SEMICLASSICAL SAMPLING AND DISCRETIZATION OF CERTAIN LINEAR INVERSE PROBLEMS

PLAMEN STEFANOV

ABSTRACT. We study sampling of functions f and their images Af under Fourier Integral Operators A at rates sh with s fixed and h a small parameter. We show that the Nyquist sampling limit of Af and f are related by the canonical relation of A using semiclassical analysis. We apply this analysis to the Radon transform in the parallel and the fan-beam coordinates. We explain and illustrate the optimal sampling rates for Af, the aliasing artifacts, and the effect of averaging (blurring) the data Af. We prove a Weyl type of estimate on the minimal number of sampling points to recover f stably in terms of the volume of its semiclassical wave front set.

1. INTRODUCTION

The classical Nyquist–Shannon sampling theorem says that a function $f \in L^2(\mathbf{R}^n)$ with a Fourier transform $\hat{f}$ supported in the box $[-B, B]^n$ can be uniquely and stably recovered from its samples $f(sk)$, $k \in \mathbf{Z}^n$ as long as $0 < s \le \pi/B$. More precisely, we have

$$f(x) = \sum_{k\in\mathbf{Z}^n} f(sk)\chi\Big(\frac{1}{s}(x - sk)\Big), \quad \chi(x) := \prod_{j=1}^{n} \operatorname{sinc}(\pi x^j) \tag{1}$$

and

$$\|f\|^2 = s^n \sum_{k\in\mathbf{Z}^n} |f(sk)|^2, \tag{2}$$

where $\|\cdot\|$ is the L^2 norm, see, e.g., [24] or [11]. If $s < \pi/B$ (strictly) then we have oversampling and one can replace the sinc function in (1) by a faster decaying one, see Theorem 3.1 below. For practical purposes, there are two major inconveniences: we need infinitely many samples and f has to be real analytic; in particular, it cannot be compactly supported unless it is zero. The stability (2) allows us to resolve those difficulties by doing approximate recovery for approximately band limited functions. Let us say that f is "essentially supported" in some box $[-R, R]^n$ in the sense that $f = f_0 + f_1$ with $\|f_1\| \le \varepsilon_1 \ll 1$ and f_0 being "essentially B-band limited" in the sense that the L^2 norm of $\hat{f}_0$ outside that frequency box is bounded by some $0 < \varepsilon_2 \ll 1$. Then (1) recovers f_0 up to an error small with ε_2, see [24], by sampling f_0 (not f). The effect of replacing f by f_0 can be estimated in terms of ε_1 as well.

There are generalizations of the sampling theorem to non-rectangular but still periodic grids (or, more precisely, lattices), see, e.g., [19, 27] or to some non-uniform ones, see, e.g., [18] but the latter theory is not as complete when $n \ge 2$. The version presented above is equivalent to viewing $\mathbf{R}^n$ as a product of n copies of $\mathbf{R}$. In particular, it is invariant under translations and dilations and has a natural extension to actions of linear transformations. On the other hand, the conditions are sharp both for uniqueness and for stability. If the sampling rate (Nyquist) condition is violated, there is non-uniqueness and if we still use (1), we get aliasing. There are also versions for f belonging

Partially supported by the National Science Foundation under grant DMS-1600327.

to spaces different than L^2, and for approximate sampling viewed as a minimizing problem [31], etc. The proof of the sampling theorem is equivalent to thinking about f as the inverse Fourier transform of $\hat{f}$, the latter compactly supported. Therefore the samples $f(sk)$ are essentially the Fourier coefficients of $\hat{f}$ extended as a $2\pi/s$ periodic function in each variable (which also explains the Nyquist limit condition), see Theorem 2.2.

The main goal of this work is to study the effect of sampling the data at a certain rate for a class of linear inverse problems. This class consists of problems of inverting a Fourier Integral Operator (FIO): find f if

$$Af = g \tag{3}$$

with g given (so far, noiseless) and A is an FIO of a certain class. There are many examples: inversion of the Euclidean X-ray and the Radon transforms, for which the sampling problem is well studied, see, e.g., the references in [24, Ch. III]; inversion of the geodesic X-ray transform and more general Radon transforms [23]; thermo and photo-acoustic tomography with a possibly variable speed [29], etc. A large class of integral geometry operators are in fact FIOs, as first noticed by Guillemin [12, 13]. The solution operators of hyperbolic problems are also FIOs in general. On the other hand, many non-linear inverse problems have a linearization of this kind, like the boundary rigidity problem or various problems of recovery coefficients in a hyperbolic equation from boundary measurements.

We study the following types of questions.

(i) Sampling Af: Given an essential frequency bound of f (the lowest possible "detail"), how fine should we sample the data Af for an accurate enough recovery? This question, posed that way, includes the problem of inverting A in the first place, in addition to worrying about sampling. The answer is specific to A which could be associated to a canonical graph or not, elliptic or not, injective or not. Then a reformulation of the first question is — if f is approximately band limited, is also Af approximately band limited, with what limit, and then what sampling rate will recover reliably Af? The problem of recovery of f after that depends on the specific A.

(ii) Resolution limit on f given the sampling rate of Af. Suppose we have fixed the sampling rate of Af (not necessarily uniformly sampled). In applications, we may not be able to sample too densely. What limit does this pose on the smallest detail of f we can recover? The answer may depend on the location and on the direction of those details.

(iii) Aliasing. Above, if f has detail smaller than that limit, there will be aliasing. How will the aliasing artifacts look like? Aliasing is well understood in classical sampling theory but the question here is what kind of artifacts an aliased Af would create in the reconstructed f.

(iv) Averaged measurements/anti-aliasing. Assume we cannot sample Af densely enough or assume that f is not even approximately band limited. Then the data would be undersampled. The next practical question is — can we blur the data *before we sample* pointwise to avoid aliasing and then view this as an essentially properly sampled problem but for a blurred version of f? Equivalently, we sample by taking local averages now instead of doing it pointwise. This is a standard technique in imaging and in signal processing but here we want to relate to reconstruction of f to the blurring of the data Af. In X-ray tomography, for example, this would mean replacing the X-ray with thin cylindrical packets of rays and/or using detectors which could average over small neighborhoods and possibly overlap. One could also vibrate the sample during the scan. In thermo and photoacoustic tomography, one can take detectors which average over some small areas. The physical detectors actually do exactly that in order to collect good signal which brings us to another motivation — physical measurements are actually already averaged and we want to understand what this does to the reconstructed f.

To answer the sampling question (i), one may try to estimate the essential support (the "band") of $\widehat{Af}$ given that of $\hat{f}$ and then apply some of the known sampling theorems. This is a possible approach for each particular problem but will also require the coefficients of A, roughly speaking, to be also band limited, and the band limit of Af would depend of that of f and on A. The operators of interest have singular Schwartz kernels however. Also, it may not be easy to get sharp constants. One might prove that if f is approximately B band limited, then Af is, say, CB approximately band limited. The success of this approach would depend heavily on having a sharp constant C. Proving that there exists some $C > 0$ even with some rough estimate on it would not be helpful because the required sample step s would have to be scaled as s/C. In some symmetric cases, such direct approach can and has been done, for example for the Euclidean X-ray/Radon transform, see the references in Section 6. All those results are asymptotic and the analysis and even the formulations of the results are not rigorous. Our main interest however is in inverse problems without symmetries (coming from differential equations with variable coefficients, for example). Note that one may try to use interpolation by various functions, like splines, for example but the same problem exists there since the bounds of the error depend on a priori bounds of some higher order derivatives, and the constants in those estimates matter.

To overcome this difficulty, we look at the problem as an asymptotic one. We think of the highest frequency of f (in some approximate sense) as a large parameter and we are interested in the optimal sampling rate for $g = Af$ when that upper bound gets higher and higher; which would force the sampling step $s > 0$ to get smaller and smaller. To model that, we rescale the dual variable ξ to ξ/h, where $0 < h \ll 1$ is a small parameter; which would rescale the sampling rate to sh. Then we assume that $|\xi|$ is bounded by some constant B which we call a semiclassical band limit of f. We think of f as a family, depending on h. The natural machinery for this is the semi-classical calculus. The frequency content of f locally is described by its semi-classical wave front set $\mathrm{WF}_h(f)$: if the latter consists (essentially) of (x, ξ) with $|\xi| \leq B$ (then we say that f is semiclassically B-band limited), then $\hat{f}$ is essentially supported in B/h. In classical terms, this means $|\hat{f}(\xi)| = O(h^\infty)$ for $|\xi| \geq B/h$, i.e., $\hat{f}$ is essentially supported in $|\xi| \leq B/h$ as $h \ll 1$. One can think of B/h as the upper bound of $|\xi|$ for $\xi \in \operatorname{supp} \hat{f}$ (up to a small error). One can also handle $O(h)$ errors instead of $O(h^\infty)$ by replacing WF_h with is semiclassical Sobolev version. If A is a semiclassical FIO (h-FIO), then $\mathrm{WF}_h(f)$ is mapped to $\mathrm{WF}_h(Af)$ by the canonical relation of A which is a multi-valued map, in general. This is also true, away from the zero frequencies, if A is a classical FIO. That property is sharp when A is elliptic (which happens for most stable problems). Therefore, we can estimate sharply $\mathrm{WF}_h(Af)$ and apply an appropriate sampling theorem. The canonical relation is typically described by some properties of the geometry of the problem, as we will demonstrate on some examples.

