

Continuous Curvelet Transform: II. Discretization and Frames

Emmanuel J. Candès and David L. Donoho

Abstract

We develop a unifying perspective on several decompositions exhibiting directional parabolic scaling. In each decomposition, the individual atoms are highly anisotropic at fine scales, with effective support obeying the parabolic scaling principle $length \approx width^2$. Our comparisons allow to extend Theorems known for one decomposition to others.

We start from a Continuous Curvelet Transform $f \mapsto \Gamma_f(a, b, \theta)$ of functions $f(x_1, x_2)$ on R^2 , with parameter space indexed by scale $a > 0$, location $b \in R^2$, and orientation θ . The transform projects f onto a curvelet $\gamma_{ab\theta}$, yielding coefficient $\Gamma_f(a, b, \theta) = \langle f, \gamma_{ab\theta} \rangle$; the corresponding curvelet $\gamma_{ab\theta}$ is defined by parabolic dilation in polar frequency domain coordinates. We establish a reproducing formula and Parseval relation for the transform, showing that these curvelets provide a continuous tight frame.

The CCT is closely related to a continuous transform introduced by Hart Smith in his study of Fourier Integral Operators. Smith's transform is based on true affine parabolic scaling of a single mother wavelet, while the CCT can only be viewed as true affine parabolic scaling in euclidean coordinates by taking a slightly different mother wavelet at each scale. Smith's transform, unlike the CCT, does not provide a continuous tight frame. We show that, with the right underlying wavelet in Smith's transform, the analyzing elements of the two transforms become increasingly similar at increasingly fine scales.

We derive a discrete tight frame essentially by sampling the CCT at dyadic intervals in scale $a_j = 2^{-j}$, at equispaced intervals in direction, $\theta_{j,\ell} = 2\pi 2^{-j/2} \ell$, and equispaced sampling on a rotated anisotropic grid in space. This frame is a complexification of the 'Curvelets 2002' frame constructed by Emmanuel Candès et al. [1, 2, 3]. We compare this discrete frame with a composite system which at coarse scales is the same as this frame but at fine scales is based on sampling Smith's transform rather than the CCT. We are able to show a very close approximation of the two systems at fine scales, in a strong operator norm sense.

Smith's continuous transform was intended for use in forming molecular decompositions of Fourier Integral Operators (FIO's). Our results showing close approximation of the curvelet frame by a composite frame using true affine parabolic scaling at fine scales allow us to cross-apply Smith's results, proving that the discrete curvelet transform gives sparse representations of FIO's of order zero. This yields an alternate proof of a recent result of Candès and Demanet about the sparsity of FIO representations in discrete curvelet frames.

Key Words. Curvelets, Parabolic Scaling, Fourier Integral Operator, Tight Frame.

Acknowledgements. This work has been partially supported by DMS 00-77261, DMS 01-40698 (FRG), DMS 98-72890 (KDI), DARPA and AFOSR. DLD would like to thank the Mathematical Institute of the University of Leiden for hospitality during Fall 2002.

1 Introduction

An important role in modern harmonic analysis is played by *parabolic dilations*

$$\phi_a(x_1, x_2) = \phi_1(a^{1/2}x_1, ax_2),$$

so called because they leave invariant the parabola $x_2 = x_1^2$. It is also useful to choose the coordinate system in which the dilation is applied, resulting in **directional parabolic dilations** of the form

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

where again R_θ is rotation by θ radians. Such dilations can be used to generate vaguely ‘wavelet-like’ decompositions, where unlike classical wavelets, resulting ‘wavelets’ are highly directionally-oriented; at fine scales they are increasingly long compared to their width: width \approx length².

The motivation for decompositions based on parabolic dilations comes from several sources. Starting in the 1970’s they were used in harmonic analysis, for example by Fefferman [9] to study the L^p boundedness of Bochner-Riesz summation and later by Seeger, Sogge, and Stein [13] to study the boundedness of Fourier Integral operators. More recently, Hart Smith [14] used parabolic scaling to define function spaces preserved by Fourier Integral Operators, while Candès and Donoho [3] used parabolic scaling to define frame decompositions of image-like objects which are smooth apart from edges; see also [7]. So parabolic dilations are useful in representing operators and singularities along curves.

In this paper, we discuss four recent types of decompositions based on parabolic scaling and describe the similarities and relationships between them.

We start from a continuous curvelet transform $f \mapsto \Gamma_f(a, b, \theta)$ of functions $f(x_1, x_2)$ on R^2 , with continuous parameter space indexed by scale $a > 0$, location $b \in R^2$, and orientation θ . Associated to each parameter triple is an analyzing element $\gamma_{ab\theta}$ generating coefficients $\Gamma_f(a, b, \theta) = \langle f, \gamma_{ab\theta} \rangle$. Each element $\gamma_{ab\theta}$ is defined by what we call polar-coordinate parabolic scaling; its Fourier transform is supported on a wedge in the polar domain, and in the spatial domain is smooth and of rapid decay away from an a by \sqrt{a} rectangle with minor axis pointing in direction θ . We establish a reproducing formula and a Parseval relation for the transform, showing that these elements provide a continuous tight frame.

The CCT is first compared to a continuous curvelet-like transform used by Hart Smith in his study of Fourier Integral Operators. In our reformulation of Smith’s transform, the difference is principally that Smith’s transform is based on affine parabolic scaling of a single mother wavelet, while the CCT uses a slightly different mother wavelet at each specific scale. The impact: Smith’s transform, unlike the CCT, does not provide a continuous tight frame. We show that, with the right mother wavelet in Smith’s transform, the analyzing elements become increasingly similar to the CCT at fine scales.

The CCT is then compared to a discrete curvelet tight frame recently developed by Candès at al. [1, 3, 2]. We take the viewpoint that this new frame can be viewed as essentially sampling the CCT at dyadic intervals in scale $a_j = 2^{-j}$, at equispaced intervals in angle $\theta_{j,\ell} = 2\pi \cdot 2^{-j/2} \cdot \ell$ and on a rotated equispaced grid in space $b_{k_1, k_2}^{(j, \ell)} = R_{\theta_{j,\ell}}(2^{-j}k_1, 2^{-j/2}k_2)$. More precisely, We show that for two slightly different curvelet systems $\gamma_{ab\theta}^0$ and $\gamma_{ab\theta}^1$, an appropriate sampling $(\psi_{j,k,\ell}) = (\gamma_{a_j, b_{k_1, k_2}^{(j, \ell)}, \theta_{j, \ell}}^r)$ (setting $r = 0$ or 1 according as the scale j is even or odd) yields a tight frame. Thus the coefficients $\alpha_{j,k,\ell} = \Gamma^r(a_j, b_{k_1, k_2}^{(j, \ell)}, \theta_{j, \ell})$ obey the discrete Parseval relation $\|f\|_2^2 = \sum |\alpha_{j,k,\ell}|^2$. This is not quite an equispaced sampling of the CCT, but rather an interleaving at alternate scales of equispaced samplings of two (very slightly) different CCT’s.

As with the CCT, the curvelet frame elements are not quite parabolic dilations all of a single generating function; there is a slight variation in the generating function from one scale to the next. As in the continuous case, we are able to show that the wavelets involved in the discrete tight frame are very close to affine scalings of a single mother wavelet. In fact we consider a system of analyzing elements made by true affine parabolic scaling, and show that, if one ‘splices’ the curvelet frame at low frequencies to the true affine parabolic system at high frequencies, one gets a discrete frame which has essentially the same properties as the discrete curvelet frame.

Hart Smith’s transform was constructed to form molecular decompositions of Fourier Integral Operators (FIO’s). Smith gave a Lemma implying that if a frame were based on true parabolic scalings of a single wavelet, it would provide a sparse representation for FIO’s of order zero. Because the curvelet frame is so close to a frame based on true parabolic scaling, we are able to use Smith’s lemma to infer that curvelets give a sparse representation of FIO’s of order 0. This yields an alternate proof of a recent result of Candès and Demanet [2] about the sparsity of FIO representations in curvelet tight-frames.

Contents

Section 2 constructs a continuous curvelet transform based on a polar parabolic scaling, providing a Calderón reproducing formula, (i.e. exact reconstruction) and a Parseval relation for that transform. Section 3 discusses our reformulation of Hart Smith’s transform based on true parabolic scaling. Section 4 samples the CCT to produce a frame and explains that this is a complexified version of the discrete curvelet frame in [1, 3, 2]. Section 5 compares the complexified discrete curvelet frame with a sampling of Smith’s transform and shows that the properties are extremely close at fine scales. Section 6 shows how to use this similarity at fine scales to apply results on FIO sparsity relevant to a sampling of Smith’s transform, obtaining results about sparsity of FIOs in the discrete curvelet frame. Section 7 concludes with a discussion.

2 Transform based on Polar Parabolic Scaling

We define a CCT with a continuous scale/location/direction parameter space; compare [4]. 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 positive 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:

$$\int_0^\infty W(r)^2 \frac{dr}{r} = 1, \quad (1)$$

$$\int_{-1}^1 V(t)^2 dt = 1. \quad (2)$$

We use these windows in the frequency domain to construct a family of complex-valued waveforms 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 $\gamma_{a,0,0}$:

$$\gamma_{ab\theta}(x) = \gamma_{a,0,0}(R_\theta(x - b)),$$

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 the $\hat{\gamma}$ 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. However, note that the element $\gamma_{a,0,0}$ 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, and so ignore low frequency adjustment terms. These elements become increasingly needle-like at fine scales.

