmetric. Suppose they have covariance matrices $C_1$ and $C_2$, pseudo-covariance matrices $P_1$ and $P_2$, respectively. We can still check that they have zero mean and $C_1 = \sigma^2$ and $C_2 = \sigma^2$, where $V$ is defined in the first step. By the definition of the 2D mother wave packet of type $(\epsilon, m)$ and the same process in the proof of Theorem 2.12 and 2.13 we can obtain a almost similar result:

1. The covariance matrix of $(W_e(a, b), W_e^*(a, b))$ is

$$V_1 = \begin{pmatrix} C_1 & P_1 \\ P_1^* & C_1^* \end{pmatrix},$$

and the distribution of $W_e(a, b)$ is described by

$$e^{-\frac{1}{2}(a_1, a_2)V_1^{-1}(a_1, a_2)^T},$$

which is well approximated by

$$e^{-\frac{c_1|a_1|^2 - \sigma^2}{|a_1|^2 - P_1^* P_1}} = e^{-\sigma^2|a_1|^2} \left( 1 + O\left( \frac{\epsilon |a_1|^2}{\sigma^2|a|^m(1-s)} \right) \right).$$

2. The covariance matrix of $(W_e(a, b), \partial_{b_i} W_e(a, b), W_e^*(a, b), \partial_{b_i} W_e^*(a, b))$ is

$$V_2 = \begin{pmatrix} C_2 & P_2 \\ P_2^* & C_2^* \end{pmatrix},$$

and its distribution is described by the joint probability density

$$e^{-\frac{1}{2}(a_1, a_2, a_3, a_4)^T V_2^{-1}(a_1, a_2, a_3, a_4)^T},$$

which is well approximated by

$$e^{-\sigma^2|a|^2 V^{-1} e^{-\frac{1}{2}(a_1, a_2, a_3, a_4)^T P_1 (a_1, a_2, a_3, a_4)^T}},$$

$$\pi^3 \sqrt{\det V} e^{-\frac{1}{2}(a_1, a_2, a_3, a_4)^T P_1 (a_1, a_2, a_3, a_4)^T \sqrt{\det V}^2 + O\left( \frac{\epsilon|a|^m}{\sigma^2|a|^{m-2}(1-s)} \right)}.$$

where $P_\epsilon$ is a matrix with 2-norm bounded by $O\left( \frac{\epsilon}{\sigma^2|a|^{m-2}(1-s)} \right)$. The only different result is the determinant error bound $O\left( \frac{\epsilon}{|a|^{m-2}(1-s)} \right)$. Since the matrix $V$ here has positive eigenvalues bounded above by $O(|a|^2)$, $O(|a|^2 s)$ and $O(|a|^{2(s-1)})$, then $C_2$ has positive eigenvalues bounded above by $O(\sigma^2|a|^2)$, $O(\sigma^2|a|^{2s})$ and $O(\sigma^2|a|^{2(s-1)})$. Because every entry in $P_2$ is bounded by $O\left( \frac{\sqrt{\epsilon}}{|a|^{m-4}(1-s)} \right)$, then determinant error bound comes from

$$O\left( |a|^2| a|^2 | a|^{2s} |a|^{2(s-1)} \right) = O\left( \frac{\epsilon}{|a|^{m-2}(m+6)s} \right).$$

51
By the same argument in the first step, we can show that there exist a diagonal matrix
\( D = \text{diag}\{D_{11}, D_{22}, D_{33}\} \) and a unitary matrix \( U \) such that
\( V^{-1} = U^*DU \). Furthermore, \( D_{11} \gtrsim |a|^{2(1-s)} \), \( D_{22} \simeq |a|^{-2s} \), and \( D_{33} \simeq |a|^{-2} \). Part (i) is still true by previous theorems.

To conclude Part (ii) to (v), we still need to estimate the probability of those events defined
in the first step, i.e., \( P(G_1) \), \( P(G_2) \), \( P(G_1 \cap G_3) \), \( P(G_2 \cap G_3) \), \( P(H_k) \) and \( P(J_k) \). By the
calculations above, we have

\[
P(G_1) = \int_{|z_1|<|a|^{-s}M_1^{1/(2+2q)}} \frac{e^{-\frac{1}{2}(z_1,z_1)V_1^{-1}(z_1,z_1)^T}}{\pi \sqrt{\det V_1}} dz_1
\]

\[
= \int_{|z_1|<|a|^{-s}M_1^{1/(2+2q)}} \frac{e^{-\sigma^{-2}|z_1|^2}}{\pi \sigma^2} \left( 1 + O \left( \frac{\epsilon |z_1|^2}{\sigma^2 a^m(1-s)} \right) \right) dz_1
\]

\[
= 1 - e^{-|a|^{-2s}M_1^{-q/(1+q)}} + O \left( \frac{\epsilon}{|a|^{m(1-s)}} \right),
\]

and similarly

\[
P(G_2) = 1 - e^{-M_1^{-q/(1+q)}} + O \left( \frac{\epsilon}{|a|^{m(1-s)}} \right).
\]

Hence, we can conclude (ii) and (iii) follows the same proof in the first step. Next, we
look at the last two parts of this theorem.

Recall that we have defined a transform \( z' = Uz \) and introduced notations \( \delta_1 = |a|^{-s}M_1^{1/(2+2q)} \), \( \delta_2 = M_1^{1/(2+3q)} \), \( \delta_3 = (1 + |a|^{1-s}) M_1^{1/(2+3q)} \), \( d_1 = \min\{\delta_2 \sqrt{2}, \delta_3 \} \), and \( d_2 = \min\{\delta_2 \sqrt{2}, \delta_3 \} \) in the first step. Let

\[
g(z) = -\frac{1}{2}(z_1^*, z_2^*, z_3^*, z_1, z_2, z_3)P(z_1, z_2, z_3, z_1^*, z_2^*, z_3^*)^T,
\]

and

\[
\tilde{g}(z') = g(U^*z').
\]
Using the same notations and a similar argument, we have

