Available online at www.sciencedirect.com
SCIENCE DIRECT
ELSEVIER
Appl. Comput. Harmon. Anal. 19 (2005) 162–197
Applied and Computational Harmonic Analysis
www.elsevier.com/locate/acha

Continuous curvelet transform
I. Resolution of the wavefront set

Emmanuel J. Candès [a], David L. Donoho [b,*]

[a] *Applied and Computational Mathematics, California Institute of Technology, Pasadena, CA 91125, USA*
[b] *Statistics, Stanford University, Stanford, CA 94305, USA*

Received 5 November 2003; accepted 16 February 2005

Available online 31 March 2005

Communicated by Guido L. Weiss

Abstract

We discuss a Continuous Curvelet Transform (CCT), a transform $f \mapsto \Gamma_f(a,b,\theta)$ of functions $f(x_1,x_2)$ on R^2 into a transform domain with continuous scale $a>0$, location $b \in R^2$, and orientation $\theta \in [0,2\pi)$. Here $\Gamma_f(a,b,\theta) = \langle f, \gamma_{ab\theta}\rangle$ projects f onto analyzing elements called *curvelets* $\gamma_{ab\theta}$ which are smooth and of rapid decay away from an a by $\sqrt{a}$ rectangle with minor axis pointing in direction θ. We call them curvelets because this anisotropic behavior allows them to 'track' the behavior of singularities along curves. They are continuum scale/space/orientation analogs of the discrete frame of curvelets discussed in [E.J. Candès, F. Guo, New multiscale transforms, minimum total variation synthesis: applications to edge-preserving image reconstruction, Signal Process. 82 (2002) 1519–1543; E.J. Candès, L. Demanet, Curvelets and Fourier integral operators, C. R. Acad. Sci. Paris, Sér. I (2003) 395–398; E.J. Candès, D.L. Donoho, Curvelets: a surprisingly effective nonadaptive representation of objects with edges, in: A. Cohen, C. Rabut, L.L. Schumaker (Eds.), Curve and Surface Fitting: Saint-Malo 1999, Vanderbilt Univ. Press, Nashville, TN, 2000]. We use the CCT to analyze several objects having singularities at points, along lines, and along smooth curves. These examples show that for fixed (x_0,θ_0), $\Gamma_f(a,x_0,\theta_0)$ decays rapidly as $a \to 0$ if f is smooth near x_0, or if the singularity of f at x_0 is oriented in a different direction than θ_0. Generalizing these examples, we show that decay properties of $\Gamma_f(a,x_0,\theta_0)$ for fixed (x_0,θ_0), as $a \to 0$ can precisely identify the wavefront set and the H^m-wavefront set of a distribution. In effect, the wavefront set of a distribution is the closure of the set of (x_0,θ_0) near which $\Gamma_f(a,x,\theta)$ is not of rapid decay as $a \to 0$; the H^m-wavefront set is the closure of those points (x_0,θ_0) where the 'directional parabolic square function'

* Corresponding author.
E-mail address: donoho@stanford.edu (D.L. Donoho).

1063-5203/$ – see front matter
doi:10.1016/j.acha.2005.02.003

$S^m(x,\theta) = \left(\int |\Gamma_f(a,x,\theta)|^2 \frac{da}{a^{3+2m}}\right)^{1/2}$ is not locally integrable. The CCT is closely related to a continuous transform pioneered by Hart Smith in his study of Fourier Integral Operators. Smith's transform is based on strict affine parabolic scaling of a single mother wavelet, while for the transform we discuss, the generating wavelet changes (slightly) scale by scale. The CCT can also be compared to the FBI (Fourier–Bros–Iagolnitzer) and Wave Packets (Cordoba–Fefferman) transforms. We describe their similarities and differences in resolving the wavefront set.

Keywords: Wavelets; Curvelets; Wave packets; Directional wavelets; Analysis of singularities; Singular support; Wavefront set; Parabolic scaling

1. Introduction

Standard wavelet transforms for two-dimensional functions $f(x_1,x_2)$ have only very crude capabilities to resolve directional features. The usual orthogonal wavelet transforms have wavelets with primarily vertical, primarily horizontal and primarily diagonal orientations. However, many persons have remarked on the need for transforms exhibiting a wide range of orientations for use with certain classes of objects f, for example those f which model images. Already in the 1980's, vision researchers (Adelson et al. [17] and Watson [23]) were inspired by a biological fact: the visual cortex, although multiscale—*like* the wavelet transform—is highly multi-orientation—*unlike* the wavelet transform. This led them to new transforms such as 'steerable pyramids' and 'cortex transforms' which offered increased directional representativeness. Since then, a wide range of directional transform ideas have been proposed.

In this paper we construct yet another directional wavelet transform, this time with a continuous transform domain; we call it the continuous curvelet transform (CCT). The result of applying the CCT to a function $f(x_1,x_2)$ is a function $\Gamma_f(a,b,\theta)$, where the scale a, location b and direction θ run through continuous ranges. By itself, the possession of a directional parameter is not very impressive. However, we claim that the scale/space/direction domain mapped out by this transform is much more finely resolved than the corresponding parameter domain defined using the 'obvious' directional wavelet scheme: there are important and natural directional phenomena which the 'obvious' approach misses completely, but which are fully revealed using $\Gamma_f(a,b,\theta)$.

1.1. The 'obvious' way to get a directional transform

Starting from the standard continuous wavelet transform, there is an 'obvious' way to create a directional wavelet transform: one takes a classical admissible wavelet ψ which is centered on the origin, 'stretches' it preferentially in one direction, say according to $\tilde{\psi}(x_1,x_2) = \psi(10x_1, x_2/10)$, so it has an elongated support (in this case, one hundred times longer than its width), considers each rotation $\psi_\theta(x) = \tilde{\psi}(R_\theta x)$ of that wavelet, and takes the generated scale-location family $\psi_{a,b,\theta}(x) = \psi_\theta((x-b)/a)/a$. This would provide a wavelet transform with strongly-oriented wavelets and a directional parameter, and it is very easy to see that it would offer an exact reconstruction formula and a Parseval-type relation. (Conceptually, it would be nothing new, as this is a continuous version of ideas such as the steerable pyramids and cortex transform.) We use this as a straw man for comparison against a more subtle notion of directional transform.

1.2. Parabolic scaling

In harmonic analysis since the 1970's there have been several important applications of decompositions based on *parabolic dilations*

$$f_a(x_1, x_2) = f_1\big(a^{1/2}x_1, ax_2\big),$$

so-called because they leave invariant the parabola $x_2 = x_1^2$. In the above equation the dilation is always twice as powerful in one fixed direction as in the orthogonal one. Decompositions also can be based on *directional parabolic dilations* of the form

$$f_{a,\theta}(x_1, x_2) = f_a\big(R_\theta(x_1, x_2)'\big),$$

where again R_θ is rotation by θ radians. The directional transform we define uses curvelets $\gamma_{ab\theta}$ which are essentially the result of such directional parabolic dilations. This means that at fine scales they are increasingly long compared to their width: width $\approx$ length^2.

The motivation for decomposition into parabolic dilations comes from two directions.

- *Representation of operators.* Already in the 1970's they were used in harmonic analysis by Fefferman [11] to study boundedness of spherical summation operators. Later Seeger, Sogge, and Stein [16] to study the boundedness of certain Fourier integral operators. More recently, Hart Smith [19,20] proposed parabolic scaling in defining molecular decompositions of Fourier integral operators. In these settings, parabolic scaling provided an efficient tool for establishing boundedness of operators. From an applications viewpoint, the most accessible observation is that parabolic scaling provides the sparsest representation of such operators [3].
- *Representation of functions.* Candés and Donoho [4] proposed parabolic scaling for use in representing functions $f(x_1, x_2)$ by highly anisotropic directional elements. To motivate this, they considered piecewise smooth functions with piecewise smooth edges. In that setting, parabolic scaling provides the sparsest representation of such functions; see also [9].

For further discussion, see Section 9 below.

1.3. Analysis of singularities

In this paper, we focus on the analysis of singularities. Suppose we have an object $f(x_1, x_2)$ which is smooth apart from a singularity along a planar curve η—for example, η could trace out a circle in the plane, and f could be discontinuous along η, a specific case being $f(x) = 1_{\{|x| \leqslant 1\}}$ which is discontinuous at the unit circle. The study of such objects can be motivated by potential imaging applications where η represents an 'edge' in the 'image' f.

In the analysis of singularities, we show in Section 5 below that the usual continuous wavelet transform will resolve the *singular support* of f. Using an appropriate wavelet ψ to start with, the classic continuous wavelet transform $CW_f(a, b) = \langle \psi_{a,b}, f\rangle$ will signal the location of the singularity through its asymptotic behavior as the scale $a \to 0$. For each fixed location x_0, $CW_f(a, x_0)$ typically will tend to zero rapidly for x_0 outside the singularity, and typically will tend to zero slowly 'on' the singularity. Thus the locations of slow decay for the wavelet transform are the points where f is singular [15, Chapter 3].

Now suppose we have a pretender to the title 'directional transform,' with parameters (a, b, θ), and we study the asymptotic behavior as $a \to 0$ for (b, θ) fixed. The 'obvious' directional wavelet transform described above will typically have rapid decay for b away from the singularity, but will, as we show in Section 6 below, have slow decay in *many* irrelevant directions θ at points b on the singularity. Thus the asymptotic behavior of the transform as $a \to 0$ for b fixed is unable to indicate clearly the true underlying directional phenomenon, which is a singularity having a precise orientation at a specific location. While a directional parameter has been added into play, it does not seem to be of much value.

The transform we define has the property that if the singularity is a curve, then for fixed (x_0, θ_0), $\Gamma_f(a, x_0, \theta_0)$ will tend to zero rapidly as $a \to 0$ unless (x_0, θ_0) matches both the location and orientation of the singularity. Thus, for example, suppose that $f = f_1 + f_2$ is the superposition of two functions with singularities only along curves η_i intersecting transversally at the point x_0. Then we expect to see two θ_i—corresponding to the orientations of the curves η_i carrying the singularities—exhibiting slow decay for $\Gamma_f(a, x_0, \theta_i)$ as $a \to 0$. And indeed, when the singularities of the f_i are well behaved, the points of slow decay for the CCT are the expected (*space*, *orientation*) pairs.

1.4. Microlocal analysis

In effect, we are saying that Γ_f is compatible with standard notions of *microlocal analysis* [14, Volume I, Chapter VIII], [22, Chapter 1], [10, Chapter 1]. One of the central notions in microlocal analysis is that of the *wavefront set* of a distribution. To a distribution f, we associate a parameter space (called the cosphere bundle $S^*(\mathbf{R}^2)$ [19]) consisting of all pairs (x, θ) where x is a spatial variable and θ is an orientation variable. The wavefront set is a subset of this parameter space summarizing the nonsmooth behavior of f. Formally, it is defined in Section 5 below. Informally, it is the collection of location/direction pairs (x, θ) where local windowing ϕf produces an object localized near x which is not smooth in direction θ. Various fundamental results in partial differential equations study and apply the notion of wavefront set; for example, it is used to make precise the notion of *propagation of singularities* of the solution of a partial differential equation over time [14, Volume I, Chapter VIII, Section 8.3], [10, Chapter 5]. Theorems 5.2 and 5.3 provide formal results whose informal meaning is that the wavefront set can be 'read off' from the CCT as the points of slow decay as $a \to 0$.

This connection between the wavefront set and the behavior of a directional transform does not exist if we use the 'obvious' approach to directional wavelet transform discussed above, as we prove below in Section 6. That directional transform has a parameter space with scale, location, and orientation, just like the CCT, but it 'smears out' singularities so that it has slow decay even at points outside the wavefront set. Thus there is an important distinction between the resolving power of the CCT and the DWT; the former correctly resolves the wavefront set while the latter does not. The sharp directional focusing provided by parabolic scaling accounts for the difference.

1.5. Contents

Section 2 constructs the CCT using parabolic scaling, providing a Calderón reproducing formula, (i.e. exact reconstruction) and a Parseval relation for that transform. Section 3 discusses some important localization properties of the curvelets $\gamma_{ab\theta}$. Section 4 studies the use of Γ_f for the analysis of several simple objects with singularities. Section 5 formally defines the wavefront set and states the result showing that decay properties of the CCT precisely resolves the wavefront set. Section 5 also states the result showing that decay of a square function based on the CCT precisely measures the microlocal Sobolev regularity.

The final sections of the paper consider several closely related transforms, including classical wavelet transforms, Hart Smith's transform [19], the FBI transform [1,8,18], and the Wave Packet transform [7], all of which are closely connected to the transform we define. Finally, the conclusion describes relationships to certain forthcoming articles. Many of the proofs are in Appendix A.

2. Continuous curvelet transform

We briefly describe the CCT developed in detail in [6]. We work throughout in $\mathbf{R}^2$, with spatial variable x, with ξ a frequency-domain variable, and with r and ω polar coordinates in the frequency-domain. We start with a pair of windows $W(r)$ and $V(t)$, which we will call the 'radial window' and 'angular window,' respectively. These are both smooth, nonnegative and real-valued, with W taking positive real arguments and supported on $r \in (1/2, 2)$ and V taking real arguments and supported for $t \in [-1, 1]$. These windows will always obey the *admissibility* conditions

$$\int_0^\infty W(ar)^2 \frac{\mathrm{d}a}{a} = 1 \quad \forall r > 0, \tag{1}$$

$$\int_{-1}^{1} V(u)^2 \,\mathrm{d}u = 1. \tag{2}$$

We use these windows in the frequency domain to construct a family of analyzing elements with three parameters: scale $a > 0$, location $b \in \mathbf{R}^2$ and orientation $\theta \in [0, 2\pi)$ (or $(-\pi, \pi)$ according to convenience below). At scale a, the family is generated by translation and rotation of a basic element γ_{a00}

$$\gamma_{ab\theta}(x) = \gamma_{a00}\big(R_\theta(x - b)\big),$$

where R_θ is the 2-by-2 rotation matrix effecting planar rotation by θ radians. The generating element at scale a is defined by going to polar Fourier coordinates (r, ω) and setting

$$\hat{\gamma}_{a00}(r, \omega) = W(a \cdot r) \cdot V(\omega/\sqrt{a}\,) \cdot a^{3/4}, \quad 0 < a < a_0.$$

Thus the support of each $\hat{\gamma}_{ab\theta}$ is a polar 'wedge' defined by the support of W and V, the radial and angular windows, applied with scale-dependent window widths in each direction. In effect, the scaling is parabolic in the polar variables r and ω, with ω being the 'thin' variable. In accord with the use of the terminology *curvelet* to denote families exhibiting such parabolic scaling [2–5], we call this system of analyzing elements curvelets. However, note that the curvelet γ_{a00} is not a simple affine change-of-variables acting on $\gamma_{a,'0,0}$ for $a' \neq a$. We initially omit description of the transform at coarse scales. Note that these curvelets are highly oriented and they become very needle-like at fine scales.

Equipped with this family of curvelets, we can define a *continuous curvelet transform* Γ_f, a function on scale/location/direction space defined by

$$\Gamma_f(a, b, \theta) = \langle \gamma_{ab\theta}, f \rangle, \quad a < a_0,\ b \in \mathbf{R}^2,\ \theta \in [0, 2\pi).$$

Here and below, a_0 is a fixed number—the coarsest scale for our problem. It is fixed once and for all, and must obey $a_0 < \pi^2$ for the above construction to work properly. $a_0 = 1$ seems a natural choice. In [6] we prove

Theorem 1. *Let $f \in L^2$ have a Fourier transform vanishing for $|\xi| < 2/a_0$. Let V and W obey the admissibility conditions* (1)–(2). *We have a Calderón-like reproducing formula, valid for such high-frequency functions*:

$$f(x) = \int \Gamma_f(a, b, \theta)\gamma_{ab\theta}(x)\,\mu(\mathrm{d}a\,\mathrm{d}b\,\mathrm{d}\theta), \tag{3}$$

and a Parseval formula for high-frequency functions

$$\|f\|_{L^2}^2 = \int \left|\Gamma_f(a, b, \theta)\right|^2 \mu(\mathrm{d}a\,\mathrm{d}b\,\mathrm{d}\theta); \tag{4}$$

in both cases, μ denotes the reference measure $\mathrm{d}\mu = \frac{\mathrm{d}a}{a^3}\,\mathrm{d}b\,\mathrm{d}\theta$.

The transform extends to functions containing low frequencies; see again [6] for the proof of the following.