This discussion leads us to the next problem we need to study:

(v) Semi-classical sampling. Understand sampling theory for functions $f_h(x)$ with compact $\mathrm{WF}_h(f)$. Above, f is actually $g = Af$ but when we are not talking about sampling $g = Af$, we use the notation f instead of g for generic functions.

We have $\mathrm{WF}_h(f) \subset \Omega \times (-B, B)^n$ with some domain Ω and $B > 0$, and one can formulate a semi-classical equivalent of the classical sampling theorem in this case, see Corollary 3.1. More generally, if $(-B, B)^n$ there is replaced by a parallelotope (an image of a cube under a linear invertible transform); or more generally by a set which has disjoint images under translations by some lattice, we have semiclassical sampling theorems in the reciprocal lattice, see Theorem 3.2; which is a semiclassical version of the result in [27]. A natural question is if we can reduce the number of the sampling points by sampling non-uniformly since $\mathrm{WF}_h(f)$ does not need to be of a product type or even if it is, the phase component may leave a lot of "holes" when tiling $\mathbf{R}^n$.

One way to do this is to use a partition of unity in the x-space and choose different sampling sets locally; but this does not lead to optimal sampling sets all the time. The optimal non-uniform sampling question is largely open even though a lot of progress has been achieved in the classical case. We prove in Section 5 an asymptotic lower bound on the number of non-uniform sampling points needed to sample stably f with $\mathrm{WF}_h(f)$ in a given compact subset $\mathcal{K}$ of $T^*\mathbf{R}^n$. It is of Weyl type and equal to $(2\pi h)^{-n}\,\mathrm{Vol}(\mathcal{K}^{\mathrm{int}})$. This generalizes a theorem by Landau [18] to our setting where the sampling is classical and the number of the sampling points in any Ω is estimated by below by $(2\pi)^{-n}\,\mathrm{Vol}(\Omega\times\mathcal{B})$, if $\mathrm{supp}\,\hat f\subset\mathcal{B}$.

We return to question (i). We can say now that knowing $\mathrm{WF}_h(Af)$ for all f semiclassically B-band limited f determines a sampling rate for Af. Indeed, let $\Sigma_h(Af)$ be the semiclassical frequency set of Af defined as the projection of $\mathrm{WF}_h(Af)$ onto the dual variable. Then an upper bound of the size, or even the shape of $\Sigma_h(f)$ determines a sharp sampling rate for Af, see Theorem 3.2. Since $\mathrm{WF}_h(Af)$ is $2n$-dimensional and $\Sigma_h(Af)$ is n-dimensional, the latter discards useful information about the x-localization of $\mathrm{WF}_h(Af)$, as discussed in the previous paragraph. We formulate results about non-uniform but still a union of locally uniform sampling lattices allowing us to use coarser sampling where the frequencies cannot reach their global maximum. We demonstrate how this works for the Radon transform. Note that the a priori assumptions on the localization of $\mathrm{WF}_h(f)$ could be pretty simple — it could be a product; but its image under the canonical relation of A may not be a product (but is always contained in a product, possibly much larger).

To answer (ii), assuming that we sample Af at a rate requiring, say $\Sigma_h(Af)$ to be supported in $\{\eta;\ |\eta_j|\le B,\ \forall j\}$, we need to map this box back by the inverse of the canonical relation of A.

The aliasing question (iii) admits a neat characterization. Aliasing is well understood in principle as frequencies ξ shifting (or "folding") in the Fourier domain. This can be regarded as a h-FIO, call it S, in our setting, see Section 3.3. If A is associated with a local canonical diffeomorphism, then inverting A with aliased measurements results in $A^{-1}SA$, which is an h-FIO with a canonical relation a composition of the three, see Section 4.3. While classical aliasing shifts frequencies but preserve space localization (see, e.g., Figure 1); we get a new effect here: the "inversion" $A^{-1}SA$ does not preserve the space localization and can shift parts of the image, see Figure 5.

The averaged measurements/anti-aliasing problem (iv) can be resolved as follows. If the a priori estimate of the size of $\mathrm{WF}_h(f)$ is too high (or infinite) for our sampling rate, we can blur the data before sampling, i.e., apply an anti-aliasing filter, which is routinely done in signal and image processing. To do this, take some $\phi\in C^\infty$ decaying fast enough, set $\phi_h=h^{-m}\phi(\cdot/h)$ (here m is the dimension where we collect the measurements, and h^{-m} is a normalization factor), and consider $\phi_h * Af$. That convolution can be seen to be a semiclassical pseudo-differential operator (ΨDO) (a Fourier multiplier) with a semiclassical symbol $\hat\phi$. One can also use a more general semiclassical ΨDO. By Egorov's theorem, $\phi_h * Af = AP_hf + O(h^\infty)f$, where P_h is a zeroth order semiclassical ΨDO with a principal symbol obtained by $\hat\phi$ pulled back by the canonical relation of A. The essential support of the full symbol is also supported where the principal one is. Therefore, if A is associated to a diffeomorphism at least (but not only), one can choose ϕ_h so that P_h plays the role of the low-pass filter needed for proper sampling if the rate of the latter on the image side is given. If the inverse problem is well posed in a certain way, we will recover P_hf stably and the latter will be a semiclassical ΨDO applied to f cutting off higher frequencies which can be viewed as f regularized. We can even choose P_h as desired, and then compute ϕ which is general may change from one sample to another. This brings us to the more general question of sampling Q_hAf, where Q_h is an h-ΨDO limiting the frequency content, instead of being just a convolution. The analysis is then similar.

We want to mention that numerical computations of FIOs by discretization is an important problem by itself, see, e.g., [1, 3, 4, 7] which we do not study here. The emphasis of this paper however is different: how the sampling rate and/or the local averaging of the FIO affect the amount of microlocal data we collect and in turn how they could limit or not its microlocal inversion.

For prior works on this subject, we refer to the study of the discretization of the Radon transform, see [2, 6, 25, 28] form sampling theory viewpoint; see also [24, Ch. III]; see also sections 6, 7 in the present paper. Recently, Katsevich studied the same problem for Radon types of transforms in [16,17,17]. His approach is not based on the sampling theory and is concerned with the accuracy of recovery of jumps across smooth surfaces.

The plan of this paper is as follows. In section 2, we recall some basic facts about semiclassical analysis and show that classical ΨDOs and FIOs map semiclassical wave front sets the way the map classical ones, i.e., by their canonical relation. In section 3, we recall the classical sampling theorem and formulate a semiclassical version of it, which is question (v) formulated above. We also characterize aliasing as a semiclassical FIO. In section 4, we explain how the theorems so far answer the basic questions (i)–(iv). We prove a semiclassical version of Landau's theorem for non-uniform sampling in section 5, showing that the number of sampling points for stable reconstruction must be at least equal to the phase volume occupied by the function we sample. The theory we develop is applied to the weighted Radon transform in the plane in section 6 for the parallel parameterization, and in section 7 for the fan-bean one. Applications to Thermo and Photo-acoustic Tomography are presented in section 8. We finish with a brief discussion in section 9.

Acknowledgments. The author thanks François Monard for the numerous discussions on both the theoretical and on the numerical aspects of this project; Yang Yang for allowing him to use his code for generating the numerical examples in Figure 20; and Maciej Zworski and Kiril Datchev for discussions on the relationship between classical and semiclassical ΨDOs and FIOs.

2. Preliminaries. Action of ΨDOs and FIOs on the semiclassical wave front set

We start with some preliminary notions in semiclassical analysis. Our main reference for the semiclassical calculus is [33], see also [10]. A few words about the notation first. All norms $\|\cdot\|$ are in $L^2(\mathbf{R}^n)$, unless indicated otherwise. The statement $f = O(h^N)$ in the space $\mathcal{H}$ means that $\|f\|_{\mathcal{H}} \leq Ch^N$ for some constant C and $0 < h \ll 1$. Similarly, $f = O(h^\infty)$ in $\mathcal{H}$ means that the above inequality holds for every $N > 0$ with $C = C_N$. We set $\langle\xi\rangle = (1 + |\xi|^2)^{1/2}$. If $\mathcal{H} = \mathcal{S}$ is the Schwartz class, then that inequality has to hold for every seminorm in $\mathcal{S}$ with a constant dependent on the seminorm. Next, $D = -\mathrm{i}\partial$ as usual.