\[ P(G_1 \cap G_3) \]
\[ = \int_{\{|z_1|<\delta_1,|z_2|+|z_3|<\delta_2^2\}} \frac{e^{-\frac{1}{2}(z_1^T,z_2^T,z_3)^T V_2^{-1}(z_1,z_2,z_3)^T}}{\pi^3 \sqrt{\det V_2}} dz_1 dz_2 dz_3 \]
\[ \geq \int_{\{|z_1|<\delta_1,|z_2|<\frac{\delta_2}{\sqrt{2}},|z_3|<\frac{\delta_2}{\sqrt{2}}\}} \frac{\pi^3 \sigma^6 (\det V)^{\frac{1}{2}} + O\left(\frac{\epsilon}{|a|^{m-2-(m+2)s}}\right)}{e^{-\frac{1}{2}(z_1^T,z_2^T,z_3)^T V_2^{-1}(z_1,z_2,z_3)^T}} dz_1 dz_2 dz_3 \]
\[ = \int_{\{|z_1|<\delta_1,|z_2|<\frac{\delta_2}{\sqrt{2}},|z_3|<\frac{\delta_2}{\sqrt{2}}\}} \frac{e^{-M_a^{-1}(D_{11}|z_1|^2+D_{22}|z_2|^2+D_{33}|z_3|^2)} e^g(z')}{\pi^3 \sigma^6 \det V} dz_1' dz_2' dz_3' + O\left(\frac{\epsilon}{|a|^{m-2-(m+2)s}}\right) \]
\[ \geq \int_{\{|z_1'|<d_1,|z_2'|<d_1,|z_3'|<d_1\}} \frac{e^{-M_a^{-1}(D_{11}|z_1'|^2+D_{22}|z_2'|^2+D_{33}|z_3'|^2)} e^g(z')}{\pi^3 \sigma^6 \det V} dz_1' dz_2' dz_3' + O\left(\frac{\epsilon}{|a|^{m-2-(m+2)s}}\right) \]
\[ = \int_{\{|z_1'|<d_1,|z_2'|<d_1,|z_3'|<d_1\}} \frac{e^{-M_a^{-1}(D_{11}|z_1'|^2+D_{22}|z_2'|^2+D_{33}|z_3'|^2)} e^g(z')}{\pi^3 \sigma^6 \det V} dz_1' dz_2' dz_3' + O\left(\frac{\epsilon}{|a|^{m-2-(m+2)s}}\right) \]
\[ = \left(1 - e^{-\frac{D_{11}d_1^2}{\sigma^2}}\right) \left(1 - e^{-\frac{D_{22}d_1^2}{\sigma^2}}\right) \left(1 - e^{-\frac{D_{33}d_1^2}{\sigma^2}}\right) + O\left(\frac{\epsilon}{|a|^{m-2-(m+2)s}}\right) + O\left(\frac{\epsilon}{|a|^{m-2-(m+2)s}}\right), \]

and similarly

\[ P(G_2 \cap G_3) \]
\[ = \int_{\{|z_1|<\delta_1,|z_2|+|z_3|<\delta_2^2\}} \frac{e^{-\frac{1}{2}(z_1^T,z_2^T,z_3)^T V_2^{-1}(z_1,z_2,z_3)^T}}{\pi^3 \sqrt{\det V_2}} dz_1 dz_2 dz_3 \]
\[ \geq \left(1 - e^{-\frac{D_{11}d_1^2}{\sigma^2}}\right) \left(1 - e^{-\frac{D_{22}d_2^2}{\sigma^2}}\right) \left(1 - e^{-\frac{D_{33}d_2^2}{\sigma^2}}\right) + O\left(\frac{\epsilon}{|a|^{m-2-(m+2)s}}\right) + O\left(\frac{\epsilon}{|a|^{m-2-(m+2)s}}\right). \]

The rest of the proof is exactly the same as the one in the first step and consequently we know this theorem is also true for a mother wave packets of type \((\epsilon, m)\) with \(m \geq \max\left\{\frac{2(1+s)}{1-s}, \frac{2}{1-s} + 4\right\}. \]

\[ \square \]

C Proofs for the theorems in Section 4

C.1 Proof of Theorem 4.7

**Proof.** The first two parts have been proved for the case of mother curvelet of type \((0, m)\) for any \(m\) in [61]. It is straightforward to generalize it to a mother curvelet of type \((\epsilon, m)\) since the proof doesn’t rely on the compact support in the frequency domain. The third part is similar to the second part and the proof only differs in one inequality. We would leave the proof for reader. \[ \square \]

C.2 Proof of Theorem 4.10

**Proof.** The proof of this theorem is nearly identical to the proof of Theorem 3.10. By the definition of 2D general curvelet transform and Lemma 4.8 and 4.9 we know the following
two estimates:
\[ |W_e(a, \theta, b)| \lesssim \sqrt{\epsilon_1} a^{1-\frac{s+t}{2}}, \]
and
\[ |\nabla_b W_e(a, \theta, b)| \lesssim \sqrt{\epsilon_1} (a^{1-\frac{s+t}{2}} + a^{1-\frac{s+t}{2}}). \]  
If \((a, \theta, b) \in R_\delta\), then \(W_g(a, \theta, b) \geq a^{-\frac{s+t}{2}} \delta\). Together with Equation (17), we have
\[ |W_f(a, \theta, b)| \geq a^{-\frac{s+t}{2}} \sqrt{\epsilon}. \]

Hence, \(S_\delta \subset R_\delta \subset R_\epsilon\), where \(R_\epsilon\) is defined in Theorem 4.7 and is a subset of \(\bigcup_{2 \leq k \leq K} Z_k\). So, (i) is true by Theorem 4.7. As for (ii), since \(R_\delta \subset R_\epsilon\), then \((a, \theta, b) \in R_\delta \cap Z_k\) implies \((a, \theta, b) \in R_\epsilon \cap Z_k\) and \(a \simeq N_k\). By Theorem 4.7, we have
\[
\frac{|v_g(a, \theta, b) - N_k \nabla \phi_k(b)|}{|N_k \nabla \phi_k(b)|} \lesssim \sqrt{\epsilon},
\]
when \(N > N_0\). Hence,
\[
\frac{|v_g(a, \theta, b) - N_k \nabla \phi_k(b)|}{|N_k \nabla \phi_k(b)|} \leq \frac{|v_f(a, \theta, b) - N_k \nabla \phi_k(b)|}{|N_k \nabla \phi_k(b)|} + \frac{|v_f(a, \theta, b) - v_g(a, \theta, b)|}{|N_k \nabla \phi_k(b)|} \leq \sqrt{\epsilon} + \sqrt{\epsilon_1} \frac{\delta}{\delta} + \frac{\epsilon_1}{\delta} \frac{a^{-(s+t)/2} + a^{1-(s+t)/2}}{\delta N_k a^{-(s+t)/2}} \leq \sqrt{\epsilon} + \epsilon_1^p,
\]
when \(N > N_0\). With a similar argument, when \((a, \theta, b) \in S_\delta \cap Z_k\), we can show that
\[
\frac{|v_g(a, \theta, b) - N_k \nabla \phi_k(b)|}{|N_k \nabla \phi_k(b)|} \leq \sqrt{\epsilon} \frac{N_k^{(s+t)/2}}{N_k^{(s+t)/2}} + \frac{|\nabla_b W_e(a, \theta, b)|}{|W_g(a, \theta, b)|} + \frac{|\nabla_b W_e(a, \theta, b)|}{|W_g(a, \theta, b)|} \leq \frac{\sqrt{\epsilon}}{N_k^{(s+t)/2}} + \frac{\sqrt{\epsilon_1}}{a^{(s+t)/2}} \frac{a^{-(s+t)/2} + a^{1-(s+t)/2}}{\delta N_k} \leq \frac{\sqrt{\epsilon} + \epsilon_1^p}{N_k^{(s+t)/2}},
\]
when \(N > N_0\). Hence, (iii) is true.