Equipped with this family of high-frequency elements, we can define a *Continuous Curvelet Transform* Γ_f , a function on scale/location/direction space:

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

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.

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:

$$f(x) = \int \Gamma(a, b, \theta) \gamma_{ab\theta}(x) \mu(da db d\theta), \quad (3)$$

and a Parseval formula for high-frequency functions:

$$\|f\|_{L^2}^2 = \int |\Gamma(a, b, \theta)|^2 \mu(da db d\theta), \quad (4)$$

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

Proof. The argument is analogous to standard arguments for the Calderón reproducing formula for the ordinary continuous wavelet transform. We rehearse only the formal aspects, ignoring convergence details, which are similar to those for the usual continuous wavelet transform [10]. Consider the contribution to the reproducing formula (3) from a single scale:

$$g_{a,\theta}(x) = \int \langle \gamma_{ab\theta}, f \rangle \gamma_{ab\theta}(x) db$$

We are to show that

$$f(x) = \int_0^{a_0} \int_0^{2\pi} g_{a\theta}(x) d\theta \frac{da}{a^3}. \quad (5)$$

Now $\gamma_{ab\theta}(x) = \gamma_{a,0,\theta}(x-b)$, so

$$\begin{aligned} g_{a,\theta}(x) &= \int \gamma_{a0\theta}(x-b) \left(\int \gamma_{a0\theta}^*(y-b) f(y) dy \right) db \\ &= \int \gamma_{a0\theta}(x-b) (\tilde{\gamma}_{a0\theta}^* \star f)(b) db \\ &= ((\gamma_{a0\theta} \star \tilde{\gamma}_{a0\theta}^*) \star f)(x), \end{aligned}$$

where $\tilde{\gamma}_{a0\theta}(x) = \gamma_{a0\theta}^*(-x)$. Now on the Fourier side,

$$(\gamma_{a0\theta} \star \tilde{\gamma}_{a0\theta}^*)^\wedge(\xi) = |\hat{\gamma}_{a0\theta}(\xi)|^2$$

Hence,

$$\hat{g}_{a,\theta}(\xi) = |\hat{\gamma}_{a0\theta}(\xi)|^2 \cdot \hat{f}(\xi), \quad \forall \xi \in \mathbf{R}^2.$$

Substituting this in (5), we obtain that the Fourier transform is given by

$$\begin{aligned} \hat{f}(\xi) &= \int \hat{g}_{a\theta}(\xi) d\theta \frac{da}{a^3} \\ &= \hat{f}(\xi) \cdot \int |\hat{\gamma}_{a0\theta}(\xi)|^2 d\theta \frac{da}{a^3}, \end{aligned}$$

and so we must verify that

$$1 = \int |\hat{\gamma}_{a0\theta}(\xi)|^2 d\theta \frac{da}{a^3} \quad \forall \xi \in \text{supp } \hat{f}. \quad (6)$$

We will see that this follows from the admissibility conditions (1)-(2). Now from the definition of $\hat{\gamma}_{a0\theta}$, if we put $e_\omega = (\cos(\omega), \sin(\omega))$,

$$\hat{\gamma}_{a0\theta}(r \cdot e_\omega) = W(a \cdot r) \cdot V((\omega - \theta)/\sqrt{a}) \cdot a^{3/4}.$$

So rewrite

$$\int_0^{a_0} \int_0^{2\pi} |\hat{\gamma}_{a_0\theta}(\xi)|^2 d\theta \frac{da}{a^3} = \int_0^{a_0} \int_0^{2\pi} W(a \cdot r)^2 V((\omega - \theta)/\sqrt{a})^2 a^{3/2} d\theta \frac{da}{a^3};$$

using admissibility of V , (2), we have $\int_0^{2\pi} V((\omega - \theta)/\sqrt{a})^2 d\omega = a^{1/2}$, so (6) reduces to

$$1 = \int_0^{a_0} W(a \cdot r)^2 \frac{da}{a} \quad \forall r = |\xi| \text{ with } \xi \in \text{supp } \hat{f};$$

Now, for $r = |\xi|$ with $\xi \in \text{supp } \hat{f}$, we have $r > 2/a_0$, so a simple rescaling of variables and admissibility of W , (1), gives

$$\int_0^{a_0} W(a \cdot r)^2 \frac{da}{a} = \int_0^{a_0 r} W(a)^2 \frac{da}{a} = \int_{1/2}^2 W(a)^2 \frac{da}{a} = 1.$$

This gives (6) and completes the formal aspects of the proof of (3).

We now consider the proof of (4).

$$\int_{\{a < a_0\}} |\langle \gamma_{ab\theta}, f \rangle|^2 \mu(da db d\theta) = \int_{\{a < a_0\}} |(\tilde{\gamma}_{a_0\theta}^* \star f)(b)|^2 db d\theta \frac{da}{a^3}.$$

The Plancherel formula now gives

$$\begin{aligned} \int_{\{a < a_0\}} |\langle \gamma_{ab\theta}, f \rangle|^2 \mu(da db d\theta) &= \frac{1}{(2\pi)^2} \int |\hat{f}(\xi)|^2 |\hat{\gamma}_{a_0\theta}(\xi)|^2 d\xi d\theta \frac{da}{a^3} \\ &= \frac{1}{(2\pi)^2} \int |\hat{f}(\xi)|^2 \left(\int |\hat{\gamma}_{a_0\theta}(\xi)|^2 d\theta \frac{da}{a^3} \right) d\xi \\ &= \frac{1}{(2\pi)^2} \int |\hat{f}(\xi)|^2 d\xi = \|f\|_{L^2}^2, \end{aligned}$$

where we used (6). □

Remark. The reference measure is important for what follows. We prefer to think of it as

$$d\mu = \frac{db}{a^{3/2}} \frac{d\theta}{a^{1/2}} \frac{da}{a}$$

suggesting that the range of b be viewed as divided into unit cells of side a by \sqrt{a} (and so area $a^{3/2}$), the range of θ is naturally viewed as divided into unit cells (intervals) of side \sqrt{a} , and the range of $\log(a)$ has unit cells of side 1. This point of view will be very important in understanding the sparsity and discretization of the transform.

We can extend this transform to low frequencies as follows. Let f be an L^2 function, and let

$$P_1(f) = \int_{a < a_0} \Gamma(a, b, \theta) \gamma_{ab\theta}(x) \mu(da db d\theta).$$

In the frequency domain, we have

$$\widehat{P_1(f)}(\xi) = \hat{f}(\xi) \cdot \left(\int_0^{a_0} W(a \cdot |\xi|)^2 \frac{da}{a} \right) = \hat{f}(\xi) \cdot \hat{\Psi}(\xi)^2,$$

say. (The definition as a square makes sense because the integrand in parentheses is real and nonnegative, $\hat{\Psi}(\xi)^2 = \int_0^{a_0/|\xi|} |W(a)|^2 \frac{da}{a}$.) Set $\hat{\Phi}(\xi)^2 = 1 - \hat{\Psi}(\xi)^2$. Then

$$P_1(f)(x) = \Psi \star f, \quad P_0(f) = f - P_1(f) = \Phi \star f.$$

From the argument in the proof of Theorem 1, we can see that

$$\hat{\Phi}(\xi) = 0, |\xi| > 2/a_0, \quad \hat{\Phi}(\xi) = 1, |\xi| < 1/(2a_0).$$

and $0 \leq \hat{\Phi}(\xi), \hat{\Psi}(\xi) \leq 1$, while by construction

$$\hat{\Phi}(\xi)^2 + \hat{\Psi}(\xi)^2 = 1.$$

Define now the ‘father wavelet’ $\Phi_{a_0,b}(x) = \Phi(x - b)$, and note that

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

Hence we have

$$f(x) = P_0(f)(x) + P_1(f)(x),$$

valid in an L^2 sense for all functions in L^2 . Moreover,

$$\int |\langle \Phi_{a_0,b}, f \rangle|^2 db = \|\Phi \star f\|_2^2,$$

and

$$\int_{\{a < a_0\}} |\langle \gamma_{ab\theta}, f \rangle|^2 \mu(dadb d\theta) = \|\Psi \star f\|_2^2.$$

This gives the formal part of the proof for:

Theorem 2 *Let $f \in L^2(\mathbf{R}^2)$. Then*

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

and

$$\|f\|_2^2 = \int |\langle \Phi_{a_0,b}, f \rangle|^2 db + \int_0^{a_0} \int \int |\langle \gamma_{ab\theta}, f \rangle|^2 \mu(da db d\theta)$$

We can think of the ‘full CCT’ as consisting of curvelets at fine scales and isotropic father wavelets at coarse scales. For our purposes, it is only the behavior of the fine-scale elements that matters.

3 Transform based on Affine Parabolic Scaling

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

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

this metric has ellipsoidal contours with minor axis pointing in direction θ .

Suppose now that we take a single ‘mother wavelet’ φ and define an affine system

$$\varphi_{ab\theta} = \varphi(P_{a,\theta}(x - b)) \cdot \text{Det}(P_{a,\theta})^{1/2}. \quad (7)$$

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 analysing waveforms. So this transform fits in with the classical notion of wavelet family, except that the family of parabolic affine transforms is nonstandard.

Hart Smith in [14] studied essentially this construction, with two inessential differences: first, instead of working with scale a and direction θ , he worked with the frequency variable $\xi \equiv a^{-1}e_\theta$,

and second, instead of using the L^2 normalizing factor $Det(P_{a,\theta})^{1/2}$, he used the L^1 normalizing factor $Det(P_{a,\theta})$. In any event, we pretend that Smith had used the scale/location/direction parametrization and the L^2 normalization as in (7) and call

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

Hart Smith's *directional wavelet transform based on affine parabolic scaling*.