2.1. Wave Front sets.

For the sake of simplicity, we work in $\mathbf{R}^n$ but those notions are extendable to manifolds. Recall that the semiclassical Fourier transform $\mathcal{F}_h f$ of a function depending also on h is given by

$$\mathcal{F}_h f(\xi) = \int e^{-\mathrm{i}x\cdot\xi/h} f(x)\,\mathrm{d}x.$$

This is just a rescaled Fourier transform $\mathcal{F}_h f(\xi) = \hat{f}(\xi/h)$. Its inverse is $(2\pi h)^{-n}\mathcal{F}_h^*$. We recall the definition of the semiclassical wave front set of a tempered h-depended distribution first. In this definition, $h > 0$ can be arbitrary but in semiclassical analysis, $h \in (0, h_0)$ is a "small" parameter and we are interested in the behavior of functions and operators as h gets smaller and smaller. Those functions are h-dependent and we use the notation f_h or $f_h(x)$ or just f. We follow [33] with the choice of the Sobolev spaces to be the semiclassical ones defined by the norm

$$\|f\|_{H_h^s}^2 = (2\pi h)^{-n} \int \langle\xi\rangle^{2s} |\mathcal{F}_h f(\xi)|^2\,\mathrm{d}\xi.$$

Then an h-dependent family $f_h \in \mathcal{S}'$ is said to be h-tempered (or just tempered) if $\|f_h\|_{H^s_h} = O(h^{-N})$ for some s and N. All functions in this paper are assumed tempered even if we do not say so. The semiclassical wave front set of a tempered family f_h is the complement of those $(x_0, \xi^0) \in \mathbf{R}^{2n}$ for which there exists a C_0^∞ function ϕ so that $\phi(x_0) \neq 0$ so that

$$\mathcal{F}_h(\phi f_h) = O(h^\infty) \quad \text{for } \xi \text{ in a neighborhood of } \xi^0 \tag{4}$$

in L^∞ (or in any other "reasonable" space, which does not change the notion). The semiclassical wave front set naturally lies in $T^*\mathbf{R}^n$ but it is not conical as in the classical case. Note that the zero section can be in $\mathrm{WF}_h(f)$. We call the elements of $\mathrm{WF}_h(f)$ *singularities* of f even though f with finite $\mathrm{WF}_h(f)$ (see below) is actually smooth.

There is no direct relationship between the semiclassical wave front WF_h and the classical one WF (when h is fixed in the latter case), see also [33]. For example, for $f \in \mathcal{S}$ independent of h, $\mathrm{WF}_h(f) = \operatorname{supp} f \times \{0\}$ while $\mathrm{WF}(f)$ is empty. On the other hand, if g is singular and compactly supported, then for $f(x) = g(x - 1/h)$ we have $\mathrm{WF}_h(f) = \emptyset$ while $\mathrm{WF}(f)$ is non-empty for every h, see [33]. Sjöstrand proposed adding the classical wave front set to WF_h by considering the latter in $T^*\mathbf{R}^n \cup S^*\mathbf{R}^n$, where the second space (the unit cosphere bundle) represents $T^*\mathbf{R}^n$ as a conic set, i.e., each (x, ξ) with ξ unit is identified with the ray $(x, s\xi)$, $s > 0$. Their points are viewed as "infinite" ones describing the behavior as $\xi \to \infty$ along different directions. An infinite point (x_0, ξ^0) does not belong to the so extended $\mathrm{WF}_h(f)$ if we have

$$\mathcal{F}_h(\phi f_h) = O(h^\infty \langle \xi \rangle^{-\infty}) \quad \text{for } \xi \text{ in a conical neighborhood of } \xi^0 \tag{5}$$

with ϕ as above. Our interest is in functions which are localized in the spatial variable and do not have infinite singularities. In [33], it is said that a tempered f_h is localized in phase space, if there exists $\psi \in C_0^\infty(\mathbf{R}^{2n})$ so that

$$(\mathrm{Id} - \psi(x, hD)) f_h = O(h^\infty), \quad \text{in } \mathcal{S}(\mathbf{R}^n), \tag{6}$$

see the definition of h-ΨDOs below. Such functions do not have infinite singularities and are smooth. We work with functions localized in phase space and those are the functions which can be sampled properly anyway. In practical applications, this assumption is satisfied by the natural resolution limit of the data we collect, for example the diffraction limit or the fact that all physical detectors are of non-infinitesimal size. Then we can view the samples as those of already locally averaged measurements, which attenuates the very high frequencies. Very small detectors would be more affected by noise and will cause worse aliasing.

Other examples of semiclassical wave front sets are the following. If $f_h = e^{\mathrm{i}x\cdot\xi_0/h}$, then $\mathrm{WF}_h(f) = \mathbf{R}^n \times \{\xi_0\}$. The coherent state

$$f_h(x; x_0, \xi_0) = e^{\mathrm{i}x\cdot\xi_0/h - |x-x_0|^2/2h} \tag{7}$$

(to normalize for unit L^2 norm, we need to multiply by $(\pi h)^{-n/4}$) satisfies $\mathrm{WF}_h(f) = \{(x_0, \xi_0)\}$. Its real or imaginary part have wave front sets at (x_0, ξ_0) and $(x_0, -\xi_0)$. We will use such states in our numerical examples.

It is convenient to introduce the notation $\Sigma_h(f)$ for the semiclassical frequency set of f.

Definition 2.1. *For each tempered f_h localized in phase space, set*

$$\Sigma_h(f) = \{\xi;\ \exists x \text{ so that } (x, \xi) \in \mathrm{WF}_h(f)\}.$$

In other words, Σ_p is the projection of $\mathrm{WF}_h(f)$ to the second variable, i.e.,

$$\Sigma_h(f) = \pi_2 \circ \mathrm{WF}_h(f), \tag{8}$$

where $\pi_2(x,\xi)=\xi$. If $\mathrm{WF}_h(f)$ (which is always closed) is bounded and therefore compact, then $\Sigma_h(f)$ is compact.

Definition 2.2. *We say that $f_h \in C_0^\infty(\mathbf{R}^n)$ is semiclassically band limited (in $\mathcal{B}$), if (i) $\operatorname{supp} f_h$ is contained in an h-independent compact set, (ii) f is tempered, and (iii) there exists a compact set $\mathcal{B}\subset\mathbf{R}^n$, so that for every open $U\supset\mathcal{B}$, we have*

$$|\mathcal{F}_h f(\xi)| \le C_N h^N \langle\xi\rangle^{-N} \quad \text{for } \xi\notin U \tag{9}$$

for every $N>0$.

If h is fixed, this estimate trivially holds for every ξ. Its significance is in the h dependence. In particular, such functions do not have infinite singularities, and are localized in phase space. In applications, we take $\mathcal{B}$ to be $[-B,B]^n$ with some $B>0$ or the ball $|\xi|\le B$ or some other set, see, e.g., Figure 12.

As an example, for $0\neq\chi\in C_0^\infty$, $f:=\chi(x)e^{\mathrm{i}x\cdot\xi_0/h}$ is semiclassically band limited with $\mathcal{B}=\{\xi_0\}$. Indeed, $\mathcal{F}_h f(\xi)=\hat\chi((\xi-\xi_0)/h)$ decays rapidly for $\xi\neq\xi_0$ as $h\to 0$. Clearly, that decay is not uniform as $\xi\to\xi_0$ which explains the appearance of U in the definition. We could have required (9) to hold in the closure of $\mathbf{R}^n\setminus\mathcal{B}$ to avoid introducing U; then in this example, one can take $\mathcal{B}$ to be the closure of every neighborhood of ξ_0. Both definitions would work fine for the intended applications. To generalize this example, we can take a superposition of such functions with ξ_0 varying over a fixed compact set $\mathcal{B}$ to get $f=\chi\mathcal{F}_h^{-1}g$ with $g\in L^1$, $\operatorname{supp} g\subset\mathcal{B}$ to be semiclassically band limited with frequency set in $\mathcal{B}$.

Another example of semiclassically band limited functions can be obtained by taking any $f\in\mathcal{E}'(\mathbf{R}^n)$ and convolving if with $\phi_h=h^n\phi(\cdot/h)$ with $\operatorname{supp}\hat\phi\in C_0^\infty$. Then $\phi_h * f$ is semiclassically band limited with $\mathcal{B}=\operatorname{supp}\hat\phi$.

Proposition 2.1. *Let $\mathcal{B}\subset\mathbf{R}^n$ be a compact set. For every tempered $f_h\in C_0^\infty$ with support contained in an h-independent compact set, the following statements are equivalent:*

(a) f_h is semiclassically band limited,
(b) f_h is localized in phase space,
(c) $\mathrm{WF}_h(f)$ is finite and compact.