Theorem 2. *Let $f \in L^2(\mathbf{R}^2)$. There is a bandlimited purely radial function Φ in L^2 so that, if $\Phi_{a_0,b}(x) = \Phi(x - b)$*

$$f(x) = \int \langle \Phi_{a_0,b}, f\rangle \Phi_{a_0,b}(x)\,\mathrm{d}b + \int_0^{a_0} \iint \langle f, \gamma_{ab\theta}\rangle \gamma_{ab\theta}(x)\,\mu(\mathrm{d}a\,\mathrm{d}b\,\mathrm{d}\theta),$$

and

$$\|f\|_2^2 = \int \langle \Phi_{a_0,b}, f\rangle^2\,\mathrm{d}b + \int_0^{a_0} \iint \left|\langle f, \gamma_{ab\theta}\rangle\right|^2 \mu(\mathrm{d}a\,\mathrm{d}b\,\mathrm{d}\theta).$$

We can think of the 'full CCT' as consisting of fine-scale curvelets and coarse-scale isotropic father wavelets. For our purposes, it is only the behavior of the fine-scale curvelets that matters. For reference below, we let $P_0(f)$ denote the contribution of all the low frequency terms

$$P_0(f)(x) = \int \langle \Phi_{a_0,b}, f\rangle \Phi_{a_0,b}(x)\,\mathrm{d}b,$$

and note that $P_0(f) = (\Psi \star f)(x)$ for a certain window Ψ; for details, see [6].

The low-frequency window Ψ has a technical property referred to frequently in proofs below: namely the property of *rapid decay at* ∞. By this we mean that Ψ obeys estimates,

$$\Psi(x) = O\big(|x|^{-N}\big), \quad \text{as } |x| \to \infty,\ \forall N > 0.$$

In fact, all its derivatives are of rapid decay as well.

3. Localization

The CCT, initially defined for L^2 objects, can extend in an appropriate sense to general tempered distributions. In this paper we always suppose that V and W are C^∞; this will imply that $\gamma_{ab\theta}(x)$ and its derivatives are each of rapid decay as $|x| \to \infty$

$$\gamma_{ab\theta}(x) = O\big(|x|^{-N}\big) \quad \forall N > 0;$$

sharper decay information is provided in (5) below. Since each wavelet $\gamma_{ab\theta}$ is by construction bandlimited (i.e. it has compact support in the frequency domain), it must therefore be a Schwartz function. The transform coefficient $\langle \gamma_{ab\theta}, f \rangle$ is therefore defined for all tempered distributions $f \in \mathcal{D}$.

We can describe the decay properties of $\gamma_{ab\theta}$ much more precisely; roughly the 'right metric' to measure distance from b is associated with an anisotropic ellipse with sides a and $\sqrt{a}$ and minor axis in direction θ, and $\gamma_{ab\theta}$ decays as a function of distance in that metric. So, suppose we let $P_{a,\theta}$ be the parabolic directional dilation of $\mathbf{R}^2$ given in matrix form by

$$P_{a,\theta} = D_{1/a} R_{-\theta},$$

where $D_{1/a} = \mathrm{diag}(1/a, 1/\sqrt{a})$ and $R_{-\theta}$ is planar rotation by $-\theta$ radians. For a vector $v \in R^2$, define the norm

$$|v|_{a,\theta} = \big|P_{a,\theta}(v)\big|;$$

this metric has ellipsoidal contours with minor axis pointing in direction θ. Also, here and below, we use the notation $\langle a \rangle = (1 + a^2)^{1/2}$.

Relevant to all the above remarks is the following lemma concerning the 'effective support' of curvelets.

Lemma 3.1. *Suppose the windows V and W are C^∞ and of compact support. Then for $N = 1, 2, \ldots,$ and corresponding constants C_N,*

$$\big|\gamma_{ab\theta}(x)\big| \leqslant C_N \cdot a^{-3/4} \cdot \big\langle |x - b|_{a,\theta} \big\rangle^{-N} \quad \forall x. \tag{5}$$

This follows directly from arguments in a companion paper [6, Lemma 5.6].

These estimates are compatible with the view that the curvelets are affine transforms of a single mother wavelet, where the analyzing elements are of the form $\psi(P_{a,\theta}(x - b))\,\mathrm{Det}(P_{a,\theta})^{1/2}$. However, it is important to emphasize that $\gamma_{ab\theta}$ does not obey true parabolic scaling, i.e. there is not a single 'mother curvelet' γ_{100} so that

$$\gamma_{ab\theta} = \gamma_{100}\big(P_{a,\theta}(x - b)\big) \cdot \mathrm{Det}(P_{a,\theta})^{1/2}.$$

A transform based on such true parabolic scaling can of course be defined; essentially this has been done by Hart Smith [19]. We will discuss Smith's transform in Section 7, and show that for many purposes, the two approaches are the same; however, as we show in that section, the reproducing formula and discretization are significantly simpler for the CCT we have proposed here.

4. Analysis of simple singularities

To develop some intuition for the behavior of the CCT and directional analysis in general, we consider several examples where f is smooth apart from singularities, discussing the asymptotic behavior of $\Gamma_f(a,b,\theta)$ as $a \to 0$ for fixed (b,θ). This is intimately related to studying the behavior in Fourier space of $\hat{f}(\lambda \cdot e_\theta)$ as $\lambda \to \infty$ for $e_\theta = (\cos(\theta), \sin(\theta))$, for the reason that $\hat{\gamma}_{ab\theta}$ is localized in a wedge around the ray $\{\lambda e_\theta \colon \lambda > 0\}$; viewed in polar coordinates, the wedge becomes increasingly narrow as a decreases towards zero. Formalizing this, we have

Observation. Suppose that f is a distribution with Fourier transform $\hat{f}$ obeying

$$\hat{f}(\lambda \cdot e_\omega) \sim \lambda^{-\rho} A(\omega), \quad \lambda \to \infty, \tag{6}$$

for some continuous function $A(\omega)$. If $A(\theta) \neq 0$, then

$$\Gamma_f(a,0,\theta) \sim a^{\rho - 3/4} A(\theta) \cdot C_\rho, \quad \text{as } a \to 0, \tag{7}$$

where

$$C_\rho = \int_{1/2}^{2} W(r) r^{1-\rho}\, \mathrm{d}r \cdot \int_{-\infty}^{\infty} V(t)\, \mathrm{d}t.$$

To see this, we use the Parseval relation $\langle \gamma_{a0\theta}, f \rangle = (2\pi)^{-2} \langle \hat{\gamma}_{a0\theta}, \hat{f} \rangle$ to pass to the frequency domain, where, because $\hat{\gamma}_{a0\theta}$ is localized in a narrow wedge about the ray $\{\lambda e_\theta \colon \lambda > 0\}$,

$$\begin{aligned}
\int \hat{\gamma}_{a0\theta}(\xi) \hat{f}(\xi)\, \mathrm{d}\xi &\sim \int \hat{\gamma}_{a0\theta}(\xi) |\xi|^{-\rho} A(\theta)\, \mathrm{d}\xi, \quad a \to 0 \\
&\sim A(\theta) \cdot \iint W(a \cdot r) \cdot V\big((\omega - \theta)/\sqrt{a}\big) \cdot a^{3/4} \cdot r^{1-\rho}\, \mathrm{d}\omega\, \mathrm{d}r \\
&= a^\rho A(\theta) a^{-3/4} C_\rho.
\end{aligned}$$

For similar reasons, we also have

Observation. Suppose that f is a distribution with Fourier transform $\hat{f}$ obeying the inequality

$$\big|\hat{f}(\lambda \cdot e_\omega)\big| \leqslant A_{\theta,\delta} \lambda^{-\rho}, \tag{8}$$

for $|\omega - \theta| < \delta$ and $\lambda > 1/\delta$; then

$$\big|\Gamma_f(a,0,\theta)\big| \leqslant a^{\rho - 3/4} A_{\theta,\delta} \cdot C, \quad \text{as } a \to 0. \tag{9}$$

For later use, we will say that $\Gamma(a,b,\theta)$ *decays rapidly* at (b,θ) if $|\Gamma(a,b,\theta)| = O(a^N)$ as $a \to 0$ for all $N > 0$. If $\Gamma(a,b,\theta)$ does not decay rapidly, we say that it *decays slowly* at (b,θ). We will say that Γ decays at rate r if $|\Gamma(a,b,\theta)| = O(a^r)$ as $a \to 0$.

An obvious consequence: if f is smooth ($C^\infty(\mathbf{R}^2)$), then, as $\hat{f}$ is of rapid decay, $\Gamma(a,b,\theta)$ decays rapidly as $a \to 0$ for all (b,θ). In this section we will consider objects which are smooth away from singularities on curves or points, and the behavior of $\Gamma_f(a,b,\theta)$ for b 'on' and 'off' the singularity. As we will see, if the singularity lies on a curve, it may also matter whether θ is transverse to the singularity or tangential to it.

The following very useful localization principle is proven in Appendix A:

Lemma 4.1. *Given two tempered distributions f_1, f_2, with $f_1 = f_2$ in a neighborhood of b, Γ_{f_1} decays rapidly at (b,θ) if and only if Γ_{f_2} decays rapidly at (b,θ). Moreover, $\Gamma_{f_1}(a,b,\theta)$ decays at rate ρ if and only if $\Gamma_{f_2}(a,b,\theta)$ also decays at rate ρ.*

4.1. Point singularities

To begin, consider the Dirac δ, placing unit mass at the origin and none elsewhere. From $\hat{\delta}(\xi) = 1\ \forall\xi$, and (7), we have that for $b = 0$,

$$\Gamma_\delta(a,0,\theta) = a^{-3/4}\cdot C \quad \forall\theta,\ \forall 0 < a < a_0;$$

so the transform actually grows as $a \to 0$. On the other hand, for $b \neq 0$, $|\Gamma(a,b,\theta)| \to 0$ rapidly as $a \to 0$, as we can see from

$$\Gamma_\delta(a,0,\theta) = \langle \gamma_{ab\theta}, \delta\rangle = \gamma_{ab\theta}(0) = \gamma_{a0\theta}(-b)$$

and Lemma 3.1 concerning rapid decay of curvelets. That lemma implies, in particular, that if $b \neq 0$, then $\gamma_{a0\theta}(-b) \to 0$ rapidly as $a \to 0$.

In short:

- If $b \neq 0$, $\Gamma_\delta(a,b,\theta)$ tends to zero rapidly as $a \to 0$;
- If $b = 0$, $\Gamma_\delta(a,b,\theta)$ grows according to the $-3/4$ power in every direction θ.

Consider now the point singularity $\sigma_\alpha(x) = |x|^\alpha$ for $-2 < \alpha < \infty$. This is locally integrable for each α in this range, and so defines a tempered distribution, for which the directional transform can be defined. By standard rescaling arguments,

$$\hat{\sigma}_\alpha(\xi) = C_\alpha|\xi|^{-2-\alpha},$$

and so, applying once again (7), we get that if $b = 0$, we have $5/4 + \alpha$ rate asymptotics. (This makes sense compared with the previous example, because the Dirac is somehow 'close' to the case $\alpha = -2$.) On the other hand, if $b \neq 0$, we get rapid decay. For example, if (b,θ) are such that $e_\theta' b \neq 0$ then, using $\hat{\gamma}_{ab\theta}(\xi) = e^{-i\xi' b}\hat{\gamma}_{a0\theta}(\xi)$ by writing

$$\begin{aligned}
\langle \hat{\gamma}_{ab\theta}, \hat{\sigma}_\alpha\rangle &= C_\alpha \int |\xi|^{-2-\alpha}\hat{\gamma}_{ab\theta}(\xi)\,\mathrm{d}\xi = C_\alpha \int r^{-2-\alpha} W(ar) e^{-ire_\omega' b}\cdot V\big((\omega-\theta)/\sqrt{a}\big) a^{3/4} r\,\mathrm{d}\omega\,\mathrm{d}r \\
&= C a^{3/4+\alpha}\int V\big((\omega-\theta)/\sqrt{a}\big)\left(\int W(ar)e^{-ire_\omega' b} a^{-\alpha} r^{-1-\alpha}\,\mathrm{d}r\right)\mathrm{d}\omega \\
&= C a^{3/4+\alpha}\int V\big((\omega-\theta)/\sqrt{a}\big)\tilde{W}\left(\frac{e_\omega' b}{a}\right)\mathrm{d}\omega \\
&\sim C a^{5/4+\alpha}\tilde{W}\left(\frac{e_\theta' b}{a}\right), \quad a \to 0,
\end{aligned}$$

where $\tilde{W}(u) = \int_{1/2}^{2} r^{-1-\alpha} W(r) e^{-iru}\,\mathrm{d}r$ is a bandlimited function, decaying rapidly as $|u| \to \infty$. We omit the case $e'_\theta b = 0$, which goes the same until the last step. To summarize in the case $\sigma_\alpha(x) = |x|^\alpha$:

- If $b \neq 0$, $\Gamma_{\sigma_\alpha}(a, b, \theta)$ tends to zero rapidly as $a \to 0$;
- If $b = 0$, $\Gamma_{\sigma_\alpha}(a, b, \theta)$ scales according to the $5/4 + \alpha$ power in every direction θ.

Note that in both these examples we see that the behavior is the same in all directions at each b: point singularities are isotropic.

4.2. Linear singularities

Consider as a prototype of linear singularity the distribution ν acting on nice functions by integration along the x_2-axis

$$\langle \nu, f \rangle = \int f(0, x_2)\,\mathrm{d}x_2.$$

This distribution is supported on the x_2-axis, and shows no sensitivity to variations of f with x_2, but is very sensitive to variations of f with x_1. The Fourier transform $\hat{\nu}$ is a distribution supported on the ξ_1-axis $\{\xi\colon \xi_2 = 0\}$ and obeys

$$\langle \hat{\nu}, \hat{f} \rangle = \int \hat{f}(\xi_1, 0)\,\mathrm{d}\xi_1.$$

Thus $\langle \gamma_{ab\theta}, \nu \rangle = (2\pi)^{-2} \cdot \int \hat{\gamma}_{ab\theta}(\xi_1)\,\mathrm{d}\xi_1$.

Now $\hat{\gamma}_{ab\theta}$ is supported in an angular wedge $\Xi(a, \theta)$ where $|\xi| \in (1/2a, 2/a)$ and $\omega \in [\theta - \sqrt{a}, \theta + \sqrt{a}\,]$. This wedge is disjoint from the ξ_1-axis if $|\theta| > \sqrt{a}$. Hence if $\theta \neq 0$, $\langle \gamma_{ab\theta}, \nu \rangle = 0$ for all sufficiently small $a > 0$. In short, if $\theta \neq 0$, we have rapid decay.

On the other hand, if $\theta = 0$,

$$\langle \gamma_{ab\theta}, \nu \rangle = (2\pi)^{-2} \int W(ar) V(0) a^{3/4} e^{-ire'_\theta b}\,\mathrm{d}r = a^{-1/4} \tilde{W}(b_1/a),$$

where $\tilde{W}(u) = (2\pi)^{-2} \cdot V(0) \cdot \int W(r) e^{-iru}\,\mathrm{d}r$ is smooth and of rapid decay as $|u| \to \infty$. Hence $\langle \gamma_{ab\theta}, \nu \rangle \to 0$ rapidly at $\theta = 0$ for each fixed nonzero b_1. Finally, if $b = (0, b_2)$,

$$\langle \gamma_{ab\theta}, \nu \rangle = \frac{1}{(2\pi)^2} \int W(ar) V(0) a^{3/4}\,\mathrm{d}r = a^{-1/4} \tilde{W}(0).$$

Hence,

- If $(b, \theta) = ((0, x_2), 0)$, $\Gamma_\nu(a, b, \theta)$ grows like $O(a^{-1/4})$ as $a \to 0$;
- Otherwise, $\Gamma_\nu(a, b, \theta)$ is of rapid decay as $a \to 0$.

So looking for places in the (b, θ)-plane where the decay of $\Gamma_H(a, b, \theta)$ as $a \to 0$ is slow will precisely reveal the orientation and location of the singularity along the line $x_1 = 0$.