While affine parabolic scaling is conceptually a bit simpler than the scaling we have mostly studied here, it does complicate life a bit. Here is the result paralleling to Theorem 1.

Theorem 3 (*Translation of Smith, 1998 into new parametrization/normalization*). *There is a Fourier multiplier M of order 0 so that whenever f is a high-frequency function supported in frequency space $|\xi| > 2/a_0$,*

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

and

$$\|f\|_2^2 = \int |\langle \varphi_{ab\theta}, M^{1/2}f \rangle|^2 d\mu.$$

Here $d\mu = a^{-3}dbd\theta da$ and Mf is defined in the frequency domain by a multiplier formula $m(|\xi|)f(\xi)$, where the multiplier m is a symbol of order 0.

Here the multiplier $m(r)$ is a smooth function tending to a constant at infinity and with decaying derivatives; for terminology on multipliers of order 0, see [16, 8].

In short, one has to work not with the coefficients of f but with those of Mf . An alternate approach, not discussed by Smith, defines dual elements $\varphi_{ab\theta}^\# \equiv M\varphi_{ab\theta}$ and changes the transform definition to either

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

or

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

This more complicated set of formulas leads to a few annoyances which are avoided using the CCT we defined in the previous section. We will see that there are other advantages to the definition of the CCT when it comes to discretizing the transform, which are discussed elsewhere.

However, for many purposes, the two transforms have similar behavior. For an elementary example, we have :

Lemma 3.1 *Suppose that the windows V and W underlying the CCT are C^∞ , and 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 , where c is some normalizing constant, and \bar{a}_0 is the transform's coarsest scale. Then at fine scales we have the equivalence

$$\sup_{b,\theta} \|\gamma_{ab\theta} - \varphi_{ab\theta}\|_2 \rightarrow 0, \quad a \rightarrow 0.$$

Much finer notions of equivalence could be developed here; some of these will be explored in far more detail in the discrete setting in Section 5 below. The proof of Lemma 3.1 will follow completely as in Section 5's discussion of the discrete case, so the proof is omitted.

4 Discretization by Sampling

Obviously $\Gamma_f(a, b, \theta)$ is not an arbitrary continuous function of a , b , and θ . It is best thought of as broken into a collection of coherent regions, each covering a ‘unit cell’ in scale/space/orientation – where a cell has μ -measure about 1. Indeed, the transform is very smooth, and over small neighborhoods of (a, b, θ) space having μ -measure much smaller than 1 it cannot vary by much.

Consider then discretizing the CCT according to tiles $Q = Q(j, k_1, k_2, \ell)$ which obey the following desiderata:

- In tile $Q(j, k_1, k_2, \ell)$, scale a runs through a dyadic interval $2^{-j} > a \geq 2^{-(j+1)}$.
- At scale 2^{-j} , locations run through rectangularly shaped regions with aspect ratio roughly 2^{-j} by $2^{-j/2}$.
- The tile contains orientations running through $2\pi\ell/2^{j/2} \leq \theta < 2\pi(\ell + 1)/2^{j/2}$.
- The location regions are rotated consistent with the orientation $b \approx R_{\theta_\ell}(k_1/2^j, k_2/2^{j/2})$, $\theta_\ell = 2\pi\ell/2^{j/2}$.
- The tiles pack together neatly to cover the full scale/location/direction space with minimal overlap.

Note again that for such tiles $\mu(Q) \approx 1$. Over such tiles different values of $\Gamma_f(a, b, \theta)$ are roughly comparable and different curvelets $\gamma_{ab\theta}$ as well. Hence it is sensible to decompose the reproducing formula into a discrete sum of subrepresentations based on coherent regions:

$$\begin{aligned}
 f(x) &= \int \Gamma(a, b, \theta) \gamma_{ab\theta}(x) d\mu \\
 &= \sum_Q \int_Q \Gamma(a, b, \theta) \gamma_{ab\theta}(x) d\mu \\
 &= \sum_Q m_Q(x), \quad m_Q(x) = \int_Q \Gamma(a, b, \theta) \gamma_{ab\theta}(x) d\mu \\
 &= \sum_Q A_Q M_Q(x), \quad A_Q = \|\Gamma(a, b, \theta)\|_{L^2(Q)}, \tag{8}
 \end{aligned}$$

where the M_Q are L^2 normalized ‘directional molecules’ and the A_Q are amplitudes. It can be shown that each M_Q is a smooth function, has anisotropic effective support obeying parabolic scaling, and so on. It can also be shown that the coefficient amplitudes measure various norms; thus $\sum_Q A_Q^2 \asymp \|f\|_2^2$, etc. Molecular decompositions of this kind have a long history in wavelet analysis [10]; it may be expected that this type of decomposition in the curvelet setting would have many equally important applications.

Unfortunately, such molecular decompositions have the drawback that they are nonlinear in f . There are many potential advantages of a discrete decomposition which is linear and has fixed elements. We now construct such a transform, roughly the idea is to *sample* the continuous transform at a range of scales a_j , orientations $\theta_{j,\ell}$ and locations $b_{k_1, k_2}^{j, \ell}$, according to

- $a_j = 2^{-j}$, $j \geq 0$.
- $\theta_{j,\ell} = \pi/2 \cdot \ell \cdot 2^{-\lceil j/2 \rceil}$, $0 \leq \ell < L_j \equiv 4 \cdot 2^{\lceil j/2 \rceil}$.
- The locations $b_{k_1, k_2}^{j, \ell}$ run through a j, ℓ dependent grid defined by

$$b_{k_1, k_2}^{j, \ell} = R_{\theta_{j,\ell}}(k_1/2^j, k_2/2^{j/2}) \tag{9}$$

where R_θ denotes planar rotation by θ radians, and k_1, k_2 run over \mathbf{Z}^2 .

The construction goes in two stages, first building a semi-discrete transform where the spatial variable b is continuous but the other variables j, ℓ are discrete; and then discretizing the space variable.

4.1 Semi-Discrete Transform

Pick now windows $W(r)$ and $V(t)$ similar to the windows of the continuous transform – both are real, nonnegative, C^∞ , supported in $(1/2, 2)$ and in $(-1, 1)$ respectively. They should obey discrete admissibility conditions analogous to the continuous ones used above:

$$\sum_{j=-\infty}^{\infty} W^2(2^j r) = 1, \quad r \in (3/4, 3/2); \quad (10)$$

$$\sum_{\ell=-\infty}^{\infty} V^2(t - \ell) = 1, \quad t \in (-1/2, 1/2). \quad (11)$$

These conditions are basically compatible with the admissibility conditions (1)-(2) for the continuous transform. More precisely, if we have a window V satisfying the above condition (11), then it automatically satisfies the continuous admissibility condition (2), while if we have a window W satisfying (10), then it automatically also satisfies the continuous admissibility condition (1) up to a constant of proportionality:

$$\int W^2(ar) \frac{da}{a} = \log(2).$$

We also define $\tilde{a}_j^{1/2} = \frac{1}{2} \cdot 2^{-\lceil j/2 \rceil}$.

We are going to construct a family $\phi_{j,b,\ell}(x)$ analogous to our earlier construction of curvelets. At scale a_j , the family is generated by translation and rotation of a basic element $\phi_{j,0,0}$:

$$\phi_{j,b,\ell}(x) = \phi_{j,0,0}(R_{\theta_{j,\ell}}(x - b)).$$

The generating element at scale a_j is defined by going to polar Fourier coordinates (r, ω) and setting

$$\hat{\phi}_{a_j,0,0}(r, \omega) = W(a_j \cdot r) \cdot V\left(\frac{\omega}{\pi \tilde{a}_j^{1/2}}\right) \cdot a_j^{3/4}, \quad j = 0, 1, \dots$$

This is very similar to the definition of the analyzing elements of the CCT. Again, these elements are not quite affine parabolic scaling of a single wavelet. Note also that the angular width of $\hat{\phi}$'s support in frequency space is $\pi \cdot 2^{-\lceil j/2 \rceil}$. We can then define the semidiscrete transform via

$$\tilde{\Gamma}(j, b, \ell) = \langle \phi_{j,b,\ell}, f \rangle.$$

This transform has an exact reconstruction formula and a Parseval relation.

Theorem 4.1 *Let f be a high frequency function with $\hat{f}(\xi)$ vanishing for $|\xi| < 2/a_0$. Then*

$$f = \sum_j \sum_\ell \int \tilde{\Gamma}(j, b, \ell) \phi_{j,b,\ell} db / a_j^{3/2}, \quad (12)$$

and

$$\|f\|_2^2 = \sum_j \sum_\ell \|\tilde{\Gamma}(j, \cdot, \ell)\|_{L_2(\mathbf{R}^2)}^2 a_j^{-3/2}. \quad (13)$$

We again just give the formal elements of the proof. Define

$$g_{j,\ell}(x) = \int \langle \phi_{j,b,\ell}, f \rangle \phi_{j,b,\ell}(x) db / a_j^{3/2}.$$

Then, as in Theorem 1,

$$\begin{aligned}\hat{g}_{j,\ell}(\xi) &= |\hat{\phi}_{j,b,\ell}(\xi)|^2 \cdot \hat{f}(\xi), \quad \forall \xi \in \mathbf{R}^2 \\ &= \hat{f}(\xi) \cdot W^2(a_j|\xi|) V^2\left(\frac{\omega - \theta_{j,\ell}}{\pi \tilde{a}_j^{1/2}}\right).\end{aligned}$$

Now

$$\sum_{\ell=0}^{L_j-1} V^2\left(\frac{\omega - \theta_{j,\ell}}{\pi \tilde{a}_j^{1/2}}\right) = \sum_{-L_j/2}^{L_j/2-1} V^2(t - \ell) = 1,$$

where t is proportional to the distance from ω to the nearest among the $\theta_{j,\ell}$. Hence,

$$\sum_{\ell=0}^{L_j-1} \hat{g}_{j,\ell}(\xi) = \hat{f}(\xi) W^2(a_j|\xi|).$$

We have assumed that $\xi \geq 2/a_0$. Hence from the fact that W is supported in $(1/2, 2)$, we have

$$\sum_{j \geq 0} W^2(a_j|\xi|) = \sum_{j \geq 0} W^2(2^{-j}|\xi|) = \sum_{j=-\infty}^{\infty} W^2(2^{-j}|\xi|) = 1.$$

We conclude that

$$\hat{f}(\xi) = \sum_j \sum_{\ell} \hat{g}_{j,\ell}(\xi),$$

and the result (12) follows.