Proof. Let f_h satisfy the conditions of Definition 2.2. Then $\mathrm{WF}_h(f)$ has no infinite points. Let $\chi_1\in C_0^\infty$ be equal to 1 in a neighborhood of $\operatorname{supp} f_h$ for all $h\ll 1$, and χ_2 be supported in the bounded $U\ni\mathcal{B}$ and equal to 1 near $\mathcal{B}$. Then (6) is satisfied with $\psi=\chi_1(x)\chi_2(\xi)$. Indeed, by (9), for every s,

$$(\mathrm{Id}-\chi_2(hD))f_h = O_{H_h^s}(h^\infty). \tag{10}$$

This implies the same estimate in the Schwartz space $\mathcal{S}$ as well. Apply $\chi_1(x)$ to get (6). Therefore, (a) $\Rightarrow$ (b).

Next, assume (b). Let the compact set $\mathcal{B}\subset\mathbf{R}^n$ be such that for ψ in (6) we have $\psi(x,\xi)=0$ for all x and $\xi\notin\mathcal{B}$. With U as in Definition 2.2, we can apply a semiclassical Fourier multiplier $\mathrm{Id}-\chi_2(hD)$ with χ_2 as above to get (10). Therefore,

$$\int\langle\xi\rangle^{2s}|(1-\chi_2(\xi))\mathcal{F}_h f_h(\xi)|^2\,\mathrm{d}\xi = O(h^\infty).$$

Set $g(\xi)=\langle\xi\rangle^s(1-\chi_2(\xi))\mathcal{F}_h f_h(\xi)$. Using Sobolev embedding and the fact that $-\mathrm{i}\partial_\xi$ corresponds to x/h via $\mathcal{F}_h$, we get $\|g\|_{L^\infty}=O(h^\infty)$. This proves (9). Therefore, (b) $\Rightarrow$ (a).

Assume (c). Using (5) and a partition of unity, we get (9), i.e., (a) holds. On the other hand, (a) implies (c) directly. □

Let f_h be semiclassically band limited and let B be so that $\Sigma_h(f) \subset \{|\xi| < B\}$. Then $f_h = \chi_1(x)\chi_2(hD)f_h + O_{\mathcal{S}}(h^\infty)$ with $\chi_{1,2}$ as in the proof of Proposition 2.1. We can assume $\operatorname{supp}\chi_2 \subset \{|\xi| < B\}$. Apply the operator $\langle hD\rangle^s$ to f_h, then on $\mathrm{WF}_h(f)$, we get that $\langle hD\rangle^s\chi_1(x)\chi_2(D)$ has full symbol $\langle\xi\rangle^s$ up to the negligible class. Therefore, knowing $\|f_h\|$ for a semiclassically band limited f_h allows us to control the semiclassical Sobolev norms of every order $s \geq 0$ as well:

$$\|f_h\|_{H_h^s} \leq \langle B\rangle^s \|f_h\| + O_s(h^\infty). \tag{11}$$

2.2. **h-ΨDOs.** We define the symbol class $S^{m,k}(\Omega)$, where $\Omega \subset \mathbf{R}^n$ is an open set, as the smooth functions $p(x,\xi)$ on $\mathbf{R}^{2n}$, depending also on h, satisfying the symbol estimates

$$|\partial_x^\alpha \partial_\xi^\beta p(x,\xi)| \leq C_{K,\alpha,\beta} h^k \langle\xi\rangle^{m-|\beta|} \tag{12}$$

for x in any compact set $K \subset \Omega$. The negligible class $S^{-\infty,\infty}$ is the intersection of all $S^{m,k}$. Given $p \in S^{m,k}(\Omega)$, we write $P = P_h = p(x, hD)$ with

$$Pf(x) = (2\pi h)^{-n} \iint e^{\mathrm{i}(x-y)\cdot\xi/h} p(x,\xi) f(y)\,\mathrm{d}y\,\mathrm{d}\xi, \tag{13}$$

where the integral has to be understood as an oscillatory one. If we stay with functions localized in phase space, the factor $\langle\xi\rangle^m$ is not needed and we can work with symbols compactly supported in ξ. Then the corresponding classes are denoted by $S^k(\Omega)$ and k is called an order. One can always divide by h^k; so understanding zero order operators is enough.

2.3. **Classical ΨDOs and semiclassical wave front sets.** We begin with an informal discussion about the relationship between classical and semiclassical ΨDOs. Let us denote by ξ the dual variable in the classical ΨDO calculus, and by η the dual variable in the semiclassical case. Formally, by (13) (with ξ replaced by η there), we have $\eta = h\xi$ and after this substitution, we seem to get a classical ΨDO. The problem is that classical symbols do not need to be smooth or even defined for ξ in a compact set; say for $|\xi| \leq C$ with some C. Then $\eta = h\xi$ maps this to $|\eta| \leq Ch$. Semiclassical symbols however need to be defined and smooth for every η in a neighborhood of the semiclassical wave front set of the function we want to study. We see that the zero section $\eta = 0$ needs to be excluded. If we try to rectify this problem by multiplying a classical symbols $p(x,\xi)$ by $\chi(\xi)$ with $\chi \in C_0^\infty$, $\chi = 0$ near $\xi = 0$ and $\chi = 1$ for large ξ, we run into the problem that $\partial_\eta\chi(\eta/h) \sim h^{-1}$ and the symbol estimates (12) are not satisfied for $\chi(\eta/h)p(x,\eta/h)$ near $\eta = 0$.

This shows that the zero section needs to be treated separately. Even when that problem does not exists, classical ellipticity does not necessarily mean semiclassical one. For example, $-\Delta$ is a classical elliptic ΨDO with symbol $|\xi|^2$ while $|\eta|^2/h^2$ is the semiclassical symbol of the same operator but $-\Delta$ is not semiclassically elliptic anymore (in $S^{2,-2}$).

Proposition 2.2. *Let K be a smoothing operator, and let $f_h \in \mathcal{E}'(\mathbf{R}^n)$ be tempered. Then $\mathrm{WF}_h(Kf) \subset \mathbf{R}^n \times \{0\}$.*

Proof. Let $\phi, \psi \in C_0^\infty$ such that $\psi = 1$ near some $\xi_0 \neq 0$ but $\psi = 0$ near 0. Then

$$\begin{aligned}
\psi\mathcal{F}_h(\phi Kf)(\xi) &= \psi(\xi) \int e^{-\mathrm{i}x\cdot\xi/h}\phi(x)Kf(x)\,\mathrm{d}x \\
&= \psi(\xi) \iint e^{-\mathrm{i}x\cdot\xi/h}\phi(x)K(x,y)f(y)\,\mathrm{d}x\,\mathrm{d}y \\
&= \psi(\xi) \iint e^{-\mathrm{i}x\cdot\xi/h}\phi(x)\left[(1-h^2\Delta_y)^m K(x,y)\right](1-h^2\Delta_y)^{-m} f(y)\,\mathrm{d}x\,\mathrm{d}y.
\end{aligned}$$

For $m \gg 1$, $(1-h^2\Delta)^{-m} f \in L^2$ with a norm $O(h^{-N})$ for some N. Fix one such m. We have $L^k e^{-\mathrm{i}x\cdot\xi/h} = e^{-\mathrm{i}x\cdot\xi/h}$ for every k with $L = \mathrm{i}h|\xi|^{-2}\xi\cdot\nabla_x$. Integrating by parts with L^k, we get $\psi\mathcal{F}_h(\phi Kf) = O(h^\infty)$. Therefore, every (x,ξ) with $\xi \neq 0$ is not in $\mathrm{WF}_h(Kf)$. □

Theorem 2.1. *Let P be a properly supported ΨDO of order m. Then for every f_h localized in phase space,*

$$\mathrm{WF}_h(Pf) \setminus 0 \subset \mathrm{WF}_h(f) \setminus 0.$$

If P is elliptic, then the inclusion above is an equality.

Proof. By the proposition above, the property of the theorem is invariant under adding a smoothing operator as it should be. Let $p(x,\xi)$ be the symbol of P, so that $P = p(x,D)$ modulo a smoothing operator. Then P is a formal h-ΨDO with a symbol $p(x,\xi/h)$. The latter is a semiclassical symbol away from every neighborhood of $\xi = 0$. Indeed, for x in every compact set K,

$$|\partial_x^\alpha \partial_\xi^\beta p(x,\xi/h)| \le C_{\alpha,\beta,K} h^{-|\beta|} |\xi/h|^{m-|\beta|} = C_{\alpha,\beta,K} h^{-m} |\xi|^{m-|\beta|} \quad \text{for } |\xi/h| > 1$$

and in particular,

$$|\partial_x^\alpha \partial_\xi^\beta p(x,\xi/h)| \le C_{\alpha,\beta,K} h^{-m} |\xi|^{m-|\beta|} \quad \text{for } |\xi| \ge \varepsilon > 0$$

for every $\varepsilon > 0$ and $0 < h \le \varepsilon$. This shows that $p(x,\xi/h)$ is a semiclassical symbol of order $(m,-m)$ restricted to $|\xi| > \epsilon$. Note that the smallness requirement on h depends on ε. To complete the proof, we need to resolve this problem.

We show next that for every fixed $\varepsilon > 0$, if $\mathrm{WF}_h(f) \subset B_\varepsilon(0)$, then $\mathrm{WF}_h(Pf) \subset B_\varepsilon(0)$ as well. This is a weaker version of what we want to prove but it is valid near $\xi = 0$.