C.3 Proof of Theorem 4.11

Proof. The proof of this theorem is similar to the proof of Theorem 3.11 but notations are much heavier. Some new notations are introduced to simplify the statement:

1. Let \(T_{\theta j}\) denote the \(j\)th element of \(R_\theta A_\alpha \xi + a \cdot u_\theta\).
2. Let \(n = (n_1, n_2)^T = (a \cos \theta, a \sin \theta)^T\).
3. Let \( \sigma^2 = \int_{\mathbb{R}^d} |\tilde{\omega}(\xi)|^2 \tilde{c} (R_{\theta} A_\omega \xi + a \cdot u_\theta) \, d\xi \) and \( \Xi = (\Xi_1, \Xi_2)^T \) be a random vector with a joint probability density function \( \sigma^{-2} |\tilde{\omega}(\xi)|^2 \tilde{c} (R_{\theta} A_\omega \xi + a \cdot u_\theta) \).

4. Let \( g_1(\Xi, n) = a^{t-1} \Xi_1 n_1 - a^{s-1} \Xi_2 n_2, \ g_2(\Xi, n) = a^{t-1} \Xi_1 n_2 - a^{s-1} \Xi_2 n_1, \) and \( g(\Xi, n) = (g_1(\Xi, n), g_2(\Xi, n))^T. \)

5. Let \( \hat{g}_1(\Xi) = a^t \Xi_1, \ \hat{g}_2(\Xi) = a^s \Xi_2, \) and \( \hat{g}(\Xi) = (\hat{g}_1(\Xi), \hat{g}_2(\Xi))^T. \)

We would also prove the case for a mother curvelet of type \((0, m)\) first. The proof for a mother curvelet is of type \((\epsilon, m)\) in the second step is similar, but need to deal with nearly circularly symmetric Gaussian variable. The main difficulty is to estimate the asymptotic behavior of the eigenvalues of the covariance matrices, which will be addressed in Step 1. The trick to deal with nearly circularly symmetric Gaussian variable is exactly the same as used in Theorem 3.11. Hence, we only sketch out Step 2.

Step 1:

Since \( w_{\alpha \omega \theta} \in L^1 \cap C^m \) and \( \nabla_b w_{\alpha \omega \theta} \in L^1 \cap C^{m-1} \cap L^1 \cap C^{m-1}, \) we know \( W_{\beta}(a, \theta, b) = W_f(a, \theta, b) + W_e(a, \theta, b) \) and \( \nabla_b W_{\beta}(a, \theta, b) = \nabla_b W_f(a, \theta, b) + \nabla_b W_e(a, \theta, b) \) are Gaussian random variables. By the properties of zero mean stationary Gaussian processes and the geometric supports of curvelets in the frequency domain, we know \( W_e(a, \theta, b) \) and \( (\nabla_b W_e(a, \theta, b), \partial_{\beta_1} W_e(a, \theta, b), \partial_{\beta_2} W_e(a, \theta, b)) \) have zero pseudo-covariance matrices. Furthermore, the variance of \( W_e(a, \theta, b) \) is \( \sigma^2 \) and the covariance matrix of

\[
(W_e(a, \theta, b), \partial_{\beta_1} W_e(a, \theta, b), \partial_{\beta_2} W_e(a, \theta, b))
\]

is \( \sigma^2 V, \) where \( V \) is an invertible and self-adjoint matrix given below:

\[
\begin{pmatrix}
1 & -2\pi i \mathbb{E} [g_1(\Xi, n) + n_1] & -2\pi i \mathbb{E} [g_2(\Xi, n) + n_2] \\
2\pi i [g_1(\Xi, n) + n_1] & 4\pi^2 \mathbb{E} [(g_1(\Xi, n) + n_1)^2] & 4\pi^2 \mathbb{E} [(g_1(\Xi, n) + n_1)(g_2(\Xi, n) + n_2)] \\
2\pi i [g_2(\Xi, n) + n_2] & 4\pi^2 \mathbb{E} [(g_1(\Xi, n) + n_1)(g_2(\Xi, n) + n_2)] & 4\pi^2 \mathbb{E} [(g_2(\Xi, n) + n_2)^2]
\end{pmatrix}.
\]

Hence, \( W_e(a, \theta, b) \) and \( (W_e(a, \theta, b), \partial_{\beta_1} W_e(a, \theta, b), \partial_{\beta_2} W_e(a, \theta, b)) \) are circularly symmetric and their distributions are described by the probability density functions

\[
e^{-\sigma^{-2}|z_1|^2 / \pi \sigma^{-2}}
\]

and

\[
e^{-\sigma^{-2}z^T V^{-1} z / \pi^2 \sigma^{6} \det V},
\]

where \( z = (z_1, z_2, z_3)^T.\) Part (i) is true by previous theorems. To prove Part (ii) to (v), we need to define the following events

\[
G_1 = \{ |W_e(a, \theta, b)| < a^{-(s+t)/2} M_\alpha^{1/(2+2q)}, \}
\]

\[
G_2 = \{ |W_e(a, \theta, b)| < M_\alpha^{1/(2+2q)}, \}
\]

\[
G_3 = \{ |\nabla_b W_e(a, \theta, b)| < M_\alpha^{1/(2+2q)} \left( a^{(t-s)/2} + a^{1-(s+t)/2} \right), \}
\]

\[
H_k = \left\{ \frac{|\nabla_b a, \theta, b| - N_k \nabla_b \phi_k(b)}{|N_k \nabla_b \phi_k(b)|} \leq \sqrt{c} + M_\alpha^{p/(1+q)} \right\},
\]

55
and

\[
J_k = \left\{ \frac{|v_1(a, \theta, b) - N_k \nabla_\theta \phi_k(b)|}{|N_k \nabla_\theta \phi_k(b)|} \leq N_k^{- (s+t)/2} \left( \sqrt{e} + M_a^{p/(1+q)} \right) \right\},
\]

for \(1 \leq k \leq K\). Next, we are going to estimate the probability \(P(G_1), P(G_2), P(G_1 \cap G_3), P(G_2 \cap G_3), P(H_k)\) and \(P(J_k)\). By the calculations above, we have

\[
P(G_1) = \int_{|z_1| < a^{-(s+t)/2} M_a^{1/(2+2q)}} \frac{e^{-\sigma^2 |z_1|^2}}{\pi \sigma^2} dz_1 \geq 1 - e^{-a^{-(s+t)} M_a^{-q/(1+q)}}
\]

and similarly

\[
P(G_2) = \int_{|z_1| < M_a^{1/(2+2q)}} \frac{e^{-\sigma^2 |z_1|^2}}{\pi \sigma^2} dz_1 \geq 1 - e^{-M_a^{-q/(1+q)}}.
\]

