# Computing Fourier integral operators with caustics

**Peter Caday**

Rice University, 6100 Main St., Houston, TX 77005

E-mail: pac5@rice.edu

**Abstract.** Fourier integral operators (FIOs) have widespread applications in imaging, inverse problems, and PDE. An implementation of a generic algorithm for computing FIOs associated with canonical graphs is presented, based on a recent paper of de Hoop et al. Given the canonical transformation and principal symbol of the operator, a preprocessing step reduces application of an FIO approximately to multiplications, pushforwards and forward and inverse discrete Fourier transforms, which can be computed in $O(N^{n+(n-1)/2} \log N)$ time for an $n$-dimensional FIO. The same preprocessed data also allows computation of the inverse and transpose of the FIO, with identical runtime. Examples demonstrate the algorithm's output, and easily extendible MATLAB/C++ source code is available from the author.

AMS classification scheme numbers: 65M99, 35S30, 65R10, 65T99

Submitted to: *Inverse Problems*

## 1. Introduction

Fourier integral operators have found use in representing solutions of PDEs and in inverse problems. For instance, they model the solution operators of hyperbolic equations, and a variety of integral transforms, such as the Radon transform and its generalizations. With these applications in mind, generic algorithms for computing FIOs would be a useful computational tool for researchers. Many FIOs of interest are associated with a canonical transformation $\chi\colon (x,\xi) \mapsto (y,\eta)$, and these will be the operators considered in this paper. For short, we will call these operators *graph FIOs*. Throughout, we consider FIOs of Hörmander type, with homogeneous phase functions.

Under a *no-caustics* assumption, any graph FIO can be represented microlocally as an oscillatory integral

$$(Af)(y) = \int_{\mathbb{R}^n} e^{\mathrm{i}S(y,\xi)} a(y,\xi) \hat{f}(\xi)\,\mathrm{d}\xi. \tag{1}$$

Direct integration of (1) is possible, but inefficient ($O(N^{2n})$ time).

It has been known for a long time, stemming from the work of Cordoba and Fefferman [1], that FIOs associated with canonical transformations preserve space-frequency localization. Since then, a number of authors have studied these localization properties in detail. Smith [2, 3] proved that map a wave packet concentrated in space and frequency at $(x,\xi)$ to one centered at $\chi(x,\xi)$. This sparsity was noted later also by Candès and Demanet [4, 5] in regard to curvelets: they proved that graph FIOs have essentially sparse representations with respect to the curvelet frame. Later, Guo and Labate [6] showed sparsity for shearlets, another space- and frequency-localized frame. Cordero et al. [7] have recently established super-exponential off-diagonal decay for a variety of evolution operators in a Gabor frame.

In a related vein, Tataru [8] defined and studied a class of FIOs directly in terms of operators on a continuous frame of (space- and frequency-localized) coherent states whose kernels exponentially decay away from the graph of the canonical transformation. Schrödinger-class FIOs have also been shown to be sparse, with respect to a Gabor frame, due to the work of Cordero, Gröchenig, Nicola, and Rodino [9–12].

These results all suggest that wave packets (and other space-frequency localized frames) might provide a useful approach to computing graph FIOs. In [13], Andersson, de Hoop, and Wendt designed a $O(N^{(3n-1)/2}\log N)$ wave packet-based algorithm for computing graph FIOs, building on [14, 15]. As well, Candès, Demanet, and Ying have given two other algorithms for efficiently computing graph FIOs in dimension 2: an $O(N^{2.5}\log N)$ algorithm [16] based on dividing frequency space into properly-sized wedges, and in [17] an intriguing $O(N^2 \log N)$ algorithm with a butterfly structure.

These three algorithms have several common elements: each involves splitting the exponential term of (1) into oscillatory and non-oscillatory parts and constructing a low-rank approximation for the non-oscillatory parts separating the $y$ and $\xi$ variables. Each has a different approach to constructing low-rank approximations: Andersson et al. use *prolate spheroidal wave functions*. Candès et al. in [16] offer two approaches: a randomized

algorithm and a deterministic method using Taylor series. Finally, [17] employs a polynomial approximation using Chebyshev grids.

All three algorithms also require the FIO to be represented in the local form (1), which is possible only in the absence of *caustics*, points where the FIO's canonical relation $\Lambda'$ cannot be locally parameterized by $(y, \xi)$ coordinates. Even away from caustics, parameterizing $\Lambda'$ by $(y, \xi)$ coordinates, instead of the natural $(x, \xi)$ coordinates, is often messy. De Hoop et al. [18] address the first problem, extending Andersson et al.'s algorithm to overcome caustics, using appropriate changes of coordinates $(x, \xi) \mapsto (\tilde{x}, \tilde{\xi})$.

In this paper, we develop a modified version of de Hoop et al.'s algorithm for computing graph FIOs. The key features of the algorithm are its ease of use and automatic handling of caustics. As input, it requires only an FIO's canonical transformation and principal symbol, which are often much more natural objects than the integral form (1). If unknown, the principal symbol can typically be taken as 1 without affecting the singularities of $Af$. We also present companion algorithms that compute the inverse and transpose operators without inverting the canonical transformation explicitly.

We begin in §2 with a brief descriptions of the FIOs we consider and a review of their properties. §3 describes the forward algorithm, which is split into two stages: local (calculating local integral forms) and global (avoiding caustics). The inverse and transpose algorithms are also described. In §4 we test the algorithm on a variety of FIOs, including the Radon transform and wave propagators. Proofs from the body of the paper follow in §5. A MATLAB implementation, with documentation, is available from the author.

## 2. Notation and Definitions

We consider Hörmander's now-classical Fourier integral operators associated with homogeneous phase functions and polyhomogeneous symbols of order $(1, 0)$. Locally, these FIOs are operators $A\colon C_0^\infty(X) \to \mathcal{D}'(Y)$ taking the form

$$(Af)(y) = \int_X \int_{\mathbb{R}^N} e^{i\phi(x,y,\theta)} a(x, y, \theta) f(x) \,\mathrm{d}\theta \,\mathrm{d}x. \tag{2}$$

Here $\phi \in C^\infty(X \times Y \times (\mathbb{R}^N \setminus 0))$ is a real-valued phase function, homogeneous of degree 1 in $\theta \in \mathbb{R}^N$, that is nondegenerate in the sense that $dd_\theta\phi$ has full rank on the critical set $\{d_\theta\phi = 0\}$. The amplitude $a$ is a classical polyhomogeneous symbol of some order $m \in \mathbb{R}$; that is, it satisfies bounds

$$\left|\partial_x^\alpha \partial_y^\beta \partial_\theta^\gamma a\right| < C_{\alpha,\beta,\gamma}(1 + |\theta|\,)^{m-|\gamma|} \tag{3}$$

locally uniformly on $X \times Y$. Polyhomogeneity means that $a$ has an asymptotic expansion

$$a \sim \sum_{j=0}^{\infty} a_{-j} \tag{4}$$

in homogeneous symbols $a_{-j}$ of strictly decreasing order tending to $-\infty$. The *order* of $A$ as an FIO is defined to be $m + \frac{1}{2}N - \frac{1}{4}(\dim X + \dim Y)$, and determines its Sobolev mapping properties if $A$ is a graph FIO.

Despite their technical nature, operators of the form (2) are abundant in applications, the exponential kernels often arising from Fourier transform representations. One of the most important objects connected with such an FIO is its canonical relation, which describes how the wavefront sets of the output $Af$ and input $f$ are related. First, we define its *Lagrangian*, a subset of $T^*Y \times T^*X$ determined by the critical points (in $\theta$) of the phase function $\phi$:

$$\Lambda = \{(y, d_y\phi; x, d_x\phi) \mid (x, y, \theta) \in \ker d_\theta\phi\} \subset T^*Y \times T^*X. \tag{5}$$

It can be shown that $\Lambda$ is indeed a Lagrangian submanifold of $T^*Y \times T^*X$ with respect to the sum of the pullbacks of the standard symplectic forms $\sigma_X, \sigma_Y$ on $T^*X$ and $T^*Y$ respectively. The actual relation between $\mathrm{WF}(Af)$ and $\mathrm{WF}(f)$ is given by a slight modification of $\Lambda$, the *canonical relation*

$$\Lambda' = \{(y, \eta, x, -\xi) \mid (y, \eta, x, \xi) \in \Lambda\} \subset T^*Y \times T^*X. \tag{6}$$

We will consider graph FIOs, operators whose canonical relation $\Lambda'$ is a the graph of some symplectomorphism $\chi\colon T^*X \to T^*Y$. In other words, $A$ maps a singularity in $f$ to at most one singularity in $Af$.

Usually, an FIO can be written only locally as an oscillatory integral (2). Globally, an FIO associated with a Lagrangian $\Lambda$ may be defined as a (locally finite) sum of such oscillatory integrals whose Lagrangians are contained in $\Lambda$. This definition also has an invariant form (which we will not need here, however) in terms of the actions of certain pseudodifferential operators [20, §25.1]. We let $I^m(\chi) = I^m(Y \times X, \chi)$ denote the space of Fourier integral operators of order $m$ with canonical transformation $\chi$.

Comparing the local and global views of FIOs, we may think of a phase function $\phi$ as locally representing the global object $\Lambda$. We may now ask the natural question: what is the global version of the amplitude $a$? Any Lagrangian $\Lambda$ may be parameterized by infinitely many phase functions $\phi$; given an FIO $A$ associated with $\lambda$, and a particular $\phi$ parameterizing $\Lambda$, an amplitude $a$ may be chosen so that the oscillatory integral (2) is equivalent to $A$ modulo smoothing operators, for $\mathrm{WF}(f)$ supported in a suitable neighborhood.

For pseudodifferential operators, the leading order term of the amplitude (the *principal symbol*) is invariant under coordinate changes, and can be obtained by applying $A$ to an oscillatory function $\exp[it\psi(x)]$ and considering its asymptotic behavior as $t \to \infty$. The same recipe can be applied to an FIO [30, §2.3]. However, in order to make the leading order amplitude $a_0$ invariant under changes of $\phi$, it is necessary to associate it with a smooth section of a particular bundle over $\Lambda$. This bundle is the tensor product of the half-density bundle on $\Lambda$ with the Maslov bundle, which accounts for multiplications by powers of $i$ under changes in $\phi$ [20, §25.1]. We will not need the precise definition of these bundles, and in fact will avoid their use by (discontinuously) trivializing them in §3.2.4. In this way, the principal symbol of a graph FIO reduces to a genuine function.

## 3. Description of the Algorithm

In this section, we develop an algorithm for numerically computing FIOs in the class $I^m(\chi)$ described above. Let $X, Y$ be compact sets in $\mathbb{R}^n$, with $n > 2$. Our inputs will be distribution $f \in \mathcal{D}'(X)$ and a graph FIO $A: C_0^\infty(X) \to \mathcal{D}'(Y)$ of order $m$. Since it is often convenient to specify $A$'s global nature rather than its integral representations, we assume that $A$ is given only in terms of its canonical transformation $\chi: T^*X \to T^*Y$ and principal symbol $\sigma$. These only characterize $A$ modulo operators of order $m-1$; however, the algorithm will compute $A$ modulo an FIO of order $m - \frac{1}{2}$, so this is not a problem. As noted above, we assume $A$ is polyhomogeneous, so that $\sigma$ is homogeneous of order $m$ in $\xi$.

Microlocally, $A$ can be described as an oscillatory integral (1), but only in the absence of caustics‡. Our algorithm is presented in two stages: the *local algorithm*, modified from [13], for approximating these microlocal integral representations, and the *global algorithm* which breaks an FIO into a sum of integrals of this kind with coordinate changes, following [18].

### *3.1. Local Algorithm*

Before describing the local algorithm, we recall the microlocal representation for graph FIOs. Our goal is to write, for some choice of $S$ and $a$,

$$(Af)(y) = \frac{1}{(2\pi)^n} \int e^{\mathrm{i}S(y,\xi)} a(y,\xi) \hat{f}(\xi)\, \mathrm{d}\xi. \tag{7}$$

We suppose local coordinates $x, y$ on $X$ and $Y$ are given, inducing coordinates $(y, \eta, x, \xi)$ on $T^*Y \times T^*X \supset \Lambda'$. If the projection $\Lambda' \to (y, \xi)$ is a local diffeomorphism, by Hörmander [19, Prop. 21.2.18] there is a generating function $S(y, \xi)$, homogeneous in $\xi$ of order 1, such that $S(y, \xi) - x \cdot \xi$ is a phase function for $\Lambda$ near $\gamma_0$. Then, if $\mathrm{WF}'(A)$ is supported in a sufficiently small neighborhood of $\gamma_0$, there is an amplitude $a(y, \xi)$ so that (7) holds, modulo $C^\infty$ [20, Prop. 25.1.5].

Hence, the condition for a representation of the form (7) to exist for $f$ microlocally supported near $\gamma_0$ is that $\gamma_0$ is not a caustic:

**Definition.** $\gamma_0$ is a *caustic* if the projection $\pi: \Lambda' \to (y, \xi)$ fails to be a local diffeomorphism at $\gamma_0$, or equivalently, $\det \partial y / \partial x = 0$.

The equivalence follows from $(x, \xi)$ being global coordinates on $\Lambda'$. Fortunately, it is always possible to choose local coordinates for $X$ in which $\gamma_0$ is not a caustic [20, Prop. 25.3.3].

Note that $S$ is uniquely defined by $\chi$. If $\gamma_0$ is not a caustic, $(y, \xi)$ are local coordinates for $\Lambda'$, so we can write $x = x(y, \xi)$ and $\eta = \eta(y, \xi)$. Since the Lagrangian associated with $S - x \cdot \xi$ is $\{y, \frac{\partial S}{\partial y}, \frac{\partial S}{\partial \xi}, -\xi\}$, we have.

$$\frac{\partial S}{\partial \xi} = x(y, \xi), \qquad \frac{\partial S}{\partial y} = \eta(y, \xi). \tag{8}$$

‡ It is interesting to note that the *generalized metaplectic operators* considered by Cordero et al. [11] always possess global integral representations.

We will also write $y = y(x, \xi)$ for the $y$-component of $\chi$, and write $\partial y/\partial x$, $\partial y/\partial \xi$ for its first derivatives.

With these preliminaries done, assume $\gamma_0$ is not a caustic and $A$ takes the form (7). We now approximate this integral.

*3.1.1. Cone Decomposition* Following Candès et al. [16], (7) is decomposed as a sum of integrals over conical regions $\Xi_1, \ldots, \Xi_r$ in frequency space. If these cones are narrow enough, then on each of them the exponential term of the integrand splits into an *oscillatory component* representing a change of coordinates, and a *nonoscillatory component* representing a pseudodifferential operator.

We begin by defining the cones. Choose $r$ approximately uniformly distributed unit covectors $\nu_1, \ldots \nu_r \in \mathbb{S}^{n-1}$, and let $\Xi_i = \{\xi \in \mathbb{R}^n \setminus \{0\} : |\frac{\xi}{|\xi|} - \nu_i| < \delta\}$, for some $\delta$ large enough that the $\Xi_i$ cover $\mathbb{R}^n \setminus \{0\}$. We will see later how $r$ should be chosen. Then, let $\{\rho_i(\xi)\}$ be a partition of unity subordinate to the $\Xi_i$, and let $f_i = \rho_i(D)f$ be the result of multiplying $f$ by $\rho_i$ in the frequency domain.

Focusing on the $i^{\text{th}}$ cone, we consider $Af_i$. Since $\hat{f}_i$ is supported on the cone $\Xi_i$, we can assume $\xi \in \Xi_i$. Choose a rotated coordinate basis $(\xi', \xi''_1, \ldots, \xi''_{n-1})$ in which $\nu_i = (1, 0, \ldots, 0)$. Expanding $S(y, \xi) = \xi' S(y, 1, \xi''/\xi')$ in a Taylor series about $\nu_i$ and using Euler's theorem gives

$$S(y,\xi) = \xi \cdot \frac{\partial S}{\partial \xi}(y, \nu_i) + \frac{1}{2}\frac{\xi''^{\top} \cdot \frac{\partial^2 S}{\partial \xi''^2}(y, \nu_i) \cdot \xi''}{\xi'} + O\left(\frac{|\xi''|^3}{|\xi'|^2}\right). \tag{9}$$

Now, suppose the cones are narrow enough that $|\xi''| = O(\sqrt{|\xi'|})$; this is the *parabolic scaling* condition for wave packets. Only in the numerical setting, where the maximum frequency is limited, is this possible. If we have an $N \times \cdots \times N$ grid of fixed size, then $|\xi| = O(N)$, and in order to get $|\xi''| = O(\sqrt{|\xi'|})$, we must have $\delta = O(N^{-1/2})$, requiring $r = O(N^{(n-1)/2})$ cones.

With this parabolic scaling assumption, the second and third terms of (9) are bounded, so these parts of $\exp[\mathrm{i}S(y, \xi)]$ do not oscillate. We group them instead with the amplitude $a(y, \xi)$ in (7). The first term of (9), however, *is* oscillatory, and it corresponds to a change of coordinates $\frac{\partial S}{\partial \xi}(y, \nu_i) \mapsto y$; for, if we were to remove the remaining terms as well as the amplitude (and integrate over all $\xi$), we would have

$$\frac{1}{(2\pi)^n} \int e^{\mathrm{i}\xi \cdot \frac{\partial S}{\partial \xi}(y,\nu_i)} \hat{f}(\xi)\, \mathrm{d}\xi = f\left(\frac{\partial S}{\partial \xi}(y, \nu_i)\right) = f(x(y, \nu_i)). \tag{10}$$

Now, let us collect in the symbol $b_i$ all the nonoscillatory terms and the frequency cutoff $\rho_i$:

$$b_i(y, \xi) = a(y, \xi) e^{\mathrm{i}(S(y,\xi) - \xi \cdot x(y,\nu_i))} \rho_i(\xi). \tag{11}$$

Let $B_i$ be the corresponding pseudodifferential operator,

$$(B_i f)(x') = \int e^{\mathrm{i}x' \cdot \xi}\, b_i(y(x', \nu_i), \xi)\, \hat{f}(\xi)\, \mathrm{d}\xi. \tag{12}$$

Then $Af_i$ can be decomposed as a change of coordinates composed with $B_i$:

$$\begin{aligned} Af_i(y) &= \int e^{\mathrm{i}S(y,\xi)} a(y,\xi)\hat{f}_i(\xi)\, d\xi = \int e^{\mathrm{i}x(y,\nu_i)\cdot\xi} b_i(y,\xi)\hat{f}(\xi)\,\mathrm{d}\xi \\ &= (B_i f)(x(y,\nu_i)). \end{aligned} \tag{13}$$

Adding up the contributions from each cone, we have

$$A = \sum_{i=1}^{r} B_i \circ T_{\nu_i}, \qquad \text{where } T_{\nu_i}(y) = x(y,\nu_i). \tag{14}$$

**Remark.** This approaches differs somewhat from Andersson et al.'s algorithm, in which frequency space is subdivided both angularly and radially in a dyadic parabolic decomposition, achieving the parabolic scaling condition. In our discrete setting, however, the radial subdivisions are not needed for parabolic scaling.