Since P is properly supported, with $\phi \in C_0^\infty$ as in the previous proof, we may assume that in ϕPf, the function f is supported in a fixed compact set. Then by a compactness argument, for every $\varepsilon' > \varepsilon$, we have $\mathcal{F}_h f(\xi) = O(h^\infty)$ for $|\xi| \ge \varepsilon'$. Then

$$\begin{aligned}
\mathcal{F}_h(\phi Pf)(\eta) &= (2\pi h)^{-n} \iiint e^{-\mathrm{i}x\cdot\eta/h + \mathrm{i}(x-y)\cdot\xi} \phi(x) a(x,\xi) f(y)\,\mathrm{d}y\,\mathrm{d}\xi\,\mathrm{d}x \\
&= (2\pi h^2)^{-n} \iiint e^{-\mathrm{i}x\cdot\eta/h + \mathrm{i}(x-y)\cdot\xi/h} \phi(x) a(x,\xi/h) f(y)\,\mathrm{d}y\,\mathrm{d}\xi\,\mathrm{d}x \\
&= (2\pi h^2)^{-n} \iint e^{\mathrm{i}x\cdot(\xi-\eta)/h} \phi(x) a(x,\xi/h) \mathcal{F}_h f(\xi)\,\mathrm{d}\xi\,\mathrm{d}x.
\end{aligned}$$

The integration above can be restricted to $|\xi| \le \varepsilon'$ with $\varepsilon' > \varepsilon$ fixed, and this will result in an $O(h^\infty)$ error. For the phase function $\mathrm{i}\Phi/h = \mathrm{i}x\cdot(\xi-\eta)/h$ we have $\Phi_x = \xi - \eta$, i.e. the zeros are on the diagonal $\xi = \eta$ in the fiber variable. The following operator preserves $\exp(\mathrm{i}\Phi/h)$:

$$L = -\mathrm{i}\frac{\xi-\eta}{|\xi-\eta|^2}\cdot h\nabla_x.$$

Since $|\xi| \le \varepsilon'$, if we restrict η to $|\eta| > \varepsilon''$ with $\varepsilon'' > \varepsilon'$, and integrate by parts, we get $O(h^\infty|\xi|^{-\infty})$ above. This proves that $\mathrm{WF}_h(Pf)$ is included in $\mathbf{R}^n \times B_{\varepsilon''}(0)$. Since $\varepsilon'' > \varepsilon$ can be taken as close to ε as we wish, this proves the claim.

Now, using h-ΨDO cutoffs, we express f as $f = f_1 + f_2$, with $\mathrm{WF}_h(f_1)$ included in $|\xi| \le \varepsilon$ and $\mathrm{WF}_h(f_2)$ included in $|\xi| \ge \varepsilon/2$. By the claim above, $\mathrm{WF}_h(Pf_1)$ is included in $|\xi| \le \varepsilon$ as well. For $\mathrm{WF}_h(Pf_2)$, by what we proved earlier, $\mathrm{WF}_h(Pf_2) \subset \mathrm{WF}_h(f_2)$. Therefore, $\mathrm{WF}_h(f) \subset \mathrm{WF}_h(f) \cup \mathbf{R}^n \times B_\varepsilon(0)$ for every $\varepsilon > 0$ which proves that $\mathrm{WF}_h(f) \subset \mathrm{WF}_h(f) \cup \mathbf{R}^n \times \{0\}$. In particular, this proves the first part of the theorem.

To prove the second part, if P is elliptic, there is a parametrix Q of order $-m$ so that $QPf = f + Kf$, where K is smoothing. Then we apply the first part of the proof. □

For future reference, note that we also proved that for every $\varepsilon > 0$, every classical ΨDO is also a semiclassical one restricted to f with $\mathrm{WF}_h(f)$ not containing ξ with $|\xi| \le \varepsilon$.

2.4. **Classical FIOs and semiclassical wave front sets.**

Theorem 2.2. *Let A be an FIO in the class $I^m(\mathbf{R}^{n_2}, \mathbf{R}^{n_1}, \Lambda)$, where $\Lambda \subset T^*(\mathbf{R}^{n_1} \times \mathbf{R}^{n_2}) \setminus 0$ is a Lagrangian manifold and $m \in \mathbf{R}$. Then for every f_h localized in phase space,*

$$\mathrm{WF}_h(Af) \setminus 0 \subset C \circ \mathrm{WF}_h(f) \setminus 0, \tag{14}$$

where $C = \Lambda'$ is the canonical relation of A.

Proof. The statement holds for h-FIOs, see, e.g, [14, 20]. When A is a classical FIO, A can be written locally as

$$Af(x) = (2\pi)^{-(n_1+n_2+2N)/4} \iint e^{\mathrm{i}\phi(x,y,\theta)} a(x,y,\theta) f(y)\, \mathrm{d}y\, \mathrm{d}\theta,$$

modulo a smoothing operator, where a is an amplitude of order $m + (n_1 + n_2 - 2N)/4$ and ϕ is a non-degenerate phase function, see [15, Chapter 25.1] and $\theta \in \mathbf{R}^N$. As in the proof of Theorem 2.1 above, we can express Af as an oscillatory integral with a phase function $\phi(x, y, \theta)/h$ and an amplitude $a(x, y, \theta/h)$. The latter is a semiclassical amplitude for $|\theta| > \varepsilon > 0$ and $0 < h < \varepsilon$. The rest of the proof is as the proof of Theorem 2.1 using the non-degeneracy of the phase. □

As above, we also proved that for every $\varepsilon > 0$, every classical FIO is also a semiclassical one restricted to f_h with $\mathrm{WF}_h(f)$ not containing ξ with $|\xi| \le \varepsilon$. Finally, if F has a left parametrix which is also an FIO, then (14) is an equality. This happens, for example, if C is locally a graph of a diffeomorphism and A is elliptic. It also happens when C satisfies the clean intersection condition, which is the case for the geodesic X-ray transform in dimensions $n \ge 3$ with no conjugate points.

Finally, we want to emphasize that while $\mathrm{WF}_h(f)$ is just a set, the theorem (as typical for such statements) gives us more than recovery of that set. If we know $Af = m$ up to an $O_{\mathcal{S}}(h^\infty)$ error, and f_1 and f_2 are two possibly different solutions, then $A(f_1 - f_2)$ has an empty semiclassical wave front set; and if A is, say elliptic and associated to a local diffeomorphism, then $\mathrm{WF}_h(f_1 - f_2)$ can only be contained in the zero section, i.e., we can recover f microlocally, away from $\xi = 0$; not just $\mathrm{WF}_h(f)$.

3. Sampling theorems

We present here a version of the classical sampling theorem, and a semiclassical extension in the spirit of [27]. This addresses the semiclassical sampling problem (v) in the Introduction.

3.1. **Sampling on periodic grids.** We start with a version of the classical Nyquist–Shannon sampling theorem which allows for oversampling. Below, $B > 0$ is a fixed constant. At this stage, for simplicity, we do not use the freedom to apply linear transformations to assume that $\operatorname{supp} f$ is included in a set more general than a box satisfying the tiling property, as in [27]. We do that in the semiclassical version of that theorem below.

Theorem 3.1. *Let $f \in L^2(\mathbf{R}^n)$. Assume that $\operatorname{supp} \hat{f} \subset [-B, B]^n$ and let $0 < s \le \pi/B$. Let $\hat{\chi} \in L^\infty$ be supported in $[-\pi, \pi]^n$ and equal to 1 on $(\pi/B).\operatorname{supp} \hat{f}$. If $s \le \pi/B$, then f can be reconstructed by its samples $f(sk)$, $k \in \mathbf{Z}^n$ by*

$$f(x) = \sum_{k \in \mathbf{Z}^n} f(sk) \chi\left(\frac{1}{s}(x - sk)\right), \tag{15}$$

and

$$\|f\|^2 = s^n \sum_{k\in\mathbf{Z}^n} |f(sk)|^2. \tag{16}$$

Proof. Since $\hat{f}$ is supported in $[-B,B]^n$, it is also supported in $[-\pi/s,\pi/s]^n$. Then we can take the $2\pi/s$ periodic extension $\hat{f}_{\text{ext}}$ of $\hat{f}$ in all variables and then the (inverse) Fourier series of that extension to get

$$\hat{f}_{\text{ext}}(\xi) = s^n \sum_k f(sk) e^{-\mathrm{i}s\xi\cdot k}. \tag{17}$$

Multiply this by $\hat{\chi}(s\xi)$ to get

$$\hat{f}(\xi) = s^n \hat{\chi}(s\xi) \sum_k f(sk) e^{-\mathrm{i}s\xi\cdot k}. \tag{18}$$

Take the inverse Fourier Transform to get (15). Equality (16) is just Parseval's equality applied to (17). □

Note that when $\hat{\chi}$ is the characteristic function of $[-\pi,\pi]$ in 1D, we get $\chi(x) = \operatorname{sinc}(\pi x)$, where $\operatorname{sinc}(x) := \sin x/x$. In higher dimensions, we get a product of such functions. When $\operatorname{supp}\hat{f} \subset (-B,B)$, one can choose $\hat{\chi} \in C_0^\infty$, which makes the series (15) rapidly convergent. Even a piecewise linear $\hat{\chi}$ will increase the convergence rate: if for $n=1$ we choose $\hat{\chi}$ to have a trapezoidal graph by defining it as linear in $[\delta,1]$ for some $\delta\in(0,1)$ (and continuous everywhere), then

$$\chi(x) = \frac{\cos x - \cos(\delta x)}{(1-\delta)x^2}$$

which is $O_\delta(x^{-2})$ instead of just $O(|x|^{-1})$ as sinc, with a constant getting large when δ is close to 1. Theorem 3.1 says that the sampling rate should not exceed π/B, which is known as the Nyquist limit. The theorem can be extended to classes of non L^2 functions and then we need the sampling rate to be strictly below the Nyquist one. If $f=\sin(x)$ for example, $B=1$ is the sharpest band limit and a sampling rate of π would yield zero values. On the other hand, that function is not in L^2. Sampling with a smaller step recovers that f uniquely in the corresponding class.