We are now ready to conclude (\(ii\)) and (\(iii\)). If \((a, \theta, b) \in R_{\delta_a}\), then

\[
|W_\varepsilon(a, \theta, b) + W_f(a, \theta, b)| \geq a^{-(s+t)/2} \left( M_a^{1/2-p}/(1+q) + \sqrt{\varepsilon} \right).
\]

If \((a, \theta, b) \notin \bigcup_{1 \leq k \leq K} Z_k\), then by Lemma 4.9

\[
|W_f(a, \theta, b)| \leq a^{-(s+t)/2} \varepsilon.
\]

Equation (28) and (29) lead to \(|W_\varepsilon(a, \theta, b)| \geq a^{-(s+t)/2} M_a^{1/(2+2q)}\). Hence,

\[
P \left( (a, \theta, b) \notin \bigcup_{1 \leq k \leq K} Z_k \right) \leq P \left( |W_\varepsilon(a, \theta, b)| \geq a^{-(s+t)/2} M_a^{1/(2+2q)} \right) = 1 - P(G_1).
\]

This means that if \((a, \theta, b) \in R_{\delta_a}\) then \((a, \theta, b) \in \bigcup_{1 \leq k \leq K} Z_k\) with a probability at least

\[
P(G_1) \geq 1 - e^{-a^{-(s+t)} M_a^{-q/(1+q)}} = 1 - e^{-O \left( N_k^{- (s+t)} M_a^{-q/(1+q)} \right)},
\]

since \(a \simeq N_k\) if \((a, \theta, b) \in Z_k\). So, (\(ii\)) is true. A similar argument applied to \((a, \theta, b) \in S_{\delta_b}\) shows that \((a, \theta, b) \in \bigcup_{1 \leq k \leq K} Z_k\) with a probability at least \(P(G_2) \geq 1 - e^{-M_a^{-q/(1+q)}}\). Hence, (\(iii\)) is proved.

Because \(V\) is invertible and self-adjoint, there exist a unitary matrix \(U\) and a diagonal matrix \(D\) such that \(V^{-1} = U^*DU\). For \(z \in \mathbb{C}^3\), let \(z' = Uz\). Introduce notations \(\delta_1 = a^{-(s+t)} M_a^{1/(2+2q)}, \delta_2 = M_a^{1/(2+2q)}, \delta_3 = (a^{(s-t)/2} + a^{1-(s+t)/2}) M_a^{1/(2+2q)}, \)

\(d_1 = \min \{ \frac{\delta_1}{\sqrt{2}}, \frac{\delta_3}{2} \}, \text{ and } d_2 = \min \{ \frac{\delta_2}{\sqrt{2}}, \frac{\delta_3}{2} \}\). Similar to the proof in Theorem 3.11 by a simple property of high dimensional polydisk, we have

\[
P(G_1 \cap G_3) = \int_{\{|z_1| < \delta_1, |z_2| + |z_3| < \delta_3\}} \frac{e^{-\sigma^2 z^* V^{-1} z}}{\pi^3 \sigma^6 \det V} dz_1 dz_2 dz_3
\]

\[
\geq \left( 1 - e^{-\frac{D_{11} d_1^2}{\sigma^2}} \right) \left( 1 - e^{-\frac{D_{22} d_1^2}{\sigma^2}} \right) \left( 1 - e^{-\frac{D_{33} d_1^2}{\sigma^2}} \right),
\]

and similarly

\[
P(G_2 \cap G_3) = \int_{\{|z_1| < \delta_2, |z_2| + |z_3| < \delta_3\}} \frac{e^{-\sigma^2 z^* V^{-1} z}}{\pi^3 \sigma^6 \det V} dz_1 dz_2 dz_3
\]

\[
\geq \left( 1 - e^{-\frac{D_{11} d_2^2}{\sigma^2}} \right) \left( 1 - e^{-\frac{D_{22} d_2^2}{\sigma^2}} \right) \left( 1 - e^{-\frac{D_{33} d_2^2}{\sigma^2}} \right).
\]
Next, we are going to estimate the asymptotic behavior of $D_{11}$, $D_{22}$ and $D_{33}$ as $a$ increases. This relies on the estimates of $D_{11}D_{22}D_{33}$, $D_{11} + D_{22} + D_{33}$ and $D_{11}^{-1} + D_{22}^{-1} + D_{33}^{-1}$ as follows. Careful algebraic calculation shows that

$$
\det V = \frac{1}{16\pi^4} \left( g_1(\Xi, n) + n_1 \right)^2 E \left[ (g_2(\Xi, n) + n_2)^2 \right] - E^2 \left[ (g_1(\Xi, n) + n_1) (g_2(\Xi, n) + n_2) \right] + 2E \left[ g_1(\Xi, n) + n_1 \right] E \left[ g_2(\Xi, n) + n_2 \right] E \left[ (g_1(\Xi, n) + n_1) (g_2(\Xi, n) + n_2) \right] - E^2 \left[ g_2(\Xi, n) + n_2 \right] E \left[ (g_1(\Xi, n) + n_1)^2 \right] - E^2 \left[ g_1(\Xi, n) + n_1 \right] E \left[ (g_2(\Xi, n))^2 \right] = E \left[ (g_1(\Xi))^2 \right] E \left[ (g_2(\Xi))^2 \right] - E^2 \left[ (g_1(\Xi)) (g_2(\Xi)) \right] + 2E \left[ g_1(\Xi) \right] E \left[ g_2(\Xi) \right] E \left[ (g_1(\Xi)) (g_2(\Xi)) \right] - E^2 \left[ g_2(\Xi) \right] E \left[ (g_1(\Xi))^2 \right] - E^2 \left[ g_1(\Xi) \right] E \left[ (g_2(\Xi))^2 \right] = a^{2(s+t)} (E \left[ \Omega_1^2 \right] E \left[ \Omega_2^2 \right] - E^2 \left[ \Omega_1 \Omega_2 \right] + 2E \left[ \Omega_1 \right] E \left[ \Omega_2 \right] E \left[ \Omega_1 \Omega_2 \right] - E^2 \left[ \Omega_2 \right] E \left[ \Omega_1^2 \right] - E^2 \left[ \Omega_1 \right] E \left[ \Omega_2^2 \right] ) \approx a^{2(s+t)}
$$

Hence,

$$
D_{11}D_{22}D_{33} = \det (V^{-1}) \gtrsim a^{-2(s+t)}.
$$

Similarly, we know

$$
\text{tr} (V^{-1}) = \frac{1}{\det V} \left( 16\pi^4 E \left[ (g_1(\Xi, n) + n_1)^2 \right] E \left[ (g_2(\Xi, n) + n_2)^2 \right] - 16\pi^4 E^2 \left[ (g_1(\Xi, n) + n_1) (g_2(\Xi, n) + n_2) \right] + 4\pi^2 E \left[ (g_1(\Xi, n) + n_1)^2 \right] - 4\pi^2 E^2 \left[ (g_1(\Xi, n) + n_1) \right] + 4\pi^2 E \left[ (g_2(\Xi, n) + n_2)^2 \right] - 4\pi^2 E^2 \left[ (g_2(\Xi, n) + n_2) \right] \right) \approx \frac{1}{a^{2s+2s}} \approx \frac{1}{\det V}.
$$