*3.1.2. Applying the pseudodifferential operator $B_i$* Equation (14) reduces the computation of graph FIOs microlocally to calculating coordinate changes and pseudodifferential operators. This section addresses the approximate computation of the associated pseudodifferential operators $B_i$. While there are generic algorithms (e.g. [21]) for computing ΨDOs, we follow here Andersson et al. [13]. We note as well that two alternative algorithms for this problem are given in [16], one deterministic and one randomized.

To compute $B_i$, the approach is to find *low-rank separated representations* of $b_i(x',\xi)$, expansions of the form $b_i(x',\xi) \approx \sum_k F_{ik}(y)G_{ik}(\xi)$, which are easily (and efficiently) approximated using fast Fourier transforms:

$$b_i(x',D)u \approx \sum_k F_{ik}(x')\,\mathrm{IDFT}[G_{ik}(\xi)\,\mathrm{DFT}[u]]. \tag{15}$$

Note that due the periodicity of the DFT, appropriate zero-padding of $u$ is required when the domain is not periodic.

*Approximating $B_i$* We start by neglecting the third-order and higher terms in (9). This truncation is equivalent to making an error of order $O(|\xi''|^3/|\xi'|^2) = O(|\xi|^{-1/2})$ in the generating function $S$. Since $|\exp[\mathrm{i}x] - 1| \le x$, we contribute an error of order $m - \frac{1}{2}$ in $\xi$ to the amplitude $b(x',\xi)$, where $m$ is the order of the FIO. Therefore, by truncating we perturb $A$ by an FIO which is half an order smoother, and replace $B_i$ by the simpler ΨDO $\overline{B}_i$ with symbol

$$\bar{b}_i(x',\xi) = \rho_i(\xi)a(y(x',\nu_i),\xi)E_i(x',\xi), \tag{16}$$

where

$$E_i(x',\xi) = \exp\left[\frac{\mathrm{i}\xi''^{\top}\cdot\frac{d^2S}{d\xi''^2}(x',\nu_i)\cdot\xi''}{2\xi'}\right]. \tag{17}$$

Setting aside for now the amplitude term $a(y,\xi)$, we are faced with constructing a low-rank separated representation of $E_i$, namely, functions $F_{ik}(x')$, $G_{ik}(\xi)$ for the approximation

$$E_i(x',\xi) \approx \sum_k F_{ik}(x')G_{ik}(\xi). \tag{18}$$

Note that the exponent of $E_i$ can be rewritten as a dot product of $\mathbb{R}^{(n-1)^2}$-valued functions in $x'$ and $\xi$:

$$E_i(x', \xi) = \exp[\mathrm{i}\tilde{\mathbf{u}}(x') \cdot \tilde{\mathbf{v}}(\xi)], \tag{19}$$

where

$$\tilde{\mathbf{u}}_{st}(x') = \frac{\partial^2 S}{\partial \xi_s'' \, \partial \xi_t''}(x', \nu), \qquad \tilde{\mathbf{v}}_{st}(\xi) = \frac{\xi_s'' \xi_t''}{2\xi'} \qquad (1 \le s, t < n). \tag{20}$$

To avoid undue proliferation of indices, $i$ has been omitted from $\tilde{\mathbf{u}}$ and $\tilde{\mathbf{v}}$.

*PSWFs* The problem of finding low-rank separated representations of (19) is elegantly handled by *prolate spheroidal wave functions* (PSWFs), a class of special functions developed by Slepian, Pollak, and Landau [22–25]. Originally developed to study bandlimited functions, their relevance here is that they are the eigenfunctions of the finite Fourier transform operator $\mathcal{E}_c \colon L^2([-1,1]) \to L^2([-1,1])$ with kernel $K_c(u,v) = \exp(\mathrm{i}cuv)$. As eigenfunctions, they form a complete basis for $L^2([-1,1])$ and allow us to construct exactly the desired separated representation (18) for $\exp(\mathrm{i}\tilde{\mathbf{u}}\tilde{\mathbf{v}})$.

**Theorem 1** (Slepian et al. [22])**.** *For all $c > 0$, there exists a complete orthonormal basis for $L^2([-1,1])$ of real eigenfunctions $\{\psi_j^{(c)}\}_{j=1}^{\infty}$ of $\mathcal{E}_c$. If $\lambda_j^{(c)}$ are the corresponding eigenvalues, $|\lambda_1^{(c)}| \ge |\lambda_2^{(c)}| \ge \cdots$, then*

$$\exp(\mathrm{i}cxy) = \sum_{j=0}^{\infty} \lambda_j^{(c)} \psi_j^{(c)}(x) \psi_j^{(c)}(y), \tag{21}$$

*and convergence is uniform.*

For computation we require a good approximate low-rank representation, where the infinite sum of (21) is replaced by a finite one. Fortunately, the $\lambda_j$ decay very rapidly for sufficiently large $j$, while the $\psi_j$ grow comparatively slowly, so we may simply truncate the infinite sum (21). Specifically,

**Theorem 2** (Landau and Widom [26])**.** *Let $N_\lambda(c, \epsilon)$ be the number of $\lambda_j$ greater than $\epsilon$ in absolute value. Then*

$$N_\lambda(c, \epsilon) = \frac{2c}{\pi} + \left( \frac{1}{\pi^2} \log \frac{1-\epsilon}{\epsilon} \right) \log c + o(\log c). \tag{22}$$

**Theorem 3** (Widom [27])**.** *For fixed $c$, as $j \to \infty$*

$$\lambda_j^{(c)} \sim \left( \frac{ec}{4(j + \frac{1}{2})} \right)^{2j+1}. \tag{23}$$

**Theorem 4** (Bonami and Karoui [28])**.** *If $j \ge 2c/\pi$, then*

$$\left\| \psi_j^{(c)} \right\|_\infty \le \varkappa \sqrt{j+1}, \text{ where } \varkappa = \frac{5}{4} \sqrt{\frac{\pi\sqrt{5}}{2}}. \tag{24}$$

Theorem 2 shows that $\lambda_j^{(c)}$ is near 1 for $j \ll 2c/\pi$, passes through $\frac{1}{2}$ for $j \approx 2c/\pi$, then rapidly approaches zero. Theorem 3 shows that the convergence is eventually superexponential. Bonami and Karoui [29] have also found a quantitative version of Theorem 3 that gives explicit bounds on $\lambda_j^{(c)}$. Combining Theorems 3 and 4 gives

**Corollary 5.** *For fixed c,*

$$\left|\exp(\mathrm{i}cxy) - \sum_{j=0}^{J} \lambda_j^{(c)} \psi_j^{(c)}(x) \psi_j^{(c)}(y)\right| = O\left(\frac{ec}{4(J+\frac{1}{2})}\right)^{2J}. \tag{25}$$

All that remains is to convert (19–20) into the proper form. Since the domain $X$ is compact, $\tilde{\mathbf{u}}$ is bounded. As for $\tilde{\mathbf{v}}$, it is bounded on $\Xi_i$ by construction (§3.1.1). Scaling $\tilde{\mathbf{u}}$ and $\tilde{\mathbf{v}}$ by their largest infinity-norms,

$$E_i(x', \xi) = \exp[\mathrm{i}c\mathbf{u}(x') \cdot \mathbf{v}(\xi)], \tag{26}$$

where

$$\mathbf{u}(x') = \frac{\tilde{\mathbf{u}}(x')}{\sup_{x'} \|\tilde{\mathbf{u}}(x')\|_\infty}, \qquad \mathbf{v}(\xi) = \frac{\tilde{\mathbf{v}}(\xi)}{\sup_{\xi} \|\tilde{\mathbf{v}}(\xi)\|_\infty}, \tag{27}$$

$$c = \sup_{x'} \|\tilde{\mathbf{u}}(x')\|_\infty \cdot \sup_{\xi} \|\tilde{\mathbf{v}}(\xi)\|_\infty. \tag{28}$$

For a multidimensional version of (21), take tensor products. For $\mathbf{j} = (j_1, \ldots, j_N) \in \mathbb{N}^N$, let $\psi_{\mathbf{j}}^{(c)}(x) = \psi_{j_1}^{(c)}(x_1) \cdots \psi_{j_N}^{(c)}(x_N)$. Then, if $x, y \in L^2([-1,1]^N)$,

$$\exp[icx \cdot y] = \prod_{1 \le \ell \le N} \exp[icx_\ell y_\ell] = \sum_{\mathbf{j} \in \mathbb{N}^N} \lambda_{\mathbf{j}}^{(c)} \psi_{\mathbf{j}}^{(c)}(x) \psi_{\mathbf{j}}^{(c)}(y). \tag{29}$$

Combining (29) with (21) gives

$$E_i(x', \xi) = \sum_{\mathbf{j} \in \mathbb{N}^N} \lambda_{\mathbf{j}}^{(c)} \psi_{\mathbf{j}}^{(c)}(\mathbf{u}(x'))\, \psi_{\mathbf{j}}^{(c)}(\mathbf{v}(\xi)). \tag{30}$$

Letting $\mathbf{J}(c, \epsilon)$ be the set of $\mathbf{j} = (j_i) \in \mathbb{N}^N$ such that $|j_1 \cdots j_N \lambda_{\mathbf{j}}^{(c)}| \le \epsilon$, we have the low-rank approximation

$$\left|\exp[\mathrm{i}cx \cdot y] - \sum_{\mathbf{j} \in \mathbf{J}(c,\epsilon)} \lambda_{\mathbf{j}}^{(c)} \psi_{\mathbf{j}}^{(c)}(x)\, \psi_{\mathbf{j}}^{(c)}(y)\right| = O(\epsilon). \tag{31}$$

For fixed $c$, $|\mathbf{J}(c, \epsilon)| = o(\log^N \epsilon)$ in view of Theorems 3 and 4, while for fixed $\epsilon$, $|\mathbf{J}(c, \epsilon)| = O(c^N)$ in view of Theorem 2.

**Remark.** Note that [18], as well as the current implementation of the algorithm, use the unit ball $\overline{\mathbb{B}}^N$ instead of the unit hypercube $[-1,1]^N$ as the domain for $\tilde{\mathbf{u}}$, $\tilde{\mathbf{v}}$. The only difference is that PSWFs are replaced by Slepian's generalized prolate spheroidal functions [25].

*Amplitude* Now that we have an effective algorithm for computing the $E_i(x', \xi)$ component of the ΨDO, we turn to the amplitude term $a(y(x', \nu), \xi)$. Again, we can do a Taylor expansion around the center of the cone. Given $\xi$, we let $\xi_0 = \nu|\xi|$; then

$$a(y, \xi) = a(y, \xi_0) + (\xi - \xi_0) \cdot \frac{\partial a}{\partial \xi}(y, \xi_0) + (\xi - \xi_0)^\top \frac{\partial^2 a}{\partial \xi^2}(y, \xi^*(y))(\xi - \xi_0) \tag{32}$$

for some $\xi^*(y)$ on the line segment between $\xi_0$ and $\xi$. The most basic possible approximation is the zeroth-order approximation

$$a(y, \xi) \approx a(y, \xi_0) = |\xi|^m \, a(y, \nu_i). \tag{33}$$

(By assumption, the principal symbol is homogeneous; hence $a$ is as well.) Because $|\xi - \xi_0| = O(|\xi|^{1/2})$ and $\partial a / \partial \xi = O(|\xi|^{m-1})$, this approximation has $O(|\xi|^{m-1/2})$ error, and therefore perturbs $A$ by an FIO of order one-half lower. This is acceptable as it coincides with the error we made when approximating $S(y, \xi)$ earlier. Hence we compute the ΨDO $\overline{\overline{B}}_i$ with symbol

$$\bar{\bar{b}}_i(x', \xi) = \rho_i(\xi) \; |\xi|^m \; a(y(x', \nu_i), \nu_i) \, E_i(x', \xi). \tag{34}$$

Computation of $\overline{\overline{B}}_i$ is via the low-rank separated approximation $\overline{\overline{B}}_{i,\epsilon}$ with symbol

$$\bar{\bar{b}}_{i,\epsilon}(x', \xi) = \rho_i(\xi) \; |\xi|^m \; a(y(x', \nu_i), \nu_i) \sum_{\mathbf{j} \in \mathbf{J}(c,\epsilon)} \lambda_{\mathbf{j}}^{(c)} \psi_{\mathbf{j}}^{(c)}(\mathbf{u}(x')) \, \psi_{\mathbf{j}}^{(c)}(\mathbf{v}(\xi)). \tag{35}$$

Before proceeding further we make a few notes on $\tilde{\mathbf{u}}$ and $\tilde{\mathbf{v}}$.

**Notes.**

(i) Since $\partial S / \partial \xi = x(y, \xi)$ by (8) we can write

$$\tilde{\mathbf{u}}_{st}(x') = \nu_s''^\top \frac{\partial x(y, \xi)}{\partial \xi} \nu_t'', \tag{36}$$

where $\nu_s''$ is the unit vector in the $\xi_s''$ direction. We can write $\tilde{\mathbf{u}}$ more directly in terms of the canonical map $\chi = (y(x, \xi), \eta(x, \xi))$ as:

$$\tilde{\mathbf{u}}_{st} = -\nu_s''^\top \left( \frac{\partial y(x, \xi)}{\partial x} \right)^{-1} \left( \frac{\partial y(x, \xi)}{\partial \xi} \right) \nu_t'' \Big|_{x = x(y,\xi)}. \tag{37}$$

(ii) Since $\tilde{\mathbf{u}}_{st} = \tilde{\mathbf{u}}_{ts}$ and $\tilde{\mathbf{v}}_{st} = \tilde{\mathbf{v}}_{ts}$, we can restrict to $s \le t$ in (19–20), and multiply one of them by 2 for $s \neq t$. This has the advantage of reducing the dimension of $\tilde{\mathbf{u}}$, $\tilde{\mathbf{v}}$ to $\binom{n-1}{2}$, decreasing the number of terms needed in the low-rank approximation (31).

(iii) The supremum in the definition of $\mathbf{u}$ in (28) can be taken over the $x$-support of $a(y(x, \nu), \nu)$ instead of all $X$, in view of the amplitude approximation (33).

*3.1.3. Local approximation formula* Now that we have reasonable approximations (30), (33) for the exponential and amplitude parts of $B_i$, we can substitute them into equation (14) for the original FIO $A$. Recalling that $\hat{f}_i = \rho_i \hat{f}$ and attaching $i$ indices to $c$, $\mathbf{u}$, and $\mathbf{v}$, we have

$$Af(y) \approx \sum_{i=1}^{r} \sum_{\mathbf{j} \in \mathbf{J}(c,\epsilon)} \lambda_{\mathbf{j}}^{(c_i)} \psi_{\mathbf{j}}^{(c_i)}(\mathbf{u}_i(x')) \, \mathcal{F}^{-1}_{\xi \to x'} \Big[ \psi_{\mathbf{j}}^{(c_i)}(\mathbf{v}_i(\xi)) \, |\xi|^m \rho_i(\xi) \hat{f}(\xi) \Big] a(y, \nu_i) \Big|_{x' = x(y,\nu_i)}. \tag{38}$$

Here $\mathcal{F}^{-1}_{\xi\to x'}$ represents the inverse Fourier transform.

Now the local approximation formula (38) is independent of $\gamma_0$, depending only on $\chi$ and $a$. Hence, while (7) was originally valid only for $\mathrm{WF}'(A)$ in a small neighborhood of $\gamma_0$, by a partition of unity the local approximation formula is valid whenever $\mathrm{WF}'(A)$ lies entirely in an open subset $U \subseteq \Lambda'$ that can be globally parameterized by $(y,\xi)$.

We can further relax the parametrization assumption. First, replace the pullback in (38) with a pushforward (in the sense of distributions) by the inverse map, $x \mapsto y(x,\nu_i)$ and a Jacobian factor. Using the pushforward is also advantageous practically, as it allows us to avoid inverting the canonical transformation. Then, dropping the prime on $x'$,

$$\begin{aligned} Af(y) \approx \mathcal{A}_\epsilon f(y) = \sum_{i=1}^{r} (y(\cdot,\nu_i))_* \left|\det \frac{\partial y}{\partial x}(x,\nu_i)\right| \\ \cdot \sum_{\mathbf{j}\in\mathbf{J}(c_i,\epsilon)} \lambda_j^{c_i}\psi_j^{c_i}(\mathbf{u}_i(x))\,\mathcal{F}^{-1}_{\xi\to x}\Big[\psi_j^{c_i}(\mathbf{v}_i(\xi))\,|\xi|^m \rho_i(\xi)\hat{f}(\xi)\Big] a(x,\nu_i). \end{aligned} \tag{39}$$

Here $a(x,\nu_i) = a(y(x,\nu_i),\nu_i)$.