We now consider (13). Arguing as in Theorem 1,

$$\begin{aligned}\sum_j \sum_{\ell} \int |\tilde{\Gamma}(j, b, \ell)|^2 db / a_j^{3/2} &= \sum_{j,\ell} \|g_{j,\ell}\|_2^2 \\ &= (2\pi)^{-2} \int |\hat{f}(\xi)|^2 \left(\sum_{j,\ell} |W(a_j|\xi|)|^2 |V(\frac{\omega - \theta_{j,\ell}}{\pi \tilde{a}_j^{1/2}})|^2 \right) d\xi \\ &= (2\pi)^{-2} \int |\hat{f}(\xi)|^2 d\xi = \|f\|_2^2.\end{aligned}$$

4.2 Tight Frame

We now go to the final step and obtain a full tight frame, by sampling the semidiscrete decomposition. We define frame coefficients

$$\alpha_{j,(k_1,k_2),\ell} = \tilde{\Gamma}(j, b_{k_1,k_2}^{j,\ell}, \ell),$$

for $j \geq 0$, $(k_1, k_2) \in \mathbf{Z}^2$, and $0 \leq \ell < L_j$; here the $b_{k_1,k_2}^{j,\ell}$ are as in (9). Abusing notation, the corresponding frame elements are

$$\phi_{j,k,\ell} \equiv \phi_{j,b_{k_1,k_2}^{j,\ell},\ell}.$$

We note that the spatial sampling uses a different rectangular grid for each different orientation and that it has a different spacing in each of the two orthogonal directions, consistent with parabolic scaling.

Theorem 4.2 *Let f be a highpass L^2 function. Then we have*

$$f = \sum_{j,k,\ell} \alpha_{j,k,\ell} \phi_{j,k,\ell},$$

in $L^2(\mathbf{R}^2)$, and

$$\|f\|_2^2 = \sum_{j,k,\ell} |\alpha_{j,k,\ell}|^2.$$

The proof is merely to verify that

$$g_{j,\ell} = \sum_k \alpha_{j,k,\ell} \phi_{j,k,\ell}$$

and

$$\|g_{j,\ell}\|_2^2 = \sum_k |\alpha_{j,k,\ell}|^2.$$

We verify this for $\ell = 0$ only, as the other cases follow by rotation of coordinates. We make two remarks, formalized as Lemmas. Together, these lemmas imply the case $\ell = 0$ and therefore yield the full proof.

Lemma 4.1 *For $j \geq 1$, $\hat{\phi}_{j,0,0}(\xi)$ has support bounded inside a rectangle which if translated to the origin, would fit inside the rectangle*

$$[-\pi 2^j, \pi 2^j] \times [-\pi 2^{j/2}, \pi 2^{j/2}].$$

This is proven by simple inspection of the region $\{|\xi| : 2^{j-1} \leq |\xi| \leq 2^{j+1}, |\omega| < \pi \tilde{a}_j^{1/2}\}$.

Lemma 4.2 *Suppose that $\Phi \in L^2(\mathbf{R}^2)$ is a bandlimited function with*

$$\text{supp}(\hat{\Phi}) \subset [-\pi A, \pi A] \times [-\pi B, \pi B].$$

Suppose that $g \in L^2(\mathbf{R}^2)$ is defined in the frequency domain by

$$\hat{g}(\xi) = |\hat{\Phi}(\xi)|^2 \hat{f}(\xi).$$

Then set $\Phi_{k_1,k_2}(x) = \Phi(x_1 - k_1/A, x_1 - k_2/B)$. We have

$$g(x) = \sum_k \langle \Phi_{k_1,k_2}, f \rangle \Phi_{k_1,k_2}(x).$$

and

$$\|g\|_2^2 = \sum_k |\langle \Phi_{k_1,k_2}, f \rangle|^2.$$

This Lemma is well-known and frequently used throughout wavelet theory and filterbank theory; compare [6, 12, 10].

4.3 Interpretation

The frame that we have just constructed is almost identical to the discrete curvelets frame proposed by Candès and coauthors [1, 2, 3]; that frame is set up so that the frame elements are real valued; it can be produced from this one by averaging together terms at $\theta_{j,\ell}$ with those at $-\theta_{j,\ell}$. Thus we see that there is an intimate connection between the discrete curvelets frame and the CCT. For simplicity, we will also call the frame constructed here a curvelet frame.

We have written above that the frame is produced by equispaced sampling of the CCT, but that is not strictly correct. While in almost all respects the $\tilde{\Gamma}$ and Γ are the same, there are two important discrepancies: first, as mentioned earlier, the W window in the continuous case obeys a slightly different normalization than the W window in the discrete case, so we could at best expect $\Gamma \propto \tilde{\Gamma}$; but more seriously, $\tilde{\Gamma}$ dilates in the polar angular variable using $\tilde{a}_j^{1/2}$ rather than $a_j^{1/2}$. Now these two quantities are identical for even j , but not for odd j . One way to look at this is as follows. It is as if we have two different continuous transforms $\Gamma^1(a, b, \theta)$ and $\Gamma^2(a, b, \theta)$, with $\Gamma = \Gamma^1$ as we have discussed so far, but with Γ^2 based on an angular window $V^2(\cdot) = V(\sqrt{2}\cdot)$. Thus Γ^2 uses a slightly different generating curvelet at each scale than Γ^1 . Then we have for appropriate constants c_i ,

$$\tilde{\Gamma}(j, k, \ell) = \begin{cases} c_1 \cdot \Gamma^1(a_j, b_{k_1, k_2}^{j, \ell}, \theta_{j, \ell}) & j \text{ even} \\ c_2 \cdot \Gamma^2(a_j, b_{k_1, k_2}^{j, \ell}, \theta_{j, \ell}) & j \text{ odd} \end{cases}.$$

In essence, the curvelet frame is an interleaving of sampling from two different frames at alternate scales. For an alternate approach, see the discussion section below.

5 Comparison of Frames

The discrete curvelet frame is not the result of affine changes of variables to a single generating element. In this section we consider a sense in which ‘at fine scales’ the curvelet frame is very close to such a system.

Consider then a curvelet in standard position and orientation: $k = 0$ and $\ell = 0$. In the Fourier domain, it is given by

$$\hat{\phi}_{j,0,0} = W(a_j r) V(\omega / \pi \tilde{a}_j^{1/2}) a_j^{3/4}. \quad (14)$$

Now define an affiliated wavelet based on true parabolic scaling i.e. *not using polar variables*:

$$\hat{\varphi}_{j,0,0}(\xi) = W(a_j \xi_1) V\left(\frac{\xi_2}{\xi_1} \frac{1}{\pi \tilde{a}_j^{1/2}}\right) a_j^{3/4}. \quad (15)$$

Note that, by construction, the $\hat{\varphi}_{j,0,0}(\xi)$ are all true affine images of a single generator:

$$\hat{\varphi}_{j',0,0}(\xi) = \hat{\varphi}_{j,0,0}(d_{j',j} \xi_1, e_{j',j} \xi_2) f_{j',j}$$

where

$$d_{j',j} = \frac{a_{j'}}{a_j}; \quad e_{j',j} = \frac{a_{j'} / \tilde{a}_{j'}^{1/2}}{a_j / \tilde{a}_j^{1/2}} \quad f_{j',j} = \frac{a_{j'}^{3/4}}{a_j^{3/4}}.$$

Such a relationship would not be true in the curvelet family, where the generator is (slightly) different at each different scale. Now visual comparison of (14)-(15) suggests that, each pair of corresponding elements in the two families are close. Indeed, the arguments to the corresponding V and W are almost the same. Letting Ξ_j denote the support of $\hat{\phi}_{j,0,0}$ we have,

$$\left\| \frac{r}{\xi_1} - 1 \right\|_{L^\infty(\Xi_j)} \rightarrow 0, \quad j \rightarrow \infty.$$

$$\left\| \frac{\xi_2 / \xi_1}{\omega} - 1 \right\|_{L^\infty(\Xi_j)} \rightarrow 0, \quad j \rightarrow \infty.$$

By smoothness of V and W we immediately see that the two families match up, element for element, at very fine scales, proving

Lemma 5.1

$$0 = \lim_{J \rightarrow \infty} \sup_{j \geq J} \sup_{k, \ell} \|\phi_{j,k,\ell} - \varphi_{j,k,\ell}\|_2;$$

In fact, a much stronger matching-up of the two systems occurs. Consider then the curvelet frame $\Phi = \{\phi_{j,k,\ell}\}$ and the *composite system* Φ_J defined so that at coarse scales it uses elements from the curvelet frame, and at fine scales it uses elements obeying true affine parabolic scaling:

$$\Phi_J = \{\phi_{j,k,\ell} : j \leq J\} \cup \{\varphi_{j,k,\ell} : j > J\}.$$

It turns out that, for large J , the two systems are nearly equivalent.