Therefore,

$$
D_{11} + D_{22} + D_{33} = \text{tr} (V^{-1}) \approx a^{2(s+2s)} \det V.
$$

Note that

$$
\text{tr} (V) = 1 + 4\pi^2 E \left[ (g_1(\Xi, n) + n_1)^2 \right] + 4\pi^2 E \left[ (g_2(\Xi, n) + n_2)^2 \right] \approx a^2,
$$

then

$$
D_{11}^{-1} + D_{22}^{-1} + D_{33}^{-1} \approx a^2.
$$

Equation (30), (31) and (32) imply $D_{11} \gtrsim a^{2-2t}$, $D_{22} \approx a^{-2s}$ and $D_{33} \approx a^{-2}$. Therefore,

$$
P (G_1 \cap G_3) \geq \left( 1 - e^{-\frac{D_{11}q}{a^2}} \right) \left( 1 - e^{-\frac{D_{22}q}{a^2}} \right) \left( 1 - e^{-\frac{D_{33}q}{a^2}} \right) = \left( 1 - e^{-O(a^{2-s-2t}M^{-q/(1+q)})} \right) \left( 1 - e^{-O(a^{-3s-t}M^{-q/(1+q)})} \right) \left( 1 - e^{-O(a^{-s-t}M^{-q/(1+q)})} \right).
$$
A similar argument leads to

$$P(G_2 \cap G_3) \geq \left(1 - e^{-\frac{B_{11}d_1^2}{\sigma^2}}\right) \left(1 - e^{-\frac{D_{22}d_2^2}{\sigma^2}}\right) \left(1 - e^{-\frac{D_{33}d_3^2}{\sigma^2}}\right)$$

$$= \left(1 - e^{-O\left(a^{2-2t}M_\alpha^{-q/(1+q)}\right)}\right) \left(1 - e^{-O\left(a^{-2s}M_\epsilon^{-q/(1+q)}\right)}\right) \left(1 - e^{-O\left(a^{-2s}M_\alpha^{-q/(1+q)}\right)}\right).$$

By Theorem 4.10 if \((a, \theta, b) \in R_{\delta_a} \cap Z_k\) for some \(k\), then

$$P(H_k) \geq P(H_k|G_1 \cap G_3) P(G_1 \cap G_3) = P(G_1 \cap G_3).$$

Note that \(a \simeq N_k\) when \((a, \theta, b) \in Z_k\), then

$$P(H_k) \geq \left(1 - e^{-O\left(a^{2-s-3t}M_\alpha^{-q/(1+q)}\right)}\right) \left(1 - e^{-O\left(a^{-3s-t}M_\alpha^{-q/(1+q)}\right)}\right) \left(1 - e^{-O\left(a^{-s-t-2}M_\alpha^{-q/(1+q)}\right)}\right)$$

$$= \left(1 - e^{-O\left(N_k^{2-s-3t}M_\alpha^{-q/(1+q)}\right)}\right) \left(1 - e^{-O\left(N_k^{-3s-t}M_\alpha^{-q/(1+q)}\right)}\right) \left(1 - e^{-O\left(N_k^{-s-t-2}M_\alpha^{-q/(1+q)}\right)}\right).$$

Similarly, if \((a, \theta, b) \in S_{\delta_b} \cap Z_k\) for some \(k\), then

$$P(J_k) \geq P(G_2 \cap G_3) \geq \left(1 - e^{-O\left(N_k^{2-t}M_\alpha^{-q/(1+q)}\right)}\right) \left(1 - e^{-O\left(N_k^{-2s}M_\alpha^{-q/(1+q)}\right)}\right) \left(1 - e^{-O\left(N_k^{-2s}M_\alpha^{-q/(1+q)}\right)}\right).$$

These arguments prove (iv) and (v).

Step 2: we now prove the case for a mother curvelet of type \((\epsilon, m)\) such that \(m \geq \max\left\{2^{(1+\epsilon)}, \frac{2}{1-t}, \frac{2}{1-t} + 4\right\}\). Sufficiently large \(m\) keeps our approximation errors small enough.

Now \(W_\epsilon(a, \theta, b)\) and \((W_\epsilon(a, \theta, b), \partial_{b_1}W_\epsilon(a, \theta, b), \partial_{b_2}W_\epsilon(a, \theta, b))\) have nearly zero Pseudo-Covariance matrices and they are nearly circularly symmetric. Suppose \(W_\epsilon(a, \theta, b)\) and \((W_\epsilon(a, \theta, b), \partial_{b_1}W_\epsilon(a, \theta, b), \partial_{b_2}W_\epsilon(a, \theta, b))\) have covariance matrices \(C_1\) and \(C_2\), pseudo-Covariance matrices \(P_1\) and \(P_2\), respectively. We can still check that they have zero mean and \(C_1 = \sigma^2\) and \(C_2 = \sigma^2V\), where \(V\) is defined in the first step. By the definition of the 2D mother wave packet of type \((\epsilon, m)\) and the same process in the proof of Theorem 2.12 and 2.13, we can obtain a almost similar result:

1. The covariance matrix of \((W_\epsilon(a, \theta, b), W_\epsilon^*(a, \theta, b))\) is

$$V_1 = \begin{pmatrix} C_1 & P_1 \\ P_1^* & C_1^* \end{pmatrix},$$

and the distribution of \(W_\epsilon(a, \theta, b)\) is described by

$$e^{-\frac{1}{2}(z_1^*; z_1) V_1^{-1}(z_1, z_1)^T}{\pi^{\sqrt{\det V_1}}},$$

which is well approximated by

$$e^{-\frac{C_1|z_1|^2 - \Re(P_1 z_1^2)}{C_1^2 - P_1 P_1^*}} e^{-\frac{|z_1|^2}{2\sigma^2}} = e^{\frac{\epsilon |z_1|^2}{\sigma^2\epsilon(1-t)}} \left(1 + O\left(\frac{\epsilon |z_1|^2}{\sigma^2\epsilon(1-t)}\right)\right).$$

58
2. The covariance matrix of \((W_e(a, \theta, b), \partial_b W_e(a, \theta, b), W_e^* (a, \theta, b), \partial_b W_e^* (a, \theta, b))\) is

\[
V_2 = \begin{pmatrix} C_2 & P_2 \\ P_2^* & C_2^* \end{pmatrix},
\]

and its distribution is described by the joint probability density

\[
e^{-\frac{1}{2}(z_1^*, z_2^*, z_3^*, z_1, z_2, z_3)V_2^{-1}(z_1^*, z_2^*, z_3^*, z_1, z_2, z_3)^T} \pi^3 \sqrt{\det V_2},
\]

which is well approximated by

\[
e^{-\sigma^2 z^* V^{-1} z} e^{-\frac{1}{2}(z_1^*, z_2^*, z_3^*, z_1, z_2, z_3)P_e(z_1, z_2, z_3, z_1^*, z_2^*, z_3^*)^T} \pi^3 \sqrt{(\det V)^2 + O\left(\frac{\epsilon}{\sigma a^{(m-4)(1-t)}}\right)},
\]

where \(P_e\) is a matrix with 2-norm bounded by \(O\left(\frac{\epsilon}{\sigma a^{(m-4)(1-t)}}\right)\). Different to Theorem 3.11, here we have two scaling parameters \(t\) and \(s\) with \(s < t\). To understand those error bounds above intuitively, we could say that \(s\) is shrinking the support of a wave packet in the frequency domain in the angular direction to make it a curvelet and hence to increase the probability of a good estimate, as we have seen that smaller parameters resulting in better robustness. Hence, Most of the error bound above is determined by \(t\), the larger one.