Next, drop the assumption that $(y,\xi)$ parameterizes $U$ and only assume $U$ contains no caustics; in other words, $\det \partial y/\partial x \neq 0$. Equation (39) still makes sense; it only depends on $\Lambda$ via $\mathbf{u}$, which by (37) can be defined directly in terms of the canonical transformation $\chi$.

Now, $\det \partial y/\partial x \neq 0$ implies that $U$ can locally be parameterized by $(y,\xi)$. Covering $U$ by open sets that can be parameterized by $(y,\xi)$ and taking a partition of unity, we find that (39) is valid for all of $U$. Specifically, if $1 = \sum \alpha_i$ is the partition of unity and $U_i$ are the open sets, approximation (39) is valid with $a$ replaced by $a\alpha_i$ for each $i$. Summing over all $i$, the formula is also valid for $a$ by linearity.

Summarizing, we have developed the following approximation.

**Theorem 6** (Local approximation algorithm)**.** *Let $A \in I^m(Y\times X)$ be an FIO associated with the canonical transformation $\chi\colon T^*X \to T^*Y$, such that $\det \partial y/\partial x \neq 0$ on $\mathrm{WF}'(A)$.*

*There exists an error operator $K \in I^{m-1/2}(Y\times X,\chi)$ such that for small $\epsilon > 0$,*

$$\frac{\|Af - Kf - \mathcal{A}_\epsilon f\|_2}{\|f\|_{H^m}} = O(\epsilon), \tag{40}$$

*where $\mathcal{A}_\epsilon f$ is the approximation defined in (39).*

The error operator $K$, of course, corresponds to the higher-order terms in the Taylor expansions of amplitude and phase that were neglected.

*Proof.* With $\mathcal{A}_\epsilon$ defined as in (39), and recalling (14), we have

$$Af(y) = \sum_{i=1}^{r} T_{\nu_i} \circ B_i, \tag{41}$$

$$\mathcal{A}_0 f(y) = \sum_{i=1}^{r} T_{\nu_i} \circ \overline{\overline{B}}_i, \tag{42}$$

$$\mathcal{A}_\epsilon f(y) = \sum_{i=1}^{r} T_{\nu_i} \circ \overline{\overline{B}}_{i,\epsilon}. \tag{43}$$

As we noted earlier, the symbols $b_i$ and $\overline{\overline{b}}_i$ of $B_i$ and $\overline{\overline{B}}_i$, respectively, differ by a symbol of order $m-\frac{1}{2}$, and $K = A - \mathcal{A}_0$ is an FIO of order $m-\frac{1}{2}$. All that remains for us is to estimate $\mathcal{A}_0 - \mathcal{A}_\epsilon$. First, we estimate $\overline{\overline{B}}_i - \overline{\overline{B}}_{i,\epsilon}$:

$$(\overline{\overline{B}}_i - \overline{\overline{B}}_{i,\epsilon})f(x') = \sum_{\mathbf{j}\notin\mathbf{J}(c_i,\epsilon)} \lambda_{\mathbf{j}}^{(c_i)}\psi_{\mathbf{j}}^{(c_i)}(\mathbf{u}(x'))\mathcal{F}^{-1}\left[\psi_{\mathbf{j}}^{(c_i)}(\mathbf{v}(\xi))\rho_i(\xi)\,|\xi|^{\,m}\hat{f}(\xi)\right]. \quad (44)$$

Since

$$\begin{aligned}\left\|\mathcal{F}^{-1}\left[\psi_{\mathbf{j}}^{(c_i)}(\mathbf{v}(\xi))\rho_i(\xi)\,|\xi|^{\,m}\hat{f}(\xi)\right]\right\|_2 &= \left\|\psi_{\mathbf{j}}^{(c_i)}(\mathbf{v}(\xi))\rho_i(\xi)\,|\xi|^{\,m}\hat{f}(\xi)\right\|_2 \\ &\le \left\|\psi_{\mathbf{j}}^{(c_i)}\right\|_\infty \|\rho_i\|_\infty \left\||\xi|^{\,m}\hat{f}(\xi)\right\|_2 \\ &\le \left\|\psi_{\mathbf{j}}^{(c_i)}\right\|_\infty \|f\|_{H^m},\end{aligned} \quad (45)$$

we have

$$\left\|\lambda_{\mathbf{j}}^{(c_i)}\psi_{\mathbf{j}}^{(c_i)}(\mathbf{u}(x'))\mathcal{F}^{-1}\left[\psi_{\mathbf{j}}^{(c_i)}(\mathbf{v}(\xi))\rho_i(\xi)\,|\xi|^{\,m}\hat{f}(\xi).\right]\right\|_2 \le \left|\lambda_{\mathbf{j}}^{(c_i)}\right|\left\|\psi_{\mathbf{j}}^{(c_i)}\right\|_\infty^2 \|f\|_{H^m}. \quad (46)$$

□

### *3.2. Global Algorithm*

In the previous section, we derived a local approximation (39) for $A$, valid as long as $\mathrm{WF}'(A)$ contains no caustics. Our goal in this section is to derive a global approximation algorithm for $A$ by expressing $A$ as a sum of oscillatory integrals without caustics to which (39) applies. To handle caustics, de Hoop et al. [18] use appropriate changes of coordinates in $X$, which they call *singularity-resolving diffeomorphisms*.

Throughout, let $x$ denote a fixed choice of coordinates on $X$, and $\tilde{x}$ another choice of coordinates. Let $(x,\xi)$ and $(\tilde{x},\tilde{\xi})$ denote the associated coordinates on $T^*X$.

*3.2.1. First Draft of Algorithm* By proposition 25.3.3 in Hörmander [20], for each $\gamma_0 \in \Lambda'$ there is always a local choice of coordinates $\tilde{x}$ for $X$ in which $\Lambda'$ does not have caustics in a neighborhood of $\gamma_0$. This gives us the following first draft of an algorithm for computing $Af$ with caustics present.

For each $\gamma \in \Lambda'$ we get a neighborhood $U_\gamma$ with associated caustic-free coordinates $\tilde{x}^{(\gamma)}$. We cover $\Lambda'$ by the $U_\gamma$, which we can assume to be conic (since $\partial y/\partial x$ is homogeneous in $\xi$). Since $X, Y$ are compact, the cosphere bundle $S^*X \times S^*Y$ is also compact, so $\Lambda' \cap (S^*X \times S^*Y)$ is compact, and we can choose a finite subcover $U_{\gamma_1} \cap (S^*X \times S^*Y), \ldots, U_{\gamma_r} \cap (S^*X \times S^*Y)$. Then $U_{\gamma_1}, \ldots, U_{\gamma_r}$ covers $\Lambda$; choosing a partition of unity $\rho_j$ subordinate to this subcover, we can write $Af = A_1 f + \cdots + A_r f$, where $A_j$ is an FIO with canonical transformation $\chi$ associated with the principal symbol $\rho_j\sigma$. (Recall $\sigma$ is the principal symbol of $A$.) Each $A_j f$ can now be computed by theorem 6.

*3.2.2. Quadratic Coordinate Changes* While the algorithm just described is valid, it does not specify how to find caustic-free coordinates. To address this, we identify a specific finite family of simple coordinate changes that are sufficient for any $A$. If $\varphi : \mathbb{R}^n \to \mathbb{R}^n$ is a diffeomorphism, and $\tilde{x} = \varphi(x)$, then

$$\frac{\partial y}{\partial \tilde{x}} = \frac{\partial y}{\partial x}(d\varphi)^{-1} + \sum_{k=1}^{n} \frac{\partial y}{\partial \xi_k} \frac{\partial (d\varphi)^\top}{\partial x^k} \xi. \tag{47}$$

The simplest possible useful diffeomorphisms are therefore those with quadratic terms.

Let $t \in \mathbb{R} \setminus \{0\}$ be a fixed constant and $J \subsetneq \{1, \ldots, n\}$, $i \in \{1, \ldots, n\} \setminus J$. Define the diffeomorphisms $\beta_{i,J} \colon \mathbb{R}^n \to \mathbb{R}^n$ as follows:

$$\beta_{i,J}(x_1, \ldots, x_n) = \left( x_1, \ldots, x_{i-1}, x_i - \frac{t}{2} \sum_{j \in J} x_j^2, x_{i+1}, \ldots, x_n \right). \tag{48}$$

It turns out that in all cases, at least one of these quadratic coordinate changes avoids a caustic.

**Proposition 7.** *For all $\gamma_0 \in \Lambda'$, there exist $J \subsetneq \{1, \ldots, n\}$ and $i \in \{1, \ldots, n\} \setminus J$ such that $\det \partial y / \partial \tilde{x}^{(i,J)}(\gamma_0) \neq 0$, where $\tilde{x}^{(i,J)} = \beta_{i,J}(x)$.*

A proof is given in §5. There are $1 + n(2^{n-1} - 1)$ of these diffeomorphisms; so 3 in dimension 2 and 10 in dimension 3.

Note that while proposition 7 guarantees the $\beta_{i,J}$ always suffice for avoiding caustics, we allow ourselves a generic set of coordinates $\tilde{x}^{(1)}, \ldots, \tilde{x}^{(Q)}$ such that no point on $\gamma$ is simultaneously a caustic with respect to all $\tilde{x}^{(j)}$. Proposition 7 then gives us one such coordinate set.

*3.2.3. Splitting the FIO* With the coordinate set chosen, define

$$U_k = \{(x, \xi) \in T^*X : \det \partial y / \partial \tilde{x}^{(k)} \neq 0\}, \tag{49}$$

and choose a partition of unity $\{\alpha_k\}_{k=1}^{Q}$ subordinate to the $U_k$. As before, $A$ can be written as $A_1 + \cdots + A_r$, where $A_k$ is an FIO associated with the canonical transformation $\chi$, with principal symbol $\alpha_k \sigma$. Each $A_k f$ can then be approximated via theorem 6.

Although it is only necessary that $\alpha_k = 0$ when $\det \partial y / \partial \tilde{x}^{(k)} = 0$, it is advantageous numerically for $\alpha_k$ to be zero when $\partial y / \partial \tilde{x}^{(k)}$ is close to singular. Equation (37) shows that $\|\tilde{\mathbf{u}}\|$ can be large when $\partial y / \partial \tilde{x}^{(k)}$ is nearly singular. Large $\tilde{\mathbf{u}}$, in turn, increase $c$ (eq. (28)), which increases the number of significant terms in the PSWF decompositions (eq. (23)). Applying more restrictive cutoffs $\alpha_k$ allows us, potentially, to avoid computations on cones $\Xi_i$ for which $c$ is large, reducing computation time.

One way to choose the $\alpha_k$ to achieve this goal is as follows: compute the minimum singular values $\mu_k(x, \xi)$ of $\partial y / \partial \tilde{x}^{(k)}$, and let $\mu(x, \xi) = \max_k \mu_k(x, \xi)$. Then, choose $\tilde{\alpha}_k \in C^\infty$ such that

$$\tilde{\alpha}_k(x, \xi) = \begin{cases} 0, & \frac{\mu_k}{\mu} < 1 - \tau_w, \\ 1, & \frac{\mu_k}{\mu} > 1 - \beta\tau_w, \end{cases} \tag{50}$$

where $\tau_w, \beta \in (0,1)$ are parameters controlling the tightness of the partition and smoothness of cutoffs respectively. By forcing $\alpha_k = 1$ when $\frac{\mu_k}{\mu}$ lies in a neighborhood of 1, we are assured that $\tilde{\alpha}_k$ is smooth. Finally, define $\alpha_k = \tilde{\alpha}_k/(\tilde{\alpha}_1 + \cdots + \tilde{\alpha}_Q)$ to complete the partition of unity construction.

*3.2.4. Amplitude Calculation* The last major ingredient for the algorithm we have neglected to discuss is how to construct the amplitude $a$ for (7) from the principal symbol. While this is standard FIO theory, special consideration is required to describe the principal symbol. Namely, the principal symbol is defined as a symbol on the half-density bundle on $T^*(X \times Y)$ tensored with the Maslov bundle; in order to use it in an algorithm these two bundles should be trivialized.

For graph FIOs, the half-density bundle has a natural global trivialization [20, following (25.3.2)′] in which the pullback of $|dy|^{1/2}|d\eta|^{1/2}$ by the projection $\pi_Y \colon \Lambda \to T^*Y$ is taken as the unit section. This half-density is invariant under changes of coordinates. Note also $|dy|^{1/2}|d\eta|^{1/2} = |dx|^{1/2}|d\xi|^{1/2}$, since $\Lambda$ is Lagrangian with respect to the sum of the symplectic forms on $T^*X$ and $T^*Y$, and the volume forms are $n^{\text{th}}$ powers of the symplectic forms.

It is also possible to globally trivialize the Maslov bundle [30, lemma 4.1.3], but there is not a natural choice of unit section in general. Several special cases do have natural unit sections, such as when the projection of $\Lambda$ onto $X \times Y$ has constant rank [30, p. 96], which for us is equivalent to $\partial y/\partial \xi$ having constant rank. Other cases having natural global trivializations are when $\chi$ is homotopic to the identity [31, ex. 3.7] or when $\mathrm{WF}'(A) \subseteq T^*X \backslash 0$ is simply-connected [31, p.284]. One example of the latter case is the inversion of the circular Radon transform, where $\mathrm{WF}'(A) = \{(r, s, \rho, \sigma) \mid |\sigma| \leq |\rho|,\, r > 0\}$ is the disjoint union of two simply-connected regions.

If we allow for discontinuous sections of the Maslov bundle, however, natural global trivializations are possible. Following Hörmander [20, before (25.1.9)], suppose two local representations of a Fourier integral distribution $u$ in local coordinates $x^{(1)}$, $x^{(2)} \in \mathbb{R}^n$ are given,

$$u(x^{(1)}) = (2\pi)^{-(n+2N_1)/4} \int_{\mathbb{R}^{N_1}} e^{i\phi_1(x^{(1)},\theta_1)} a_1(x^{(1)}, \theta_1)\, d\theta_1, \tag{51}$$

$$u(x^{(2)}) = (2\pi)^{-(n+2N_2)/4} \int_{\mathbb{R}^{N_2}} e^{i\phi_2(x^{(2)},\theta_2)} a_2(x^{(2)}, \theta_2)\, d\theta_2. \tag{52}$$

If $u$ has order $m$, Hörmander shows

$$a_2 d_{C_2}^{1/2} - e^{\pi i s/4} a_1 d_{C_1}^{1/2} \in S^{m-1}(\Lambda, \Omega_\Lambda^{1/2}), \tag{53}$$

where

$$s = \operatorname{sgn} d^2_{\theta_1}\phi_1(x^{(1)}, \theta_1) - \operatorname{sgn} d^2_{\theta_2}\phi_2(x^{(2)}, \theta_2), \tag{54}$$

and $d_{C_i}$ is the quotient density on $C_i = \ker d_{\theta_i}\phi_i$ defined by $d_{\theta_i}\phi_i$ with respect to the Euclidean volume densities on $\mathbb{R}^d \times \mathbb{R}^{N_i}$. Then, $C_i$ can be identified with $\Lambda$ via the standard isomorphism; similarly, the signature terms of (54) can be interpreted as functions on $\Lambda$.

Recalling our choice of unit section for the half-density bundle, $|dx|^{1/2}|d\xi|^{1/2}$, and dividing by it, we can get an expression for the principal symbol from (53):

$$\sigma = a_i \frac{d_{C_i}^{1/2}}{|dx|^{1/2}|d\xi|^{1/2}} \exp\left[\frac{i\pi}{4} \operatorname{sgn} d^2_{\theta_i}\phi_i\right] \in S^m/S^{m-1}. \tag{55}$$

In our situation, we have two local forms for graph FIOs in different coordinates. Here the $\theta_i$ are the dual variables $\xi^{(i)}$ to the $x^{(i)}$, and $\phi_i(x, y, \xi) = S_i(y, \xi) - x \cdot \xi$. For graph FIOs, Hörmander shows

$$d_{C_i} = |dy|\,|d\xi| = \left|\det \frac{\partial y}{\partial x}\right| |dx|\,|d\xi|\,.$$

Dividing out by the unit section $|dx|^{1/2}|d\xi|^{1/2}$ for $\Lambda$ leaves

$$\sigma = a_i \left|\det \frac{\partial y}{\partial x^{(i)}}\right|^{1/2} \exp\left[\frac{i\pi}{4} \operatorname{sgn} d^2_{\xi_i}\phi_i\right].$$

Finally, since $d_\xi \phi = d_\xi S - x = x(y, \xi) - x$, we have $d^2_\xi \phi = \frac{\partial x}{\partial \xi} = -(\frac{\partial y}{\partial x})^{-1}(\frac{\partial y}{\partial \xi})$. Dropping the subscripts,

$$\sigma = a \left|\det \frac{\partial y}{\partial x}\right|^{1/2} \exp\left[-\frac{i\pi}{4} \operatorname{sgn}\left(\frac{\partial y}{\partial x}\right)^{-1}\left(\frac{\partial y}{\partial \xi}\right)\right]. \tag{56}$$

For $F \in I^m(\chi)$, denote by $\sigma_F$ the principal symbol of $F$ defined by (56).

Note that $\sigma$ is discontinuous unless $\partial y/\partial \xi$ has constant rank. From a numerical standpoint, this would indicate that care should be taken where $\partial y/\partial \xi$ drops rank; in practice, this discontinuous trivialization appears to work satisfactorily. For pseudodifferential operators, $\sigma$ agrees with the usual principal symbol for pseudodifferential operators, pulled back from $\Lambda$ to $T^*X$.