Theorem 5.1 *For all sufficiently large J , Φ_J is a frame. In fact the frame bounds tend to 1 as $J \rightarrow \infty$. Let f be an L^2 function, and let α be the coefficient sequence generated by the curvelet frame Φ and let $\alpha^{(J)}$ be the coefficient sequence generated by the composite frame Φ_J . Then*

$$0 = \lim_{J \rightarrow \infty} \sup_{f \neq 0} \frac{\|\alpha(f) - \alpha^{(J)}(f)\|_2}{\|f\|_2}.$$

Dually, let α be a coefficient sequence and let $f(\alpha)$ be the synthesis of f using Φ and $f_J(\alpha)$ using Φ_J . Then

$$0 = \lim_{J \rightarrow \infty} \sup_{\alpha \neq 0} \frac{\|f(\alpha) - f_J(\alpha)\|_2}{\|\alpha\|_2}.$$

Finally, suppose that f is a function with sparse curvelet transform: for $0 < p \leq 1$ we have $\|\alpha(f)\|_p < \infty$. Then f also has a sparse Φ_J frame transform $\|\alpha^{(J)}(f)\|_p < \infty$ - and vice versa. In fact, with constants that depend on p only,

$$\|\alpha(f)\|_p \asymp \|\alpha^{(J)}(f)\|_p.$$

This result justifies the effective equivalence of our notion of parabolic scaling to traditional parabolic scaling. Either system gives sparse coefficients if and only if the others sequence does. It will play a key role in deducing the FIO representation theorem in the next section. We carry out the proof over the next two subsections.

5.1 Gram Matrices of the Two Frames

For notational simplicity let $Q = (j, k, \ell)$ denote a scale/location/orientation triple. Consider the tight frame Gram matrix

$$M^\#(Q, Q') = \langle \phi_Q, \phi_{Q'} \rangle$$

and the cross-frame matrix

$$M_J(Q, Q') = \langle \phi_Q, \varphi_{Q'} \rangle.$$

The cross-frame matrix relates coefficients α in the Φ frame to coefficients $\alpha^{(J)}$ in the Φ_J system. Thus, if $f = \Phi_J(\alpha^{(J)})$ then $f = \Phi(\alpha)$ where $\alpha = M_J \alpha^{(J)}$. We also observe that because Φ is a tight frame, $M^\#$ is Hermitian and idempotent

$$M^\# = (M^\#)^H; \quad (M^\#)^2 = M^\#. \tag{16}$$

Here M^H means the Hermitian transpose of M .

For such matrices, and for $0 < p \leq 1$ define the p -norm by the maximum ℓ^p norm of any row or column:

$$\|M\|_p = \max \left(\left(\sup_Q \sum_{Q'} |M(Q, Q')|^p \right)^{1/p}, \left(\sup_{Q'} \sum_Q |M(Q, Q')|^p \right)^{1/p} \right)$$

(Of course, for $0 < p < 1$ this is not actually a norm but instead a quasi-norm; it does not obey the triangle inequality but instead the p -triangle inequality.) For this norm and $p \leq 1$, we observe that if N and M are matrices

$$\|M \cdot N\|_p \leq \|M\|_p \cdot \|N\|_p,$$

as can be seen by systematic application of the p -triangle inequality for vectors $\sum_i |u_i + v_i|^p \leq \sum_i |u_i|^p + \sum_i |v_i|^p$. A further useful observation is that the usual matrix norm

$$\|M\| \equiv \sup_{x \neq 0} \frac{\|Mx\|}{\|x\|}$$

is controlled by the p -norm:

$$\|M\| \leq \|M\|_p.$$

Thus convergence in p -norm implies usual norm convergence of matrices.

Using these facts and sharpening the analysis beyond the pointwise convergence, we get the following showing strong convergence:

Lemma 5.2 *For all $J > J_0$, $\|M_J\|_p < \infty$. Moreover,*

$$\|M^\# - M_J\|_p \rightarrow 0, \quad J \rightarrow \infty.$$

For sufficiently large J , the matrix M_J^\dagger can be defined by the convergent infinite series

$$M_J^\dagger = M_J \sum_{t \geq 1} (M^\# - M_J M_J^H)^t;$$

this obeys $\|M_J^\dagger\|_p < \infty$, and is a generalized right-hand inverse

$$M_J M_J^\dagger = M^\#. \tag{17}$$

Finally,

$$\|M^\# - M_J^\dagger\|_p \rightarrow 0, \quad J \rightarrow \infty.$$

This of course implies everything stated in Theorem 5.1, and much more. The proof depends on an analysis, given in the next section, where we use smoothness of V and W to refine the comparison of $\hat{\phi}_{j,0,0}$ and of $\varphi_{j,0,0}$ as in Lemma 3.1 so as to yield the following:

Lemma 5.3 *For $p \in (0, 1]$ there is $C_p < \infty$ so*

$$\max_Q \sum_{Q'} |\langle \phi_Q, \phi_{Q'} \rangle|^p < C_p. \tag{18}$$

For $p \in (0, 1]$ there is a sequence $(\epsilon_{J,p})$ with $\epsilon_{J,p} \rightarrow 0$ as $J \rightarrow \infty$ so that letting φ_Q denote the Q -th element of the frame Φ_J ,

$$\max_{Q'} \sum_Q |\langle \phi_Q, \phi_{Q'} - \varphi_{Q'} \rangle|^p < \epsilon_{J,p} \tag{19}$$

Also, for the same $p \in (0, 1]$ and the same sequence $(\epsilon_{J,p})$

$$\max_Q \sum_{Q'} |\langle \phi_Q, \phi_{Q'} - \varphi_{Q'} \rangle|^p < \epsilon_{J,p} \tag{20}$$

Together with Hermitian symmetry of $M^\#$, (18) immediately implies the sparsity of the frame Gramian:

$$\|M^\#\|_p^p < C_p < \infty,$$

Using the p -triangle inequality $\|M_J\|_p^p \leq \|M^\#\|_p^p + \|M^\# - M_J\|_p^p$ we get the uniform boundedness of M_J :

$$\|M_J\|_p^p \leq C_p + \epsilon_{J,p}.$$

Hence, $\|M_J\|_{\ell_p} < \infty$ and for large J , $\|M_J\| < \infty$ as well. It follows that, if $\alpha = M_J\alpha^{(J)}$,

$$\|\alpha\|_2 = \|M_J\alpha^{(J)}\| \leq \|M_J\| \cdot \|\alpha^{(J)}\|_2,$$

which is half what we need to show that Φ_J is a frame. The Lemma also implies the convergence

$$\|M_J - M^\#\|_p \rightarrow 0, \quad J \rightarrow \infty,$$

which gives

$$\|M_J\| \rightarrow \|M^\#\| = 1;$$

hence the upper frame bound constant for Φ_J tends to 1 with increasing J .

We now turn to parts of the Theorem concerning lower frame bounds; this means we must study the existence and properties of the pseudo-inverse M^\dagger .

Suppose now that we have a coefficient sequence $\alpha^{(J)}$ which can synthesize f using the Φ_J frame: $f = \Phi_J(\alpha^{(J)}) = \sum_Q \alpha_Q^{(J)} \varphi_Q$. We wish to convert this to a coefficient sequence α that synthesizes the same function, only using the Φ frame: $f = \sum \alpha_Q \phi_Q$ where $\alpha_Q = \langle f, \phi_Q \rangle$. This job is accomplished by $\alpha = M_J\alpha^{(J)}$. The purpose of the matrix M_J^\dagger is to go in the other direction. Given coefficients α that synthesize f through the frame Φ , find coefficients that synthesize that same function through the frame Φ_J . This means to ‘invert’ M_J , i.e. to solve $M_J\alpha^{(J)} = \alpha$, which we will write as $\alpha^{(J)} = M_J^\dagger\alpha$. Now we are interested in solving this equation only when we are given α satisfying $\alpha = M^\#\alpha$. Our solution should therefore satisfy

$$MM^\dagger\alpha = M^\#\alpha.$$

The result M^\dagger therefore obeys (17) above.

To obtain such a matrix, we consider an iterative scheme based on simple relaxation for solving for $\alpha^{(J)}$ given α . Set for short $M = M_J$ and set $\delta^{(0)} = \alpha$, where α is in $\text{range}(M^\#)$. Then put

$$A^{(1)} = M^H\delta^{(0)}, \quad \delta^{(1)} = \delta^{(0)} - MA^{(1)}.$$

In essence, we are using M^H to ‘guess’ an element $A^{(1)}$ which may be close to $\alpha^{(J)}$. Then we compute the implied approximation to $\alpha^{(J)}$ which such a ‘guess’ would generate, and get the approximation error $\delta^{(1)}$. Now as $\delta^{(0)} \in \text{range}(M^\#)$, we have $\delta^{(1)} \in \text{range}(M^\#)$ as well. We can continue this iteration, getting $A^{(2)}$, $\delta^{(2)}$, etc., where for clarity we spell out

$$A^{(2)} = M^H\delta^{(1)}, \quad \delta^{(2)} = \delta^{(1)} - MA^{(2)};$$

and later terms in the iteration are defined analogously. Note that with sufficient control on $\|M - M^\#\|_p$, we can show that this iteration converges geometrically. We can formalize this:

Lemma 5.4 *Fix $0 < p \leq 1$, and suppose $\|M - M^\#\|_p < (1/4)^{1/p}/\|M^\#\|_p$. Then*

$$\|\delta^{(k)}\|_2 \leq (3/4)^k \|\delta^{(0)}\|_2,$$

and so both series $\sum_{k \geq 0} \delta^{(k)}$ and $\sum_{k \geq 0} A^{(k)}$ are absolutely summable.