Since the matrix \(V\) here has positive eigenvalues bounded above by \(O(a^2)\), \(O(a^{2s})\) and \(O(a^{2(t-1)})\), then \(C_2\) has positive eigenvalues bounded above by \(O(\sigma^2 a^2)\), \(O(\sigma^2 a^{2s})\) and \(O(\sigma^2 a^{2(t-1)})\). Because every entry in \(P_2\) is bounded by \(O\left(\frac{\epsilon}{a^{m(1-t)}}\right)\), then determinant error bound comes from

\[
O\left(\frac{a^2 a^2 a^{2s} a^{2(t-1)}}{a^{m-4)(1-t)}}\right) = O\left(\frac{\epsilon}{a^{m-2-(m+2)4t}}\right).
\]

By the same argument in the first step, we can show that there exist a diagonal matrix \(D = \text{diag}\{D_{11}, D_{22}, D_{33}\}\) and a unitary matrix \(U\) such that \(V^{-1} = U^* D U\). Furthermore, \(D_{11} \geq a^{2(1-t)}\), \(D_{22} \simeq a^{-2s}\), \(D_{33} \simeq a^{-2}\). Part (i) is still true by previous theorems. To conclude Part (ii) to (v), we still need to estimate the probability of those events defined in the first step, i.e., \(P (G_1)\), \(P (G_2)\), \(P (G_1 \cap G_3)\), \(P (G_2 \cap G_3)\), \(P (H_k)\) and \(P (J_k)\). By the calculations above, we have

\[
P(G_1) = \int_{|z_1| < a^{-(s+t)/2} M_a^{1/(2+2q)}} e^{-\frac{1}{2}(z_1^*, z_1) V_1^{-1}(z_1^*, z_1)^T} \pi \sqrt{\det V_1} d z_1
\]

\[
= \int_{|z_1| < a^{-(s+t)/2} M_a^{1/(2+2q)}} e^{-\sigma^2 |z_1|^2} \pi \sigma^2 \left(1 + O\left(\frac{\epsilon |z_1|^2}{a^{m(1-t)}}\right)\right) d z_1
\]

\[
= 1 - e^{-a^{-(s+t)} M_a^{-q/(1+q)}} + O\left(\frac{\epsilon}{a^{m(1-t)}}\right),
\]

and similarly

\[
P(G_2) = 1 - e^{-M_a^{-q/(1+q)}} + O\left(\frac{\epsilon}{a^{m(1-t)}}\right).
\]
Hence, we can conclude (ii) and (iii) follows the same proof in the first step. Next, we look at the last two parts of this theorem.

Recall that we have defined a transform \( z' = Uz \) and introduced notations \( \delta_1 = a^{-(s+t)/2}M_a^{1/2(2+2q)}, \delta_2 = M_a^{1/(2+2q)}, \delta_3 = (a^{(t-s)/2} + a^{-(s+t)/2}) M_a^{1/(2+2q)}, d_1 = \min\left\{ \frac{\delta_1}{\sqrt{2}}, \frac{\delta_2}{\sqrt{2}} \right\}, \) and \( d_2 = \min\left\{ \frac{\delta_2}{\sqrt{2}}, \frac{\delta_3}{\sqrt{2}} \right\} \) in the first step. Let

\[
g(z) = -\frac{1}{2}(z_1^*, z_2^*, z_3^*, z_1, z_2, z_3)P_t(z_1, z_2, z_3, z_1^*, z_2^*, z_3^*)^T,
\]

and

\[
\tilde{g}(z') = g(U^*z').
\]

Using the same notations and a similar argument, we have

\[
P(G_1 \cap G_3) = \int_{\{|z_1|<\delta_1, |z_2|+|z_3|<\delta_2^3\}} e^{-\frac{1}{2}(z_1^2, z_2^2, z_3^2, z_1, z_2, z_3)\left[ \frac{1}{2}V_2^{-1}(z_1, z_2, z_3, z_1^*, z_2^*, z_3^*)^T d_1 d_2 d_3 \right.} \left. \frac{1}{\pi^3 \sqrt{|\det V_2|}} d_1 d_2 d_3 \rangle \langle e^{-\sigma^2 z^T V^{-1} z} g(z) \right)
\]

\[
\geq \int_{\{|z_1|<\delta_1, |z_2|<\frac{\delta_1}{\sqrt{2}}, |z_3|<\frac{\delta_2}{\sqrt{2}}\}} e^{-\frac{1}{2}M_a^{-1} D(D_11|z_1|^2 + D_22|z_2|^2 + D_33|z_3|^2)} e^{\tilde{g}(z')} \left. \frac{1}{\pi^3 \sigma^6 \sqrt{|\det V|} (|\det V|)^2 + O \left( \frac{\epsilon}{a(m-2-mt-2s)} \right) \right) d_1 d_2 d_3 + O \left( \frac{\epsilon}{a(m-2-mt-2s)} \right)
\]

\[
\geq \int_{\{|z_1|<\delta_1, |z_2|<\frac{\delta_1}{\sqrt{2}}, |z_3|<\frac{\delta_2}{\sqrt{2}}\}} e^{-\frac{1}{2}M_a^{-1} D(D_11|z_1|^2 + D_22|z_2|^2 + D_33|z_3|^2)} e^{\tilde{g}(z')} \left. \frac{1}{\pi^3 \sigma^6 \sqrt{|\det V|} (|\det V|)^2 + O \left( \frac{\epsilon}{a(m-2-mt-2s)} \right) \right) d_1 d_2 d_3 + O \left( \frac{\epsilon}{a(m-2-mt-2s)} \right)
\]

\[
\geq \int_{\{|z_1|<\delta_1, |z_2|<\frac{\delta_1}{\sqrt{2}}, |z_3|<\frac{\delta_2}{\sqrt{2}}\}} e^{-\frac{1}{2}M_a^{-1} D(D_11|z_1|^2 + D_22|z_2|^2 + D_33|z_3|^2)} e^{\tilde{g}(z')} \left. \frac{1}{\pi^3 \sigma^6 \sqrt{|\det V|} (|\det V|)^2 + O \left( \frac{\epsilon}{a(m-2-mt-2s)} \right) \right) d_1 d_2 d_3 + O \left( \frac{\epsilon}{a(m-2-mt-2s)} \right)
\]