Safarov [31] has developed in much more depth a slightly different discontinuous trivialization of the principal symbol bundle, and proves formulas for the singular principal symbols of products and adjoints of graph FIOs. His singular symbol differs from (56) in that the signature is shifted by the rank of $\partial x/\partial \xi$.§

### *3.3. Forward Algorithm*

Combining everything so far, we can present the final approximation:

$$\begin{aligned} Af(y) \approx \mathcal{A}_\epsilon f(y) = \sum_{k=1}^{Q} \sum_{i=1}^{r_k} & \left(y \circ \beta_k^{-1}(\cdot, \tilde{\nu}_{i,k})\right)_* \\ & \left[\left|\det \frac{\partial y}{\partial \tilde{x}^{(k)}}(\tilde{x}^{(k)}, \tilde{\nu}_{i,k})\right| \cdot \sum_{\mathbf{j} \in \mathbf{J}(c_{i,k},\epsilon)} \lambda_{\mathbf{j}}^{c_{i,k}} \psi_{\mathbf{j}}^{c_{i,k}}(\mathbf{u}_{i,k}(\tilde{x}^{(k)})) \cdot \right. \\ & \left. \mathcal{F}^{-1}_{\tilde{\xi} \to \tilde{x}^{(k)}}\left[\psi_{\mathbf{j}}^{c_{i,k}}(\mathbf{v}_{i,k}(\tilde{\xi})) \cdot |\tilde{\xi}|^m \rho_{i,k}(\tilde{\xi}) \widehat{\beta_k^* f}(\tilde{\xi})\right] \cdot (a_k \circ \beta_k^{-1})(\tilde{x}^{(k)}, \tilde{\nu}_{i,k})\right], \end{aligned} \tag{57}$$

§ Safarov also considers FIOs acting on half-densities, eliminating the need for the half-density factor in (56).

**Pseudocode 1** Approximate $Af$

Precompute $\alpha_k$, $a_k$, $c_{i,k}$, $\mathbf{u}_{i,k}$, and PSWFs.
**for** each coordinate change $k = 1$ **to** $Q$ **do**
  Pull back $f$ by $\beta_k$.
  Zero pad $\beta_k^* f$, and apply DFT.
  **for** each cone $i = 1$ **to** $r_k$ **do**
    **for** each significant PSWF $j \in \mathbf{J}_{c,e}$ **do**
      Multiply $\widehat{\beta_k^* f}$ by $\rho_{i,k}|\tilde{\xi}|^m$ and the PSWF $\psi_j^{c_{i,k}}(\mathbf{v}_{i,k}(\tilde{\xi}))$.
      Apply inverse DFT.
      Multiply by $\lambda_j^{c_{i,k}} \psi_j^{c_{i,k}}(\mathbf{u}_{i,k})$.
      Add to current sum.
    **end for**
    Apply pushforward (with Jacobian factor) by $y \circ \beta_k^{-1}$.
    Add to current sum.
  **end for**
**end for**
Add to current sum.

where

$$a_k = \alpha_k \sigma_k \left( \det \frac{\partial y}{\partial x} \right)^{-1/2} \exp \left[ \frac{i\pi}{4} \operatorname{sgn} \left( \frac{\partial y}{\partial x} \right)^{-1} \left( \frac{\partial y}{\partial \xi} \right) \right]. \tag{58}$$

Here $k$ indexes the choices of coordinates, $i$ the cones, and $\mathbf{j}$ the PSWFs. The pseudocode listing 1 indicates how the global approximation formula translates to an algorithm.

**Theorem 8.** *Suppose the values $f$ are given on an $L \times \cdots \times L$ grid in $X$, and the values of $Af$ are to be calculated on an $M \times \cdots \times M$ grid in $Y$. Assume the bandwidths $c_{i,k}$, the PSWFs $\psi_{\mathbf{j}}^{(c_{i,k})}$, and eigenvalues $\lambda_{\mathbf{j}}^{(c_{i,k})}$ are precomputed.*

*Computing $\mathcal{A}_\epsilon f$ using approximation (57) and DFTs requires worst-case time*

$$O(L^{(3n-1)/2} \log L \log^{n(n-1)/2} \epsilon + L^{(n-1)/2} M^n). \tag{59}$$

*Proof.* After the precomputation steps, the global algorithm reduces to applying the local algorithm $Q = O(1)$ times. Assume each pullback $\beta_k^* f$ is computed on a grid of size $O(L)$. In the local algorithm, the bulk of the computational work comes from computing the PSWF contributions and the pushforwards. For each of the $O(L^{(n-1)/2})$ cones, there are $\mathcal{J}(c, \epsilon) = O(\log^{n(n-1)/2})$ significant PSWF terms, each requiring $O(L^n \log L)$ work (pointwise multiplications and an inverse DFT). The pushforward step in each cone requires $O(L^n + \kappa M^n)$ time, where $\kappa$ is a constant depending only on the FIO; specifically, $\kappa$ is the maximum number of preimages of a point in $Y$ under $\tilde{x}^{(i)} \mapsto y(\tilde{x}^{(i)}, \tilde{\xi}^{(i)})$ for each $i$ and all $\tilde{\xi}^{(i)}$. □

### *3.4. Inverse and Transpose*

This section develops accompanying algorithms for computing $A^\top g$ and $A^{-1}g$ (assuming $A$ is elliptic) given exactly the same information as before: the canonical transformation $\chi$ and the principal symbol $\sigma$ of $A$. Theoretically such algorithms are superfluous, of course, since $A^{-1}$ and $A^\top$ are graph FIOs associated with the inverse canonical transformation $\chi^{-1}$.

However, computing $\chi^{-1}$ may be difficult or inefficient, and if $\chi^{-1}$ is not a graph, it must be decomposed as a union of graphs. Moreover, the algorithms we develop come essentially for free: $A^{-1}$ and $A^\top$ can be computed with the same information and precomputation steps used previously to compute $A$. Throughout $A$ is as before, and $g = g(y)$ an arbitrary distribution on $Y$.

*3.4.1. Transpose* Recall that the overall structure of the local algorithm for $Af$ involved writing $A$ as a sum of ΨDO's with coordinate changes. Originally, from (14) we have $Af(y) = \sum_{i=1}^r (B_i f_i)(x(y,\nu_i))$. Later, we approximated the ΨDO $B_i$ by $\overline{\overline{B}}_i$, which had a convenient separated representation, and interpreted $x(y,\nu_i)$ as a pushforward by $x \mapsto y(x,\nu_i)$, along with a Jacobian factor. Hence the local forward algorithm can be expressed as:

$$Af(y) \approx \sum_{i=1}^{r} y(\cdot,\nu_i)_* \left[ (\overline{\overline{B}}_i f_i) \cdot \det \frac{\partial y}{\partial x} \right]. \tag{60}$$

The global algorithm is similar, but with coordinate changes $\beta_k \colon x \mapsto \tilde{x}^{(k)}$, and a microlocal partition of unity $\alpha_k \colon T^*X \to [0,1]$. We can express it as

$$Af(y) \approx \sum_{k=1}^{Q} \sum_{i=1}^{r_k} y \circ \beta_k^{-1}(\cdot,\tilde{\nu}_{i,k})_* \left[ \overline{\overline{B}}_{i,k}((\beta_k^{-1})^* f) \cdot \det \frac{\partial y}{\partial \tilde{x}^{(k)}} \right]. \tag{61}$$

where $\overline{\overline{B}}_{i,k}$ is the same as $\overline{\overline{B}}_i$ in $(\tilde{x}^{(k)}, \tilde{\xi}^{(k)})$ coordinates, except that the symbol is multiplied by the microlocal partition of unity, $\alpha_k$. Now we transpose; if $f \in C_0^\infty(X)$, $g \in C_0^\infty(Y)$, then

$$\langle Af, g \rangle \approx \sum_{k=1}^{Q} \sum_{i=1}^{r_k} \left\langle y \circ \beta_k^{-1}(\cdot,\tilde{\nu}_{i,k})_* \left[ \overline{\overline{B}}_{i,k}((\beta_k^{-1})^* f) \cdot \det \frac{\partial y}{\partial \tilde{x}^{(k)}} \right], g \right\rangle \tag{62}$$

$$= \sum_{k=1}^{Q} \sum_{i=1}^{r_k} \left\langle \overline{\overline{B}}_{i,k}((\beta_k^{-1})^* f) \cdot \det \frac{\partial y}{\partial \tilde{x}^{(k)}}, y \circ \beta_k^{-1}(\cdot,\tilde{\nu}_{i,k})^* g \right\rangle \tag{63}$$

$$= \sum_{k=1}^{Q} \sum_{i=1}^{r_k} \left\langle f, (\beta_k^{-1})_* \overline{\overline{B}}_{i,k}^\top \left( \det \frac{\partial y}{\partial \tilde{x}^{(k)}} \right) y \circ \beta_k^{-1}(\cdot,\tilde{\nu}_{i,k})^* g \right\rangle, \tag{64}$$

so

$$A^\top g \approx \sum_{k=1}^{Q} \sum_{i=1}^{r_k} (\beta_k^{-1})_* \overline{\overline{B}}_{i,k}^\top \left( \det \frac{\partial y}{\partial \tilde{x}^{(k)}} \right) y \circ \beta_k^{-1}(\cdot,\tilde{\nu}_{i,k})^* g. \tag{65}$$

Recalling the separated representation of the symbol $\overline{\overline{b}}_{i,k}$ (the superscripts on $\tilde{x}$ and $\tilde{\xi}$ are omitted for clarity),

$$\overline{\overline{b}}_{i,k} = \sum_{\mathbf{j} \in \mathbb{N}^N} (\rho_{i,k} a_k)(\tilde{x}, \tilde{\nu}_{i,k}) \lambda_{\mathbf{j}} \psi_{\mathbf{j}}^{c_{i,k}}(\mathbf{u}(\tilde{x})) \psi_{\mathbf{j}}^{c_{i,k}}(\mathbf{v}(\tilde{\xi})), \tag{66}$$

we have

$$\overline{\overline{B}}_{i,k} = \sum_{\mathbf{j}\in\mathbb{N}^N} (\rho_{i,k}a_k)(\tilde{x},\tilde{\nu}_{i,k})\lambda_{\mathbf{j}}\psi_{\mathbf{j}}^{c_{i,k}}(\mathbf{u}(\tilde{x}^{(k)}))\circ\mathcal{F}^{-1}\circ\psi_{\mathbf{j}}^{c_{i,k}}(\mathbf{v}(\tilde{\xi}^{(k)}))\circ\mathcal{F}, \tag{67}$$

where the first and third operators in the composition are multiplication operators. Taking the transpose,

$$\overline{\overline{B}}_{i,k}^{\top} = \sum_{\mathbf{j}\in\mathbb{N}^N} \mathcal{F}^{-1}\circ\psi_{\mathbf{j}}^{c_{i,k}}(\mathbf{v}(\tilde{\xi}^{(k)}))\circ\mathcal{F}\circ(\rho_{i,k}a_k)(\tilde{x},\tilde{\nu}_{i,k})\lambda_{\mathbf{j}}\psi_{\mathbf{j}}^{c_{i,k}}(\mathbf{u}(\tilde{x}^{(k)})). \tag{68}$$

Finally, we can truncate the infinite sum in (68) as before, and equations (65) and (68) then give us an formula for the transpose operator analogous to our earlier formula (57) for $A$.

**Remark.** As an alternative approach, we can develop this representation by writing $A^{\top}g$ locally as an oscillatory integral, similar to (7) but with the roles of the domain and codomain variables switched:

$$(A^{\top}g)(x) = \int e^{i(S(y,\xi)-x\cdot\xi)}a^{\top}(y,\xi)g(y)\,dy\,d\xi. \tag{69}$$

where $a^{\top}$ is an amplitude function. Separating out the first term in the Taylor expansion of the phase as before, we can write $A^*$ locally as a sum of changes of coordinates and pseudodifferential operators, then perform a second-order Taylor approximation of the phase and expand the exponential using PSWFs. This leads once more to formulas (65) and (68).

*3.4.2. Inverse* The major application of the transpose algorithm is computing the inverse of an elliptic Fourier integral operator. Since $(A^{-1})^{\top}$ is an FIO with canonical transformation $\chi$, all that is left is to compute its principal symbol. The rough intuition is that the principal symbol of $A^{-1}$ should be the reciprocal of that of $A$, and the principal symbols of $(A^{-1})^{\top}$ and $A^{-1}$ should coincide. With the trivializations of the half-density and Maslov bundles utilized here, this is exactly true.

**Proposition 9.** *If $A \in I^m(\chi)$ is elliptic, $\sigma_{A^{-1}} = \sigma_A^{-1}$.*

**Lemma 10.** *For any $A \in I^m(\chi)$, the principal symbols of $A$ and $A^{\top}$ are identical.*

Hence, the inverse can be computed using the transpose algorithm applied to $(A^{-1})^{\top}$. The proofs are deferred to section 5.

## 4. Examples

In this section, we compute several FIOs using a MATLAB/C++ implementation of the algorithm described in section 3.

### *4.1. Radon transform*

The 2D Radon transform is a prototypical FIO. With respect to standard coordinates, every point is a caustic, so the use of coordinate changes is essential.

Let $\theta$ be the angular coordinate on $\mathbb{S}^1$, and let $\hat{\theta} = (\cos\theta, \sin\theta) \in \mathbb{R}^2$. Then

$$Rf(s,\theta) = \int_{x\cdot\hat{\theta}=s} f(x)\, d\ell(x) = \frac{1}{2\pi}\int_{\mathbb{R}^2}\int e^{i\varsigma(x\cdot\hat{\theta}-s)} f(x)\, d\varsigma\, dx. \tag{70}$$

Letting $\varsigma, \vartheta, \xi$ be the dual variables to $s, \theta, x$, the canonical relation of $R$ is

$$\Lambda' = \left\{ (s,\theta,\varsigma,\vartheta;x,\xi) \;\middle|\; s = \pm x\cdot\frac{\xi}{|\xi|},\ \hat{\theta} = \pm\frac{\xi}{|\xi|},\ \varsigma = \pm|\xi|\,,\ \vartheta = \mp x\cdot\xi^{\perp} \right\}. \tag{71}$$

$\Lambda'$ is a 1–to–2 map, equal to the disjoint union of two canonical graphs $\Lambda_\pm$, one for each choice of sign in (71). Hence $R$ is the sum of two FIOs $R_\pm$ associated with the two canonical graph, with principal symbols equal to the restrictions of $\sigma_R$ to $\Lambda_\pm$.

The canonical relation (71) shows that $\operatorname{rank}\partial(s,\theta)/\partial x = 1$, so that every point on $\Lambda_\pm$ is a caustic.

*Principal Symbol* Writing $R$ in the standard form (52) with $\theta = \sigma$, we have $n = 4$, $N = 1$, $\phi = \varsigma(x\cdot\hat{\theta} - s)$, so

$$Rf(s,\theta) = \frac{1}{2\pi}\int_{\mathbb{R}^2}\int_{\mathbb{R}} e^{i\varsigma(x\cdot\hat{\theta}-s)} f(x)\, d\varsigma\, dx \tag{72}$$

$$= \frac{1}{(2\pi)^{3/2}}\int_{\mathbb{R}^2}\int_{\mathbb{R}} e^{i\varsigma(x\cdot\hat{\theta}-s)}\sqrt{2\pi} f(x)\, d\varsigma\, dx \tag{73}$$

The principal symbol is by (55)

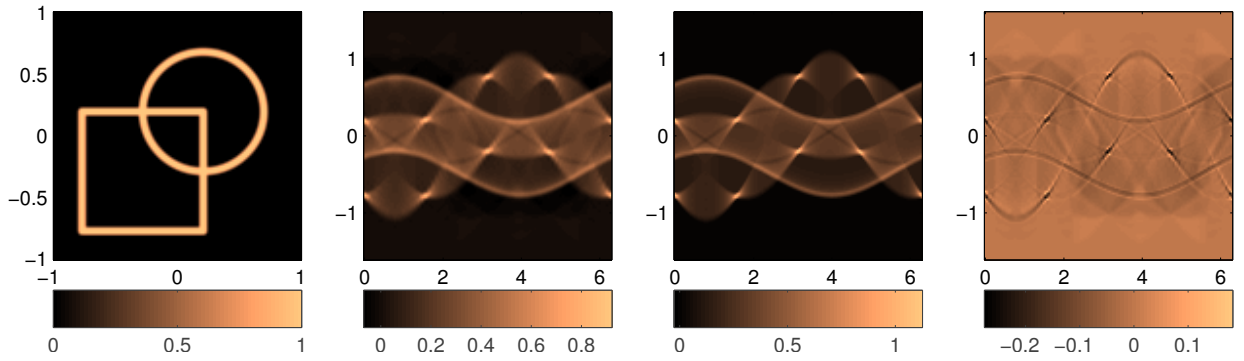
$$\sigma = \sqrt{2\pi}\left(\frac{d_C}{|dx||d\xi|}\right)^{1/2} \exp\left[\frac{i\pi}{4}\operatorname{sgn} d_\varsigma^2\phi\right]. \tag{74}$$

Since $d_\varsigma^2\phi = 0$, the Maslov factor is 1. All that remains to compute $d_C$, the quotient volume form defined by $d_\varsigma\phi$. Now $d_\varsigma\phi$ maps the vector field $\frac{\partial}{\partial s}$ on $X\times Y\times\mathbb{R} = \{(x,s,\theta,\varsigma)\}$ to a unit-volume basis in its image, so since $|ds||d\theta||dx||d\varsigma|$ is a volume form on $C$, we get $d_C = |d\theta||dx||d\varsigma|$. We push $d_C$ forward to $\Lambda'$, where $\theta, \varsigma$ form a polar coordinate representation for $\xi$ (up to sign). Hence $|d\xi| = |\varsigma|\cdot|d\theta||d\varsigma|$, giving us

$$d_C = |d\theta|\,|dx|\,|d\varsigma| = \frac{|dx||d\xi|}{|\xi|}, \tag{75}$$

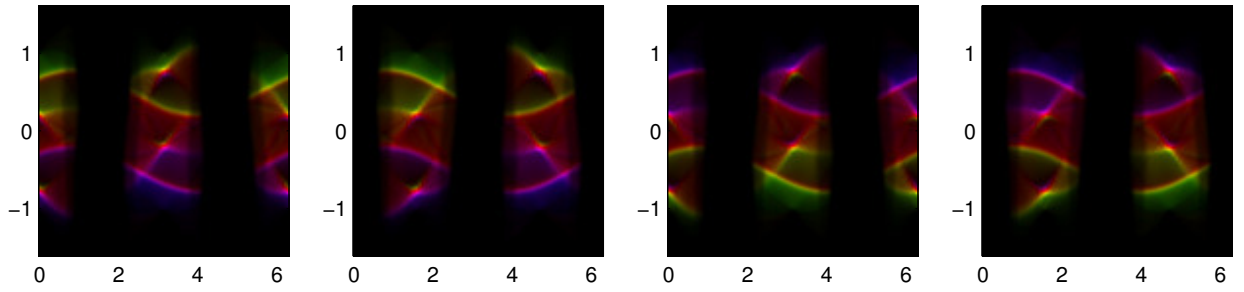
and finally

$$\sigma_R = \sqrt{2\pi}\left(\frac{d_C}{|dx||d\xi|}\right)^{1/2} = \sqrt{2\pi}\cdot|\xi|^{\,-1/2}. \tag{76}$$



(a) $f$ (input) (b) Computed $\mathcal{A}_\epsilon f$ (c) Reference $Rf$ (d) Error, $\mathcal{A}_\epsilon f - Rf$

Figure 1: Comparison of generic FIO algorithm with reference Radon algorithm. $\theta$ and $s$ are displayed on the $x$ and $y$ axes, respectively.