The same considerations apply to other linear singularities; consider the planar Heaviside $H(x) = 1_{\{x_1 \geqslant 0\}}$. As $\nu = \frac{\partial}{\partial x_1} H$, we have $\hat{H}(\xi) = (i\xi_1)^{-1}\hat{\nu}(\xi)$, and so

$$\langle \gamma_{ab\theta}, H \rangle = \int_{-\infty}^{\infty} (i\xi)^{-1}\hat{\nu}(\xi)\hat{\gamma}_{ab\theta}(\xi_1, 0)\,\mathrm{d}\xi_1.$$

Applying the argument from the earlier case of ν, we have that

- If $\theta \neq 0$, $\Gamma_H(a,b,\theta)$ will be zero as soon as $|\theta| > \sqrt{a}$ and so decays rapidly as $a \to 0$;
- If $\theta = 0$, and b is not of the form $(0, x_2)$, $\Gamma_H(a,b,\theta)$ decays rapidly as $a \to 0$;
- If $\theta = 0$, and b is of the form $(0, x_2)$, $\Gamma_H(a,b,\theta)$ decays as $C \cdot a^{3/4}$.

Once again, looking for places in the (b,θ) plane where $\Gamma_H(a,b,\theta)$ decays slowly as $a \to 0$ will precisely reveal the orientation of the singularity along the line $x_1 = 0$.

Comparing the last two examples, we see that where the decay is slow, the *rate* of decay reveals the *strength* of the singularity. In comparing the asymptotic behavior of the CCT for ν with that for H, we have for $(b,\theta) = ((0,x_2),0)$ the growth $\Gamma_\nu \sim Ca^{-1/4}$ as $a \to 0$ versus the decay $\Gamma_H \sim C'a^{3/4}$; this reflects H's role as a weaker singularity than ν. Recalling $\nu = \frac{\partial}{\partial x_1} H$, the difference of 1 in the exponents of the rates as $a \to 0$ is well calibrated to the intrinsic 'order' of the two objects, which must differ by 1 (as $\frac{\partial}{\partial x_1}$ is of order 1).

4.3. Polygonal singularities

Now consider the 'corner' singularity $L(x_1, x_2) = 1_{\{x_1>0\}} \cdot 1_{\{x_2>0\}}$, which has linear singularities along the positive x_1- and x_2-axes, and a point singularity at $(0,0)$. As L is a direct product, its Fourier transform $\hat{L}(\xi) = C/(\xi_1 \cdot \xi_2)$; but this can be written using polar Fourier variables as

$$\hat{L}(\xi) = C \cdot r^{-2} \frac{1}{\cos(\omega)\sin(\omega)}.$$

Consider first the case $b = 0$. By (7), if θ is not one of the Cartesian directions $\{0, \pm\pi/2, \pi\}$,

$$\Gamma_L(a,0,\theta) \sim C \cdot a^{2-3/4} \cdot \frac{1}{\cos(\theta)\sin(\theta)}, \quad a \to 0.$$

The analysis in the direction of the compass points is a bit more subtle; one can show

$$\Gamma_L(a,0,\theta) \sim C' \cdot a^{3/2-3/4}, \quad a \to 0,\ \theta \in \{0, \pm\pi/2, \pi\}; \tag{10}$$

note that this is what one gets (up to a constant) merely by using the L^∞ nature of L, so it seems pointless to give details here. Consider now the case where $b \neq 0$ is in the positive half of the x_1- or x_2-axes.

In the vicinity of the point $b = (x_1, 0)$ with $x_1 > 0$, L coincides with the Heaviside $1_{\{x_2 \geqslant 0\}}$. In the vicinity of the point $b = (0, x_2)$ with $x_2 > 0$, L coincides with the Heaviside $1_{\{x_1 \geqslant 0\}}$. Lemma 4.1 shows that the decay properties of $\Gamma_L(a,b,\theta)$ are completely local at b. It follows that the decay properties of $\Gamma_L(a,b,\theta)$ at such b are given by those of Heavisides in x_1 or x_2 depending on b. From our earlier analysis of the Heaviside (in x_1), we conclude that $\Gamma_L(a,b,\theta)$ has rapid decay unless $(b,\theta) = ((x_1,0), \pm\pi/2)$, or unless $(b,\theta) = ((0,x_2),\pi)$ or $((0,x_2),0)$. In these latter two cases, we conclude that $\Gamma_L(a,b,\theta) \sim C' \cdot a^{3/4}$.

Finally, if b is not in the positive half of the x_1- or x_2-axes, $\Gamma_L(a,b,\theta)$ decays rapidly. Indeed, apply Localization Lemma 4.1: L agrees locally with the constant function 1 or the constant function zero, and of course $\Gamma_1(a,b,\theta)=0$ for all $a<a_0$.

In short,

- If $b=0$ and θ is not aligned with the axes, then $\Gamma_L(a,b,\theta)=O(a^{5/4})$;
- If $b=0$ and θ is aligned with the axes, then $\Gamma_L(a,b,\theta)=O(a^{3/4})$;
- If $b\neq 0$ but b is on the singularity and θ is aligned with the singularity, then $\Gamma_L(a,b,\theta)=O(a^{3/4})$;
- Otherwise, $\Gamma_L(a,b,\theta)$ decays rapidly.

Now we can consider more general corner singularities, defined by wedges

$$L_{\theta_1,\theta_2}(x_1,x_2)=1_{\{e'_{\theta_1}x\geqslant 0\}}\cdot 1_{\{e'_{\theta_2}x\geqslant 0\}}; \tag{11}$$

the analysis will be qualitatively similar to the analysis above, with various obvious translations, replacing the positive x_1- and x_2-axes by more general rays. Moreover, if we translate such wedges so the corner is somewhere besides 0, the role played by $b=0$ will simply translate in the obvious way.

If we now consider the indicator of a polygon P, we note that locally we are in the setting of one of the wedges (11), and so the decay of Γ_P will be rapid away from the boundary of P, and also rapid on the boundary away from the vertices of the polygon and away from the direction normal to the boundary; otherwise the decay will be slow, in a way similar to the analysis of L above.

For example, consider the object S which is the indicator of the square $-1\leqslant x_1,x_2\leqslant 1$. This is the gluing together of four translated and perhaps rotated copies of the corner singularity L, and the directional wavelet transform has decay properties obtained by gluing together the asymptotic behavior of each of those copies. As a result,

- If b is not on the boundary of the square, we have rapid decay as $a\to 0$;
- If b lies in the boundary of the square there are two cases:
  - At the corners $b\in(\pm 1,\pm 1)$, we have decay at rate $A(\theta)a^{5/4}$, with coefficient $A(\theta)=1/(\sin(\theta)\cos(\theta))$, except where $A=+\infty$, in which case the decay is at rate $a^{3/4}$.
  - On the sides, we have rapid decay as long as the direction θ is not normal to the boundary of the square, in which case we have decay at rate $a^{3/4}$.

Hence we have rapid decay at (b,θ) pairs away from the position/orientation of the singularity, but slow decay in all directions at the corners and still slower decay on the sides, in directions normal to the sides of the square. Again, the position and strength of the singularities are reflected in the $a\to 0$ asymptotics of $\Gamma_S(a,b,\theta)$.

4.4. Curvilinear singularities

We now consider some objects with singularities along curves. Let B be the indicator of the unit disk $D=\{|x|\leqslant 1\}$. Note that B is singular along the boundary $\partial D=\{|x|=1\}$, and that the singularity at $x\in\partial D$ has unit normal pointing in direction $x/\|x\|$. Let $\omega(x)$ be the angle in $[-\pi,\pi)$ corresponding to this, so $e_{\omega(x)}=x/\|x\|$.

The localization lemma implies that $\Gamma_B(a,b,\theta)$ decays rapidly as $a \to 0$ unless b lies in ∂D. We will also show that, even if b lies in the boundary of D, $\Gamma_B(a,b,\theta)$ decays rapidly as $a \to 0$ unless $\theta = \omega(b)$. In short, the decay is slow precisely where the singularities of B lie, and only in the precise direction normal to those singularities. A similar pattern holds for other objects with singularities along the boundary of the disk, such as $B_\alpha(x) = (1-x^2)_+^\alpha$, for $\alpha > 0$: $\Gamma_{B_\alpha}(a,b,\theta)$ decays rapidly as $a \to 0$ unless b is in the boundary of the disk and $\theta = \omega(b)$.

The pattern holds much more generally. Consider the indicator function of the set C: $f = 1_C(x)$, where C is assumed convex, with smooth boundary having nonvanishing curvature. Then $\Gamma(a,b,\theta)$ tends to zero rapidly with a unless $b \in \partial C$, and unless θ is a direction normal to ∂C at b.

In the next section, we will put these conclusions in a larger context, having to do with wavefront sets of distributions. For the moment, we simply sketch the reasons for these facts in the case $B = 1_D(x)$.

As B is radial, $\hat{B}(\xi)$ is also radial, and

$$\hat{B}(\lambda \cdot e_0) = \beta(\lambda) \equiv \int_{-1}^{1} \sqrt{1-t^2} e^{i\lambda t}\, \mathrm{d}t; \tag{12}$$

this is related to the Bessel function J_1; in fact $\beta(\lambda) = C \cdot J_1(\lambda)/\lambda$; [21, p. 338]. It is consequently well understood, and using oscillatory integral techniques as in, e.g., [21,22], one can show the following:

Lemma 4.2. *Let $\beta(\lambda)$ be as in* (12). *Then, for a constant c_0,*

$$\beta(\lambda) \sim c_0 \lambda^{-3/2}\big(e^{i\lambda} + e^{-i\lambda}\big), \quad \lambda \to \infty, \tag{13}$$

and, for $m = 1, 2, \ldots,$ and constants c_m,

$$\left(\frac{\partial}{\partial \lambda}\right)^m \beta(\lambda) \sim c_m \lambda^{-3/2}\big(e^{i\lambda} \pm_m e^{-i\lambda}\big), \quad \lambda \to \infty,$$

where the sign in $\pm_m$ depends on m.

This shows that $\hat{B}(\lambda e_0) = \beta(\lambda)$ is slowly decaying as $\lambda \to \infty$, with oscillations at definite frequencies (± 1). The presence of oscillations in β signals the presence of the cutoffs ± 1 in the defining integral; going back to the original setting, they signal the presence of singularities in B at $\pm e_\theta$.

Now consider the behavior of Γ at (x_0, θ_0) where x_0 is a point in the boundary of the disk, and θ_0 is the boundary normal at that point: $\theta_0 = \omega(x_0)$. Without loss of generality, consider $x_0 = (1,0)$ so that $\theta = 0$. We will see that

$$\Gamma_B(a, x_0, \theta_0) \sim C a^{3/4}, \quad a \to 0.$$

$$\langle \hat{\gamma}_{a10}, \hat{B} \rangle = \iint W(ar) V(\omega/\sqrt{a}) e^{-i\xi'(1,0)} a^{3/4} \beta(r) r \, \mathrm{d}\omega \, \mathrm{d}r. \tag{14}$$

The oscillatory factor $e^{-i\xi'(1,0)} = e^{-ir\cos(\omega)}$ depends nonlinearly on the polar coordinates and must be carefully handled. So define

$$U_a(u) = \int_{-1}^{1} V(t) e^{-i\frac{u}{a}(\cos(\sqrt{a}t)-1)}\, \mathrm{d}t.$$

A simple change of variables gives

$$\langle \hat{\gamma}_{a10}, \hat{B} \rangle = \int_{1/2a}^{2/a} W(ar)U_a(ar)a^{5/4}e^{ir}\beta(r)r\,\mathrm{d}r. \tag{15}$$

Defining $u = ar$ and $\eta_a(u) = W(u)U_a(u)u$, we are led to consider

$$a^{-3/4}\int_{1/2}^{2} \eta_a(u)e^{i\frac{u}{a}}\beta\left(\frac{u}{a}\right)\mathrm{d}u;$$

defining $\zeta_a(u) = a^{-3/2}e^{i\frac{u}{a}}\beta(\frac{u}{a})$ and rescaling, we consider

$$I(a) = \int_{1/2}^{2} \eta_a(u)\zeta_a(u)\,\mathrm{d}u,$$

which is related to the original question by the relation $\Gamma(a, x_0, \theta_0) = a^{3/4}I(a)$.

Now (η_a) is a family of smooth compactly-supported functions which is uniformly in C^∞. Indeed the varying factor $U_a(u)$ has for integrand the form $V(t)\exp\{-ih_a(u,t)\}$, where $h_a \equiv \frac{u}{a}(\cos(\sqrt{a}t) - 1)$ is an equicontinuous family of smooth functions over the range $0 < a < a_0$, $u \in (1/2, 2]$, and $t \in [-1, 1]$. In the limit, $h_a(u,t) \to ut^2$ as $a \to 0$. In fact defining $U_0(u) = \int_{-1}^{1} V(t)e^{-iut^2}\,\mathrm{d}t$, we see that $U_a \to U_0$ in the norm of $C^k[1/2, 2]$ for every $k = 1, 2, \ldots$. Hence $\eta_a \to \eta_0$ in each C^k as well.

Define also the family of functions $\tilde{\zeta}_a(u) = u^{-3/2}(1 + e^{i2u/a})$. Then according to Lemma 4.2, there is ε_a tending to zero with a so that

$$\left|\zeta_a(u) - \tilde{\zeta}_a(u)\right| \leqslant \varepsilon_a.$$

Hence, by the uniform bound $\|\eta_a\|_{L^\infty} < C$ for all $a < a_0$,

$$\left|\int_{1/2}^{2} \eta_a(u)\big(\zeta_a(u) - \tilde{\zeta}_a(u)\big)\,\mathrm{d}u\right| \leqslant C' \cdot \varepsilon_a.$$

Now

$$\int_{1/2}^{2} \eta_a(u)\tilde{\zeta}_a(u)\,\mathrm{d}u = \int_{1/2}^{2} \eta_a(u)u^{-3/2}\,\mathrm{d}u + \int_{1/2}^{2} \eta_a(u)u^{-3/2}e^{i2u/a}\,\mathrm{d}u = T_1(a) + T_2(a).$$

Obviously $T_1(a)$ tends to $T_1(0) \equiv \int_{1/2}^{2} \eta_0(u)u^{-3/2}\,\mathrm{d}u$. On the other hand, $T_2(a)$ can be interpreted as an evaluation of the Fourier transform of $F_a(u) = \eta_a(u)u^{-3/2}$ at frequency $\lambda = 2/a$. Now the family of functions $\{F_a\colon 0 < a < a_0\}$ has all its u-derivatives bounded uniformly in a and so the corresponding Fourier transforms decay rapidly, uniformly in a. Hence $T_2(a) = \hat{F}_a(2/a)$ decays rapidly as $a \to 0$.

Combining the above, we have

$$\Gamma(a, x_0, \theta_0) = a^{3/4}I(a) \sim a^{3/4}T_1(a) \sim a^{3/4}T_1(0), \quad a \to 0.$$

We now consider the case where $x_0 \in \partial D$, but we are looking at a direction θ which is not normal to the singularity: $\theta_0 \neq \omega(x_0)$. Repeating the steps leading to (15), but modified for the present case, we are led to define

$$U_{a,\theta}(u) = \int_{-1}^{1} V(t) e^{-i \frac{u}{a} \cos(\theta + \sqrt{a} t)} \, \mathrm{d}t.$$

We note that

$$\langle \hat{\gamma}_{a,1,\theta}, \hat{B} \rangle = \int_{1/2a}^{2/a} W(ar) U_{a,\theta}(ar) a^{5/4} g(r) r \, \mathrm{d}r.$$

We will show that

$$\sup_{u \in [1/2,2]} U_{a,\theta}(u) = O\big(a^N\big), \quad a \to 0, \tag{16}$$

which will force rapid decay of the associated curvelet coefficient.

We note that, for a small, and $\pi - \sqrt{a} > |\theta| > \sqrt{a}$, $U_{a,\theta}$ is an oscillatory integral. Recall the following standard fact about oscillatory integrals; again see [21, p. 331].

Lemma 4.3. *Let $A(t)$ be in $C^\infty(\mathbf{R})$ and let $\Phi(t)$ be a C^1 function with*

$$\|\Phi'\| \geqslant \eta > 0 \tag{17}$$

everywhere. Then

$$\left| \int A(t) e^{i\lambda\Phi(t)} \, \mathrm{d}t \right| \leqslant C_{N,\eta} \lambda^{-N}, \quad \lambda > 0,$$

where $C_{N,\eta}$ is uniform in Φ satisfying (17).