\[
= \left( 1 - e^{-\frac{D_{11}^2}{\sigma^2}} \right) \left( 1 - e^{-\frac{D_{22}^2}{\sigma^2}} \right) \left( 1 - e^{-\frac{D_{33}^2}{\sigma^2}} \right) + O \left( \frac{\epsilon}{a(m-4)(1-t-2)} \right) + O \left( \frac{\epsilon}{a(m-2-mt-2s)} \right),
\]

and similarly

\[
P(G_2 \cap G_3) = \int_{\{|z_1|<\delta_2, |z_2|+|z_3|<\delta_3^3\}} e^{-\frac{1}{2}(z_1^2, z_2^2, z_3^2, z_1, z_2, z_3)\left[ \frac{1}{2}V_2^{-1}(z_1, z_2, z_3, z_1^*, z_2^*, z_3^*)^T d_1 d_2 d_3 \right.} \left. \frac{1}{\pi^3 \sqrt{|\det V_2|}} d_1 d_2 d_3 \rangle \langle e^{-\sigma^2 z^T V^{-1} z} g(z) \right)
\]

\[
\geq \int_{\{|z_1|<\delta_2, |z_2|<\frac{\delta_1}{\sqrt{2}}, |z_3|<\frac{\delta_2}{\sqrt{2}}\}} e^{-\frac{1}{2}M_a^{-1} D(D_11|z_1|^2 + D_22|z_2|^2 + D_33|z_3|^2)} e^{\tilde{g}(z')} \left. \frac{1}{\pi^3 \sigma^6 \sqrt{|\det V|} (|\det V|)^2 + O \left( \frac{\epsilon}{a(m-2-mt-2s)} \right) \right) d_1 d_2 d_3 + O \left( \frac{\epsilon}{a(m-2-mt-2s)} \right)
\]

\[
= \left( 1 - e^{-\frac{D_{11}^2}{\sigma^2}} \right) \left( 1 - e^{-\frac{D_{22}^2}{\sigma^2}} \right) \left( 1 - e^{-\frac{D_{33}^2}{\sigma^2}} \right) + O \left( \frac{\epsilon}{a(m-4)(1-t-2)} \right) + O \left( \frac{\epsilon}{a(m-2-mt-2s)} \right).
\]

The rest of the proof is exactly the same as the one in the first step and consequently we know this theorem is also true for a mother wave packets of type \((\epsilon, m)\) with \(m\) larger than \(\max\left\{ \frac{2(1+s)}{1-t}, \frac{2}{1-t} + 4 \right\} \).
References

[1] http://users.ece.utexas.edu/~bevans/projects/rfc/software/

[2] G. Ariel, B. Engquist, N. M. Tanushev, and R. Tsai. Gaussian beam decomposition of high frequency wave fields using expectation-maximization. *J. Comput. Phys.*, 230(6):2303–2321, Mar. 2011.

[3] G. Ariel, B. Engquist, N. M. Tanushev, and R. Tsai. Gaussian beam decomposition of high frequency wave fields using expectation-maximization. *J. Comput. Physics*, pages 2303–2321, 2011.

[4] F. Auger and P. Flandrin. Improving the readability of time-frequency and time-scale representations by the reassignment method. *Signal Processing, IEEE Transactions on*, 43(5):1068–1089, 1995.

[5] F. Auger, P. Flandrin, Y.-T. Lin, S. McLaughlin, S. Meignen, T. Oberlin, and H.-T. Wu. Time-frequency reassignment and synchrosqueezing: An overview. *Signal Processing Magazine, IEEE*, 30(6):32–41, Nov 2013.

[6] B. Boashash and S. Member. Estimating and interpreting the instantaneous frequency of a signal. In *Proceedings of the IEEE*, pages 520–538, 1992.

[7] P. J. Brockwell and R. A. Davis. *Introduction to Time Series and Forecasting*. Springer, 2nd edition, Mar. 2002.

[8] M. Brown and R. Clapp. (t, x) domain, patternbased ground roll removal.

[9] T. T. Cai. Adaptive wavelet estimation: A block thresholding and oracle inequality approach. *Ann. Statist*, 27:898–924, 1998.

[10] T. T. Cai and H. H. Zhou. A data-driven block thresholding approach to wavelet estimation. *The Annals of Statistics*, 37(2):569–595, 04 2009.

[11] E. Candès, L. Demanet, D. Donoho, and L. Ying. Fast discrete curvelet transforms. *Multiscale Model. Simul.*, 5(3):861–899, 2006.

[12] E. J. Candès and D. L. Donoho. Continuous curvelet transform. I. Resolution of the wavefront set. *Appl. Comput. Harmon. Anal.*, 19(2):162–197, 2005.

[13] E. J. Candès and D. L. Donoho. Continuous curvelet transform. II. Discretization and frames. *Appl. Comput. Harmon. Anal.*, 19(2):198–222, 2005.

[14] E. J. Candès and D. L. Donoho. Continuous curvelet transform. II. Discretization and frames. *Appl. Comput. Harmon. Anal.*, 19(2):198–222, 2005.

[15] E. J. Candès, P. R. Charlton, and H. Helgason. Detecting highly oscillatory signals by chirplet path pursuit. *Applied and Computational Harmonic Analysis*, 24(1):14 – 40, 2008.

[16] E. Chassande-Mottin, F. Auger, and P. Flandrin. Time-frequency/time-scale reassignment. In *Wavelets and signal processing*, Appl. Numer. Harmon. Anal., pages 233–267. Birkhäuser Boston, Boston, MA, 2003.
[17] E. Chassande-Mottin, I. Daubechies, F. Auger, and P. Flandrin. Differential reassignment. *Signal Processing Letters, IEEE*, 4(10):293–294, 1997.

[18] E. Chassande-Mottin and P. Flandrin. On the timefrequency detection of chirps1. *Applied and Computational Harmonic Analysis*, 6(2):252–281, 1999.

[19] E. Chassande-Mottin and A. Pai. Best chirplet chain: Near-optimal detection of gravitational wave chirps. *Physical Review D*, 73(4), 2006.

[20] C. K. Chui, Y.-T. Lin, and H.-t. Wu. Real-time dynamics acquisition from irregular samples – with application to anesthesia evaluation. *arXiv:1406.1276 [math.NA]*, 2014.

[21] 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 [math.NA]*, Submitted on 20 Nov 2012.

[22] I. Daubechies, J. Lu, and H.-T. Wu. Synchrosqueezed wavelet transforms: an empirical mode decomposition-like tool. *Appl. Comput. Harmon. Anal.*, 30(2):243–261, 2011.