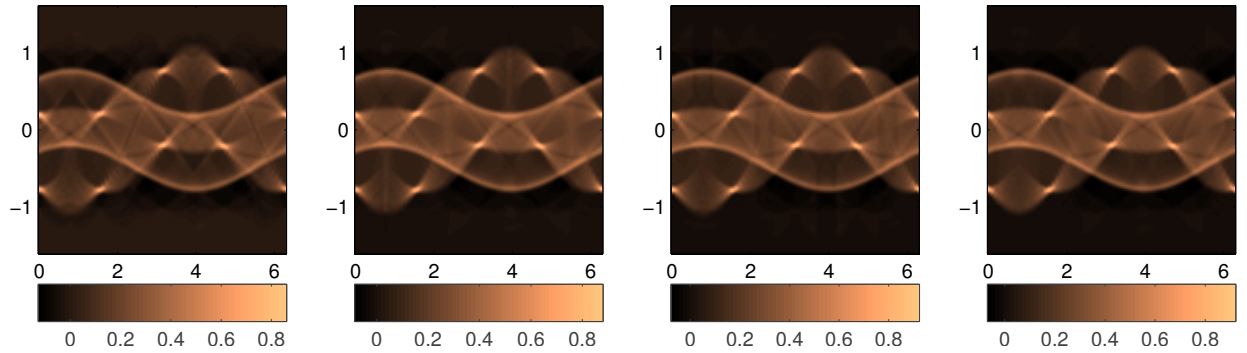
(a) $R_+$, $x_1 \mapsto x_1 + \frac{t}{2}x_2^2$ (b) $R_+$, $x_2 \mapsto x_2 + \frac{t}{2}x_1^2$ (c) $R_-$, $x_1 \mapsto x_1 + \frac{t}{2}x_2^2$ (d) $R_-$, $x_2 \mapsto x_2 + \frac{t}{2}x_1^2$

Figure 2: Contributions to $Rf$ from each local integral representation. Color and brightness represent complex phase and magnitude, respectively.

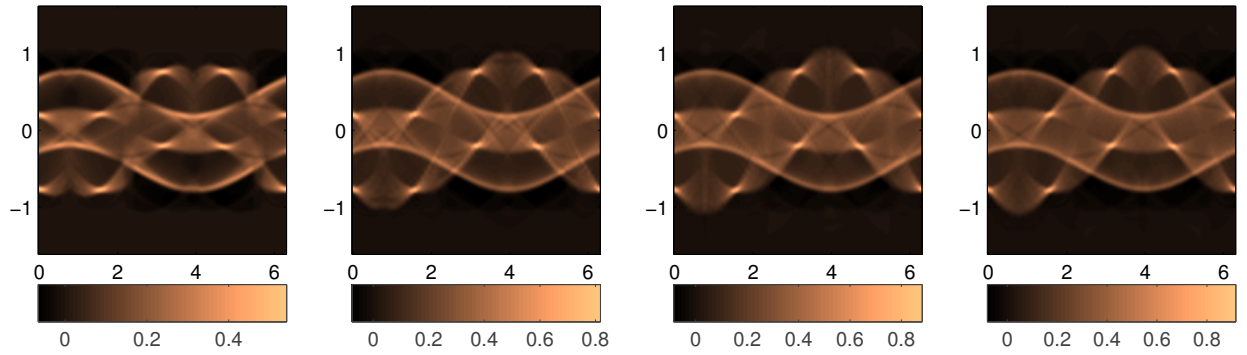
*Example Output* Figure 1 compares the Radon transform computed by the new algorithm with a naïve $O(N^3)$ reference implementation (fast $O(N^2 \log N)$ algorithms for computing the Radon transform are also available).

To handle caustics in $R = R_+ + R_-$, two changes of coordinates, $\beta_{1,(2)}\colon x_1 \mapsto x_1 + \frac{t}{2}x_2^2$ and $\beta_{2,(1)}\colon x_2 \mapsto x_2 + \frac{t}{2}x_1^2$ are used (see (48)). Since $\det(\frac{dy}{dx}) \equiv 0$, we do not use the original coordinates. Figure 2 shows the contributions to $R_+f$ and $R_-f$ from each choice of coordinates.

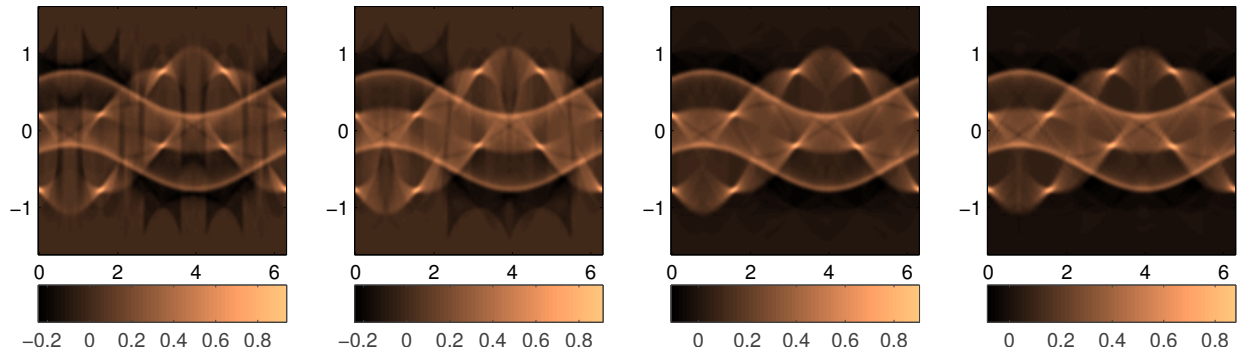
There are a number of parameters in the algorithm. Figures 3 and 4 compares the results with varying zero-padding factor, number of frequency cones, and amount of distortion in the coordinate changes (choice of $t$ in (48)). Increasing the number of frequency cones decreases the error in the approximation of the amplitude and phase. A larger zero-padding factor in the DFT decreases periodicity artifacts, which in this example appear as smooth errors outside the support of the true $Rf$. Finally, a larger distortion parameter $t$ creates more diverse coordinate choices, potentially improving caustic avoidance. In all cases, these benefits typically require increased runtime.



(a) Zero-padding factor 1 (b) Zero-padding factor 2 (c) Zero-padding factor 3 (d) Zero-padding factor 4

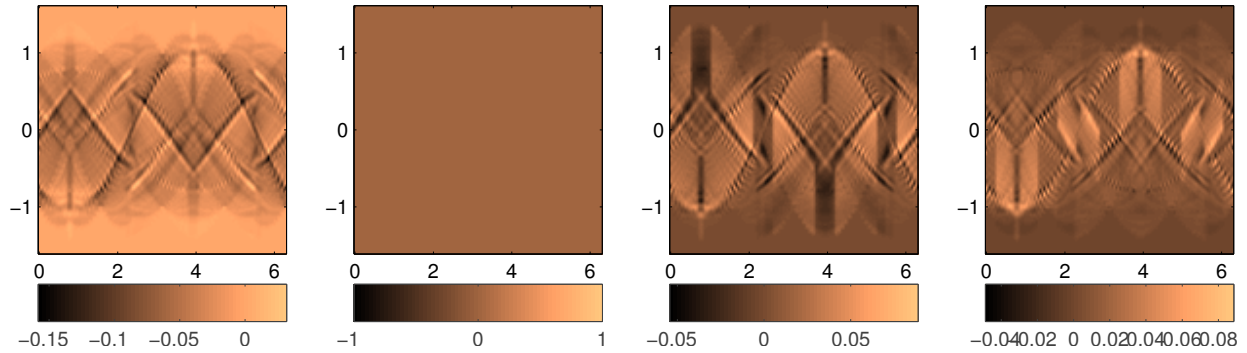


(e) 24 cones (f) 56 cones (g) 112 cones (h) 448 cones

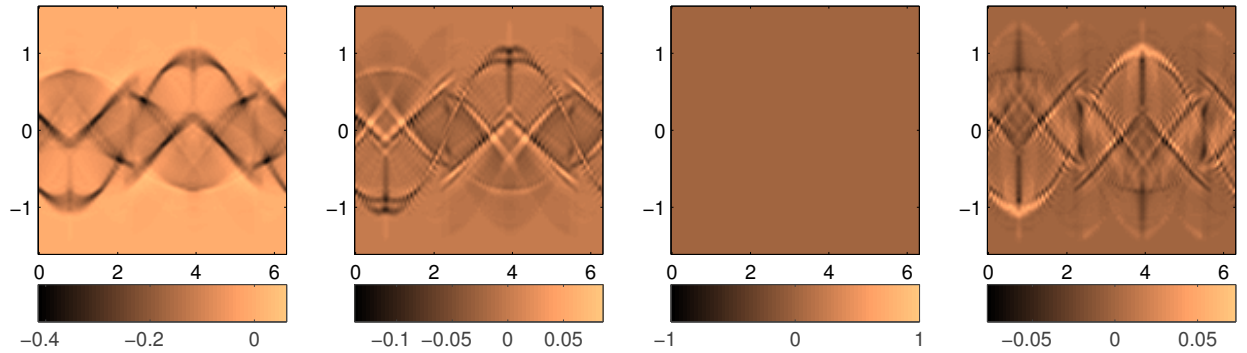


(i) $t = 0.25$ (j) $t = 0.5$ (k) $t = 2$ (l) $t = 3$

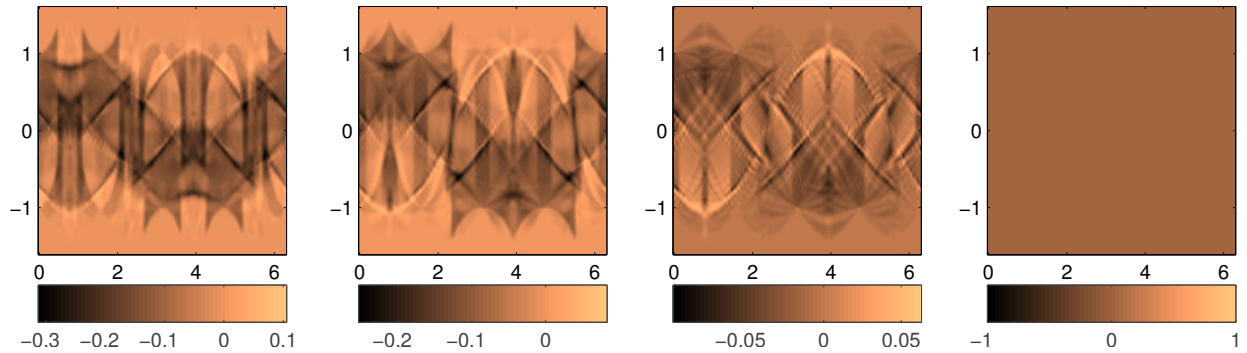
Figure 3: Adjusting algorithm parameters. Top row: zero-padding factor for DFT; middle row: number of cones; bottom row: bending coefficient for diffeomorphisms. The base parameters were a zero-padding factor of 2, $t = 3$, and 112 cones.



(a) Zero-padding factor 1 (b) Zero-padding factor 2 (c) Zero-padding factor 3 (d) Zero-padding factor 4



(e) 24 cones (f) 56 cones (g) 112 cones (h) 448 cones



(i) $t = 0.25$ (j) $t = 0.5$ (k) $t = 2$ (l) $t = 3$

Figure 4: Relative errors in Figure 3, compared to $\mathcal{A}_\epsilon$ computed for the base parameters: a zero-padding factor of 2, 112 cones, and $t = 3$. Top row: zero-padding factor for DFT; middle row: number of cones; bottom row: bending coefficient for diffeomorphisms.

### *4.2. Wave propagation*

The initial value problem for the scalar wave equation with variable speed is a prime example of caustic formation:

$$\begin{cases} (\partial_t^2 - c^2\Delta)u &= 0, \qquad (x,t) \in X \times \mathbb{R}_+, \\ u(x,0) &= g(x), \\ u_t(x,0) &= h(x). \end{cases} \tag{77}$$

In this section, we will apply the algorithm to a propagator operator for a similar equation.

It is well-known that this solution operator is a sum of two FIOs, and that the singularities of the initial data $f$ follow the bicharacteristics of the wave operator $\Box_c = \partial_t^2 - c^2\Delta$. Furthermore, it is well-known that the bicharacteristics project to geodesics in $X$, and that a singularity at $(x,\xi) \in T^*X$ in the Cauchy data $(g,h)$ propagates to two singularities in $u(\cdot,t)$ at $(\gamma_{x,\xi}(\pm t), \gamma'_{x,\xi}(\pm t))$, where $\gamma_{x,\xi}(s)$ is the unit-speed geodesic with $\gamma_{x,\xi}(0) = x$, $\gamma'_{x,\xi}(0) = \xi/|\xi|$.

Factoring $\partial_t^2 - c^2\Delta$ as $(\partial_t - iP)(\partial_t + iP)$ [32], where $P = \sqrt{-c^2\Delta}$ is a pseudodifferential square root, allows us to write any solution to (77) as a sum of solutions to the two half-wave equations

$$\begin{cases} (\partial_t \pm iP)u &= 0, \qquad (x,t) \in X \times \mathbb{R}_+, \\ u(x,0) &= f_\pm(x). \end{cases} \tag{78}$$

for suitable $f_\pm$ depending on $g, h$. The solution operator for (78) can be written as $S_\pm = e^{\mp itP}$. Restricting $S_+$ to $t = T$ gives the operator $S_T = e^{-iTP}$; this is what we will compute.

By propagation of singularities, as previously mentioned, $S_T$ has canonical relation

$$\chi(x,\xi) = \left(\gamma_{x,\xi}(T), \gamma'_{x,\xi}(T)\right). \tag{79}$$

For the principal symbol, since $P$ has a real symbol, it is self-adjoint, as $S_T$ is unitary: $S_T^* S_T = e^{-iTP^*} e^{iTP} = I$. By the principal symbol calculus, $\sigma_{S_T^* S_T} = |\sigma_{S_T}|^2$, so $|\sigma_{S_T}| = 1$. When $T$ is small enough that there are no caustics, we can solve (78) using a Lax parametrix construction, taking $S_+$ of the form

$$(S_+ f)(t,x) = \int e^{i\phi(x,t,\xi)} a(x,t,\xi) \hat{f}(\xi)\, d\xi. \tag{80}$$

It is well known that $\phi$ satisfies an eikonal equation, and writing $a$ in an asymptotic expansion $a \sim \sum a_{-j}$, where $a_{-j} \in S^{-j}$, gives transport equations for each term $a_{-j}$. The equation for $a_0$ ensures that $a_0$ is real as long as $\phi$ is defined (i.e. before the formation of caustics). Evaluating at $t = T$ implies that the principal symbol of $S_T$ without the Maslov factor, $\tilde{\sigma}_{S_T} = a d_C^{1/2}/(|dx|^{1/2}|d\xi|^{1/2})$, is real. Further, since $|\sigma_{S_T}| = 1$ and $\tilde{\sigma}_{S_0} = 1$, we must have $\tilde{\sigma}_{S_T} = 1$ for all $T$ by continuity in $t$.

The Maslov factor arising from the Lax parametrix is $-\operatorname{sgn}(\partial y_T/\partial x)^{-1}(\partial y_T/\partial \xi)$, where $y_T(x,\xi)$ is the $y$-component of the canonical transformation at $(x,\xi)$. For $T$ sufficiently small, we have $\operatorname{sgn}(\partial y_T/\partial x)^{-1}(\partial y_T/\partial \xi) = \operatorname{sgn}(\partial y_0/\partial x)^{-1}(\partial y_0/\partial \xi) = 1$ by continuity, giving a principal symbol of $\sigma_{S_T} = \exp[-i\pi/4]$. For larger $T$, we can find the principal symbol using Proposition 13.

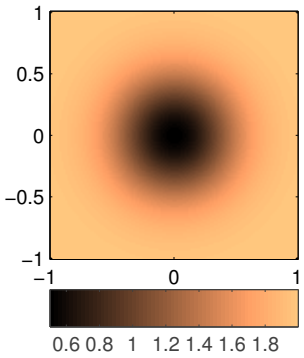



Figure 5: Lensing wave speed for examples.

*Results* To implement $S_T$, the canonical transformation $\chi$ was computed by solving the geodesic equation with RK4 time stepping.