In our case we need more than the lemma itself; we need the proof idea, which introduces the differential operator

$$(Df)(t) = \frac{\mathrm{d}}{\mathrm{d}t} \left(\frac{f(t)}{i\lambda\Phi'(t)} \right).$$

Then repeated integration-by-parts gives

$$\int A(t) e^{i\lambda\Phi(t)} \, \mathrm{d}t = \int \big(D^N A\big)(t) e^{i\lambda\Phi(t)} \, \mathrm{d}t.$$

Consider now applying this argument to the integral defining $U_{a,\theta}(u)$ with $A(t) = V(t)$, $\Phi(t) = u\cos(\theta + \sqrt{a} t)$ and $\lambda = a^{-1}$. Then

$$(Df)(t) = \frac{\mathrm{d}}{\mathrm{d}t} \left(\frac{f(t)}{i a^{-1/2} u \cdot \sin(\theta + \sqrt{a} t)} \right).$$

Hence

$$U_{a,\theta}(u) = \int_{-1}^{1} \big(D^N A\big)(t) e^{-ia^{-1}\Phi(t)} \, \mathrm{d}t,$$

and so

$$\left|U_{a,\theta}(u)\right| \leqslant 2\left\|D^N A\right\|_\infty.$$

Now, consider $\dot{V} = DV$, then

$$\dot{V}(t) = \sqrt{a}\frac{V'(t)}{iu\sin(\theta+\sqrt{a}t)} + \sqrt{a}\frac{V(t)\cos(\theta+\sqrt{a}t)}{iu\sin(\theta+\sqrt{a}t)^2}$$

as $u \in [1/2, 2]$, there are constants C_1, C_2 so that

$$\|\dot{V}\|_\infty \leqslant \sqrt{a}\cdot\left(C_1\|V'\|_\infty + C_2\|V\|_\infty\right).$$

A similar argument can be used to control $\|\frac{\mathrm{d}}{\mathrm{d}t}V\|_\infty$ in terms of $\sqrt{a}\cdot(C_1\|V''\|_\infty + C_2\|V'\|_\infty)$. Applying this estimate repeatedly to $V \in C^\infty(\mathbf{R})$, we get

$$\left\|D^N A\right\|_\infty \leqslant \sqrt{a}^N \cdot C_N,$$

and so (16) follows.

5. Microlocal analysis

We now put the calculations of the previous section in a larger context, using microlocal analysis; the subject is developed in numerous places; see, for example, [10,14].

Definition 5.1. The *singular support* of a distribution f, $\operatorname{sing\,supp}(f)$, is the set of points x_0 where, for every smooth 'bump' function $\phi \in C^\infty$, $\phi(x_0) \neq 0$, localized to a ball $B(x_0,\delta)$ near x_0, the windowed function ϕf has a Fourier transform $\widehat{\phi f}(\xi)$ which is not of rapid decay as $|\xi| \to \infty$.

Here of course, rapid decay means $\hat{f}(\xi) = O(|\xi|^{-N})$ for all $N > 0$.

Thus, in our earlier examples:

- $\operatorname{sing\,supp}(\delta) = \{0\}$,
- $\operatorname{sing\,supp}(\nu) = \{(0,x_2)\colon x_2 \in \mathbf{R}\}$,
- $\operatorname{sing\,supp}(S) = \{(x_1,x_2)\colon \max(|x_1|,|x_2|) = 1\}$,
- $\operatorname{sing\,supp}(B) = \{x \in \partial D\}$.

We observe that, in all these examples:

$$\operatorname{sing\,supp}(f) = \left\{x\colon \left(\sup_\theta\left|\Gamma_f(a,x,\theta)\right|\right) \text{ decays slowly as } a \to 0\right\}.$$

A slightly weaker statement is true in general. Say that Γ_f *decays rapidly near* x_0 if, for some neighborhood $\mathcal{B}$ of x_0

$$\left|\Gamma_f(a,b,\theta)\right| = O\left(a^N\right), \quad \text{as } a \to 0,$$

with the $O()$ term uniform in θ and in $b \in \mathcal{B}$.

Theorem 5.1. *Let*

$$\mathcal{R} = \{x_0\colon\ \Gamma_f \textit{ decays rapidly near } x_0 \textit{ as } a \to 0\}.$$

Then sing supp(f) *is the complement of* $\mathcal{R}$.

The proof is given in Appendix A.

Definition 5.2. The *wavefront set* of a distribution f, $WF(f)$, is the set of points (x_0, θ_0) where $x_0 \in$ sing supp(f) and, for every smooth 'bump' function $\phi \in C^\infty$, $\phi(x_0) \neq 0$, localized to a ball $B(x_0, \delta)$ near x_0, the windowed function ϕf has a Fourier transform $\widehat{\phi f}(\xi)$ which is not of rapid decay in any wedge defined in polar coordinates by $|\omega - \theta_0| < \delta$.

Here of course, rapid decay in a wedge means $\hat{f}(\lambda e_\omega) = O(|\lambda|^{-N})$ for all $N > 0$, uniformly in $|\omega - \theta_0| < \delta$.

In our earlier examples:

- $WF(\delta) = \{0\} \times [0, 2\pi)$,
- $WF(\nu) = \{((0, x_2), 0)\colon x_2 \in \mathbf{R}\}$,
- $WF(B) = \{((\cos(\theta), \sin(\theta)), \theta)\colon \theta \in [0, 2\pi)\}$.

In short, in our earlier examples:

$$WF(f) = \bigl\{(x_0, \theta_0)\colon \Gamma_f(a, x_0, \theta_0) \text{ decays slowly as } a \to 0\bigr\}.$$

A slightly weaker statement is true in general. Say that Γ_f *decays rapidly near* (x_0, θ_0) if, for some neighborhood $\mathcal{N}$ of (x_0, θ_0)

$$\bigl|\Gamma_f(a, b, \theta)\bigr| = O\bigl(a^N\bigr), \quad \text{as } a \to 0,$$

with the $O()$ term uniform over $(b, \theta) \in \mathcal{N}$. The following result is proved in Appendix A.

Theorem 5.2. *Let*

$$\mathcal{R} = \bigl\{(x_0, \theta_0)\colon \Gamma_f \textit{ decays rapidly near } (x_0, \theta_0) \textit{ as } a \to 0\bigr\}.$$

Then $WF(f)$ *is the complement of* $\mathcal{R}$.

In short, the $a \to 0$ asymptotics of the CCT precisely resolve the wavefront set. This fact may be compared to the wavelet transform, where the asymptotics precisely resolve the singular support (see the next section).

The CCT also measures notions of microlocal Sobolev regularity.

Definition 5.3. A distribution f is microlocally in the L^2 Sobolev space H^s at (x_0, θ_0), written $f \in H^s(x_0, \theta_0)$, if, for some smooth 'bump' function $\phi \in C^\infty(\mathbf{R}^2)$, $\phi(x_0) \neq 0$, localized to a ball $B(x_0, \delta)$ near x_0, and for some smooth bump function $\beta \in C^\infty_{\mathrm{per}}[0, 2\pi)$ obeying $\beta(\theta_0) = 1$ and localized to a ball near θ_0, the space/direction localized function $f_{\phi,\beta}$ defined in polar Fourier coordinates by $\beta(\omega)\widehat{\phi f}(r\cos(\omega), r\sin(\omega))$ belongs to the weighted L^2 space $L^2((1 + |\xi|^2)^{s/2}\,\mathrm{d}\xi)$.

The following is also proved in Appendix A.

Theorem 5.3. *Let $S_2^m(x,\theta)$ denote the (normal-approach, parabolic-scaling) square function*

$$S_2^m(x,\theta) = \left(\int_0^{a_0} \left|\Gamma_f(a,x,\theta)\right|^2 a^{-2m} \frac{\mathrm{d}a}{a^3} \right)^{1/2}.$$

The distribution $f \in H^m(x_0,\theta_0)$ if and only if for some neighborhood $\mathcal{N}$ of (x_0,θ_0),

$$\int_{\mathcal{N}} \left(S_2^m(x,\theta)\right)^2 \mathrm{d}x\, \mathrm{d}\theta < \infty.$$

In short, microlocal regularity is determined by an L^2 condition on the decay of the directional wavelet transform.

As an example, $B_\alpha(x) = (1-|x|^2)_+^\alpha$ is Hölder(α) at $x_0 \in \partial D$; is in every $H^s(x_0,\theta_0)$ whenever $x_0 \notin \partial D$ and is in $H^s(x_0,\theta_0)$ for $s < \alpha + 1/2$ when (x_0,θ_0) aligns with the boundary of the disk. More revealingly, if we have the spatially variable exponent $\beta(\theta) = (1+\sin(\theta/2))/2$, then

$$f(x) = \begin{cases} (1-|x|^2)_+^{\beta(\omega(x))}, & x \neq 0, \\ 1, & x = 0, \end{cases}$$

is in $H^s((\cos(\theta),\sin(\theta)),\theta)$, where $\theta \in [0,2\pi)$, for $s < \beta(\theta) + 1/2$. As the strength of the singularity changes spatially, the measured regularity changes in a matching way.

6. Comparison to classical wavelets

We briefly remark on how the two main results above compare to what can be done with 'classical wavelets.'

Suppose we define a classical wavelet φ by taking the same window W as for curvelets, and working in the frequency domain via

$$\hat{\varphi}(\xi) = c \cdot W\left(|\xi|\right), \tag{18}$$

where c is a normalization constant. Then we translate and dilate, producing the family of wavelets with typical element $\varphi_{ab}(x) = \varphi((x-b)/a)/a$. The (classical) wavelet transform

$$CW_f(a,b) = \langle \varphi_{ab}, f \rangle, \quad a > 0,\ b \in \mathbf{R}^2,$$

has a Calderón reproducing formula

$$f = \int CW_f(a,b)\varphi_{ab}\, \mu(\mathrm{d}a\, \mathrm{d}b),$$

and Parseval relation

$$\|f\|^2 = \int \left|CW_f(a,b)\right|^2 \mu(\mathrm{d}a\, \mathrm{d}b),$$

where now the reference measure $\mu(\mathrm{d}a\, \mathrm{d}b) = a^{-3}\, \mathrm{d}a\, \mathrm{d}b$, and c has been chosen in (18) to make these identities valid.

Say that CW_f *decays rapidly near* x_0 if, for some neighborhood $\mathcal{B}$ of x_0

$$\left|CW_f(a,b)\right| = O\left(a^N\right), \quad \text{as } a \to 0,$$

with the $O()$ term uniform in $b \in \mathcal{B}$.

Theorem 6.1. *Let the defining window W be C^∞. Define the set of rapid decay via*

$$\mathcal{R} = \{x\colon\ CW_f \text{ decays rapidly near } x \text{ as } a \to 0\}.$$

Then $\operatorname{sing\,supp}(f)$ *is the complement of* $\mathcal{R}$.

In short, *the classical wavelet transform resolves the singular support*. However, it cannot resolve the wavefront set, as there is no directional parameter to even make such a question admissible.

Now consider the 'obvious' way to define the directional transform based on the 'stretching' of classical wavelets so that they become strongly directional, as described in the Introduction. With φ as just defined, set

$$\tilde{\varphi}(x_1, x_2) = \varphi(10x_1, x_2/10).$$

This defines a wavelet which is strongly oriented; then define

$$\tilde{\varphi}_{ab\theta} = c \cdot \tilde{\varphi}\big(R_\theta(x-b)/a\big)/a,$$

where c is a normalizing constant. For the 'obvious' directional transform

$$\widetilde{DW}_f(a,b,\theta) = \langle \tilde{\varphi}_{ab\theta}, f\rangle, \quad a > 0,\ b \in \mathbf{R}^2,\ \theta \in (0, 2\pi]$$

we have a reproducing formula and a Parseval relation formally very similar to those we have seen for Γ.

Note that this construction captures the spirit of many existing directional transform constructions [17,23]. We can ask for this transform whether it resolves the wavefront set, namely whether the set of points (x_0, θ_0) of slow decay (or perhaps its closure) is interesting for microlocal analysis. However, in general this will not be the case.

We illustrate this through an example: the linear singularity ν discussed in Section 4. We ask about the set of points (x, θ) where $\widetilde{DW}_\nu(a, x, \theta)$ has rapid decay as $a \to 0$. To do the required calculation, we let $\tilde{W}(r, \omega)$ denote the Fourier transform of $\tilde{\varphi}$ expressed in polar coordinates. Consider the situation at $b = (0,0)$, which lies on the singularity

$$\widetilde{DW}_\nu(a, 0, \theta) = \int \hat{\tilde{\varphi}}_{a0\theta}(\xi_1, 0)\, \mathrm{d}\xi_1 = \int \tilde{W}(a\xi_1, 0-\theta)a\, \mathrm{d}\xi_1 = \int \tilde{W}(u, 0-\theta)\, \mathrm{d}u = A(\theta),$$

say, where $A(\theta)$ is a smooth function. In particular, A changes smoothly in the vicinity of 0, and so does not sharply distinguish behavior in the direction of the singularity from behavior in other directions. More to the point: in general, assuming the original window $W > 0$ on its support, $A > 0$ and so $\widetilde{DW}_\nu(a, 0, \theta)$ is of slow (i.e. no) decay in every direction. By contrast, the directional transform $\Gamma_\nu(a, 0, \theta)$ based on curvelets is of rapid decay in all directions except for $\theta \in \{0, \pi\}$.

To summarize: the wavelet transform resolves the singular support of distributions, but the 'obvious' directional wavelet transform does not resolve the wavefront set.

Hence, the CCT provides a finer notion of directional analysis than schemes based on 'classical' wavelet constructions.

7. Comparison to Hart Smith's transform

Recall now the parabolic rescaling transformation $P_{a,\theta}$ of Section 3.

Suppose now that we take a single 'wavelet' φ and define an affine system

$$\varphi_{ab\theta} = \varphi\big(P_{a,\theta}(x-b)\big) \cdot \mathrm{Det}(P_{a,\theta})^{1/2}. \tag{19}$$

Classically, the term 'wavelet transform' has been understood to mean that a single waveform is operated on by a family of affine transformations, producing a family of analyzing waveforms. However, classically, the scaling involved behaves like a equally in all spatial variables. In this case variables in direction θ scale differently from those in the perpendicular direction. Hart Smith in [19] defined essentially this continuous transform, with two inessential differences: first, instead of working directly with scale a and direction θ, he did equivalent work using the frequency variable $\xi \equiv a^{-1}e_\theta$, and second, instead of using the L^2 normalization $\mathrm{Det}(P_{a,\theta})^{1/2}$, he used the L^1 normalization $\mathrm{Det}(P_{a,\theta})$. In any event, we pretend that Smith had used our notation and normalization as in (19) and call

$$\bar{\Gamma}_f(a,b,\theta) = \langle \varphi_{ab\theta}, f\rangle.$$

Hart Smith's *affine parabolic scaling transform*.

While affine parabolic scaling is conceptually a bit simpler than the polar-variables scaling we have mostly studied here, it does complicate life a bit. The reconstruction formula can be written so: let f be a high-frequency function, then there is a Fourier multiplier M so that

$$f = \int \langle \varphi_{a,b,\theta}, Mf\rangle \varphi_{ab\theta}\,\mathrm{d}\mu$$

and

$$\|f\|_2^2 = \int \big|\langle \varphi_{a,b,\theta}, Mf\rangle\big|^2\,\mathrm{d}\mu.$$

Here $d\mu = a^{-3}\,\mathrm{d}b\,\mathrm{d}\theta\,\mathrm{d}a$ and Mf is defined in the frequency domain by a multiplier formula $m(|\xi|)\hat{f}(\xi)$, where $m(r)$ is such that $\log m(\exp(u))$ is C^∞ and $\log m(\exp(u)) \to 0$, as $u \to +\infty$, together with all its derivatives.

In short, one has to work not with the coefficients of f but with those of Mf. Equivalently, one defines dual elements $\varphi^\sharp_{ab\theta} \equiv M\varphi_{ab\theta}$ and changes the transform definition to either

$$f = \int \langle \varphi^\sharp_{a,b,\theta}, f\rangle \varphi_{ab\theta}\,\mathrm{d}\mu$$

or

$$f = \int \langle \varphi_{a,b,\theta}, f\rangle \varphi^\sharp_{ab\theta}\,\mathrm{d}\mu.$$

This more complicated set of formulas leads to a few annoyances which are avoided using the transform that we have defined in Section 2 above. There are other advantages to our definition of Γ when it comes to discretizing the transform, which are discussed in [6].