Remark 3.1. The subspace L^2_B of L^2 consisting of functions f with $\operatorname{supp}\hat{f}\subset[-B,B]^n$ is a Hilbert space itself. If we consider the samples $f(sk)$ as elements of $\ell^2_s(\mathbf{Z}^n)$ with measure s^n (i.e., sequences $a=\{a_k\}_{k\in\mathbf{Z}^n}$ with $\|a\|^2 = s^n\sum_k |a_k|^2$), then the theorem implies that the map

$$L^2_B \ni f \Longrightarrow f(sk) \in \ell^2_s(\mathbf{Z}^n)$$

is unitary; i.e., not just an isometry but a bijective one because for every choice of the sampled values in $\ell^2_s(\mathbf{Z}^n)$ there is unique $f\in L^2_B$ with those values given by (15). Therefore, the set of the samples is not an overdetermined one for a stable recovery. Also, (15) is just an expansion of f in an orthogonal basis.

There are several direct generalizations possible. First, one can have different band limits for each component: $|\xi_j|\le B_j$ on $\operatorname{supp}\hat{f}$. Then we need $s_j \le \pi/B_j$. One can have frequency support in non-symmetric intervals but since we are interested mostly in real valued functions, we keep them symmetric.

The next observation is that the set containing $\operatorname{supp}\hat{f}$ does not have to be a box. Of course, if it is bounded, it is always included in some box but more efficient sampling can be obtained if we know that the group generated by the shifts $\xi_j \mapsto \xi_j + 2B_j$, $j=1,\dots,n$ maps $\operatorname{supp}\hat{f}$ into mutually

disjoint sets. Then we take the periodic extension of $\hat f$ w.r.t. that group. The requirement for $\hat\chi$ then is to have disjoint images under those shifts and to be equal to 1 on $\operatorname{supp}\hat f$; then the step from (17) to (18) works in the same way. Note that $\operatorname{supp}\hat\chi$ may not be contained in $\prod[-B_j, B_j]$.

Finally, one may apply general non-degenerate linear transformation as in [27]. It is well known that the change $x = Wy$ with W an invertible matrix triggers the change $\xi = (W^*)^{-1}\eta$ of the dual variables. Let f be band limited in $\mathcal{B}$. Assume that W is such that the images of $\mathcal{B}$ under the translations $\xi \mapsto \xi + 2\pi(W^*)^{-1}k$, $k \in \mathbf{Z}^n$ are mutually disjoint. Ideally, those shifts would cover the whole $\mathbf{R}^n$ (the tiling property). Then in the y variables, the shifts of $\operatorname{supp}\hat f$ by the group $2\pi\mathbf{Z}^n$ are mutually disjoint (i.e., for $g(y) := f(Wy)$, we have that for $\hat g(\eta)$). Then g is band limited in a box with a half-side $B = \pi$ and then for every $s \in (0, 1]$, f is stably determined by its samples $f(sWk)$, $k \in \mathbf{Z}^n$. Moreover, the reconstruction formula (15) still holds with the requirement that $\chi = 1$ on $\operatorname{supp}\hat f$ and $\operatorname{supp}\chi$ has disjoint images under the translations by that group; there is an additional $|\det W|$ factor (see also Theorem 3.2) coming from the change of the variables.

We present next a semiclassical version of the sampling theorem. It can be considered as an approximate rescaled version of the classical theorem when the conditions are approximately held, with an error estimate. We present a version with a uniform estimate of the error and in the corollary below, we formulate a simplified corollary without that uniformity but which requires f_h to be semiclassically band limited (only), i.e., it applies to a single f_h.

Theorem 3.2. *Assume that $\Omega \subset \mathbf{R}^n$, $\mathcal{B} \subset \mathbf{R}^n$ is open and bounded. Let $f_h \in C_0^\infty(\Omega)$ satisfy*

$$\|(\mathrm{Id} - \psi(x, hD))f_h\|_{H_h^m} = O_m(h^\infty)\|f_h\|, \quad \forall m \gg 0 \tag{19}$$

with some $\psi \in C_0^\infty(\mathbf{R}^{2n})$ so that $\operatorname{supp}_\xi \psi \subset \mathcal{B}$. Let $\hat\chi \in L^\infty(\mathbf{R}^n)$ be so that $\operatorname{supp}\hat\chi \subset \mathcal{B}$ and $\hat\chi = 1$ near $\operatorname{supp}_\xi \psi$.

Assume that W is an invertible matrix so that the images of $\mathcal{B}$ under the translations $\xi \mapsto \xi + 2\pi(W^*)^{-1}k$, $k \in \mathbf{Z}^n$, are mutually disjoint. Then for every $s \in (0, 1]$,

$$f_h(x) = |\det W| \sum_{k\in\mathbf{Z}^n} f_h(shWk)\chi\Big(\frac{\pi}{sh}(x - shWk)\Big) + O_{H^m}(h^\infty)\|f\|, \tag{20}$$

for every $m \geq 0$ and

$$\|f_h\|^2 = |\det W|(sh)^n \sum_{k\in\mathbf{Z}^n} |f_h(shWk)|^2 + O(h^\infty)\|f\|^2. \tag{21}$$

Remark 3.2. We allow $\hat\chi$ to be non-smooth here, and in particular, χ being a product of sinc functions is allowed. When discussing aliasing as an FIO however, we take $\hat\chi$ to be smooth. Next, the number of non-zero terms in (20) is $O(h^{-n})$ because $f_h \in C_0(\Omega)$ by assumption. Even without that assumption, (19) would imply that there are only that many significant samples; the rest would contribute to the error terms only.

Proof. As in the remark above, we can make the change of variables $x = Wy$; then the dual variables η is related to ξ by $\eta = W^*\xi$. In the new variables, the images of $\mathcal{B}$ under the translations $\eta \to \eta + 2\pi k$ do not intersect. Therefore, it is enough to consider the case $W = \mathrm{Id}$.

Set

$$g_h(x) := \mathcal{F}_h^{-1}\hat\chi\mathcal{F}_h f_h. \tag{22}$$

Since $\mathcal{F}_h g_h$ is the classical Fourier transform of the rescaled $h^n g_h(hx)$, by the previous theorem and the remark after it, with $B = \pi$,

$$h^n g_h(hx) = \sum_k h^n g_h(shk)\chi((x - sk)/s).$$

Replace hx by x to get (20) and (21) for g_h without the error terms, i.e.,

$$g_h(x) = \sum_{k\in\mathbf{Z}^n} g_h(shk)\chi\left(\frac{1}{sh}(x-shk)\right), \quad \|g_h\|^2 = (sh)^n \sum_{k\in\mathbf{Z}^n} |g_h(shk)|^2. \tag{23}$$

To estimate the error, write

$$\begin{aligned} r_h := f_h - g_h &= \mathcal{F}_h^{-1}(1-\hat\chi)\mathcal{F}_h f_h = (\mathrm{Id}-\hat\chi(hD))f_h \\ &= (\mathrm{Id}-\hat\chi(hD))\left(\psi(x,hD)f_h + \tilde r_h\right), \end{aligned}$$

where $\tilde r_h$ satisfies (19).