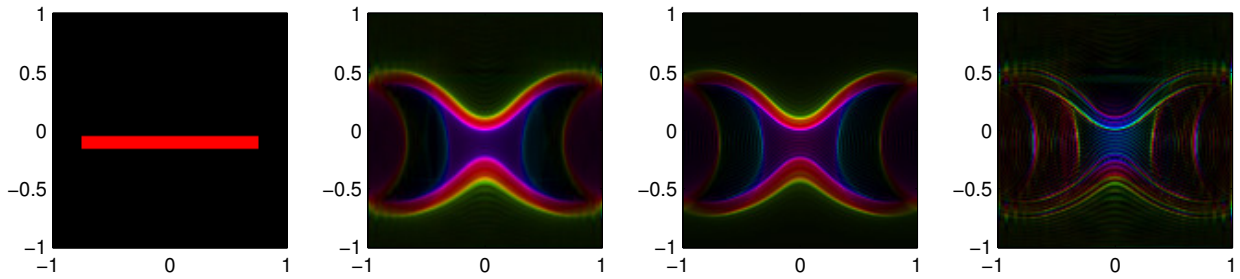
Figures 6 to 10 compares the results of the algorithm with a reference algorithm that computes $S_T$ using a pseudospectral method in space and RK4 in time, approximating $P$ by its leading order term $c|D|$. The sound speed has a Gaussian low-speed lens, as in [18], centered at the origin (Figure 5).

In Figure 6, solutions are calculated at $T = 0.3$, where no caustics have yet formed. The new algorithm compares favorably with the reference algorithm, with fewer high-frequency artifacts, and the complex phases are also correct.

**Remark.** Note that the FIO approach handles infinite domains naturally, without the need for absorbing boundary conditions (Figure 7). The pushforwards by the canonical transformation simply map escaping wavefronts outside the domain. Periodic domains may also be obtained by treating the codomain of the pushforwards as periodic, and performing no zero-padding (Figure 6c). Dirichlet or Neumann boundary conditions on arbitrary domains are likewise possible; the canonical transformation would then be modified so as to reflect wavefronts. This flexibility is a definite advantage of the FIO approach.

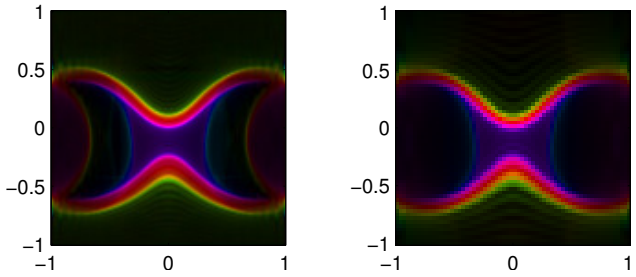
We now increase the time to $T = 0.6$, after the formation of caustics (Figure 8). Five coordinate choices were used here: $x_1 \mapsto x_1 \pm \frac{1}{2}x_2^2$ and $x_2 \mapsto x_2 \pm \frac{1}{2}x_1^2$, together with the identity. Figure 9 shows the contributions to $S_T f$ arising from each coordinate choice. These contributions are computed automatically. As an algorithm for finding the correct Maslov factor is not yet implemented, a phase difference is visible in the caustic in the error plot (d). However, the wavefronts are reconstructed correctly.

Figure 10 illustrates the results of a set of coordinates that does not cover all caustics. In this example, just two coordinate changes were used: $x_1 \mapsto x_1 + \frac{1}{2}x_2^2$ and $x_2 \mapsto x_2 + \frac{1}{2}x_1^2$. The result is an incomplete wavefront set (Figure 10(b)).



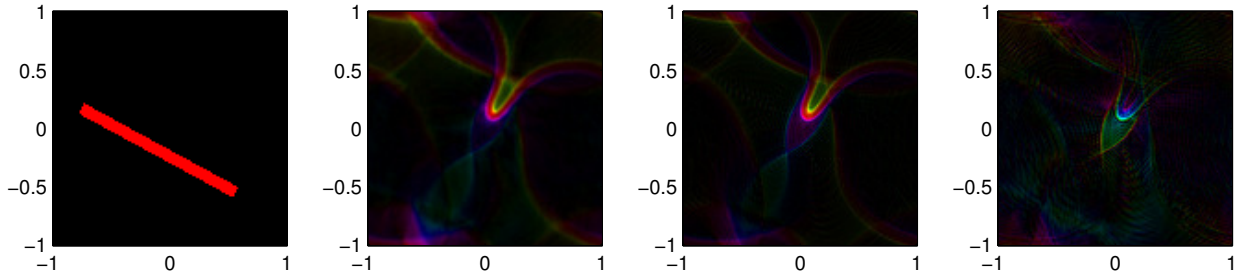
(a) $f$ (input) (b) Computed $\mathcal{A}_\epsilon f$ (c) Reference $S_T f$ (d) Error, $\mathcal{A}_\epsilon f - Rf$

Figure 6: Comparison of generic FIO algorithm with reference half-wave propagator algorithm, no caustics. Color represents the phase and brightness the magnitude of these complex-valued functions.



(a) $\mathcal{A}_\epsilon f$, periodic grid (b) $\mathcal{A}_\epsilon f$, infinite grid

Figure 7: Generic FIO algorithm with periodic and infinite grids. Absorbing boundary conditions are not required to simulate an infinite grid.



(a) $f$ (input) (b) Computed $\mathcal{A}_\epsilon f$ (c) Reference $S_T f$ (d) Error, $\mathcal{A}_\epsilon f - Rf$

Figure 8: Comparison of generic FIO algorithm with reference algorithm, caustics present.

### *4.3. SAR Reflection Operator*

In synthetic aperture radar (SAR) imaging, a plane or satellite flies along a path $\gamma$, emitting radiation and measuring the echoes returned from the ground. Under a set of simplifying assumptions (flat terrain and an undirected beam), the forward (imaging) operator can be

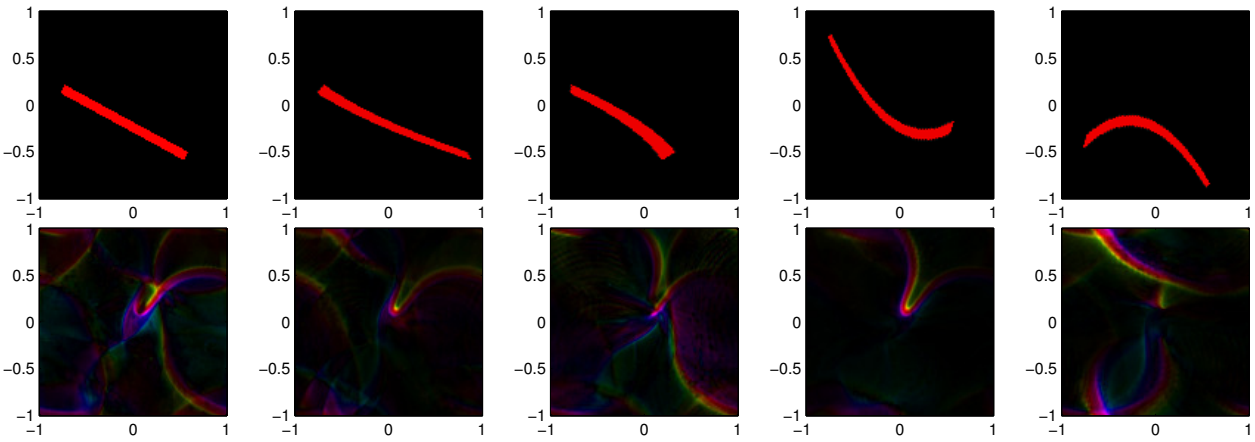
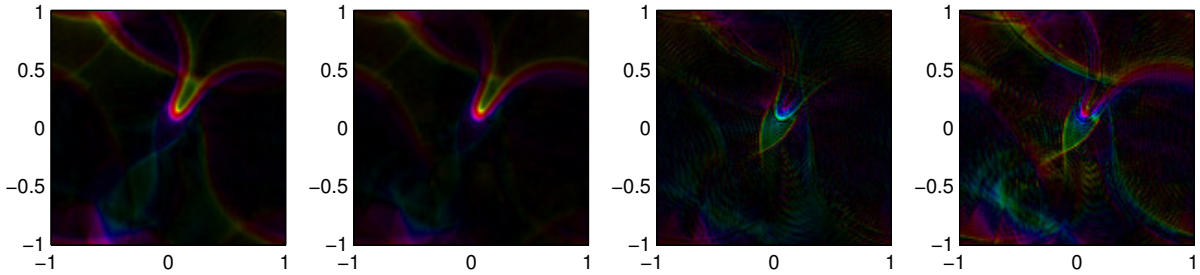


Figure 9: Coordinate changes and contributions of each to the computed $\mathcal{A}_\epsilon f$ for the setup of Figure 8.



(a) $\mathcal{A}_\epsilon f$, $Q = 5$ (b) $\mathcal{A}_\epsilon f$, $Q = 2$ (c) Error, $Q = 5$ (d) Error, $Q = 2$

Figure 10: Insufficient coordinate changes: (a) five coordinate choices, a complete set; (b) two coordinate choices, showing an incomplete wavefront set; (c) and (d) errors of each.

modeled by the circular Radon transform $R_\gamma$:

$$R_\gamma f(r, s) = \int_{|x-\gamma(s)|=r} f(x)\, d\ell. \tag{81}$$

Stefanov and Uhlmann [33] showed that $R_\gamma$ is microlocally noninjective: after localizing $\gamma$ near some $s_0$ and restricting the domain to a sufficiently small set $X \subseteq \mathbb{R}^2 \setminus \gamma$, there is an FIO $U$ such that $R_\gamma \circ U = R_\gamma$ modulo smoothing operators. Intuitively, $U$ may be thought of as reflection across the curve $\gamma$. The canonical relation of this operator is

$$\begin{aligned}\Lambda' = \Big\{(x', \xi', x, \xi) \;\Big|\; \exists s, t\colon\quad & x + t\xi = \gamma(s) = x' + t\xi', \\ & \dot\gamma(s) \cdot \xi = \dot\gamma(s) \cdot \xi',\ \dot\gamma(s)^\perp \cdot \xi = -\dot\gamma(s)^\perp \cdot \xi'\Big\}.\end{aligned} \tag{82}$$

In other words, $\Lambda'$ is the graph of the map which sends singularities $(x, \xi)$ to their mirror images (about the tangent line to $\gamma$) on the opposite side of $\gamma$ (Figure 11). [34] gives a backpropagation algorithm for approximating $U$ and shows that $U$ has principal symbol 1, modulo a Maslov factor. The precise Maslov factor is given by the following proposition.

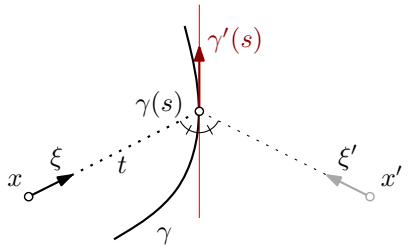



Figure 11: Canonical relation of $U$: reflection across a curve

**Proposition 11.** *The principal symbols of $R_\gamma$ and $U$ are:*

$$\sigma_{R_\gamma} = \sqrt{\frac{2\pi r}{|\xi||\sin\theta|}}, \tag{83}$$

$$\sigma_U = \exp\left[-\frac{i\pi}{4}\operatorname{sgn}\left(\xi\cdot\gamma''(s)\right)\right], \tag{84}$$

*where $\theta$ is the angle between $\xi$ and $\gamma'(s)$.*

Geometrically, the Maslov factor in $\sigma_U$ indicates whether $\gamma$ is concave toward or away from $x$ at $\gamma(s)$. The proof is given in section 5.

We now compare the current algorithm with the reference wave-based algorithm described in [34]. To compute the canonical transformation of $U$, we must locate the intersection points for the ray through $(x,\xi)$ with $\gamma$. To do so, $\gamma$ is represented as the zeros of a defining function $\Gamma(x)$, and the intersection points are found using recursive subdivision.

Figure 12 shows a simple example where $\gamma$ is a parabola. The support of $f$ is in the caustic-free region of $\Lambda_U$, so no coordinate changes are required. The new algorithm compares favorably to the reference wave-based algorithm, with a clear wavefront set and no high-frequency artifacts.

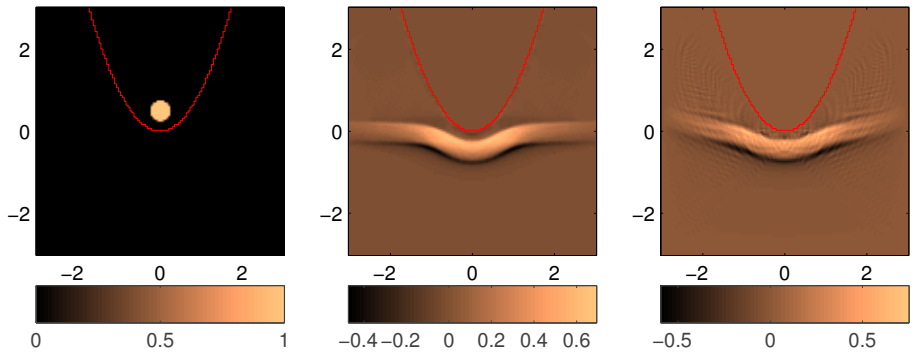
In Figure 13, we compare $R_\gamma f$ with $R_\gamma U f$ for the $Uf$ computed by both algorithms. By definition, $R_\gamma f = R_\gamma U f \bmod C^\infty$ for the correct $Uf$. While the new algorithm produces more smooth error above the support of the true $R_\gamma f$ than the reference algorithm, the wavefront set is better preserved and there is less high-frequency noise. As well, smooth error is inevitable, since Agranovsky and Quinto [35] prove that $R_\gamma$ is injective for curved $\gamma$; hence, $R_\gamma f = R_\gamma U f$ cannot hold exactly.

## 5. Proofs

### *5.1. Sufficiency of Quadratic Coordinate Changes*

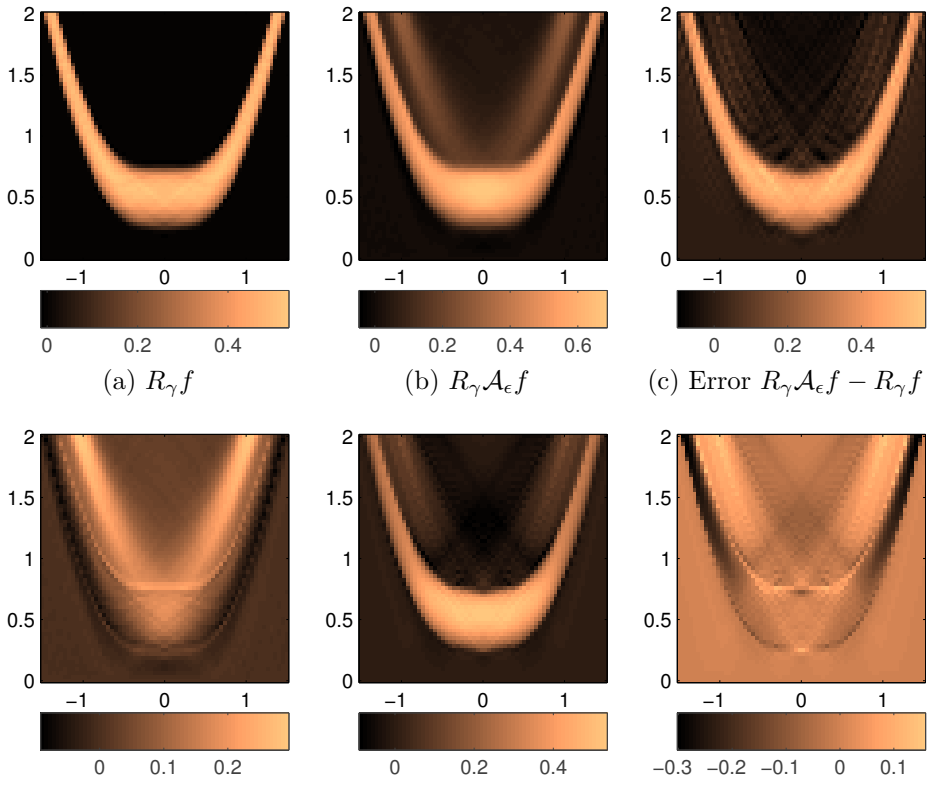
**Proposition 7.** *For all $\gamma_0 \in \Lambda'$, there exist $J \subsetneq \{1,\dots,n\}$ and $i \in \{1,\dots,n\}\setminus J$ such that $\det \partial y/\partial \tilde{x}^{(i,J)}(\gamma_0) \neq 0$, where $\tilde{x}^{(i,J)} = \beta_{i,J}(x)$.*

For the proof of proposition 7, we use the following lemma:




(a) $f$ (input) (b) $\mathcal{A}_\epsilon f$ (c) Reference $Uf$

Figure 12: Comparison of the new algorithm with a reference algorithm for computing the SAR reflection operator $U$.




(a) $R_\gamma f$ (b) $R_\gamma \mathcal{A}_\epsilon f$ (c) Error $R_\gamma \mathcal{A}_\epsilon f - R_\gamma f$

(d) $R_\gamma Uf$ (reference) (e) $R_\gamma Uf$ (reference, fine grid) (f) Error $R_\gamma Uf - R_\gamma f$

Figure 13: Applying the operator $R_\gamma$ to each function in Figure 12, for points on $\gamma$ with $|x| \leq 1.5$. For the exact reflection operator $U$, $R_\gamma f = R_\gamma Uf$ mod $C^\infty$. In (e), $Uf$ is computed using the reference algorithm on a grid half the size and spacing of that used for (c).