However, for the purposes of this paper, the two transforms are equally valuable:

Lemma 7.1. *Suppose that the mother wavelet generating the Smith transform $\bar{\Gamma}$ has the frequency-domain representation*

$$\hat{\varphi}_{a00}(\xi) = cW(a\xi_1)V\left(\frac{\xi_2}{\sqrt{a}\xi_1}\right)a^{3/4}, \quad a < \bar{a}_0,$$

for the same windows V and W underlying the construction of Γ, where c is some normalizing constant, and $\bar{a}_0$ is the transform's coarsest scale. The following two properties are equivalent:

- *Γ_f is of rapid decay near (x, θ).*
- *$\bar{\Gamma}_f$ is of rapid decay near (x, θ).*

The following two properties are equivalent:

- *The square function $S_2^m(x, \theta)$ based on Γ_f is square-integrable in a neighborhood of (x_0, θ_0).*
- *The square function $\bar{S}_2^m(x, \theta)$ based on $\bar{\Gamma}_f$ is square-integrable in a neighborhood of (x_0, θ_0).*

In particular, Smith's transform resolves the wavefront set and the H^s wavefront set.

This lemma can be proven by adapting estimates in Section 5 of [6]. We omit details.

8. Comparison to the FBI and wave packet transforms

The idea of using wavelet-like transforms to perform microlocal analysis goes back to Bros and Iagolnitzer [1] and, independently, Cordoba and Fefferman [7], who both defined transforms with implicitly a kind of scaling related to parabolic scaling, and used these to attack various questions in microlocal analysis. For the sake of brevity, we modify the transform definitions in a way that enables an easy comparison with what we have discussed above.

As in the case of the CCT, we adopt the parameter space (a, b, θ); we pick a smooth radial window $W(x)$ and define a collection of analyzing elements according to

$$\phi_{ab\theta} = \exp\{ia^{-1}e'_\theta(x-b)\}W\big((x-b)/\sqrt{a}\big)/\sqrt{a}.$$

These can be viewed as Gabor functions where the frequency and the window size are linked by the quadratic relation:

$$\text{window size}^2 = \text{spatial frequency}.$$

We have an oscillatory waveform supported in an isotropic window of radius $\sqrt{a}$ centered at b, and with a dominant frequency $\xi = a^{-1}e_\theta$. This waveform makes $O(a^{-1/2})$ oscillations within its effective support, and the wavecrests are aligned normal to e_θ. Effectively this is a packet of waves. Define then the *wave packet transform*

$$WP_f(a, b, \theta) = \langle \phi_{ab\theta}, f \rangle.$$

We ignore for now issues of reconstruction and stability; these obviously can be dealt with by tools such as: introducing a cutoff-scale a_0, using the 'low frequency' trick of Section 2, and the 'multiplier trick' of the previous section.

We simply point out that *WP* resolves the wavefront set: the wavefront set is the closure of the set of points (x_0, θ_0) where $WP(a, x_0, \theta_0)$ is of slow decay. The first result of this type appears to have been given by P. Gérard [8,13], who proved that the FBI transform resolves the wavefront set. The FBI transform is usually defined in terms of complex variables and a precise definition would take us far afield; it can be, for present purposes, very roughly described as a wave packet transform WP_f using a Gaussian window $W(x) = \exp\{-|x|^2\}$.

9. Discussion

We have studied here only the use of polar parabolic scaling, where our basic analyzing element involves the frequency-domain formula $W(ar) \cdot V((\omega - \theta)/\sqrt{a})$. It is obvious that one could modify the transform by changing the scaling exponent on a in the V-factor; the square root is by no means mandatory. In fact choose any $\beta \in (0, 1)$, and consider a directional wavelet-like transform DWT^β generated from translations of the basic wavelet

$$\hat{\varphi}^{\beta}_{a,0,\theta}(\xi) = W(ar) \cdot V\big((\omega - \theta)/a^\beta\big)a^{(2-\beta)/2}, \tag{20}$$

where of course $\beta = 1/2$ is the case studied in this paper. In some sense as $\beta \to 0$ the elements behave more like classical wavelets (being more isotropic) and as $\beta \to 1$ the elements behave more like ridgelets, being poorly localized in one dimension and well localized in the other dimension. By adapting the arguments of this paper, one can show that every such 'directional wavelet transform' will resolve the wavefront set correctly.

The rationale for preferring the CCT among all these transforms is a quantitative one. We view 'resolution of the wavefront set' as saying that the scale/location/direction plane is *qualitatively sparse*, becoming very small at fine scales, except near the locations and directions of singularities.

For a quantitative notion of sparsity, we suggest to study objects with singularities of a given fixed type, for example, discontinuities along C^2 curves, and consider the size as measured by inequalities of the form

$$\mu\big\{(a, b, \theta)\colon \big|\Gamma_f(a, b, \theta)\big| > \varepsilon\big\} \leqslant C\varepsilon^{-1/p}, \quad \varepsilon < \varepsilon_0. \tag{21}$$

In such inequalities, the smaller p, the more quantitatively sparse the scale/location/direction plane. It turns out (compare [5]) that the CCT obeys such an inequality for every $p > 2/3$. Define the sparsity index p^* of a transform plane as the infimum of those p for which we have inequality (21) for each piecewise smooth function with discontinuities along piecewise C^2 curves. It turns out that for directional wavelet transforms DWT^β based on the nonparabolic scaling laws, the sparsity index p^* with $\beta \neq 1/2$ in (20) is worse than it is for parabolic scaling $\beta = 1/2$. In fact,

$$p^*(\beta) = \begin{cases} \frac{1-\beta}{1-\beta/2}, & 0 < \beta \leqslant 1/2, \\ \frac{1}{2-\beta}, & 1/2 \leqslant \beta < 1, \end{cases}$$

which of course achieves its best value $2/3$ at $\beta = 1/2$. We prove this in a manuscript in preparation.

A different reason for preferring parabolic scaling arises in representing Fourier integral operators (FIOs), and was Hart Smith's original reason for defining the affine parabolic scaling transform [19]. Smith considered the kernel K_T of the FIO T:

$$K_T(a, b, \theta; a,' b,' \theta') = \langle \varphi_{ab\theta}, T\varphi_{a'b'\theta'}\rangle.$$

He demonstrated that, using a suitable anisotropic distance measure, K_T decays rapidly as the arguments diverge from the diagonal. This rapid decay means intuitively that the parabolic scaling elements $\varphi_{ab\theta}$ quasi-diagonalize FIOs. The idea that parabolic scaling provides the 'right' representation for FIOs has been made much more precise by Candès and Demanet [3] who worked with a discrete curvelets frame, and showed that FIOs can be represented as matrices in that frame which are almost-diagonal in an appropriate sense. Such almost-diagonalization is a nontrivial property of a frame representation and apparently is uniquely due (among scaling principles) to parabolic scaling.

In the sequel [6], we look at representations of FIOs from the CCT viewpoint. We show how Smith's result on decay of K_T implies decay of the comparable kernel for the CCT:

$$K_T'(a,b,\theta;a,'b,'\theta') = \langle \gamma_{ab\theta}, T\gamma_{a'b'\theta'}\rangle.$$

We build a tight frame by discretizing the CCT, essentially by regular sampling of (a,b,θ) on a 'curvelet grid.' It follows that the matrix of an FIO in the CCT-derived curvelet tight frame will be sparse. Hence the polar parabolic scaling in this paper can be seen to lead straightforwardly to a sparse matrix representation of FIOs. As it turns out, the CCT-derived curvelet tight frame is simply a complexified version of the frame used by Candès and Demanet. Thus the CCT approach leads to FIO sparsity of the kind proven by Candès and Demanet, simply by applying Smith's continuum result.

Acknowledgments

This work has been partially supported by DMS 00-77261, DMS 01-40698 (FRG), DMS 98-72890 (KDI), DARPA ACMP and AFOSR MURI. EJC was supported by an Alfred P. Sloan Fellowship and DOE DE-FG03-02ER25529. D.L.D. thanks the Mathematical Institute at the University of Leiden for hospitality during Fall 2002.

Appendix A

A.1. Proof of Lemma 4.1

We only sketch the proof, for the case that f_1 and f_2 are both locally bounded functions

$$\langle \gamma_{ab\theta}, f_1\rangle - \langle \gamma_{ab\theta}, f_2\rangle = \int \gamma_{ab\theta}(x)(f_1 - f_2)(x)\,\mathrm{d}x.$$

Let $\mathcal{B}$ be a ball centered at b on which $f_1 = f_2$. Then

$$\left|\int \gamma_{ab\theta}(x)(f_1 - f_2)(x)\,\mathrm{d}x\right| \leqslant \int\limits_{\mathcal{B}^c} \left|\gamma_{ab\theta}(x)\right|\mathrm{d}x \cdot \|f_1 - f_2\|_{L^\infty(\mathcal{B}^c)}.$$

Lemma 3.1 tells us that the wavelet $\gamma_{a0\theta}$ is effectively localized to a ball of radius a; in fact for any $\varepsilon > 0$,

$$\int\limits_{\{|x|>\varepsilon\}} \left|\gamma_{a0\theta}(x)\right|\mathrm{d}x = O\left(a^N\right), \quad N > 0.$$

and so, as $\mathcal{B}^c \subset \{x\colon |x-b| > \varepsilon\}$ for some $\varepsilon > 0$, $\int_{\mathcal{B}^c} |\gamma_{ab\theta}(x)|\,\mathrm{d}x = O(a^N)$ for each $N > 0$. Hence, if $\langle \gamma_{ab\theta}, f_1\rangle = O(a^\rho)$ for some specific ρ, $\langle \gamma_{ab\theta}, f_2\rangle = O(a^\rho)$ as well. This proves the lemma in case the f_i are bounded functions. To get the general case, use integration by parts sufficiently many times to obtain a distribution which is a locally bounded function.

A.2. Proof of Theorem 5.1: Resolution of the singular support

Our proof will show that, on the one hand, $\operatorname{sing\,supp}(f)^c \subset \mathcal{R}$, and on the other, $\mathcal{R} \subset \operatorname{sing\,supp}(f)^c$. For the proof we will assume that f is a bounded function: $\|f\|_\infty \leqslant M$. The same proof works for general distributions by modifying the bounding strategy, though we omit details. Our arguments depend on Lemmas A.1–A.5 which are stated and proved in a later subsection.

A.2.1. The CCT decays rapidly outside the singular support

Suppose that $x_0 \notin \operatorname{sing\,supp}(f)$. Then there is a smooth bump function $\phi \in C^\infty$ supported on a ball $\mathcal{B}$ centered at x_0, with $\phi = 1$ on a smaller ball $\mathcal{B}_0$ also centered at x_0, so that $\phi f \in C^\infty$.

We will use this to show that Γ_f decays rapidly near x_0. Now

$$\Gamma_f(a,b,\theta) = \langle \gamma_{ab\theta}, \phi f\rangle + \big\langle \gamma_{ab\theta}, (1-\phi) f\big\rangle. \tag{A.1}$$

As ϕf is C^∞ the first term on the RHS is of rapid decay, uniformly in b and θ.

Consider the second term on the RHS; recall the localization of $\gamma_{ab\theta}$, which guarantees that for $b \in \mathcal{B}_0$, and a small, $\gamma_{ab\theta}$ is very small on the support of $(1-\phi)f$. Apply Lemma A.1 below with $g = (1-\phi)f$. We conclude that for each $N > 0$, $|\langle \gamma_{ab\theta}, (1-\phi)f\rangle| \leqslant C_N \langle a^{-1}\rangle^{-N}$ valid for all θ and all $b \in \mathcal{B}_0$.

Because both terms on the RHS of (A.1) decay rapidly, uniformly over $(b,\theta) \in \mathcal{B}_0 \times [0, 2\pi)$, we have shown that Γ_f decays rapidly near x_0.

A.2.2. The function is locally C^∞ where the CCT decays rapidly

Suppose now that Γ_f decays rapidly near x_0. Then there is a ball $\mathcal{B}$ containing x_0 on which the decay is uniform over $(b,\theta) \in \mathcal{B} \times [0, 2\pi)$. Pick a function ϕ which is supported in a ball $\mathcal{B}_0 \Subset \mathcal{B}$. Let $\delta = d(\mathcal{B}_0, \mathcal{B}^c)$, and with $\eta = \delta/2$, let $\mathcal{B}_1 = \{x\colon \mathrm{d}(x, \mathcal{B}_0) < \eta\}$ be the η-enlargement of $\mathcal{B}_0$. Note that Γ_f decays rapidly, uniformly in $\mathcal{B}_1$ and even in a further η-enlargement of $\mathcal{B}_1$.

Put $g = \phi f$; decompose

$$\widehat{\phi f}(\xi) = \hat{g}(\xi) = \hat{g}_0(\xi) + \hat{g}_1(\xi) + \hat{g}_2(\xi),$$

where $\hat{g}_0 = (\phi P_0(f))\hat{\ }(\xi)$, and, setting $\mathcal{Q}_1 = (0, a_0] \times \mathcal{B}_1 \times (0, 2\pi]$ and $\mathcal{Q}_2 = (0, a_0] \times \mathcal{B}_1^c \times (0, 2\pi]$,

$$\hat{g}_i(\xi) = \int\limits_{\mathcal{Q}_i} \hat{\gamma}_{ab\theta}(\xi)\Gamma_g(a,b,\theta)\,\mathrm{d}\mu, \quad i = 1, 2. \tag{A.2}$$

Note first $\hat{g}_0(\xi)$ is of rapid decay as $|\xi| \to \infty$ because both $P_0(f)$ and ϕ are C^∞.

The term $\hat{g}_2(\xi)$ is of the form stipulated by Lemma A.2 below—i.e. the support of ϕf is well separated from $\mathcal{B}_1^c$—and so we conclude that it is of rapid decay as $|\xi| \to \infty$.

To bound $\hat{g}_1$, we use Lemma A.4 below, which considers an integral of the form (A.2) where the $\mathcal{B}$ factor in the $\mathcal{Q}$ region is compact, and shows that, provided Γ_g is uniformly of rapid decay in $\mathcal{Q}$, then the resulting $\hat{g}_1$ is of rapid decay as $|\xi| \to \infty$.

Therefore, we seek to establish that Γ_g is of rapid decay in $\mathcal{B}_1 \times (0, 2\pi]$. Write $f = f_0 + f_1 + f_2$, where $f_0 = P_0(f)$ and

$$f_i(x) = \int_{\mathcal{Q}_i} \gamma_Q(x)\Gamma_f(Q)\,\mathrm{d}\mu(Q), \quad i = 1, 2.$$

Here we bundle scale/location/direction parameters into the tuple $Q = (a, b, \theta)$. Then of course $\Gamma_g = \sum_{i=0}^{2} \Gamma_{\phi f_i}$ and we seek to establish rapid decay of the individual terms. $\Gamma_{\phi f_0}$ decays rapidly because of smoothness of ϕ and $P_0(f)$.

Consider then contributions from ϕf_1. Then

$$\Gamma_{\phi f_1}(Q) = \int_{\mathcal{Q}_1} \langle \phi\gamma_Q, \gamma_{Q'}\rangle \Gamma_f(Q')\,\mathrm{d}\mu(Q').$$

We recall that $|\Gamma_f(Q)| = O(a^m)$ for each $m > 0$ uniformly over $\mathcal{B}_1 \times [0, 2\pi)$. We can further subdecompose $\mathcal{Q}_1 = \mathcal{Q}_{1,0} \cup \mathcal{Q}_{1,1} \cup \mathcal{Q}_{1,2}$, where $a' > \mathrm{diam}(\mathrm{supp}(\phi))$, where $a' \leqslant \mathrm{diam}(\mathrm{supp}(\phi)) < \sqrt{a'}$, and where $\sqrt{a'} \leqslant \mathrm{diam}(\mathrm{supp}(\phi))$. We get a corresponding decomposition $\Gamma_{\phi f_1}(Q) = G_{1,0}(Q) + G_{1,1}(Q) + G_{1,2}(Q)$. At fine scales $a \ll \mathrm{diam}(\mathrm{supp}(\phi))^2$, the main contribution to $\Gamma_{\phi f_1}$ will turn out to be provided by region $G_{1,2}$ arising from similarly fine scales; we spell out the argument in that case. For that region Lemma A.3 below shows that for each $N = 1, 2, \ldots,$ and supposing $\sqrt{a}, \sqrt{a'}$ are both small compared to $\delta = \mathrm{diam}(\mathrm{supp}(\phi))$,

$$\bigl|\langle \phi\gamma_Q, \gamma'_{Q'}\rangle\bigr| \leqslant C_N \cdot \left\langle \frac{a}{a'} \right\rangle^{-N} \cdot \left\langle \frac{a'}{a} \right\rangle^{-N} \bigl\langle d(b,' b)/\sqrt{a}\bigr\rangle^{-N} \quad \forall 0 < a, a' \leqslant a_0. \tag{A.3}$$

Note that for $m > 4$ and $N > 2m + 1$

$$\int_0^{\delta} \left\langle \frac{a}{a'} \right\rangle^{-N} \cdot \left\langle \frac{a'}{a} \right\rangle^{-N} (a')^m \frac{\mathrm{d}a'}{(a')^3} \leqslant C_{m,n,\delta} \cdot a^{m-2}, \quad 0 < a < \delta.$$

Combining the last three remarks, we get that $G_{1,2}(Q)$ is of rapid decay over the region of interest,

$$\bigl|G_{1,2}(Q)\bigr| \leqslant C \left| \int_{\mathcal{Q}_1} \left\langle \frac{a}{a'} \right\rangle^{-N} \cdot \left\langle \frac{a'}{a} \right\rangle^{-N} \cdot (a')^m \,\mathrm{d}\mu(Q') \right| \leqslant C a^{m-2}. \tag{A.4}$$

This is true for all $m > 4$, so $G_{1,2}(a, b, \theta)$ is of rapid decay as $a \to 0$, uniformly over $\mathcal{B}_0 \times [0, 2\pi)$. The arguments for $G_{1,i}(Q)$, $i = 1, 2$ are similar, using other branches of Lemma A.3. We conclude that $\Gamma_{\phi f_1}(a, b, \theta)$ is of rapid decay uniformly over the region of interest.