We estimate the term including $\tilde r_h$ first. Clearly, for every m, N,

$$\|(\mathrm{Id}-\hat\chi(hD))\tilde r_h\|_{H_h^m} \le C\|\tilde r_h\|_{H_h^m} \le C_{m,N}h^N\|f_h\|. \tag{24}$$

To estimate $(\mathrm{Id}-\hat\chi(hD))\psi(x,hD)f_h$, we can use the semiclassical calculus. Since $\operatorname{supp}(1-\hat\chi)$ and $\operatorname{supp}_\xi\psi$ do not intersect, one can replace $\hat\chi$ by a smooth function $\hat\chi_1$ with the same property so that $\hat\chi_1\hat\chi = \hat\chi_1$. Then as in (24), it is enough to estimate the H_h^m norm of $(\mathrm{Id}-\hat\chi_1(hD))\psi(x,D)f_h$ because $1-\hat\chi = (1-\hat\chi_1)(1-\hat\chi)$. The operator $(\mathrm{Id}-\hat\chi_1(hD))\psi(x,D)$ is in $\Psi^{-\infty,-\infty}$ which proves an estimate of the kind (24) for that term as well. Therefore, $r_h = O_{H^m}(h^\infty)$ for every m. Using Sobolev embedding, we get $r_h(shk) = O(h^\infty)\|f\|$ for shk in any fixed neighborhood of $\bar\Omega$; and there are $O(h^{-n})$ of those. This implies that replacing $g_h(shk)$ by $f_h(shk)$ in the interpolation formula in (23) results in an $O(h^\infty)$ error in every H_h^m. We already established such an error estimate if we replace g_h by f_h on the left. The Parseval's identity (21) follows from those arguments as well. This completes the proof. □

In particular, when we are given a single semiclassically band limited f_h, we have the corollary below. Note that estimates (4), (5), (6), and (9) are not formulated uniformly in f_h; and on the other hand, estimate (4) is the standard definition of the semiclassical wave front set. Theorem 3.2 has uniform estimates of the error but requires the stronger condition (19). Also, one can have the semiclassical band limited assumptions as in the corollary below but still the more general sampling geometry as in Theorem 3.2.

Corollary 3.1. *Let f_h be semiclassically band limited with $\Sigma_h(f) \subset \prod(-B_j, B_j)$ with some $B_j > 0$. Let $\hat\chi_j \in L^\infty(\mathbf{R})$ be supported in $[-\pi,\pi]^n$ and $\hat\chi_j(\pi\xi_j/B_j) = 1$ for $\xi \in \Sigma_h(f)$. If $0 < s_j \le \pi/B_j$, then*

$$f_h(x) = \sum_{k\in\mathbf{Z}^n} f_h(s_1hk_1,\dots,s_nhk_n)\prod_j \chi_j\left(\frac{1}{s_jh}(x^j - s_jhk_j)\right) + O_{\mathcal{S}}(h^\infty)\|f\|, \tag{25}$$

and

$$\|f_h\|^2 = s_1\dots s_nh^n \sum_{k\in\mathbf{Z}^n} |f_h(s_1hk_1,\dots,s_nhk_n)|^2 + O(h^\infty)\|f\|^2. \tag{26}$$

Proof. Take $\Omega \supset \operatorname{supp} f$, $W = \operatorname{diag}(\pi/B_1,\dots,\pi/B_n)$. Then $\det W = \pi^n/(B_1\dots B_n)$. If $\delta < 1$ is close enough to 1, then the conditions of Theorem 3.2 are satisfied and we can take $\hat\chi$ there (depending on $\xi\in\mathbf{R}^n$) to be the product $\hat\chi_0(\pi\xi_1/B_1)\dots\hat\chi_0(\pi\xi_1/B_n)$, where χ_0 (depending on an 1D variable) is as in the corollary but related to $B_j = 1$. Set $\hat\chi_j(\xi) = \hat\chi_0(\xi/B_j)$; then $\chi_j(x) = B_j\chi_0(B_jx)$. Then Theorem 3.2 implies the corollary. □

In other words, if the sampling rate sh is smaller than $\pi h/B$, we have an accurate recovery. Note that the sampling rate s is rescaled by h compared to Theorem 3.1. We call s a relative sampling rate.

Remark 3.3. Let $\operatorname{supp} f \in \Omega \Subset \mathbf{R}^n$ be band limited as in the corollary, and take for simplicity $B_j = B$ for all j. Then the smallest subset in the phase space $T^*\Omega$ containing $\mathrm{WF}_h(f)$ for all such f's is $\bar\Omega \times [-B, B]^n$. Its volume with respect to the volume form $\mathrm{d}x\,\mathrm{d}\xi$ is

$$\operatorname{Vol}\left(\bar\Omega \times [-B, B]^n\right) = (2B)^n \operatorname{Vol}(\Omega).$$

The number of samples we need for f is bounded from below by

$$\#\text{ of samples} = \frac{\operatorname{Vol}(\Omega)}{(\pi h/B)^n} = (2\pi h)^{-n} \operatorname{Vol}\left(\bar\Omega \times [-B, B]^n\right) \tag{27}$$

up to an $o(1)$ relative error as it follows by comparing the volume of Ω with the number of the lattice points we can fit in it. The factor $(2\pi h)^{-n}$ is sometimes multiplied by $\mathrm{d}\xi$ to define a natural rescaled measure $(2\pi h)^{-n}\mathrm{d}\xi$ in the ξ space. Therefore, we see that the asymptotic number of samples needed to sample such f's has a sharp lower bound equal to the phase volume occupied by f in the phase space w.r.t. to that semiclassical measure. Note that this is sharp when applied to f's with $\mathrm{WF}_h(f)$ contained in the product of Ω an a box in the ξ variable. This is a version of the lower bound in [18] in our setting, which we prove in Theorem 5.1.

Finally, we prove below that in simplest case, the sampling data is not overdetermined (as a limit when $\Sigma_h(f)$ gets closer and closer to the bounding box). The next corollary can be considered as a slightly weaker semiclassical version of Remark 3.1. It also shows that the set every set $\{shk\}$ as in the corollary below, for which we either have critical sampling of undersampling, is a set of interpolation in the sense of Landau [18].

Corollary 3.2. *Let $\Omega \ni \Omega_0$ be both open. Fix $B > 0$, $s \ge \pi/B$. For $0 < h \ll 1$, let $K_0(h) \subset \mathbf{Z}^n$ be the set of those k for which $shk \in \Omega_0$. Then for every collection of complex numbers $\{y_{k,h}\}$, $k \in K_0(h)$ with $\sum_k |y_{k,h}|^2$ tempered, there exists a semiclassically band limited f_h with $\mathrm{WF}_h(f) \subset \Omega \times [-B, B]^n$ so that $f(shk) = y_{k,h}$.*

Proof. Set

$$g_h(x) = \sum_{k\in K_0(h)} y_{k,h}\chi\left(\frac{1}{sh}(x - shk)\right), \tag{28}$$

where χ is a product of sinc functions as in (1). Then $g(shk) = y_{k,h}$ and we have

$$\mathcal{F}_h g_h(\xi) = \mathbf{1}_{[-\pi,\pi]^n}(s\xi) \sum_{k\in K(h)} y_{k,h} e^{-\mathrm{i}sk\cdot\xi}$$

with $\mathbf{1}_{[-\pi,\pi]^n}$ being the characteristic function of $[-\pi, \pi]^n$. Then $\mathrm{WF}_h(g_h) \subset \mathbf{R}^n \times [-B, B]^n$. Let $\psi \in C_0^\infty(\Omega)$ be equal to 1 near $\bar\Omega_0$, and set $f_h = \psi g_h$. Then f_h has the required properties. □

3.2. **Non-linear transformations and non-uniform sampling.** Non-uniform sampling in dimensions $n \ge 1$ is not well understood unless we sample on Cartesian products of non-uniform 1D grids. We mention here some easy to obtain but not very far reaching extensions. If $x = \phi(y)$ is a diffeomorphism, one may try to obtain sampling theorems for $f(x)$ by applying the results above to $g(y) := f(\phi(y))$ if the latter happens to be band-limited. Then the sampling points y_k, $k \in \mathbf{Z}^n$ will give us sampling points $x_k = \phi(y_k)$. This is what we did above with $\phi(y) = Wy$ linear. Then we first choose V so that the translates of $\operatorname{supp}\hat f$ (or $\Sigma_h(f)$) under the group $\eta \to \eta + V\mathbf{Z}^n$ are mutually disjoint and find W from the equation $V = 2\pi(W^*)^{-1}$. Let us consider the semiclassical sampling theorem now. The analog of this for non-linear transformations is the well known property that $\mathrm{WF}_h(f)$ transforms as a set of covectors, i.e., $\xi = ((\mathrm{d}\phi)^*)^{-1}\eta$. Now, if we want to prescribe

$\mathrm{d}\phi(y)$ pointwise (at $y = \phi^{-1}(x)$ for every x) based on an a priori knowledge of $\mathrm{WF}_h(f)$, we run into the problem that the equation $\mathrm{d}\phi = \Phi$ with Φ given is not solvable unless $\mathrm{d}\Phi = 0$ (i.e., the matrix-valued map Φ is curl-free). On the other hand, we can take a partition of unity and on each covering open set, we can take a linear transformation giving is better sampling locally. In Section 6, we prove a lower bound for the number of sampling points which is sharp at least in the situation of Theorem 3.2.