**Lemma 12.** *Let $E = T_{\gamma_0}(T^*X \times T^*Y)$, and let $L = T_{\gamma_0}\Lambda'$. There is a Lagrangian subspace $M \subseteq E$ with $L \cap M = \{0\}$ of the form*

$$M_J = \operatorname{span}\left\{ \left\{ \frac{\partial}{\partial x^i} \;\middle|\; i \notin J \right\} \cup \left\{ \frac{\partial}{\partial \xi_j} \;\middle|\; j \in J \right\} \cup \left\{ \frac{\partial}{\partial \eta_1}, \ldots, \frac{\partial}{\partial \eta_n} \right\} \right\}, \quad J \subsetneq \{1, \ldots, n\}. \tag{85}$$

*Proof.* Let $\sigma$ be the symplectic form on $E$, and let $M = \operatorname{span}\{\partial/\partial\eta_1, \ldots, \partial/\partial\eta_n\}$. Since $\Lambda'$ is the graph of $\chi$, the projection $\pi\colon \Lambda' \to T^*X$ is a diffeomorphism, which implies $L \cap M = \{0\}$. Now let us enlarge $M$ in two ways: $M_1 = \operatorname{span}\{M, \frac{\partial}{\partial x^1}\}$ and $M_2 = \operatorname{span}\{M, \frac{\partial}{\partial \xi_1}\}$. Both are isotropic. Suppose $L \cap M_1, L \cap M_2$ are both nontrivial, and pick nonzero $v_1 = m_1 + a\frac{\partial}{\partial x^1}$, $v_2 = m_2 + b\frac{\partial}{\partial \xi_1}$ in $L \cap M_1$, $L \cap M_2$, respectively. Then

$$\sigma(v_1, v_2) = \sigma(m_1, m_2) + a\sigma(\tfrac{\partial}{\partial x^1}, m_2) + b\sigma(m_1, \tfrac{\partial}{\partial \xi_1}) + ab\sigma(\tfrac{\partial}{\partial x^1}, \tfrac{\partial}{\partial \xi_1}) = ab. \tag{86}$$

Since $L$ is Lagrangian, $\sigma(v_1, v_2) = 0$, so one of $a, b$ is zero, and therefore one of $v_1, v_2$ lies in $L \cap M$, a contradiction. Hence at least one of $L \cap M_1$, $L \cap M_2$ is $\{0\}$. Replace $M$ by one of the $M_i$ for which is true. Repeating the process with $\frac{\partial}{\partial x^k}$, $\frac{\partial}{\partial \xi_k}$ for $k = 2, \ldots, n$, we obtain the $M$ described in (85). There is one point to verify: that $J \neq \{1, \ldots, n\}$. Because $\Lambda'$ is conic, the radial vector $\sum(\frac{\partial}{\partial \xi_k}(\xi_0)_k + \frac{\partial}{\partial \eta_k}(\eta_0)_k)$ lies in $L = T_{\gamma_0}\Lambda'$, where $\gamma_0 = (x_0, \xi_0, y_0, \eta_0)$. Hence $M$ cannot contain all the $\frac{\partial}{\partial \xi_k}$, and $J \neq \{1, \ldots, n\}$. □

*Proof of Proposition 7.* By lemma 12, there is some $M$ of the form (85) transversal to $L$. Of all these, choose one where $J$ is minimal; i.e. no $M_{J'}$ with $J' \subsetneq J$ is transversal to $L$. Let $I = \{1, \ldots, n\} \setminus J$. Now $M_J$ is the kernel of $d\pi_J$, where $\pi_J$ is the projection of $T^*X \times T^*Y$ onto $(x_J, \xi_I, y)$. $L \cap M_J = \emptyset$ is equivalent to $d\pi_J|_{\Lambda'}$ being bijective; so $(x_J, \xi_I, y)$ are local coordinates for $\Lambda'$. Since $(x, \xi)$ are also local coordinates, this is equivalent to the non-singularity of $[\frac{\partial y}{\partial x_I}\frac{\partial y}{\partial \xi_J}]$ at $\gamma_0$.

Now we turn to $\frac{\partial y}{\partial \tilde{x}}$ for the $\beta_{i,J}$ map (where $i$ is to be determined). Applying (47) to the $\beta_{i,J}$,

$$\frac{\partial y}{\partial \tilde{x}_k} = \begin{cases} \frac{\partial y}{\partial x_k}, & k \in J \\ \frac{\partial y}{\partial x_k} + \xi_i \frac{\partial y}{\partial \xi_k}, & k \notin J. \end{cases} \tag{87}$$

Choose some $i \in I$ such that $(\xi_0)_i \neq 0$. Such an $i$ exists because otherwise the radial vector $\rho = \sum(\frac{\partial}{\partial \xi_k}(\xi_0)_k + \frac{\partial}{\partial \eta_k}(\eta_0)_k)$ of $T_{\gamma_0}\Lambda'$ would lie in $M_J$. Since $J$ is minimal, for $j \in J$, $\frac{\partial y}{\partial x_j}$ is a linear combination of $\{\frac{\partial y}{\partial x_k} : k \in I\}$. Hence, because $[\frac{\partial y}{\partial x_I}\frac{\partial y}{\partial \xi_J}]$ is nonsingular at $\gamma_0$, so is $\frac{\partial y}{\partial \tilde{x}}$. □

### *5.2. Principal Symbol Calculus*

It is well-known that the principal symbol of the composition of graph FIOs is given by multiplying them, up to factors coming from the Maslov and half-density bundle. The following proposition states how composition works with the trivialization of these bundles we are using.

**Proposition 13.** *Let $A \in I^{m_1}(\chi_1)$, $B \in I^{m_2}(\chi_2)$ be graph FIOs associated with canonical transformations $\chi_1 : T^*X \to T^*Y$, $\chi_2 : T^*Y \to T^*Z$. Then the principal symbol of the composition $BA$ is*

$$\sigma_{BA} = \sigma_A \sigma_B \exp\left[\frac{i\pi}{4}\left(\operatorname{sgn}(Q_A + Q_B) - \operatorname{sgn} Q_A - \operatorname{sgn} Q_B\right)\right], \tag{88}$$

*where $Q_A = -\partial y(x,\eta)/\partial\eta$ and $Q_B = \partial y(z,\eta)/\partial\eta$ for any choice of local coordinates $y$ on $Y$ such that $(x,\eta)$, resp. $(z,\eta)$ are local coordinates on the graph of $\chi_1$, resp. $\chi_2$.*

Note that in (88) $\sigma_B$ is implicitly pulled back to $T^*X$ via $\chi_1$. Much like Safarov's singular symbols [31, Corollary 4.3], we may write the Maslov factor in terms of a Kashiwara index [36]:

**Corollary 14.** *Under the hypotheses of Proposition 13,*

$$\sigma_{BA} = \sigma_A \sigma_B \cdot \exp\left[\frac{i\pi}{4}\kappa(d\chi_1(V_\xi), V_\eta, d\chi_2^{-1}(V_\zeta))\right], \tag{89}$$

*where $\kappa$ is the Kashiwara index, and $V_\xi \in T_\xi T^*X$, $V_\eta \in T_\eta T^*Y$, $V_\zeta \in T_\zeta T^*Z$ are the vertical subspaces over $\xi, \eta, \zeta$ respectively.*

*Proof.* Our starting point is the standard proof of FIO composition [30]. Working microlocally near some $\gamma \in T^*X$, we can write the composition $BA$ as an FIO plus a smoothing operator (which we disregard as usual), after performing a change of coordinates $y \mapsto \mathfrak{y}/H$, $H = |(\theta_1, \theta_2)|$, where the $\theta_i$ are the frequency variables for the phases of $A$ and $B$.

To simplify matters, we have restricted our attention to graph FIOs. Choose local coordinates $x, y, z$ on $X, Y, Z$ respectively, and write $(x(y,\eta), \xi(y,\eta))$ and $(z(y,\eta), \zeta(y,\eta))$ for the coordinate functions of $\chi_1^{-1}$ (defined locally near $\chi_1(x_0, \xi_0)$) and $\chi_2$ respectively. Assuming that $\det \partial x/\partial y(\chi_1(\gamma)) \neq 0$, we have that $(x,\eta)$ are local coordinates on $\Lambda_1$ near $\gamma$. Similarly, if $\det \partial z/\partial y(\chi_1(\gamma)) \neq 0$ then $(z,\eta)$ are local coordinates on $\Lambda_2$ near $\chi_1(\gamma)$. Given coordinates $x, z$ on $X, Z$, we can always choose $y$ generically so that both determinants are nonzero.

Let $S_1(x,\eta) = \eta \cdot y^{(1)}(x,\eta)$, $S_2(z,\eta) = \eta \cdot y^{(2)}(z,\eta)$. Then $\phi_1(x,y,\eta) = y \cdot \eta - S_1(x,\eta)$ is a phase function parameterizing $\Lambda_1$, and likewise $\phi_2(y,z,\eta) = S_2(z,\eta) - y \cdot \eta$ parameterizes $\Lambda_2$. Hence there exist amplitudes $a \in S^{m_1}$, $b \in S^{m_2}$ such that $A$ and $B$ can be written modulo smoothing operators as

$$(Af)(y) = (2\pi)^{-n} \iint \exp\left(i\left[y \cdot \eta - S_1(x,\eta)\right]\right) a(y,\eta) f(x)\, dx\, d\eta, \tag{90}$$

$$(Bg)(z) = (2\pi)^{-n} \iint \exp\left(i\left[S_2(z,\eta) - y \cdot \eta\right]\right) b(y,\eta) g(y)\, dy\, d\eta, \tag{91}$$

for $\mathrm{WF}(f)$, $\mathrm{WF}(g)$ supported in small neighborhoods of $\gamma_0$ and $\chi_1(\gamma_0)$, respectively. Formally composing $A$ and $B$, we proceed as in the standard proof of FIO composition, substituting

$y = \mathfrak{y}/H$ and multiplying the amplitude by a cutoff $\rho(\eta_1, \eta_2)$ near $\{\eta_1 = 0\} \cup \{\eta_2 = 0\}$, yielding

$$\begin{aligned} BAf(z) = (2\pi)^{-2n} \iiiint \exp\left(i\left[S_2(z,\eta_2) - S_1(x,\eta_1) + \mathfrak{y}\cdot(\eta_1-\eta_2)/H\right]\right) \\ \cdot a\Big(\frac{\mathfrak{y}}{H},\eta_1\Big)\, b\Big(\frac{\mathfrak{y}}{H},\eta_2\Big)\, H^{-n} f(x)\, dx\, d\eta_1\, d\mathfrak{y}\, d\eta_2. \end{aligned} \tag{92}$$

The critical set of the phase is

$$C_\phi = \{(x,z,\mathfrak{y},\eta_1,\eta_2) \mid \eta_1 = \eta_2,\ \mathfrak{y}/H = \partial_{\eta_1} S_1 = \partial_{\eta_2} S_2\}\,. \tag{93}$$

On $C_\phi$, we write $\eta = \eta_1 = \eta_2$. To compute the half-density and Maslov factor components of the principal symbol, we need the second derivative matrix $d_{(\mathfrak{y},\theta_1,\theta_2)} d\phi$ on $C_\phi$. Using Lemma 15, we find

$$d_{(\mathfrak{y},\eta_1,\eta_2)} d_{(x,z,\mathfrak{y},\eta_1,\eta_2)}\phi = \begin{bmatrix} 0 & 0 & 0 & \frac{I}{H} & -\frac{I}{H} \\ -\partial_x\partial_{\eta_1} S_1 & 0 & \frac{I}{H} & Q_A - \frac{1}{H^2}(y\eta^\top + \eta y^\top) & 0 \\ 0 & \partial_z\partial_{\eta_2} S_2 & -\frac{I}{H} & 0 & Q_B + \frac{1}{H^2}(y\eta^\top + \eta y^\top) \end{bmatrix}, \tag{94}$$

where $Q_A = -\partial y(z,\eta)/\partial\eta$, $Q_B = \partial y(x,\eta)/\partial\eta$.

*Half-density factor:* To find the half-density factor, we need the quotient volume $d_C$ on $C_\phi$. Since $d_{(\mathfrak{y},\eta_1,\eta_2)}$ maps $(\frac{\partial}{\partial x}, \frac{\partial}{\partial z}, \frac{\partial}{\partial \eta_2})$ to a basis in $\mathbb{R}^{3n}$ with volume $H^{-n}|\partial_x\partial_{\eta_1} S_1||\partial_x\partial_{\eta_2} S_2|$,

$$d_C = |dY|\,|d\eta_1|\,|H|^n\,|\partial_x\partial_{\eta_1} S_1|\,|\partial_x\partial_{\eta_2} S_2| \tag{95}$$

$$= |dy|\,|d\eta_1|\,|H|^{2n}\,|\partial_x\partial_{\eta_1} S_1|\,|\partial_x\partial_{\eta_2} S_2| \tag{96}$$

$$= |dx|\,|d\xi|\,|H|^{2n}\,|\partial_x\partial_{\eta_1} S_1|\,|\partial_x\partial_{\eta_2} S_2|\,. \tag{97}$$

Similarly, we can find from (91) the half-density terms in the principal symbols of $A$ and $B$. Denoting them $d_{C_1}$, $d_{C_2}$ respectively, we have $d_{C_1} = |dx||d\xi||\partial_x\partial_{\eta_1} S_1|$, $d_{C_2} = |dy||d\eta||\partial_x\partial_{\eta_2} S_2|$.

*Maslov factor:* Denote similarity of matrices by $\sim$. Starting from Lemma 15,

$$d^2_{(\mathfrak{y},\eta_1,\eta_2)}\phi \sim \begin{bmatrix} 0 & I & -I \\ I & Q_A & 0 \\ -I & 0 & Q_B \end{bmatrix} \sim \begin{bmatrix} 0 & I & -I \\ I & Q_A - \epsilon I & 0 \\ -I & 0 & Q_B + \epsilon I \end{bmatrix} \tag{98}$$

for any $\epsilon$. Choosing $\epsilon$ so that $Q_A - \epsilon I$ is nonsingular, we apply further similarity transforms to achieve

$$d^2_{(\mathfrak{y},\eta_1,\eta_2)}\phi \sim \begin{bmatrix} -Q_A + \epsilon I & 0 & 0 \\ 0 & Q_A - \epsilon I & 0 \\ 0 & 0 & Q_A + Q_B \end{bmatrix}. \tag{99}$$

Hence $\operatorname{sgn} d^2_{(\mathfrak{y},\eta_1,\eta_2)}\phi = \operatorname{sgn}(Q_A + Q_B)$; by comparison, the signatures in the Maslov factors for $A$ and $B$ are just $\operatorname{sgn} Q_A$ and $\operatorname{sgn} Q_B$. Combining this with (97) and the amplitude of (92), we have the following principal symbol for $BA$:

$$\sigma_{BA} = |H|^{-n} ab\rho \sqrt{|H|^{2n}\,|\partial_x\partial_{\eta_1} S_1|\,|\partial_x\partial_{\eta_2} S_2|}\, \exp\left[\frac{\pi i}{4}\operatorname{sgn}(Q_A + Q_B)\right]. \tag{100}$$

No powers of $2\pi$ occur in (100) as the power of $2\pi$ occurring in (92) is exactly the correct one for $3n$ frequency variables in $n$ dimensions. We can also drop $\rho$ from (100) since this modifies $\sigma_{BA}$ only by a term in $S^{-\infty}$. Since

$$\sigma_A = a\sqrt{|\partial_x \partial_{\eta_1} S_1|}\exp\left[-\frac{\pi i}{4}\operatorname{sgn} Q_A\right], \tag{101}$$

$$\sigma_B = b\sqrt{|\partial_x \partial_{\eta_1} S_1|}\exp\left[+\frac{\pi i}{4}\operatorname{sgn} Q_B\right], \tag{102}$$

this establishes the formula for the principal symbol microlocally near $\gamma$; we are now done as $\gamma$ was arbitrary. □

*Proof of Corollary 14.* Note that

$$d\chi_1(V_\xi) = \left\{-Q_A\eta\cdot\frac{\partial}{\partial y} + \eta\cdot\frac{\partial}{\partial\eta}\right\}, \qquad d\chi_2(V_\zeta) = \left\{Q_B\eta\cdot\frac{\partial}{\partial y} + \eta\cdot\frac{\partial}{\partial\eta}\right\}. \tag{103}$$

The corollary then follows from Proposition 13 by [31, (1.6)]. □

In the proof of Proposition 13, we needed a computational lemma that will also be necessary later on for computing the Maslov factor for the SAR reflection operator:

**Lemma 15.** *Let $\phi_1(x,y,\theta_1)$, $\phi_2(y,z,\theta_2)$ be phase functions, and $\phi(x,\mathfrak{y},z,\theta_1,\theta_2) = \phi_1(x,\mathfrak{y}/H,\theta_1) + \phi_2(\mathfrak{y}/H,z,\theta_2)$, where $H = |(\theta_1,\theta_2)|$. Let $d^2_{y,\theta_i}\phi_i = \left[\begin{smallmatrix} A_i & B_i^\top \\ B_i & C_i\end{smallmatrix}\right]$, let $A = H^{-2}(A_1+A_2)$, and let $M_i = H^{-1}\theta_i y^\top$. Then*

$$d^2_{(\mathfrak{y},\theta_1,\theta_2)}\phi = \begin{bmatrix} A & * & * \\ \frac{B_1}{H} - M_1A & C_1 - \frac{B_1M_1^\top + M_1B_1^\top}{H} + M_1AM_1^\top & * \\ \frac{B_2}{H} - M_2A & -\frac{B_2M_1^\top + M_1B_2^\top}{H} + M_2AM_1^\top & C_2 - \frac{B_2M_2^\top + M_2B_2^\top}{H} + M_2AM_2^\top \end{bmatrix}, \tag{104}$$

*where the $*$ entries are given by symmetry. Furthermore, letting $\sim$ denote similarity,*

$$d^2_{(\mathfrak{y},\theta_1,\theta_2)}\phi \sim \begin{bmatrix} A_1+A_2 & B_1^\top & B_2^\top \\ B_1 & C_1 & 0 \\ B_2 & 0 & C_2 \end{bmatrix}. \tag{105}$$

*Proof.* The Hessian (104) is by direct computation. Applying the similarity transform $V d^2_{(\mathfrak{y},\theta_1,\theta_2)} V^\top$ gives (105), where

$$V = \begin{bmatrix} H & & \\ M_1 & I & \\ M_2 & & I \end{bmatrix}. \tag{106}$$

□

Note that Proposition 13 holds for any FIOs $A$ and $B$ whose canonical relations are locally canonical graphs. An easy corollary is the principal symbol of the inverse.