Consider now contributions from ϕf_2

$$\Gamma_{\phi f_2}(Q) = \int_{\mathcal{Q}_2} \langle \phi\gamma_Q, \gamma_{Q'}\rangle \Gamma_f(Q')\,\mathrm{d}\mu(Q').$$

We partition the integration region $\mathcal{Q}_2$ into two subsets, $\mathcal{Q}_{2,1}$ where $\mathrm{d}(b,b') > \eta$ and $\mathcal{Q}_{2,2}$ where $\mathrm{d}(b,b') \leqslant \eta$, getting $\Gamma_{\phi f_2}(Q) = G_{2,1}(Q) + G_{2,2}(Q)$, say. To study $G_{2,1}$, apply (A.3) noting that $\mathrm{d}(b,b') \geqslant \eta$. Arguing in a fashion similar to (A.6) in Lemma A.2 below, we get that for large N,

$$\int\limits_{d(b,'b)>\eta} \big\langle d(b,'b)/\sqrt{a}\big\rangle^{-N} \,\mathrm{d}b' \leqslant \int\limits_{\eta}^{\infty} \langle r/\sqrt{a}\,\rangle^{-N} r\,\mathrm{d}r \leqslant C\cdot a\cdot \langle \eta/\sqrt{a}\,\rangle^{-N+2}.$$

We note that as f is bounded, Γ_f is uniformly bounded by $Ca^{3/4}$. As a result, $G_{2,1}(a,b,\theta)$ will be of rapid decay as $a \to 0$, uniformly over $\mathcal{B}_1 \times [0,2\pi)$.

Now as for $G_{2,2}$, note that if $b \in \mathcal{B}_1$ and $d(b,b') < \eta$, then $b' \in \mathcal{B}$. Hence Γ_f is of rapid decay uniformly over $\mathcal{Q}_{2,2}$. Repeating the analysis leading to (A.4) gives exactly the same conclusion.

Combining these observations, $\Gamma_{\phi f_2}(a,b,\theta)$ will be of rapid decay as $a \to 0$, uniformly over $\mathcal{B}_1 \times [0,2\pi)$. As all the Γ_{g_i} are now seen to be of rapid decay, Lemma A.4 below shows that $\hat{g}_1(\xi)$ is of rapid decay as $|\xi| \to \infty$. We can now conclude that $\hat{g}$ is of rapid decay, completing the proof. □

A.3. Localization lemmas

We collect in this section the lemmas used in Theorem 5.1 above.

Lemma A.1. *If g is supported in a set $\mathcal{B}$ and $\|g\|_\infty \leqslant M$ then for all $N > 0$,*

$$\big|\Gamma_g(a,b,\theta)\big| \leqslant C_N \cdot M \cdot a^{1/4}\big\langle \mathrm{d}(b,\mathcal{B})/\sqrt{a}\,\big\rangle^{-N}.$$

Proof.

$$\big|\langle\gamma_{ab\theta}, g\rangle\big| \leqslant \|\gamma_{ab\theta}\|_{L^1(\mathcal{B})} \cdot \|g\|_{L^\infty(\mathcal{B})} = M\|\gamma_{ab\theta}\|_{L^1(\mathcal{B})}.$$

Now recalling Lemma 3.1, we have for $N > 0$,

$$\big|\gamma_{ab\theta}(x)\big| \leqslant C_N a^{-3/4}\big\langle |x-b|_{a,\theta}\big\rangle^{-N};$$

as $|x-b|_{a,\theta} \geqslant |x-b|/\sqrt{a}$, and putting $\eta = \mathrm{d}(b,\mathcal{B})$, we get

$$\int\limits_{\mathcal{B}} \big|\gamma_{ab\theta}(x)\big|\,\mathrm{d}x \leqslant \int\limits_{\eta}^{\infty} C_N a^{-3/4}\cdot \big\langle d(x,\mathcal{B})/\sqrt{a}\,\big\rangle^{-N}\,\mathrm{d}x \leqslant C_N a^{-3/4}\cdot \int\limits_{\eta}^{\infty} \langle r/\sqrt{a}\,\rangle^{-N} r\,\mathrm{d}r$$

$$= C_N a^{1/4}\cdot \int\limits_{\eta/\sqrt{a}}^{\infty} \langle u\rangle^{-N} u\,\mathrm{d}u = a^{1/4} G_N(\eta/\sqrt{a}\,),$$

where $G_N(u) \leqslant C_N' \langle u\rangle^{-N+2}$. The lemma is proved. □

Lemma A.2. *Let g be supported in a ball $\mathcal{B}$ and let $(\mathcal{B}^\eta)^c \equiv \{x\colon\ \mathrm{d}(x,B) > \eta\}$. Suppose that $\|g\|_\infty \leqslant C$. Let $\mathcal{T} \subset [0,2\pi)$. Define*

$$\hat{g}_0(\xi) = \int\limits_0^{a_0}\int\limits_{\mathcal{T}}\int\limits_{(\mathcal{B}^\eta)^c} \Gamma_g(a,b,\theta)\hat{\gamma}_{ab\theta}(\xi)\,\mathrm{d}\mu.$$

Then $\hat{g}_0(\xi)$ is of rapid decay as $|\xi| \to 0$, with constants that depend only on C and η.

Proof. From the previous lemma,

$$\left|\Gamma_g(a,b,\theta)\right| \leqslant C_N \cdot a^{1/4}\left\langle \mathrm{d}(b,B)/\sqrt{a}\right\rangle^{-N}.$$

Now define

$$\Xi(a,\theta) \equiv \left\{\xi\colon\ 1/2 \leqslant a|\xi| \leqslant 2,\ |\omega-\theta| \leqslant \sqrt{a}\right\};$$

this is the support of $\hat{\gamma}_{ab\theta}$. As $|\hat{\gamma}_{ab\theta}(\xi)| \leqslant Ca^{3/4}1_{\Xi(a,\theta)}(\xi)$,

$$\int\limits_{(\mathcal{B}^\eta)^c} \left|\Gamma_g(a,b,\theta)\right| |\hat{\gamma}_{ab\theta}(\xi)|\,\mathrm{d}b \leqslant C\cdot a\cdot 1_{\Xi(a,\theta)}(\xi)\cdot \int\limits_{(B^\eta)^c} \left\langle d(b,B)/\sqrt{a}\right\rangle^{-N}\mathrm{d}b. \tag{A.5}$$

Now using polar coordinates,

$$\int\limits_{(B^\eta)^c} \left\langle d(b,\mathcal{B})/\sqrt{a}\right\rangle^{-N}\mathrm{d}b \leqslant \int\limits_{\eta}^{\infty} \langle r/\sqrt{a}\rangle^{-N} r\,\mathrm{d}r \leqslant a\cdot C_N'\cdot\langle \eta/\sqrt{a}\rangle^{-N+2}. \tag{A.6}$$

Recalling the definition of $\Xi(a,\theta)$, $\int 1_{\Xi(a,\theta)}(\xi)\,\mathrm{d}\theta \leqslant C\sqrt{a}$, so

$$\int\limits_0^{a_0}\int\limits_{\mathcal{T}} 1_{\Xi(a,\theta)}(\xi)a^2\langle\eta/\sqrt{a}\rangle^{-N}\,\mathrm{d}\theta\,\mathrm{d}a \leqslant C\cdot \int\limits_{1/2|\xi|}^{2/|\xi|} a^{5/2}\langle\eta/\sqrt{a}\rangle^{-N}\,\mathrm{d}a \leqslant C\cdot|\xi|^{-7/2}\cdot\langle\eta\sqrt{|\xi|/2}\rangle^{-N}.$$

As this is true for every $N > 0$, the lemma is proved. □

Lemma A.3. *Let ϕ_1 be C^∞ and supported in $B(0,1)$. Let $\phi(x) = \phi_1((x-b)/a_\phi)$. Suppose $\sqrt{a}, \sqrt{a'} \leqslant a_\phi$. Then for $N > 0$,*

$$\left|\langle\phi\gamma_{ab\theta},\gamma_{a'b'\theta'}\rangle\right| \leqslant C_N\cdot\left\langle\frac{a}{a'}\right\rangle^{-N}\cdot\left\langle\frac{a'}{a}\right\rangle^{-N}\left\langle d(b,'b)/\sqrt{a}\right\rangle^{-N}\cdot\left\langle d(\theta,'\theta)/\sqrt{a}\right\rangle^{-N}\quad \forall 0 < a, a' \leqslant a_\phi^2. \tag{A.7}$$

Suppose $\sqrt{a'} \leqslant a_\phi$, $a \leqslant a_\phi < \sqrt{a}$. Then for $N > 0$,

$$\left|\langle\phi\gamma_{ab\theta},\gamma_{a'b'\theta'}\rangle\right| \leqslant C_N\cdot\left\langle\frac{a}{a'}\right\rangle^{-N}\cdot\left\langle\frac{a'}{a}\right\rangle^{-N}\left\langle d(b,'b)/\sqrt{a}\right\rangle^{-N}\cdot\left\langle d(\theta,'\theta)/a_\phi\right\rangle^{-N}. \tag{A.8}$$

Suppose $\sqrt{a'} \leqslant a_\phi$, $a_\phi \leqslant a \leqslant a_0$. Then for $N > 0$,

$$\left|\langle\phi\gamma_{ab\theta},\gamma_{a'b'\theta'}\rangle\right| \leqslant C_N\cdot\left\langle\frac{a'}{a_\phi}\right\rangle^{-N}\left\langle d(b,'b)/a_\phi\right\rangle^{-N}. \tag{A.9}$$

Lemma A.4. *Let $\mathcal{B}$ be a compact set, and let $\mathcal{Q} = (0,a_0]\times\mathcal{B}\times\mathcal{T}$. Suppose that $G(a,b,\theta)$ is of rapid decay as $a \to 0$, uniformly in $\mathcal{Q}$. Define*

$$\hat{g}_0(\xi) = \int\limits_0^{a_0}\int\limits_{\mathcal{T}}\int\limits_{\mathcal{B}} \hat{\gamma}_{ab\theta}(\xi)G(a,b,\theta)\,\mathrm{d}\mu.$$

Then $\hat{g}_0(\xi)$ goes to zero rapidly as $|\xi| \to \infty$.

Proof. Recall that

$$\left|\hat{\gamma}_{ab\theta}(\xi)\right| \leqslant Ca^{3/4} \cdot 1_{\Xi(a,\theta)}(\xi),$$

where again $\Xi(a,\theta) \equiv \{1/2a \leqslant |\xi| \leqslant 2/a,\ |\omega - \theta| \leqslant \sqrt{a}\}$. Our hypothesis gives for each $N > 0$

$$\sup\left\{\left|G(a,b,\theta)\right|: 1/2 \leqslant a|\xi| \leqslant 2, b \in \mathcal{B}\right\} \leqslant C_n a^N, \quad 0 < a < a_0.$$

We then have for each N

$$\begin{aligned}\left|\hat{g}_0(\xi)\right| &= \left|\int_{\mathcal{Q}} \hat{\gamma}_{ab\theta}(\xi) G(a,b,\theta) a^{-3}\,\mathrm{d}b\,\mathrm{d}\theta\,\mathrm{d}a\right| \leqslant C_N a^{3/4} \cdot \int_{\mathcal{Q}} 1_{\Xi(a,\theta)}(\xi) a^{N-3}\,\mathrm{d}b\,\mathrm{d}\theta\,\mathrm{d}a \\ &\leqslant C_N \cdot |\xi|^{-N+2.25}. \qquad \square\end{aligned}$$

A.4. Proof of Theorem 5.2

There are two ways to prove this result; we briefly describe each. The more sophisticated approach is to adapt Theorem 5.3, and use the fact that the complement of the wavefront set is exactly where one is microlocally in every H^m, $\forall m$. Then one shows that every $S^m(f)$ is locally square integrable if and only if Γ_f is uniformly of rapid decay. There are two key points in this last equivalence. First, and most obviously, is the fact that, for each fixed choice of (a,θ), $\Gamma(a,b,\theta)$ is a bandlimited function of b, and so its $L^2(db)$ norm over a compact interval is comparable to its $L^\infty(db)$ norm:

$$\left\|\Gamma(a,\cdot,\theta)\right\|_{L^\infty(\mathcal{B})} \leqslant C\left\|\Gamma(a,\cdot,\theta)\right\|_{L^2(\mathcal{B})}$$

while

$$\left\|\Gamma(a,\cdot,\theta)\right\|_{L^2(\mathcal{B})} \leqslant Ca^{-3/4} \cdot \left\|\Gamma(a,\cdot,\theta)\right\|_{L^\infty(\mathcal{B})}.$$

Thus L^2 control at a certain m will guarantee uniform control at a certain m'. Second, and more subtly, if f is microlocally in every H^m at a point, one shows that the implicit sequence of neighborhoods $\mathcal{N}_m$, which seems to depend on m, can be chosen to be the same, $\mathcal{N}_0$, for every m.

The second approach is more concrete. We merely repeat the proof of Theorem 5.1 about the singular support, referring now in every instance, not to $\mathcal{B}_1 \times [0,2\pi)$ but to $\mathcal{B}_1 \times \mathcal{T}_1$ where $\mathcal{T}_1$ is a neighborhood of θ_0. We make exactly the same decomposition, e.g., into $\hat{g}(\xi) = \hat{g}_0(\xi) + \hat{g}_1(\xi) + \hat{g}_2(\xi)$, only this time the sets $\mathcal{B}_1$ and $\mathcal{B}_1^c$ are replaced by $\mathcal{B}_1 \times \mathcal{T}_1$ and $(\mathcal{B}_1 \times \mathcal{T}_1)^c$. We then sharpen the inequalities involved to add angular sensitivity, for example, with

$$\left|\langle\phi\gamma_{ab\theta}, \gamma_{a'b'\theta'}\rangle\right| \leqslant C_N \cdot \left\langle\frac{a}{a'}\right\rangle^{-N} \cdot \left\langle\frac{a}{a'}\right\rangle^{-N} \left\langle d(b,b')/\sqrt{a}\right\rangle^{-N} \cdot \left\langle d(\theta,\theta')/\sqrt{a}\right\rangle^{-N} \quad \forall 0 < a, a' \leqslant a_0. \tag{A.10}$$

A.5. Proof of Theorem 5.3

We again prove the result only under the additional assumption that f is a bounded function.