Proof. Indeed, each $\delta^{(k)}$ is in the range of $M^\#$, and

$$\delta^{(k+1)} = (I - MM^H)\delta^{(k)} = (M^\# - MM^H)\delta^{(k)}.$$

Now, writing $\Delta = M - M^\#$

$$M^\# - MM^H = (M^\#)^2 - (M^\# + \Delta)(M^\# + \Delta)^H = \Delta(M^\#)^H + M^\#\Delta^H + \Delta\Delta^H.$$

Hence, from $\|\Delta\|_p = \|M - M^\#\|_p < (1/4)^{1/p}/\|M^\#\|_p$, and $\|M^\#\|_p \geq 1$

$$\|M^\# - MM^H\|_p^p \leq 2\|M^\#\|_p^p \|\Delta\|_p^p + \|\Delta\|_p^{2p} < 1/2 + 1/16 = 3/4.$$

Hence,

$$\|\delta^{(k+1)}\|_2 \leq \|M^\# - MM^H\|_p \|\delta^{(k)}\|_2 \leq \|M^\# - MM^H\|_p \|\delta^{(k)}\|_2 \leq (3/4)^{1/p} \|\delta^{(k)}\|_2.$$

Given this geometric decay, and the boundedness $\|M^H\| \leq \|M\|_p < \infty$ we conclude that $(A^{(k)})_k = (M^H \delta^{(k)})_k$ is a summable sequence. \square

This Lemma justifies the definition $\alpha^{(J)} = A^{(1)} + A^{(2)} + \dots$. It also justifies the formal calculation

$$M\alpha^{(J)} = MA^{(1)} + MA^{(2)} + \dots = (\delta^{(0)} - \delta^{(1)}) + (\delta^{(1)} - \delta^{(2)}) + \dots = \delta^{(0)}$$

to conclude that

$$M\alpha^{(J)} = \alpha;$$

in short, the iterative scheme rigorously solves the problem of ‘inverting’ M .

Now in effect the iterative scheme is equivalent to applying the matrix M^\dagger , where

$$M^\dagger = M \sum_{t \geq 1} (M^\# - MM^H)^t;$$

By Lemma 5.4, this is well defined as soon as $\|M_J - M^\#\|_p < (1/4)^{1/p}/\|M^\#\|_p$ (which will eventually be satisfied for J large enough); the sum on the right hand side converges, because (as in Lemma 5.4) this implies $\|M^\# - MM^H\|_p \leq 3/4$, and so defines a matrix with

$$\|M^\dagger\|_p \leq \|M\|_p \left(\sum_{t \geq 1} \|(M^\# - MM^H)^t\|_p^p \right)^{1/p} \leq 3 \cdot \|M\|_p;$$

and hence $\|M^\dagger\| < \infty$. This gives the lower frame bound immediately:

$$\|\alpha^{(J)}\|_2 = \|M_J^\dagger \alpha\| \leq \|M_J^\dagger\| \|\alpha\|_2,$$

All the claims in Theorem 5.1 are now established.

5.2 Sparsity of the Gram Matrix

Here and below, we use the notation $\langle a \rangle = (1 + a^2)^{1/2}$. All the claims given in Lemma 5.3 follow from the two basic sets of estimates. First, let $\psi_{j,0,0}$ denote either of $\phi_{j,0,0}$ or $\varphi_{j,0,0}$. Then, for each $m = 1, 2, 3, \dots$ there are constants C_m so that

$$|\langle \psi_{j,0,0}, \phi_{j',k,\ell} \rangle| \leq C_m \mathbf{1}_{\{|j-j'| \leq 1\}} \cdot \mathbf{1}_{\{|\theta_{j,\ell}| \leq 10\bar{a}_j^{1/2}\}} \cdot \langle |b_{k_1, k_2}^{j,\ell}|_{a_j,0} \rangle^{-m} \quad \forall j, j', k, \ell. \quad (21)$$

In words, this says that different terms interact only if they are comparable in scale and orientation and then only if their locations are close in the metric $|v|_{a_j,0} = |(v_1/a_j, v_2/\sqrt{a_j})|_2$.

The second estimate concerns the difference between the two systems. For each $m = 1, 2, 3, \dots$, there is a sequence $\epsilon_{m,j}$ tending to zero with increasing j so that

$$|\langle \phi_{j,0,0} - \varphi_{j,0,0}, \phi_{j',k,\ell} \rangle| \leq \epsilon_{m,j} \cdot \mathbf{1}_{\{|j-j'| \leq 1\}} \cdot \mathbf{1}_{\{|\theta_{j,\ell}| \leq 10\bar{a}_j^{1/2}\}} \cdot \langle |b_{k_1, k_2}^{j,\ell}|_{a_j,0} \rangle^{-m} \quad \forall j, j', k, \ell. \quad (22)$$

The interpretation is similar to the previous one, only the point is that even when terms interact, they are small for large j .

Before developing these estimates, we remark that they immediately imply (18), (19), (20). Indeed, $|b_{k_1, k_2}^{j, \ell}|_{a_j, 0} \geq C \langle |(k_1, k_2)| \rangle$. Thus,

$$\sum_{k_1, k_2} |\langle |b_{k_1, k_2}^{j, \ell}|_{a_j, 0} \rangle^{-m}|^p \leq C \sum_{k_1, k_2} \langle |(k_1, k_2)| \rangle^{-mp}.$$

On the other hand $\langle k_1 \rangle \langle k_2 \rangle \leq 2 \langle |(k_1, k_2)| \rangle$, and picking $mp > 1$, we have for $i = 1, 2$ that $\sum_{k_i} \langle k_i \rangle^{-mp} < C_{m, p}$.

Both estimates follow from familiar principles about decay of Fourier transforms of smooth functions, after translation into a setting of parabolic scaling. We first recall the well-known basic principle.

Lemma 5.5 *Let g be a bandlimited function, with \hat{g} supported in a fixed bounded rectangle Ξ and belonging to $C_0^\infty(\Xi)$. Then for each $m = 2, 4, 6, \dots$ there are constants C_m depending only on m and $\text{diam}(\Xi)$, so that*

$$|g(b)| \leq C_m \cdot (\|\hat{g}\|_\infty + \|\hat{g}\|_{C^m}) \langle |b| \rangle^{-m}$$

Proof. This is very standard, but we reproduce it here for the convenience of some readers. From the Fourier inversion $g(b) = (2\pi)^{-2} \int e^{i\xi' b} \hat{g}(\xi) d\xi$ and the spatial-domain multiplier representation of the frequency-domain Laplacian $\Delta = \sum_i \frac{\partial^2}{\partial \xi^2}$,

$$(-|b|^2)^k g(b) = (2\pi)^{-2} \int e^{i\xi' b} (\Delta^k \hat{g}(\xi)) d\xi, \quad k = 1, 2, \dots$$

we immediately get

$$|g(b)| \leq \int |\hat{g}(\xi)| d\xi, \quad |b|^{2k} |g(b)| \leq \int |\Delta^k \hat{g}(\xi)| d\xi$$

from which

$$(1 + |b|^{2k}) |g(b)| \leq \text{Diam}(\Xi)^2 \cdot (\|\hat{g}\|_\infty + \|\hat{g}\|_{C^{2k}}),$$

and $(1 + |b|^{2k}) \geq \langle |b| \rangle^{-2k}$. □

An obvious parabolic rescaling of Lemma 5.5 gives

Lemma 5.6 *Suppose we have a sequence of functions (f_j) so that every \hat{f}_j is supported in a rectangle*

$$\Xi_j = [-C_1/a_j, C_2/a_j] \times [-C_2/a_j^{1/2}, C_2/a_j^{1/2}]$$

and each rescaled function

$$\hat{g}_j(u, v) = \hat{f}_j(u/a_j, v/\sqrt{a_j}) a_j^{-3/2}$$

obeys $\|\hat{g}_j\|_{C^m} \leq \gamma_m$, $m = 2, 4, 6, \dots$, independently of j . Then for $m = 1, 2, 3, \dots$ there are constants c_m so that

$$|\hat{f}_j(b)| \leq c_m (\gamma_0 + \gamma_m) \cdot \langle |b|_{a_j, 0} \rangle^{-m} \quad \forall b.$$

We now apply this parabolic variant of the decay principle to get the two estimates (21)-(22). To get the first estimate, we note that

$$\langle \phi_{j, 0, 0}, \phi_{j, k, \ell} \rangle = (2\pi)^{-2} \int e^{-i\xi' b_k} \hat{f}_j(\xi) d\xi,$$

where

$$\hat{f}_j(\xi) = W(a_j r) V\left(\frac{\omega}{\pi \tilde{a}_j^{1/2}}\right) W(a_{j'} r) V\left(\frac{\omega - \theta_{j', \ell}}{\pi \tilde{a}_{j'}^{1/2}}\right) a_j^{3/4} \cdot a_{j'}^{3/4}.$$