[23] I. Daubechies and S. Maes. A nonlinear squeezing of the continuous wavelet transform based on auditory nerve models. In *Wavelets in Medicine and Biology*, pages 527–546. CRC Press, 1996.

[24] A. M. De Livera and R. J. Hyndman. Forecasting time series with complex seasonal patterns using exponential smoothing. Monash Econometrics and Business Statistics Working Papers 15/09, Monash University, Department of Econometrics and Business Statistics, 2009.

[25] L. Demanet and L. Ying. Wave atoms and sparsity of oscillatory patterns. *Appl. Comput. Harmon. Anal.*, 23(3):368–387, 2007.

[26] D. L. Donoho. De-noising by soft-thresholding. *IEEE Trans. Inf. Theor.*, 41(3):613–627, may 1995.

[27] D. L. DONOHO and J. M. JOHNSTONE. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81(3):425–455, 1994.

[28] S. Fomel. Applications of plane-wave destruction filters. *Geophysics*, 2002.

[29] S. Fomel. Seismic data decomposition into spectral components using regularized nonstationary autoregression. *GEOPHYSICS*, 78(6):O69–O76, 2013.

[30] R. G. Gallager. Circularly-symmetric gaussian random vectors. *preprint*, 2008.

[31] R. Herrera, J. Han, and M. van der Baan. Applications of the synchrosqueezing transform in seismic time-frequency analysis. *Geophysics*, 79(3):V55–V64, 2014.

[32] T. Hou, Z. Shi, and P. Tavallali. Nonparametric and adaptive modeling of dynamic seasonality and trend with heteroscedastic and dependent errors. *arXiv:1210.4672 [math.ST]*, 2012.

[33] T. Hou, Z. Shi, and P. Tavallali. Convergence of a data-driven time-frequency analysis method. *arXiv:1303.7048 [math.NA]*, 2013.
[34] T. Hou, Z. Shi, and P. Tavallali. Sparse time frequency representations and dynamical systems. arXiv:1312.0202, 2013.

[35] T. Y. Hou and Z. Shi. Data-driven timefrequency analysis. Applied and Computational Harmonic Analysis, 35(2):284 – 308, 2013.

[36] 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(1971):903–995, 1998.

[37] N. E. Huang, Z. Wu, S. R. Long, K. C. Arnold, X. Chen, and K. Blank. On instantaneous frequency. Adv. Adapt. Data Anal., 1(2):177–229, 2009.

[38] W. Huang, Z. Shen, N. E. Huang, and Y. C. Fung. Engineering analysis of biological variables: An example of blood pressure over 1 day. Proc. Natl. Acad. Sci., 95, 1998.

[39] H.-H. Kuo. White noise distribution theory. Probability and stochastics series. CRC Press, Boca Raton (USA), 1996.

[40] C. Li and M. Liang. A generalized synchrosqueezing transform for enhancing signal timefrequency representation. Signal Processing, 92(9):2264 – 2274, 2012.

[41] F. Neeser and J. Massey. Proper complex random processes with applications to information theory. Information Theory, IEEE Transactions on, 39(4):1293–1302, Jul 1993.

[42] B. Picinbono. On instantaneous amplitude and phase of signals. IEEE Trans. Signal Processing, pages 552–560, 1997.

[43] S. Richwalski, K. Roy-Chowdhury, and J. C. Mondt. Multi-component wavefield separation applied to high-resolution surface seismic data. Journal of Applied Geophysics, 46(2):101–114, 2001.

[44] H. M. Singer and I. Singer. Analysis and visualization of multiply oriented lattice structures by a two-dimensional continuous wavelet transform. Phys. Rev. E, 74:031103, 2006.

[45] J. B. Tary, R. H. Herrera, J. Han, and M. van der Baan. Spectral estimation - what's next? what's new? Reviews of Geophysics, Submitted.

[46] M. Taylor. Random fields: stationarity, ergodicity, and spectral behavior, www.unc.edu/math/Faculty/met/rndfcn.pdf.

[47] G. Thakur, E. Brevdo, N. S. Fučkar, and H.-T. Wu. The synchrosqueezing algorithm for time-varying spectral analysis: robustness properties and new paleoclimate applications. Signal Processing, 93(5):1079–1094, 2013.

[48] G. Thakur and H.-T. Wu. Synchrosqueezing-based recovery of instantaneous frequency from nonuniform samples. SIAM J. Math. Analysis, 43(5):2078–2095, 2011.

[49] A. Van Den Bos. The multivariate complex normal distribution-a generalization. Information Theory, IEEE Transactions on, 41(2):537–539, Mar 1995.
[50] M. van der Baan. http://www.ualberta.ca/~vanderba/ftp/benchmark_signals.zip.

[51] M. van der Baan. Pp/ps wavefield separation by independent component analysis. *Geophys. J. Int.*, (166):339–348, 2006.

[52] A. D. Veltcheva. Wave and group transformation by a hilbert spectrum. *Coastal Engineering Journal*, 44(4), 2002.

[53] H.-T. Wu. Instantaneous frequency and wave shape functions (i). *Applied and Computational Harmonic Analysis*, 35(2):181 – 199, 2013.

[54] H.-T. Wu, Y.-H. Chan, Y.-T. Lin, and Y.-H. Yeh. Using synchrosqueezing transform to discover breathing dynamics from \{ECG\} signals. *Applied and Computational Harmonic Analysis*, 36(2):354 – 359, 2014.

[55] H.-T. Wu, P. Flandrin, and I. Daubechies. One or two frequencies? The synchrosqueezing answers. *Adv. Adapt. Data Anal.*, 3(1-2):29–39, 2011.

[56] H.-T. Wu, S.-S. Hseu, M.-Y. Bien, Y. R. Kou, and I. Daubechies. Evaluating physiological dynamics via synchrosqueezing: Prediction of ventilator weaning. *Biomedical Engineering, IEEE Transactions on*, 61(3):736–744, March 2014.

[57] Z. Wu, N. E. Huang, and X. Chen. Some considerations on physical analysis of data. *Advances in Adaptive Data Analysis*, 3(1-2):95–113, 2011.

[58] H. Yang. Synchrosqueezed wave packet transforms and diffeomorphism based spectral analysis for 1d general mode decompositions. *Applied and Computational Harmonic Analysis*, (0):–, 2014.

[59] H. Yang, J. Lu, and L. Ying. Crystal image analysis using 2d synchrosqueezed transforms. *arXiv:1402.1262 [math.NA]*, Submitted on 6 Feb 2014.

[60] H. Yang and L. Ying. Synchrosqueezed wave packet transform for 2d mode decomposition. *SIAM Journal on Imaging Sciences*, 6(4):1979–2009, 2013.

[61] H. Yang and L. Ying. Synchrosqueezed curvelet transform for two-dimensional mode decomposition. *SIAM Journal on Mathematical Analysis*, 46(3):2052–2083, 2014.

[62] C. Yarham, U. Boeniger, and F. Herrmann. *Curvelet based ground roll removal.*