A.5.1. Proof that microlocal H^m implies integrability of S^m

Suppose that for every C^∞ function ϕ identically one in some ball around x_0 and vanishing outside a (larger) ball, and for all sufficiently small $\delta > 0$, the δ-aperture cone in frequency space,

$$\mathcal{C} = \mathcal{C}_{\theta_0,\delta} = \{\lambda e_\omega \colon \lambda > 0,\ |\omega - \theta_0| < \delta\}$$

obeys the sectorial integrability

$$\int_{\mathcal{C}} \bigl|\widehat{\phi f}(\xi)\bigr|^2 |\xi|^{2m}\, d\xi < \beta_1.$$

We will show that there exist a_0, $\mathcal{B}$, and $\mathcal{T}$ so that the region $\mathcal{Q} = (0, a_0] \times \mathcal{B} \times \mathcal{T}$ obeys the integrability

$$\int_{\mathcal{Q}} \bigl|\Gamma_f(a,b,\theta)\bigr|^2 a^{-2m}\, d\mu < \beta_2, \tag{A.11}$$

where β_2 depends on β_1, Γ, $\operatorname{supp}(\phi)$, m and $\|f\|_\infty$.

Proof. Choose ϕ as guaranteed in the statement of the hypothesis. Let $\mathcal{B}$ be a ball contained in the set where $\phi = 1$, so that $d(\mathcal{B}, (\operatorname{supp}\phi)^c) \equiv \eta > 0$. We first show that

$$\int_0^{a_0}\int_0^{2\pi}\int_{\mathcal{B}} \bigl|\langle \gamma_{ab\theta}, (1-\phi) f\rangle\bigr|^2 a^{-2m}\, d\mu < \beta_{2,0}. \tag{A.12}$$

We later consider a similar integral involving ϕf, only over a smaller range of angles. Applying Lemma A.2 to $g = (1-\phi) f$, we get for each $N > 0$ and $b \in \mathcal{B}$

$$\bigl|\langle \gamma_{ab\theta}, (1-\phi) f\rangle\bigr| \leqslant C_N a^{1/4} \langle \eta/\sqrt{a}\,\rangle^{-N},$$

picking $N > m + 2$, we get (A.12).

We now wish to show that, picking a neighborhood $\mathcal{T}$ of θ_0.

$$\int_0^{a_0}\int_{\mathcal{T}}\int_{\mathcal{B}} \bigl|\langle \gamma_{ab\theta}, \phi f\rangle\bigr|^2 a^{-2m}\, d\mu < \infty. \tag{A.13}$$

The desired conclusion (A.11) will then follow.

It will be convenient to let Q denote a variable tuple (a, b, θ). Then, it is understood that references to a, b, etc., refer to appropriate components of the Q currently under consideration. Introduce notation $A(Q) \equiv \Gamma(Q) a^{-m}$ and $B(\xi) \equiv \hat{f}|\xi|^m$, we have $A(Q) = \int K_m(Q,\xi) B(\xi)\, d\xi$, where $K_m(Q,\xi) \equiv \hat{\gamma}_Q(\xi)(a|\xi|)^{-m}$. With this notation, suppose we can show, that for proper choice of $\mathcal{T}$ and a_0, and with the notation $\mathcal{Q} = (0, a_0] \times \mathcal{T} \times \mathcal{B}$, K_m is the kernel of a bounded operator

$$T_m : L^2(\mathcal{C}, d\xi) \mapsto L^2(\mathcal{Q}, d\mu). \tag{A.14}$$

Until further comment all integrations over ξ are understood to range over $\mathcal{C}$, and we drop the explicit subscript m on K_m. Define

$$A^2(a,\theta) = \int \left| \int K_m\bigl((a,b,\theta),\xi\bigr) \hat{f}(\xi)\, d\xi \right|^2 db.$$

Then $A^2(a,\theta) = \|g_{a,\theta}(b)\|_{L^2}^2$, where $g_{a,\theta}(b) = \int e^{-i\xi' b}\hat{\gamma}_{a0\theta}(\xi)(a|\xi|)^{-m}\hat{f}(\xi)\,\mathrm{d}\xi$. By Parseval,

$$\|g_{a,\theta}\|_{L^2}^2 = (2\pi)^{-2}\int \bigl|\hat{\gamma}_{a0\theta}(\xi)\bigl(a|\xi|\bigr)^{-m}\bigr|^2\bigl|\hat{f}(\xi)\bigr|^2\,\mathrm{d}\xi.$$

As $|\hat{\gamma}_{a0\theta}(\xi)(a|\xi|)^{-m}| \leqslant C_m a^{3/4} 1_{\Xi(a,\theta)}(\xi)$, we get $A^2(a,\theta) \leqslant a^{3/2}\int 1_{\Xi(a,\theta)}(\xi)|\hat{f}(\xi)|^2\,\mathrm{d}\xi$.

For a set $\mathcal{Q}$ still to be determined,

$$\int_{\mathcal{Q}} \bigl|A(Q)\bigr|^2\,\mathrm{d}\mu = \int A^2(a,\theta)a^{-3}\,\mathrm{d}\theta\,\mathrm{d}a \leqslant C_m\cdot\int a^{3/2}\biggl(\int 1_{\Xi(a,\theta)}\bigl|\hat{f}(\xi)\bigr|^2\,\mathrm{d}\xi\biggr)a^{-3}\,\mathrm{d}\theta\,\mathrm{d}a$$

$$= C_m\cdot\int\bigl|\hat{f}(\xi)\bigr|^2\biggl(\int_{\mathcal{Q}} 1_{\Xi(a,\theta)}(\xi)a^{-3/2}\,\mathrm{d}\theta\,\mathrm{d}a\biggr)\mathrm{d}\xi = C_m\cdot\int\bigl|\hat{f}(\xi)\bigr|^2 M_{\mathcal{Q}}(\xi)\,\mathrm{d}\xi,$$

say, where we have put

$$M_{\mathcal{Q}}(\xi) \equiv \biggl(\int_{\mathcal{Q}} 1_{\Xi(a,\theta)}(\xi)a^{-3/2}\,\mathrm{d}\theta\,\mathrm{d}a\biggr).$$

The desired operator boundedness will follow if we can arrange for the set $\mathcal{Q}$ to satisfy the pair of conditions:

$$\operatorname{supp}(M_{\mathcal{Q}}) \subset \mathcal{C}; \tag{A.15}$$

$$\|M_{\mathcal{Q}}\|_\infty < \infty. \tag{A.16}$$

Fix a_0 so small that $2\sqrt{a_0} < \delta$, where δ is the aperture of the cone $\mathcal{C} = \mathcal{C}_{\theta_0,\delta}$. Set $\mathcal{T} = (\theta_0 - \sqrt{a_0}, \theta_0 + \sqrt{a_0})$. Then we have the inclusion

$$a \in (0, a_0],\quad \theta \in \mathcal{T},\quad |\omega - \theta| < \sqrt{a_0} \quad\Longrightarrow\quad a^{-1}e_\omega \in \mathcal{C}.$$

It follows from this that

$$a \in (0, a_0],\quad \theta \in \mathcal{T} \quad\Longrightarrow\quad \Xi(a,\theta) \subset \mathcal{C}.$$

Define then $\mathcal{Q} = (0, a_0] \times \mathcal{B} \times \mathcal{T}$. The conclusion (A.15) follows. To get (A.16), note that $\int_{\Xi(a,\theta)}\mathrm{d}\theta\,\mathrm{d}a \leqslant Ca^{3/2}$.

A.5.2. *Proof that integrability of S^m implies microlocal H^m*

We suppose there is a region $\mathcal{Q}$ of the form $(0, a_0] \times \mathcal{B} \times \mathcal{T}$ where $(x_0, \theta_0) \subset \mathcal{B} \times \mathcal{T}$ so that

$$\int_{\mathcal{Q}} |\Gamma_f|^2 a^{-2m}\,\mathrm{d}\mu < \infty.$$

We use Lemma A.5 below to conclude that there are $\mathcal{B}_1 \Subset \mathcal{B}$ and $\mathcal{T}_1 \Subset \mathcal{T}$ forming a region $\mathcal{Q}_1 = (0, a_0] \times \mathcal{B}_1 \times \mathcal{T}_1 \subset \mathcal{Q}$ so that for all $\phi \in C^\infty$ supported in sufficiently small neighborhoods of x_0

$$\int_{\mathcal{Q}_1} |\Gamma_{\phi f}|^2 a^{-2m}\,\mathrm{d}\mu < \beta; \tag{A.17}$$

i.e. we are inferring decay of the CCT of ϕf.

We will pick a cone $\mathcal{C} = \mathcal{C}_{x_0,\theta_0}$ which is associated at high frequencies with the interior of $\mathcal{Q}_1$, i.e. so that for some $\lambda_0 > 0$,

$$\mathcal{C} \cap \{\xi\colon |\xi| > \lambda_0\} \Subset \{\xi = a^{-1}e_\theta\colon Q = (a,b,\theta) \in \mathcal{Q}\}. \tag{A.18}$$

We wish to infer finiteness of the microlocal Sobolev integral

$$\int_{\mathcal{C}} \left|\widehat{\phi f}(\xi)\right|^2 |\xi|^{2m}\,\mathrm{d}\xi < \eta, \tag{A.19}$$

where η depends only on β, the bound $\|f\|_\infty \leqslant M$, the size of $\mathcal{B}$ and the size of the support of ϕ.

We make the decomposition

$$\widehat{(\phi f)}(\xi) = \widehat{P_0(\phi f)}(\xi) + \int_{\mathcal{Q}_1} + \int_{\mathcal{Q}_2} \hat{\gamma}_{ab\theta}(\xi)\Gamma_{\phi f}\,\mathrm{d}\mu = \hat{g}_0(\xi) + \hat{g}_1(\xi) + \hat{g}_2(\xi), \tag{A.20}$$

say, where by $\mathcal{Q}_2$ we mean $(0,a_1] \times (\mathcal{T}_1 \times \mathcal{B}_1)^c$. We wish to show that

$$\int_{\mathcal{C}} \left|\hat{g}_i(\xi)\right|^2 |\xi|^{2m}\,\mathrm{d}\xi < \eta_i, \quad i = 0,1,2, \tag{A.21}$$

where we can control the η_i using β, the support properties of ϕ, the bound on f, and m.

The contribution of $\hat{g}_0$ to the integral in (A.21) can be controlled easily using the support properties of ϕ and boundedness of f; as $\mathrm{supp}(\phi f) \subset \mathcal{B}$ and $P_0()$ is a convolution operator with kernel Ψ, we can show $\|P_0(\phi f)\|_{W_2^m} \leqslant \|\Psi\|_{W_2^m} \cdot \|\phi f\|_{L^1}$, where W_2^m denotes the usual L^2 Sobolev norm of mth order, and so η_0 depends merely on m, on $\|f\|_\infty$ and $\mathrm{supp}(\phi)$.

The contribution of $\hat{g}_2$ can further be decomposed; noting

$$(\mathcal{B}_1 \times \mathcal{T}_1)^c \subset (\mathcal{B}_1 \times \mathcal{T}_1^c) \cup (\mathcal{B}_1^c \times \mathcal{T}_1) \cup (\mathcal{B}_1^c \times \mathcal{T}_1^c),$$

we can define regions $\mathcal{Q}_{2,i}$ with, e.g., $\mathcal{Q}_{2,1} = \mathcal{B}_1 \times \mathcal{T}_1^c$, etc. Using this, we can decompose $\hat{g}_2 = \sum_{j=1}^3 \hat{g}_{2j}$ according to integration over appropriate subregions in (A.20). (There is no issue with how we handle the overlapping parts as double-counting will make no difference.) We then naturally bound the individual contributions $\eta_{2,j}$, $j = 1,2,3$ to the sector integral (A.21) with $i = 2$.

By construction of Γ in (A.18), $\mathcal{T}_1^c$ is separated from the set of directions in Γ by at least a fixed angular distance, δ, say. Hence, for sufficiently small $a < \delta^2$, every $\hat{\gamma}_{ab\theta}(\xi)$ is zero whenever $\theta \in \mathcal{T}_1^c$ and $\xi \in \Gamma$. Hence, $\hat{g}_{21}(\lambda e_\omega)$ vanishes for large $\lambda > \lambda_2$, uniformly in f. This geometric fact leads to bounds on $\eta_{2,1}$ as follows:

$$\hat{g}_{21}(\lambda e_\omega) = \int_{\mathcal{Q}_{2,1}} 1_{\{a>\delta^2\}} \hat{\gamma}_{ab\theta}(\xi)\Gamma_{\phi f}\,\mathrm{d}\mu.$$

Now $|\Gamma_{\phi f}(a,b,\theta)| \leqslant Ca^{3/4}\|f\|_\infty$; while $|\gamma_{ab\theta}(\xi)| \leqslant Ca^{3/4} 1_{\Xi(a,\theta)}(\xi)$. Now from $\int_{\mathcal{B}_1} \mathrm{d}b \leqslant C$ and $\int 1_{\Xi(a,\theta)}\,\mathrm{d}\theta \leqslant Ca^{1/2}$, we conclude that for $\xi \in \Gamma$,

$$\int_{\mathcal{T}_1^c} \int_{\mathcal{B}} \left|\hat{\gamma}_{ab\theta}(\xi)\right| |\Gamma_{\phi f}| \frac{\mathrm{d}b}{a^{3/2}} \frac{\mathrm{d}\theta}{a^{1/2}} \leqslant C \cdot 1_{\{1/2 \leqslant a|\xi| \leqslant 2\}} \cdot 1_{\{a>\delta^2\}} \cdot \|f\|_\infty.$$

Hence, $|\hat{g}_{21}(\xi)| \leqslant C$ and

$$\int_{\mathcal{C}} \left|\hat{g}_{21}(\xi)\right|^2 |\xi|^{2m}\,\mathrm{d}\xi \leqslant C \cdot \int_0^{C/\delta^2} r^{2m} r\,\mathrm{d}r \leqslant C \cdot \delta^{-4m-4};$$

reviewing the argument shows $\eta_{2,1}$ depends on the support properties of ϕ, the bound on f, the separation constant δ, and m.

Lemma A.2, applied to $\hat{g}_{22}$, shows that it is of rapid decay, uniformly in f bounded by M and in the radius of $\mathcal{B}$ and support of ϕ. This leads directly to an acceptable bound for $\eta_{2,2}$.

The term $\hat{g}_{23}$ associated with $(\mathcal{T}_1^c \times \mathcal{B}_1^c)$ can be bounded by either or both of the arguments used on the pieces $\hat{g}_{21}$, $\hat{g}_{22}$. We conclude that the contribution of $\hat{g}_2$ to the microlocal Sobolev integral in (A.21) can be bounded by a η_2 controlled in terms of β, the bound on f, exactly as desired.

We now focus on the term $\hat{g}_1$. We wish to show that (A.17) implies

$$\int_{\mathcal{C}} \left|\hat{g}_1(\xi)\right|^2 |\xi|^{2m}\,\mathrm{d}\xi < \eta_1. \tag{A.22}$$

This is settled by Lemma A.6.

A.5.3. Lemmas used in Theorem 5.3

Lemma A.5. *Suppose that* $\mathcal{Q}_0 = (0, a_0] \times (\theta_0 - \delta_0, \theta_0 + \delta_0) \times B(x_0, \delta_0)$, *and*

$$\int_{\mathcal{Q}_0} \left|\Gamma_f(a, b, \theta)\right|^2 a^{-2m}\,\mathrm{d}\mu \leqslant \beta_1.$$

Then there is a smooth ϕ *with* $\operatorname{supp}(\phi) \Subset B(x_0, \delta_0)$, *so that*

$$\int_{\mathcal{Q}_0} \left|\Gamma_{\phi f}(a, b, \theta)\right|^2 a^{-2m}\,\mathrm{d}\mu \leqslant \beta_2,$$

where β_2 *depends only on* β_1, $\mathcal{Q}_0$ *and the size of* $\operatorname{supp}(\phi)$.

Proof. We choose ϕ as the dilation and translation of a standard C^∞ 'bump' function so that it is localized to the given ball of radius δ_0, with $\|\phi\|_\infty = 1$ and $\|\phi\|_2 = c \cdot \delta_0$. The CCT coefficients of ϕf can be decomposed into contributions from finer and coarser scales

$$\Gamma_{\phi f}(Q) = \int \Gamma_f(Q') K_0(Q, Q')\,\mathrm{d}\mu(Q') + \int P_0(f)(b') K_1(Q, b')\,\mathrm{d}b',$$

where $K_0(Q, Q') \equiv \langle \phi\gamma_Q, \gamma_{Q'}\rangle$ and $K_1(Q, b') \equiv \langle \phi\gamma_Q, \Phi(\cdot - b')\rangle$. We will analyze only the first term on the RHS, showing that

$$\int_{\mathcal{Q}_0} |\Gamma_{\phi f}|^2 a^{-2m}\,\mathrm{d}\mu \leqslant C \cdot \int_{\mathcal{Q}_0} |\Gamma_f|^2 a^{-2m}\,\mathrm{d}\mu; \tag{A.23}$$

the second term can be treated similarly, and the two inequalities together provide the desired result.