As we are trying to control a Fourier transform, the estimation Lemma 5.6 will be brought into play. Note that the pairs in the product $(WV) \cdot (WV)$ have disjoint support unless $|j - j'| \leq 1$ and $|\theta_{j', \ell'}| < 10\sqrt{a_j}$. Rescaling the product according to

$$\hat{g}_j(u, v) = \hat{f}_j(u/a_j, v/\sqrt{a_j}) a_j^{-3/2}$$

yields a function which can be decomposed into factors following the original product structure:

$$\hat{g}_j(u, v) = \tilde{V}_{0,j}(u, v) \tilde{W}_{0,j}(u, v) \cdot \tilde{V}_{0,j',j}(u, v) \tilde{W}_{1,j',j}(u, v).$$

Here the factors $\tilde{V}_{i,j}$ and $\tilde{W}_{i,j}$ individually belong to C^∞ , and by inspection we can see they obey bounds on derivatives independent of j and j' as soon as they are sufficiently large. For example, consider $\tilde{W}_{0,j}(u, v) = W(\sqrt{u^2 + a_j \cdot v^2})$. This is a function on a fixed domain $(u, v) \in [C_1/4, 4C_1] \times [C_2, C_2]$, for all large enough j . Now evidently, for each $m \geq 1$ we can find a constant η_m so that on this domain

$$\|\sqrt{u^2 + a_j \cdot v^2}\|_{C^m} \leq \eta_m \quad j \rightarrow \infty,$$

and of course W is smooth, so similar types of control are available for the C^m norms $\tilde{W}_{0,j}$ on this domain, valid for all sufficiently large j . Similar analyses apply to the other terms. Hence the product of those terms is C^∞ with bounds on the C^m norms independent of j, j' once they are both sufficiently large and $|j - j'| \leq 1$. Applying the estimation Lemma 5.6 gives the result (21).

The argument for (22) is similar.

$$\langle \phi_{j,0,0} - \varphi_{j,0,0}, \phi_{j',k,\ell} \rangle = (2\pi)^{-2} \int e^{-i\xi' b_k} \hat{f}_j(\xi) d\xi,$$

where

$$\hat{f}_j(\xi) = \left(W(a_j r) V\left(\frac{\omega}{\pi \tilde{a}_j^{1/2}}\right) - W(a_j \xi_1) V\left(\frac{\xi_2}{\xi_1} \frac{1}{\pi \tilde{a}_j^{1/2}}\right) \right) \cdot W(a_{j'} r) V\left(\frac{\omega - \theta_{j', \ell}}{\pi \tilde{a}_{j'}^{1/2}}\right) a_j^{3/4} \cdot a_{j'}^{3/4}.$$

which rescales as

$$\hat{g}_j(u, v) = \left(\tilde{W}_{0,j}(u, v) \tilde{V}_{0,j}(u, v) - \tilde{W}_{1,j}(u, v) \tilde{V}_{1,j}(u, v) \right) \cdot \tilde{W}_{2,j',j}(u, v) \tilde{V}_{2,j',j}(u, v).$$

Here, for example

$$\tilde{W}_{0,j}(u, v) = W(\sqrt{u^2 + a_j \cdot v^2}), \quad \tilde{W}_{1,j}(u, v) = W(u).$$

These are both functions on a fixed domain in $(u, v) \in [C_1/4, 4C_1] \times [C_2, C_2]$. Now evidently for each $m \geq 1$ on this domain we have

$$\|\sqrt{u^2 + a_j \cdot v^2} - u\|_{C^m} \rightarrow 0, \quad j \rightarrow \infty.$$

By smoothness of W we have that, for every $m = 1, 2, \dots$, there is a sequence $\eta_{m,j}^1 \rightarrow 0$ as $j \rightarrow \infty$ with

$$\|\tilde{W}_{0,j}(u, v) - \tilde{W}_{1,j}(u, v)\|_{C^m} \leq \eta_{j,m}^1.$$

We get a sequence $\eta_{j,m}^2$ giving similar control on the factors $\tilde{V}_{0,j}(u, v) - \tilde{V}_{1,j}(u, v)$ by parallel arguments. The factors $\tilde{V}_{2,j',j}(u, v) \tilde{W}_{2,j',j}(u, v)$ are handled as for the first family of estimates discussed earlier, the C^m norms being bounded by constants γ_m for all sufficiently large j' . Combining all these bounds, we get a sequence $\epsilon_{j,m} \rightarrow 0$ so that for each m

$$\|\hat{g}_j(u, v)\|_{C^m} \leq \epsilon_{j,m} \rightarrow 0 \quad j \rightarrow \infty.$$

Applying Lemma 5.6 completes the estimate (22).

6 Sparse Representation of FIO's

Parabolic-scaling decompositions have been used in earlier work to obtain representations and boundedness properties of Fourier Integral Operators. In particular, Hart Smith in [14] proved the invariance under diffeomorphisms of certain properties on molecular decompositions $f = \sum A_Q m_Q$ as in (8). His arguments concerned, as we have mentioned, a continuous transform based on affine parabolic scaling *strictu sensu*. However, using the results we have just proven on the comparison of frames Φ and Φ_J , we can draw parallel conclusions for the frame Φ , which we now do.

A (local) Fourier Integral Operator of order 0 [11, 8] is an operator T generated by

$$(Tf)(x) = \int e^{i\Phi(x,y,\xi)} a(x,y,\xi) f(y) dy d\xi; \quad (23)$$

here a belongs to the symbol class $S^0(\mathbf{R}^4 \times \mathbf{R}^2)$ of usual pseudo-differential symbols as in [11], and the phase function Φ satisfies nondegeneracy conditions

$$\det\left(\frac{\partial}{\partial x_i} \frac{\partial}{\partial \xi_j} \Phi\right) \neq 0; \quad \det\left(\frac{\partial}{\partial y_i} \frac{\partial}{\partial \xi_j} \Phi\right) \neq 0.$$

Examples of FIO's include:

- Change of variables operators $Tf(x) = f(\kappa(x))$ where κ is a diffeomorphism of \mathbf{R}^2 , in which case $a = 1$ and $\Phi(x,y,\xi) = \xi'(y - \kappa(x))$;
- Pseudodifferential operators, where $Tf(x) = \int e^{-i\xi'x} a(x,\xi) \hat{f}(\xi) d\xi$, which implicitly is of the form (23) with $\Phi(x,y,\xi) = \xi'(x - y)$, and a is a symbol of order 0.

There is some sense in which these two examples, taken together, exhaust the class of FIO's; microlocally, an FIO may be interpreted as a composition of a change-of-variables with a pseudodifferential operator.

A key notion in microlocal analysis is that of canonical transformation of phase space; we let $S^*(\mathbf{R}^2)$ denote the cosphere bundle of \mathbf{R}^2 – loosely, $\{(x_0, \theta_0) : x_0 \in \mathbf{R}^2, \theta_0 \in [0, 2\pi)\}$. If we consider a diffeomorphism $\kappa : \mathbf{R}^2 \mapsto \mathbf{R}^2$, then it maps space/codirection pairs (x_0, θ_0) into space/codirection pairs $\chi(x_0, \theta_0) = (\kappa(x_0), \kappa^* \theta_0)$, where $\kappa^* \theta_0$ is the codirection into which the codirection θ_0 based infinitesimally at x_0 is mapped under κ . In effect, a diffeomorphism κ of the base space induces a diffeomorphism χ of the phase space. More generally, a canonical transformation is a diffeomorphism of the phase space which locally behaves as if it were induced by such a global diffeomorphism of the base space.

We now show how to adapt ideas of [14] to obtain the following result.

Theorem 6.1 *Suppose T is a local Fourier integral operator of order 0 such that the Lagrangian relation of T is the graph of a homogeneous canonical transformation, and such that the distribution kernel of T vanishes outside a compact set. Let $S = (\langle \phi_Q, T \phi_{Q'} \rangle)$ denote the matrix representation of the operator T using the curvelets frame. Then the matrix is sparse:*

$$\|S\|_p < \infty \quad \forall p > 0.$$

In short, the curvelet frame sparsely represents Fourier Integral Operators of order 0. More extensive results of this kind have been developed by Candès and Demanet. Our argument here merely applies an estimate from the paper [14] in the frame Φ_J and then uses the sparsity of the change-of-frame matrices connecting Φ and Φ_J .

To state the key estimate, we use Smith's parabolic distance in phase space [14]. Given phase space pairs (x, θ) , (x', θ') , define the pseudo-distance

$$d(x, \theta; x', \theta') = |\langle e_\theta, x' - x \rangle| + |\langle e_{\theta'}, x' - x \rangle| + \min(|x' - x|, |x' - x|^2) + |\theta - \theta'|^2,$$

where $e_\theta \equiv (\cos(\theta), \sin(\theta))$. Roughly speaking, displacements in space which align with the codirections are treated much more seriously than those which are not aligned. Using these notions, Smith showed the following:

Lemma 6.1 (Smith, Lemma 3.11, 1998) Let T be as in the statement of Theorem 6.1. Let η^i , $i = 1, 2$, be bandlimited functions with $\hat{\eta}^i$ both C^∞ and supported in $|\xi - (1, 0)| < 1/4$. Let $\eta_{j,0,\ell}^i$ be affine parabolic dilation of η_i according to $R_{\theta_{j,\ell}} D_{a_j}$, and let $\eta_{j,b,\ell}^i(x) = \eta_{j,0,\ell}^i(x - b)$. Set

$$T_{j,\ell}^{j',\ell'}(b, b') = \langle \eta_{j,b,\ell}^1, T\eta_{j',b',\ell'}^2 \rangle, \quad b, b' \in \mathbf{R}^2.$$

Then, for each $N > 0$,

$$|T_{j,\ell}^{j',\ell'}(b, b')| \leq C_N \cdot \langle \frac{a_{j'}}{a_j} \rangle^{-N} \cdot \langle \frac{a_j}{a_{j'}} \rangle^{-N} \cdot \langle d((b, \theta_{j,\ell}); \chi(b', \theta_{j',\ell'})) \rangle^{-N}. \quad (24)$$

Here the constant C_N depends on N, T , and η^i , $i = 1, 2$, but not on j, j', ℓ , or ℓ' .