**Proposition 9.** *If $A \in I^m(\chi)$ is elliptic, $\sigma_{A^{-1}} = \sigma_A^{-1}$.*

*Proof.* Let $\gamma \in T^*X$. Since $\chi$ is a symplectomorphism, it is a local diffeomorphism and therefore $\chi^{-1}$ is locally a graph. Now, apply Proposition 13 to the composition $I = A^{-1}A$. The principal symbol of the identity is 1 (equal to its principal symbol as a pseudodifferential operator), so

$$1 = \sigma_A \sigma_{A^{-1}} \exp \frac{\pi i}{4} \Big[ \operatorname{sgn}(Q_{A^{-1}} + Q_A) - \operatorname{sgn}(Q_{A^{-1}}) - \operatorname{sgn}(Q_A) \Big]. \tag{107}$$

Since $A^{-1}$ is associated with the canonical relation $\chi^{-1}$, we have $Q_{A^{-1}} = -Q_A$ ($y(x,\eta) = y(z,\eta)$ in the terminology of the proof of Proposition 13). Hence $\sigma_A \sigma_{A^{-1}} = 1$. □

**Lemma 10.** *For any $A \in I^m(\chi)$, the principal symbols of $A$ and $A^\top$ are identical.*

*Proof.* This is immediate since the kernels of $A$ and $A^\top$ differ only in the ordering of $x$ and $y$, which does not affect the value of the principal symbol. □

### *5.3. SAR Reflection Operator Principal Symbol*

**Proposition 11.** *The principal symbols of $R_\gamma$ and $U$ are:*

$$\sigma_{R_\gamma} = \sqrt{\frac{2\pi r}{|\xi||\sin\theta|}}, \tag{83}$$

$$\sigma_U = \exp\left[ -\frac{i\pi}{4} \operatorname{sgn}\left( \xi \cdot \gamma''(s) \right) \right], \tag{84}$$

*where $\theta$ is the angle between $\xi$ and $\gamma'(s)$.*

*Proof of formula for $\sigma_{R_\gamma}$.* For $\theta \in \mathbb{R}$, define $\hat{\theta} = (\cos\theta, \sin\theta)$. First, define

$$Rf(r,p) = \int_0^{2\pi} r f(p + r\hat{\theta})\, d\theta. \tag{108}$$

Applying the inverse Fourier transform in $r$ and its inverse, $\mathcal{F}_r \mathcal{F}_r^{-1}$, inside the integral,

$$Rf(r,p) = \frac{1}{2\pi} \int_0^{2\pi} \int_{\mathbb{R}} \int_{\mathbb{R}} e^{i\rho(r'-r)} r' f(p + r'\hat{\theta})\, d\rho\, dr'\, d\theta. \tag{109}$$

Changing from polar coordinates,

$$Rf(r,p) = \frac{1}{2\pi} \int_{\mathbb{R}} \int_{\mathbb{R}^2} e^{i\rho(|x-p|-r)} f(x)\, d\rho\, dx. \tag{110}$$

Restricting to $\gamma$ gives us

$$R_\gamma f(r,s) = (2\pi)^{-3/2} \int_{\mathbb{R}} \int_{\mathbb{R}^2} e^{i\rho(|x-\gamma(s)|-r)} \sqrt{2\pi} f(x)\, d\rho\, dx. \tag{111}$$

$\phi$ is a phase function (in particular, smooth) for $x$ restricted to $\mathbb{R}^2 \setminus \gamma$, and

$$dd_\rho \phi = \begin{bmatrix} 1 & \frac{x-\gamma(s)}{|x-\gamma(s)|} \cdot \gamma'(s) & -\frac{x-\gamma(s)}{|x-\gamma(s)|} & 0 \end{bmatrix}. \tag{112}$$

Hence $\phi$ is also nondegenerate. So $d\rho$ maps $\frac{\partial}{\partial r}$ to a unit volume basis, implying $d_C = |d\rho||ds||dx|$. Letting $\xi = d_x\phi$, we can compute

$$\left| \frac{\partial \xi}{\partial(\rho, s)} \right| = \left| \det \begin{bmatrix} \dfrac{\xi}{|\xi|} & \dfrac{\rho\gamma'(s)}{|x-\gamma(s)|} \end{bmatrix} \right| = \left| \frac{\rho}{r} \frac{\xi}{|\xi|} \cdot \gamma'(s) \right| = \frac{|\xi|}{r} \left| \sin\theta \right|. \tag{113}$$

Hence $d_C = |\rho r^{-1} \sin\theta||dx||d\xi|$, and the Maslov factor is 1 since $d^2_\rho \phi = 0$. Therefore, the principal symbol of $R_\gamma$ is $\sqrt{2\pi r/|\xi||\sin\theta|}$, as claimed. As expected, it has degree $-1/2$.

*Proof of formula for $\sigma_U$.* Let $\Sigma'_L, \Sigma'_R$ be open sets in $T^*(\mathbb{R}^2 \setminus \gamma)$ such that rays through elements $(x,\xi) \in \Sigma'_L$, resp. $\Sigma'_R$ intersect $\gamma$ exactly once from the left, resp. right, and $\Sigma'_R = \mathcal{C}(\Sigma'_L)$. Choose a subset $\Sigma_L \subset \overline{\Sigma_L} \subset \Sigma'_L$ and let $\Sigma_R = \mathcal{C}(\Sigma_L)$. Let $\alpha_L$, $\alpha_R$ be bump functions on $T^*\mathbb{R}^2$ supported in $\Sigma'_L$, $\Sigma'_R$ respectively, and equal to 1 on $\Sigma_L$ and $\Sigma_R$ respectively. Define $R_{\gamma,L} = R_\gamma \alpha_L(x,D)$ and similarly for $R_{\gamma,R}$.

By Proposition 13, the principal symbol of $R_{\gamma,R}$ is equal to $\sigma_{R_\gamma}$ on $\Sigma_R$, which is bounded away from zero. Therefore $R_{\gamma,R}$ is elliptic on $\Sigma_R$ and hence has a microlocal inverse $R^{-1}_{\gamma,R}$ such that $R^{-1}_{\gamma,R} R_{\gamma,R} = I$ microlocally on $\Sigma_R$. By Proposition 9, the symbol of $R^{-1}_{\gamma,R}$ is the reciprocal of that of $R_{\gamma,R}$. For $f$ with $\mathrm{WF}(f) \subseteq \Sigma_L$, we have $R_\gamma f = R_{\gamma,L} f = R_{\gamma,R} R^{-1}_{\gamma,R} R_{\gamma,L} f$ modulo $C^\infty$, hence by the definition of the reflection operator [33], $U = R^{-1}_{\gamma,R} R_{\gamma,L}$ microlocally.

With this in hand, we can compute the principal symbol of $U$ via Proposition 13. Since the principal symbols of $R_{\gamma,L}$ and $R^{-1}_{\gamma,R}$ are reciprocals, all that remains is the Maslov factor. Instead of computing it from Proposition 13, we express $R_{\gamma,L}$ and $R^{-1}_{\gamma,R}$ in a different form, using (111) to write $R_{\gamma,\square}$ ($\square$ indicating either $L$ or $R$) to leading order as

$$R_{\gamma,\square} f(r,s) = (2\pi)^{-3/2} \int_{\mathbb{R}} \int_{\mathbb{R}^2} e^{i\rho(|x-\gamma(s)|-r)} \sqrt{2\pi} \tilde{\alpha}_\square(x,r,s,\rho) f(x)\, d\rho\, dx \\ \text{mod } I^{-3/2}(\mathbb{R}_+ \times I \times \mathbb{R}^2). \tag{114}$$

where $\tilde{\alpha}_\square$ are some order zero amplitudes. Specifically, choose $\tilde{\alpha}_\square$ so that the restriction of $\tilde{\alpha}$ to the critical set $C_\phi$ is equal to $\alpha_\square$ after both are pulled back to $\Lambda'$; then the principal symbols of both sides of (114) are identical. Now, composing $R^{-1}_{\gamma,R} \circ R_{\gamma,L}$ and proceeding as in the proof of Proposition 13, we obtain a Maslov factor $\exp[\frac{i\pi}{4} \operatorname{sgn} Q]$ where $Q = d^2_{\mathfrak{y},\rho_L,\rho_R} \phi$ (the subscripts on $\rho$ and other variables indicate the FIOs $R^{-1}_{\gamma,R}$ or $R_{\gamma,L}$ they originate from). The signature of $Q$ can be obtained from Lemma 15; if $\varphi_\square(r,s,\rho_\square,x)$ is the phase in (114), then we can compute

$$d^2_{r,s,\rho} \varphi_\square = \begin{bmatrix} 0 & 0 & 1 \\ 0 & k_\square & -\xi_\square \cdot \gamma'(s) \\ 1 & -\xi_\square \cdot \gamma'(s) & 0 \end{bmatrix}, \quad k_\square = -\xi_\square \cdot \gamma''(s) - \rho_\square \frac{(\frac{\xi_\square}{|\xi_\square|} \cdot \gamma'^\perp(s))^2}{|\xi|}. \tag{115}$$

Applying Lemma 15, we have

$$d^2_{\mathfrak{y},\rho_L,\rho_R} \sim \begin{bmatrix} 0 & 0 & 1 & 1 \\ 0 & k_L + k_R & -\frac{\sigma_L}{\rho_L} & -\frac{\sigma_R}{\rho_R} \\ 1 & -\frac{\sigma_L}{\rho_L} & 0 & 0 \\ 1 & -\frac{\sigma_R}{\rho_R} & 0 & 0 \end{bmatrix}. \tag{116}$$

Applying similarity transformations we can reduce the matrix to

$$d^2_{\mathfrak{y},\rho_L,\rho_R} \sim \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & k_L + k_R & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \tag{117}$$

which has signature equal to $\operatorname{sgn}[k_L + k_R]$. Now, $\rho_L = -\rho_R$ on the critical set of the phase of$R^{-1}_{\gamma_R} R_{\gamma_L}$. As well, $\gamma''(s)$ is perpendicular to $\gamma'(s)$, and $\frac{\xi_L}{|\xi_L|} \cdot \gamma'^{\perp}(s) = -\frac{\xi_R}{|\xi_R|} \cdot \gamma'^{\perp}(s)$. Hence the second terms in $k_L$, $k_R$ cancel, while the first terms are identical, leaving $k_L + k_R = -2\xi_L \cdot \gamma''(s)$. Hence we have a Maslov factor of $\exp[-\frac{\pi i}{4} \operatorname{sgn}(\xi_L \cdot \gamma''(s))]$. Dropping the $L$ gives us the formula for $\sigma_U$. □

## 6. Conclusion

We have implemented an algorithm based on that described by de Hoop et al. [18] for computing Fourier integral operators associated with canonical graphs with potential caustics, and demonstrated that it successfully computes a variety of operators based on a description of their canonical transformations and principal symbols. We have also described algorithms for computing the inverse and transpose of a graph FIO using the precomputed information from the forward algorithm.

## Acknowledgments

The present work was previously published as part of the author's Ph.D. thesis. The author would like to thank his advisor, Gunther Uhlmann, for suggesting this project and for his guidance, as well as the National Science Foundation and the PIs of the UW Research Training Group in Inverse Problems and Partial Differential Equations for their support (grant 0838212). The author is also thankful for the referees' comments, which have improved this article's exposition.

## References

[1] Córdoba A and Fefferman C 1978 *Comm. Partial Differential Equations* **3** 979–1005 ISSN 0360-5302 URL http://dx.doi.org/10.1080/03605307808820083

[2] Smith H F 1998 *Ann. Inst. Fourier (Grenoble)* **48** 797–835 ISSN 0373-0956 URL http://www.numdam.org/item?id=AIF_1998__48_3_797_0

[3] Smith H F 1998 *J. Geom. Anal.* **8** 629–653 ISSN 1050-6926 URL http://dx.doi.org/10.1007/BF02921717

[4] Candès E and Demanet L 2003 *C. R. Math. Acad. Sci. Paris* **336** 395–398 ISSN 1631-073X URL http://dx.doi.org/10.1016/S1631-073X(03)00095-5

[5] Candes E J and Demanet L 2002 Curvelets, warpings, and optimal representations of fourier integral operators Tech. rep. Citeseer

[6] Guo K and Labate D 2008 *J. Fourier Anal. Appl.* **14** 327–371 ISSN 1069-5869 URL http://dx.doi.org/10.1007/s00041-008-9018-0

[7] Cordero E, Nicola F and Rodino L 2015 *Trans. Amer. Math. Soc.* **367** 7639–7663 ISSN 0002-9947 URL http://dx.doi.org/10.1090/S0002-9947-2015-06302-8

[8] Tataru D 2004 Phase space transforms and microlocal analysis *Phase space analysis of partial differential equations. Vol. II* Pubbl. Cent. Ric. Mat. Ennio Giorgi (Scuola Norm. Sup., Pisa) pp 505–524

[9] Cordero E, Gröchenig K and Nicola F 2012 *J. Fourier Anal. Appl.* **18** 661–684 ISSN 1069-5869 URL http://dx.doi.org/10.1007/s00041-011-9214-1

[10] Cordero E, Nicola F and Rodino L 2009 *Appl. Comput. Harmon. Anal.* **26** 357–370 ISSN 1063-5203 URL http://dx.doi.org/10.1016/j.acha.2008.08.003

[11] Cordero E, Nicola F and Rodino L 2015 *J. Fourier Anal. Appl.* **21** 694–714 ISSN 1069-5869 URL http://dx.doi.org/10.1007/s00041-014-9384-8

[12] Cordero E, Nicola F and Rodino L 2015 *Rev. Mat. Iberoam.* **31** 461–476 ISSN 0213-2230 URL http://dx.doi.org/10.4171/RMI/841

[13] Andersson F, de Hoop M V and Wendt H 2012 *Multiscale Model. Simul.* **10** 111–145 ISSN 1540-3459 URL http://dx.doi.org/10.1137/100808174

[14] Duchkov A A, Andersson F and de Hoop M V 2010 *IEEE Transactions on Geoscience and Remote Sensing* **48** 3408–3423 ISSN 0196-2892

[15] de Hoop M V, Smith H, Uhlmann G and van der Hilst R D 2009 *Inverse Problems* **25** 025005, 21 ISSN 0266-5611 URL http://dx.doi.org/10.1088/0266-5611/25/2/025005

[16] Candès E, Demanet L and Ying L 2007 *SIAM J. Sci. Comput.* **29** 2464–2493 ISSN 1064-8275 URL http://dx.doi.org/10.1137/060671139

[17] Candès E, Demanet L and Ying L 2009 *Multiscale Model. Simul.* **7** 1727–1750 ISSN 1540-3459 URL http://dx.doi.org/10.1137/080734339

[18] de Hoop M V, Uhlmann G, Vasy A and Wendt H 2013 *Multiscale Model. Simul.* **11** 566–585 ISSN 1540-3459 URL http://dx.doi.org/10.1137/120889642

[19] Hörmander L 1994 *The analysis of linear partial differential operators. III* (*Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]* vol 274) (Springer-Verlag, Berlin) ISBN 3-540-13828-5 pseudo-differential operators, Corrected reprint of the 1985 original

[20] Hörmander L 1994 *The analysis of linear partial differential operators. IV* (*Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]* vol 275) (Springer-Verlag, Berlin) ISBN 3-540-13829-3 fourier integral operators, Corrected reprint of the 1985 original

[21] Bao G and Symes W W 1996 *SIAM J. Sci. Comput.* **17** 416–429 ISSN 1064-8275 URL http://dx.doi.org/10.1137/S1064827593258279

[22] Slepian D and Pollak H O 1961 *Bell System Tech. J.* **40** 43–63 ISSN 0005-8580

[23] Landau H J and Pollak H O 1961 *Bell System Tech. J.* **40** 65–84 ISSN 0005-8580

[24] Landau H J and Pollak H O 1962 *Bell System Tech. J.* **41** 1295–1336 ISSN 0005-8580

[25] Slepian D 1964 *Bell System Tech. J.* **43** 3009–3057 ISSN 0005-8580

[26] Landau H J and Widom H 1980 *J. Math. Anal. Appl.* **77** 469–481 ISSN 0022-247X URL http://dx.doi.org/10.1016/0022-247X(80)90241-3

[27] Widom H 1964 *Arch. Rational Mech. Anal.* **17** 215–229 ISSN 0003-9527

[28] Bonami A and Karoui A 2014 *C. R. Math. Acad. Sci. Paris* **352** 229–234 ISSN 1631-073X URL http://dx.doi.org/10.1016/j.crma.2014.01.004

[29] Bonami A and Karoui A 2010 (*Preprint* 1012.3881) URL http://arxiv.org/abs/1012.3881

[30] Duistermaat J J 1996 *Fourier integral operators* (*Progress in Mathematics* vol 130) (Birkhäuser Boston, Inc., Boston, MA) ISBN 0-8176-3821-0

[31] Safarov Y 2014 A symbolic calculus for Fourier integral operators *Geometric and spectral analysis* (*Contemp. Math.* vol 630) (Amer. Math. Soc., Providence, RI) pp 275–290 URL http://dx.doi.org/10.1090/conm/630/12670

[32] Demanet L and Ying L 2011 *SIAM Rev.* **53** 71–104 ISSN 0036-1445 URL http://dx.doi.org/10.1137/080731311

[33] Stefanov P and Uhlmann G 2013 *SIAM J. Appl. Math.* **73** 1596–1612 ISSN 0036-1399 URL http://dx.doi.org/10.1137/120882639

[34] Caday P 2015 *Inverse Problems* **31** 015002, 22 ISSN 0266-5611 URL http://dx.doi.org/10.1088/0266-5611/31/1/015002

[35] Agranovsky M L and Quinto E T 1996 *J. Funct. Anal.* **139** 383–414 ISSN 0022-1236 URL http://dx.doi.org/10.1006/jfan.1996.0090

[36] Cappell S E, Lee R and Miller E Y 1994 *Comm. Pure Appl. Math.* **47** 121–186 ISSN 0010-3640 URL http://dx.doi.org/10.1002/cpa.3160470202