Define $A(Q) = a^{-m}\Gamma_{\phi f}(Q)$, and $B(Q) = a^{-m}\Gamma_f(Q)$, and $K_m(Q,Q') = K_0(Q,Q')(\frac{a'}{a})^m$. Then $A(Q) = \int K_m(Q,Q')B(Q')\,\mathrm{d}\mu(Q')$. If K_m defines a bounded mapping from $L^2(\mathcal{Q}_0, d\mu)$ to itself, then (A.23) holds.

Schur's lemma [12] bounds the operator norm of K_m by the geometric mean of these quantities

$$I = \sup_Q \int \bigl|K_m(Q,Q')\bigr|\,\mathrm{d}\mu(Q'), \qquad II = \sup_{Q'} \int \bigl|K_m(Q,Q')\bigr|\,\mathrm{d}\mu(Q).$$

We develop a bound on I; II can be handled similarly. Consider the case where the scale a of Q is $a = \delta_0$. This is essentially the worst case for the supremum defining I, and the other cases can be handled similarly. Let B denote the support of ϕ. Using the localization Lemma 3.1, we have

$$\bigl|K_0(Q,Q')\bigr| \leqslant \|\phi\gamma_Q\|_{L^1(B)} \cdot \|\gamma_{Q'}\|_{L^\infty(B)} \leqslant c a^{3/4}\delta_0^2 \cdot c(a')^{-3/4}\bigl\langle |B-b'|_{a,'\theta'}\bigr\rangle^{-N},$$

where

$$|B-b'|_{a,'\theta'} \equiv \inf_{b\in B} |b-b'|_{a,'\theta'}.$$

We note that at scales $a' \gg \delta_0$,

$$\int \bigl\langle |B-b'|_{a,'\theta'}\bigr\rangle^{-N}\,\mathrm{d}b' \leqslant c \int \bigl\langle |b'|_{a,'\theta'}\bigr\rangle^{-N}\,\mathrm{d}b'.$$

Moreover, by rescaling variables,

$$\int \bigl\langle |b'|_{a,'\theta'}\bigr\rangle^{-N}\,\mathrm{d}b'/(a')^{3/2} = \int \bigl\langle |x|\bigr\rangle^{-N}\,\mathrm{d}x = c.$$

In the case $m > 5/4$, the coarse-scale integral obeys

$$\begin{aligned}\int_{\delta_0}^{a_0} \bigl|K_m(Q,Q')\bigr|\,\mathrm{d}\mu(Q') &= \int_{\delta_0}^{a_0} (a'/a)^m \bigl|K_0(Q,Q')\bigr|\,\mathrm{d}\mu(Q') \leqslant c\delta_0^2 \cdot \int_{\delta_0}^{a_0} (a'/a)^{m-3/4}(a')^{-3/2}\,\mathrm{d}a' \\ &\leqslant c \cdot \delta_0^{-m+11/4}.\end{aligned}$$

For the fine-scale integral $\int_{a'\leqslant\delta_0} |K_m(Q,Q')|\,\mathrm{d}\mu(Q')$ we use the estimate

$$\int \bigl\langle |B-b'|_{a,'\theta'}\bigr\rangle^{-N}\,\mathrm{d}b' \leqslant c \cdot \mathrm{Area}(B) = c'\delta_0^2,$$

valid for $a' < \delta_0$ to conclude if $m > 11/4$ that

$$\int_{a'\leqslant\delta_0} \bigl|K_m(Q,Q')\bigr|\,\mathrm{d}\mu(Q') \leqslant c(\delta_0)^4 \cdot a^{-m+3/4} \cdot \int_0^{\delta_0} (a')^{m-15/4}\,\mathrm{d}a' \leqslant c\delta_0^{-m+19/4}.$$

We infer that I can be controlled in terms of δ_0 alone, i.e. diam(supp(ϕ)). Similar arguments will control II. The result is a bound on the operator norm of K_m in terms of diam(supp(ϕ)).

Lemma A.6. *Let $\mathcal{C}$ be a cone in Fourier space. Let*

$$\hat{g}_0(\xi) = \int_{\mathcal{Q}_1} \Gamma_f(Q)\hat{\gamma}_Q(\xi)\,\mathrm{d}\mu.$$

Then

$$\int_{\mathcal{C}} \left|\widehat{\phi f}(\xi)\right|^2 |\xi|^{2m}\,\mathrm{d}\xi \leqslant C'_m \int_{\mathcal{Q}_1} \left|\Gamma_{\phi f}(a,b,\theta)\right|^2 a^{-2m}\,\mathrm{d}\mu.$$

Proof. Set $B(Q) = \Gamma_{\phi f}(Q)a^{-m}$, $A(\xi) = |\xi|^m \hat{g}_1(\xi)$, and $K_m(\xi, Q) = \hat{\gamma}_Q(\xi)(a|\xi|)^m$. Then $A(\xi) = \int_{\mathcal{Q}_1} K_m(\xi, Q)B(Q)\,\mathrm{d}\mu(Q)$. Then (A.22) follows if K_m is the kernel of the bounded linear operator

$$T_m : L^2\big(\mathcal{Q}_1, \mathrm{d}\mu(Q)\big) \mapsto L^2(\mathcal{C}, \mathrm{d}\xi). \tag{A.24}$$

To see this, decompose $A(\xi)$ into its contributions, scale-by-scale, $A(\xi) = \int A_a(\xi)\frac{\mathrm{d}a}{a}$, where

$$A_a(\xi) = \int B(Q)K_m(\xi, Q)a^{-2}\,\mathrm{d}b\,\mathrm{d}\theta,$$

and set $\hat{B}(a,\xi,\theta) = \int B(a,b,\theta)e^{-i\xi' b}\,\mathrm{d}b$. Then

$$\begin{aligned}
A_a(\xi) &= \int B(a,b,\theta)e^{-i\xi' b}\big(a|\xi|\big)^m \hat{\gamma}_{a0\theta}(\xi)a^{-2}\,\mathrm{d}b\,\mathrm{d}\theta = \int \hat{B}\big(a,\xi,\theta\big)\big(a|\xi|\big)^m \hat{\gamma}_{a0\theta}(\xi)a^{-2}\,\mathrm{d}\theta \\
&= \int \hat{B}\big(a,\xi,\theta\big)\big(a|\xi|\big)^m W\big(a|\xi|\big)V\left(\frac{\omega-\theta}{\sqrt{a}}\right)a^{3/4}a^{-2}\,\mathrm{d}\theta \\
&= \big(a|\xi|\big)^m W\big(a|\xi|\big)a^{-3/4}\int \hat{B}(a,\xi,\theta)V\left(\frac{\omega-\theta}{\sqrt{a}}\right)a^{-1/2}\,\mathrm{d}\theta \\
&= \big(a|\xi|\big)^m W\big(a|\xi|\big)a^{-3/4}\hat{B}(a,\xi),
\end{aligned}$$

say. Now

$$A(\xi) = \int A_a(\xi)\frac{\mathrm{d}a}{a} = \int \big(a|\xi|\big)^m W\big(a|\xi|\big)a^{-3/4}\hat{B}(a,\xi)\frac{\mathrm{d}a}{a},$$

and, as $W(ar)$ is supported in $1/2 \leqslant ar \leqslant 2$,

$$\big|A(\xi)\big|^2 \leqslant \left(\int_{1/2|\xi|}^{2/|\xi|} \big|\big(a|\xi|\big)^m W\big(a|\xi|\big)a^{-3/4}\hat{B}(a,\xi)\big|^2\frac{\mathrm{d}a}{a}\right)\cdot\left(\int_{1/2|\xi|}^{2/|\xi|}\frac{\mathrm{d}a}{a}\right).$$

Also,

$$\int_{1/2|\xi|}^{2/|\xi|} \big|\big(a|\xi|\big)^m W\big(a|\xi|\big)a^{-3/4}\hat{B}(a,\xi)\big|^2\frac{\mathrm{d}a}{a} \leqslant C_m\cdot|\xi|^{-3/2}\cdot\int_{1/2|\xi|}^{2/|\xi|}\big|\hat{B}(a,\xi)\big|^2\frac{\mathrm{d}a}{a}.$$

We conclude that for an arbitrary cone $\mathcal{C}$,

$$\int_{\mathcal{C}} \big|A(\xi)\big|^2\,\mathrm{d}\xi \leqslant C_m \int_{\mathcal{C}} |\xi|^{-3/2}\cdot\left(\int_{1/2|\xi|}^{2/|\xi|}\big|\hat{B}(a,\xi)\big|^2\frac{\mathrm{d}a}{a}\right)\mathrm{d}\xi. \tag{A.25}$$

In a moment, we will establish that the RHS of (A.25) is bounded by

$$\leqslant C \int_0^{a_1} \int_{\mathcal{T}_1} \int_{\mathcal{B}_1} \left|B(a,b,\theta)\right|^2 \frac{\mathrm{d}a\,\mathrm{d}b\,\mathrm{d}\theta}{a^3}.$$

This implies

$$\int_{\mathcal{C}} \left|A(\xi)\right|^2 \mathrm{d}\xi \leqslant C'_m \int_{\mathcal{Q}_1} \left|B(Q)\right|^2 \mathrm{d}\mu(Q), \tag{A.26}$$

and (A.24) follows. To get (A.25), recall

$$\hat{B}(a,\xi) = \int_{\mathcal{T}_1} \hat{B}(a,\xi,\theta) V\left(\frac{\omega-\theta}{\sqrt{a}}\right)\frac{\mathrm{d}\theta}{\sqrt{a}}.$$

Use Cauchy–Schwarz to write

$$\begin{aligned}\left|\hat{B}(a,\xi)\right|^2 &\leqslant \left(\int_{\mathcal{T}_1} \left|\hat{B}(a,\xi,\theta)\right|^2 1_{\{|\theta-\omega|\leqslant\sqrt{a}\}}\,\mathrm{d}\theta\right)\cdot\left(\int V^2\left(\frac{\omega-\theta}{\sqrt{a}}\right)\frac{\mathrm{d}\theta}{a}\right)\\ &\leqslant C\cdot a^{-1/2}\int_{\mathcal{T}_1}\left|\hat{B}(a,\xi,\theta)\right|^2 1_{\{|\theta-\omega|\leqslant\sqrt{a}\}}\,\mathrm{d}\theta.\end{aligned}$$

Note also that, by definition of $\hat{B}$ and Parseval,

$$\int \left|\hat{B}(a,\xi,\theta)\right|^2 \mathrm{d}\theta = (2\pi)^2 \int_{\mathcal{B}_1} \left|B(a,b,\theta)\right|^2 \mathrm{d}b.$$

Let $J_1(a,\theta,\omega) = 1_{\{|\theta-\omega|\leqslant\sqrt{a}\}}$ and $J_2(a,\xi) = 1_{\{1/2\leqslant a|\xi|\leqslant 2\}}$. Then clearly

$$\int \left|\hat{B}(a,\xi,\theta)\right|^2 J_1(a,\omega,\theta) J_2(a,\xi)\,\mathrm{d}\xi \leqslant (2\pi)^2 \int_{\mathcal{B}_1} \left|B(a,b,\theta)\right|^2 \mathrm{d}b.$$

Collecting these comments and applying them to the RHS of (A.25) gives

$$\begin{aligned}\int_{\mathcal{C}} |\xi|^{-3/2}\left(\int_{1/2|\xi|}^{2/|\xi|}\left|\hat{B}(a,\xi)\right|^2\frac{\mathrm{d}a}{a}\right)\mathrm{d}\xi &= \int_0^{a_1}\int_{\mathcal{C}}\left|\hat{B}(a,\xi)\right|^2|\xi|^{-3/2}J_2(a,\xi)\,\mathrm{d}\xi\,\frac{\mathrm{d}a}{a}\\ &\leqslant C\int_0^{a_1}\int_{\mathcal{T}_1}\int_{\mathcal{C}}\left|\hat{B}(a,\xi,\theta)\right|^2|\xi|^{-3/2}a^{-1/2}J_1(a,\omega,\theta)J_2(a,\xi)\,\mathrm{d}\xi\,\mathrm{d}\theta\,\frac{\mathrm{d}a}{a}\\ &\leqslant C'\int_0^{a_1}\int_{\mathcal{T}_1}\int_{\mathcal{B}_1}\left|B(a,b,\theta)\right|^2 a^{-2}\,\mathrm{d}b\,\mathrm{d}\theta\,\frac{\mathrm{d}a}{a}.\end{aligned}$$

This gives (A.26). □

A.6. Proof of Theorem 6.1

The argument is the same as for Theorem 5.1, only with curvelets, with their anisotropic estimates, replaced by wavelets and their isotropic estimates.

References

[1] J. Bros, D. Iagolnitzer, Support essentiel et structure analytique des distributions, Séminaire Goulaouic–Lions–Schwartz, exp. no. 19, 1975–1976.
[2] E.J. Candès, F. Guo, New multiscale transforms, minimum total variation synthesis: applications to edge-preserving image reconstruction, Signal Process. 82 (2002) 1519–1543.
[3] E.J. Candès, L. Demanet, Curvelets and Fourier integral operators, C. R. Acad. Sci. Paris, Sér. I (2003) 359–398.
[4] E.J. Candès, D.L. Donoho, Curvelets: a surprisingly effective nonadaptive representation of objects with edges, in: A. Cohen, C. Rabut, L.L. Schumaker (Eds.), Curve and Surface Fitting: Saint-Malo 1999, Vanderbilt Univ. Press, Nashville, TN, 2000.
[5] E.J. Candès, D.L. Donoho, New tight frames of curvelets and optimal representations of objects with piecewise C^2 singularities, Comm. Pure Appl. Math. (2004) 219–266.
[6] E.J. Candès, D.L. Donoho, Continuous curvelet transform II. Discretization and frames, Manuscript, 2002.
[7] A. Cordoba, C. Fefferman, Wave packets and Fourier integral operators, Comm. Partial Differential Equations 3 (1978) 979–1005.
[8] J.M. Delort, FBI Transformation: Second Microlocalization and Semilinear Caustics, Lecture Notes in Mathematics, 1992.
[9] D.L. Donoho, Sparse components of images and optimal atomic decomposition, Constr. Approx. 17 (3) (2001) 353–382.
[10] J.J. Duistermaat, Fourier Integral Operators, Progress in Mathematics, Birkhäuser, Basel, 1996.
[11] C. Fefferman, A note on spherical summation multipliers, Israel J. Math. 15 (1973) 44–52.
[12] M. Frazier, B. Jawerth, G. Weiss, Littlewood–Paley theory and the study of function spaces, in: NSF-CBMS Regional Conf. Ser in Mathematics, vol. 79, Amer. Math. Soc., Providence, RI, 1991.
[13] P. Gérard, Moyennisation et regularité deux-microlocale, Ann. Sci. École Norm. Sup. (4) 23 (1990) 89–121.
[14] L. Hörmander, The Analysis of Linear Partial Differential Operators, Springer, New York/Berlin, 1983.
[15] Y. Meyer, Ondelettes et operateurs: I, Hermann, Paris, 1990.
[16] A. Seeger, C.D. Sogge, E.M. Stein, Regularity properties of Fourier integral operators, Ann. Math. 134 (1993) 231–251.
[17] E.P. Simoncelli, W.T. Freeman, E.H. Adelson, D.J. Heeger, Shiftable multi-scale transforms (or “What’s wrong with orthonormal wavelets”), IEEE Trans. Inform. Theory, Special Issue on Wavelets 38 (2) (March 1992) 587–607.
[18] J. Sjostrand, Singularités Analytiques Microlocales Asterisque 95, 1992.
[19] H.F. Smith, A Hardy space for Fourier integral operators, J. Geom. Anal. 8 (4) (1998) 629–653.
[20] H.F. Smith, Wave equations with low regularity coefficients, Documenta Mathematica, Extra Volume ICM 1998, II, 723–730.
[21] E.M. Stein, Harmonic Analysis: Real-Variable Methods, Orthogonality, and Oscillatory Integrals, Princeton Univ. Press, Princeton, NJ, 1993.
[22] C.D. Sogge, Fourier Integrals in Classical Analysis, Cambridge Univ. Press, Cambridge, 1993.
[23] A.B. Watson, The cortex transform: rapid computation of simulated neural images, Computer Vision Graphics Image Process. 39 (3) (September 1987) 311–327.