Define a matrix $S^{i,i'}$ by sampling (b, b') -space according to the schemes we used in Section 4 above:

$$S_{Q,Q'}^{i,i'} = \left(T_{j,\ell}^{j',\ell'}(b_{k_1,k_2}^{j,\ell}, b_{k'_1,k'_2}^{j',\ell'}) \right)$$

The lemma implies that this matrix is sparse: for each $p > 0$, $\|S^{i,i'}\|_p < \infty$.

In a moment we will show how this sparsity immediately implies:

Lemma 6.2 Let T be as in the statement of Theorem 6.1. Let $S^{(J)}$ denote the matrix defined using the frame Φ_J by

$$S_{Q,Q'}^{(J)} = \langle \varphi_Q, T\varphi_{Q'} \rangle.$$

This matrix is sparse: for each $p > 0$, $\|S^{(J)}\|_p < \infty$.

There is also a corresponding part of Smith's lemma concerning low-frequency functions, which offers the expected counterpart of the above, and which we use implicitly without any comment.

Before proving this lemma, we remark that it proves Theorem 6.1. Indeed, the S matrix to be bounded in the Theorem is related to $S^{(J)}$ of the Lemma by:

$$S = M_J S^{(J)} M_J^\dagger$$

where M_J and M_J^\dagger are the change-of-frame matrices in the previous section. But, of course

$$\|S\|_p \leq \|M_J\|_p \cdot \|S^{(J)}\|_p \cdot \|M_J^\dagger\|_p$$

where finiteness of $\|M_J\|_p$ and $\|M_J^\dagger\|_p$ has been established in the last section. Hence, finiteness of $\|S^{(J)}\|_p$ implies that of $\|S\|_p$.

It remains to prove Lemma 6.2. Note that the Lemma 6.1 would provide exactly what is needed, if it could be applied to elements of the curvelet frame Φ_J with a constant C_N not depending on j, j', ℓ or ℓ' . However, Smith's Lemma as stated makes the (in our context) restrictive assumption that η^i have Fourier transforms supported in $|\xi - (1, 0)| < 1/4$, so it does not apply immediately; and the curvelet frame Φ_J does not use affine parabolic scaling, which is also an obstacle to immediate application.

Each frame element $\varphi_{j,0,\ell}$, after parabolic affine rescaling, has a Fourier transform $\hat{\varphi}^j$ which is compactly supported in a fixed rectangle in polar coordinates, $[1/2, 2] \times [-1, 1]$, say. This rectangle can be covered by a finite system of overlapping balls $B(\xi_i, 1/4)$. With such a system, we can construct a smooth finite partition of unity (w_i) such that

$$\hat{\varphi}^j(\xi) = \sum_i w_i(\xi) \hat{\varphi}^j(\xi) = \sum_i \eta^{j,i}(\xi). \quad (25)$$

Now each $\eta^{j,i}$ is localized in a small ball in frequency space as in Smith's hypothesis. However, the ball is 'centered' at ξ_i rather than $(1, 0)$. Hence the parabolic rescaling $\eta_{j,b,\ell}^{j,i}$ is centered at

scale a_j^i (rather than a_j) and angle $\theta_{j,\ell}^i$ (rather than $\theta_{j,\ell}$). Lemma 6.1 applies in this setting, with this slightly different set of angles and scales, yielding:

$$|\langle \eta_{j,b,\ell}^{j,i}, \eta_{j',b',\ell'}^{j',i'} \rangle| \leq C_{N,(j,i),(j',i')} \cdot \langle \frac{a_{j'}^{i'}}{a_j^i} \rangle^{-N} \cdot \langle \frac{a_j^i}{a_{j'}^{i'}} \rangle^{-N} \cdot \langle d((b, \theta_{j,\ell}^i); \chi(b', \theta_{j',\ell'}^{i'})) \rangle^{-N}. \quad (26)$$

Now note that

$$|\theta_{j,\ell}^i - \theta_{j,\ell}| < C\sqrt{a_j}, \quad |\log(a_j/a_j^i)| < 3.$$

We have the relations

$$\langle \frac{a_j^i}{a_{j'}^{i'}} \rangle \asymp \langle \frac{a_j}{a_{j'}} \rangle$$

and

$$\langle d((b, \theta_{j,\ell}^i); \chi(b', \theta_{j',\ell'}^{i'})) \rangle \asymp \langle d((b, \theta_{j,\ell}); \chi(b', \theta_{j',\ell'})) \rangle;$$

in both, the implied constants are independent of j, j' . Using the finiteness of the sum in (25), we sum inequalities (26) to get that

$$|\langle \varphi_{j,b,\ell}, T\varphi_{j',b',\ell'} \rangle| \leq C_{N,j,j'} \cdot \langle \frac{a_{j'}}{a_j} \rangle^{-N} \cdot \langle \frac{a_j}{a_{j'}} \rangle^{-N} \cdot \langle d((b, \theta_{j,\ell}); \chi(b', \theta_{j',\ell'})) \rangle^{-N}. \quad (27)$$

We now make the observation that, although the system Φ_J is not generated by affine parabolic scaling of a single element, it is generated by affine parabolic scaling of only $J + 1$ different elements – because there are only J different levels where we use polar parabolic scaling as opposed to true affine parabolic scaling. Hence, as the underlying estimate (24) is uniform across all pairs at fine scales generated by affine parabolic scaling, there are really only finitely many different constants $C_{N,j,j'}$ involved in this estimate, and so, taking

$$C_N^* = \max_{0 \leq j, j' \leq J} C_{N,j,j'}$$

gives the same form of inequality as (27) with C_N^* in place of $C_{N,j,j'}$. For large enough N this inequality is p -th power summable either in j, k, ℓ or j, k', ℓ' ; so we get a sum bounded independently of the row or column being summed, hence Lemma 6.2 follows.

7 Discussion

We have described several transforms and their interrelationships. There are other possibilities. For example, it is possible to further simplify the relation between continuous and discrete tight frames. One can, in fact, define a continuous transform $\tilde{\Gamma}(a, b, \theta)$ which makes a continuous tight frame, and in which simple equispaced sampling yields (up to a proportionality factor), the coefficients of a discrete tight frame. One simply defines curvelets spanning *two octaves* at once, and samples only *every other* scale. The details are easy to supply using the framework of Sections 2 and 4 above.

References

- [1] E. J. Candès and F. Guo (2002). New Multiscale Transforms, Minimum Total Variation Synthesis: Applications to Edge-Preserving Image Reconstruction. *Signal Processing* **82**, 1519–1543.
- [2] Candès, E. J.; Demanet, L. (2003). Curvelets and Fourier integral operators. *C. R. Acad. Sci. Paris, Ser. I* **336**, 395–398.

- [3] Candès, E.J. and Donoho, D.L. (2002) New Tight Frames of Curvelets and Optimal Representations of Objects with Piecewise C^2 Singularities. To appear, *Comm. Pure Appl. Math.*
- [4] Candès, E.J. and Donoho, D.L. (2003) Continuous Curvelet Transform: I. Resolution of the Wavefront Set.
- [5] Candès, E.J. and Donoho, D.L. (2000) Curvelets: a surprisingly effective nonadaptive representation of objects with edges. In *Curve and Surface Fitting: Saint-Malo 1999* Albert Cohen, Christophe Rabut, and Larry L. Schumaker (eds.) Vanderbilt University Press, Nashville, TN.
- [6] Daubechies, I. (1992) *Ten Lectures on Wavelets*. CBMS-NSF Series in Applied Mathematics, No. 61. SIAM Philadelphia.
- [7] Donoho, D.L. (2001) Sparse Components of Images and Optimal Atomic Decomposition. *Constructive Approximation*, **17**, no. 3, 353–382.
- [8] Duistermaat, J.J. (1996) *Fourier Integral Operators*. Progress in Mathematics. Basel: Birkhäuser.
- [9] Fefferman, Charles (1973). A note on spherical summation multipliers. *Israel J. Math.* **15**, 44–52.
- [10] M. Frazier, B. Jawerth, and G. Weiss (1991) *Littlewood-Paley Theory and the study of function spaces*. NSF-CBMS Regional Conf. Ser in Mathematics, **79**. American Math. Soc.: Providence, RI.
- [11] Lars Hörmander (1983) *The Analysis of Linear Partial Differential Operators*. New York and Berlin: Springer.
- [12] Meyer, Y. (1990a) *Ondelettes*. Paris: Hermann. English translation published by Cambridge University Press.
- [13] Seeger, A., C.D. Sogge, and E.M. Stein (1993) Regularity properties of Fourier Integral operators. *Ann. Math.* **134**: 231-251.
- [14] Smith, Hart F. (1998) A Hardy space for Fourier integral operators. *J. Geom. Anal.* **8**, no. 4, 629–653.
- [15] Smith, Hart F. (1998) Wave equations with low regularity coefficients, *Documenta Mathematica*, Extra Volume ICM 1998, II, 723-730.
- [16] Elias M. Stein (1993) *Harmonic analysis: real-variable methods, orthogonality, and oscillatory integrals*. Princeton University Press, Princeton, NJ.