3.3. **Aliasing.** Aliasing is a well known phenomenon in classical sampling theory when there Nyquist limit condition is not satisfied (i.e., we have undersampling). For simplicity, we will recall the basic notions when the sampling is done on a rectangular grid but the more general periodic case in Theorem 3.2 can be handled similarly.

Assume that we sample each x^j-th variable of $f(x)$ with a steps sh. Different steps s_j can be handled easily with a diagonal linear transformation. We do not assume that this steps satisfy the Nyquist condition in Corollary 3.2. Assume also that we use formula (25). Following the proof of Theorem 3.1, (17) still holds but the periodic extension $\hat{f}_{\mathrm{ext}}$ consists of a superposition of periodic shift of $\hat{f}_h$ w.r.t. to the group $(2\pi/s)\mathbf{Z}^n$:

$$\hat{f}_{\mathrm{ext}}(\xi) = \sum_{k\in\mathbf{Z}^n} \hat{f}(\xi + (2\pi/s)k). \tag{29}$$

Now they can overlap and the restriction of $\hat{f}_{\mathrm{ext}}$ to $[-\pi/s, \pi/s]^n$ is not $\hat{f}$ anymore. Then (18) needs to be corrected by replacing the l.h.s. by $\hat{\chi}(s\xi)\hat{f}_{\mathrm{ext}}(\xi)$. In the semiclassical case, the proof of Theorem 3.2 shows that the r.h.s. of (25) approximates not f_h but

$$Gf_h := \mathcal{F}_h^{-1}\hat{\chi}(s\cdot)(\mathcal{F}_h f_h)_{\mathrm{ext}}. \tag{30}$$

From now one, we assume $\hat{\chi} \in C_0^\infty$. We see that high (outside $[-\pi/s, \pi/s]^n$) frequencies ξ will be shifted by $(2\pi/s)k$ for some $k \in \mathbf{Z}^n$ so they land in that box and their amplitudes will be added to the ones with that actual frequency there. This is known as "folding" of frequencies. Note also that for f_h real valued, $\mathrm{WF}_h(f)$ and $\Sigma_h(f)$ are even, i.e., symmetric under the transform $\xi \mapsto -\xi$.

In fact, $G = \sum_{k\in\mathbf{Z}} G_k$, where each G_k is an h-FIO with a canonical relation given by the shifts

$$S_k : (x, \xi) \longmapsto (x, \xi + 2\pi k/s). \tag{31}$$

Indeed, we have

$$\begin{aligned} G_k f(x) &= (2\pi h)^{-n} \int e^{\mathrm{i}(x-y)\cdot\xi/h - 2\pi\mathrm{i}k\cdot y/sh}\hat{\chi}(s\xi)f(y)\,\mathrm{d}y\,\mathrm{d}\xi \\ &= e^{-2\pi\mathrm{i}k\cdot x/sh}\hat{\chi}(shD + 2k\pi)f. \end{aligned} \tag{32}$$

This FIO preserves the space localization (as it is clear from (31)) but shifts the frequencies, which can be viewed as changing the direction and the magnitudes of the latter. This FIO view of aliasing will be very helpful later. Note that in the classical case, when χ is a product of sinc functions, $\hat{\chi}$ is not smooth, and the aliasing then in not an FIO (classical or not). The resulting aliasing is not "completely local" then as it is evident from the fact that sinc has a slow decay at infinity. When we oversample a bit, we can choose $\hat{\chi}$ to be smooth however; then $\chi \in \mathcal{S}$ and aliasing is local in the sense of rapid decay.

In applications, other reconstructions are used, for example splines. Without a proof, we will mention that the aliasing artifacts are similar then. Regardless of the reconstruction used, aliasing is loss of information.

As an example, we plot the function f_h consisting of a sum of the real parts of a sum of two coherent states, see (7), with $h = 0.01$ and $h = 0.04$, respectively and unit ξ_0's, on the rectangle

$[-1,1]^2$. If we take $h = 0.1$ as the small parameter, then the higher frequency state requires a sampling rate sh with $s < \pi$, therefore, it needs more than $2/(\pi h) \approx 64$ sampling points in each variable. The lower frequency state requires about $1/4$ of that. In Figure 1, we sample on a 81×81 grid fist, and then on a 41×41 one next. In the first case, both patterns are oversampled, while in the second one, only one is, and the other one is undersampled. We reconstruct the image with the `lanczos3` interpolation in MATLAB, which is a truncated version of the sinc reconstruction; and since the truncation happens at zeros of the sinc function, at least jumps are avoided, and the resulting functions can be used as an approximation of the more general interpolation functions in (25). This reconstruction is shown in Figure 1b; the aliasing of the smaller pattern changes the direction and the magnitude of the frequency. In the third and the fourth plots, we show the absolute values of the Fourier transforms of both images: the oversampled one is plotted in its Nyquist box while the white box is the Nyquist box of the undersampled one next to it. The absolute value of the Fourier transform of the undersampled image is plotted last. One can see that the frequencies outside the Nyquist box in the previous case have shifted by $(0, -2B)$ and $(0, 2B)$, where $2B$ is the side of the white box.

Figure 1. Aliasing: from left to right: (a) the original $f_{\rm o}$ on a 81×81 lattice; (b) the undersampled $f_{\rm u}$ with on a 41×41 lattice: the lower left pattern is undersampled, the larger one is oversampled; (c) $|\hat{f}_{\rm o}|$ with the white box representing the Nyquist limit in the next plot; (d) $|\hat{f}_{\rm u}|$, the white box is the Nyquist limit.

4. Sampling classical FIOs

We are ready to formulate the main results of this paper concerning sampling classical FIOs. We label them by (i), (ii), (iii) and (iv) as in the Introduction. We formulate some of them as theorems; and some not, to simplify the exposition.

4.1. **(i) Sampling Af.** Let A be a classical FIO as in Theorem 2.2. If we know a priori that f is semiclassically band limited, then so is Af and we can find a sharp upper bound on its band limit. This determines a sharp sampling rate for Af by Theorem 3.2.

Theorem 4.1. *Let A be a classical FIO as in Theorem 2.2 with canonical relation C. Let f_h be semiclassically band limited. Assume that $C \circ \mathrm{WF}_h(f) \subset \Omega \times \mathcal{B}$ with Ω a bounded domain and $\mathcal{B}$ open and bounded. If W is as in Theorem 3.2, then Af can be reconstructed in Ω up to an $O(h^\infty)$ error from its values on the lattice $shWk$, $k \in \mathbf{Z}$, $0 < s \le 1$ in the sense of (20), (21).*

If we do not require uniformity of the error in Af, one can take $s = 1$.

4.2. **(ii) Resolution limit of f given the sampling rate of Af.** Assume now that we sample Af at a rate $s_j h$ in the x^j-th variable with some fixed $s_j > 0$; or on a more general periodic lattice. What resolution limit does this impose on f?

By Corollary 3.1, to avoid aliasing, $\Sigma_h(Af)$ should be in the box $[-B_j, B_j]$ with $B_j > \pi/s_j$ (we assume that we deal with real valued functions, and therefore always work in frequency sets symmetric about the origin). Then if

$$\pi_2 \circ C \circ \mathrm{WF}_h(f) \subset \prod_j (-B_j, B_j), \tag{33}$$

there is no loss (up to $O(h^\infty)$ when the data Af has been sampled. Note that the canonical relation C does not need to be an 1-to-1 map. If C is a local diffeomorphism and if A is elliptic, then (33) is sharp in the sense that if it is violated, f cannot be recovered up to $O(h^\infty)$, i.e., there will be aliasing. This happens for the 2D Radon transform, for example.

We write (33) in a different way:

$$\mathrm{WF}_h(f) \subset C^{-1} \circ \Big(\mathbf{R}^m \times \prod_j (-B_j, B_j) \Big), \tag{34}$$

assuming that A takes values in $\mathbf{R}^m$. Similarly to C, the inverse canonical relation C^{-1} does not need to be an 1-to-1 map.

Relation (34) gives easily a sharp limit (sharp when A is elliptic, associated to a local diffeomorphism) on $\Sigma_p(f)$ guaranteeing no aliasing. It actually says something more: the resolution limit on f is microlocal in nature, i.e., it may depend on the location and on the direction. We illustrate this below with the Radon transform.

4.3. **(iii) Aliasing artifacts.** When (34) is violated and when A is elliptic, associated to a local diffeomorphism, for example, there will be aliasing of Af. To understand how this affects f, if we use back-projection, i.e., a parametrix A^{-1}, we recall that the aliasing can be interpreted as an h-FIO associated to shifts of the dual variable, see (31) and (32). Then the inversion would be $A^{-1}G_k A$; and by the h-FIO calculus, away from zero frequencies, that is an h-FIO with a canonical relation $C^{-1} \circ S_k \circ C$ acting on $(x, \xi) \in \operatorname{supp} \hat\chi(s \cdot + 2k\pi)$. We will not formulate a formal theorem of this type; instead we will illustrate it in the examples in the next sections. The classical aliasing described in Section 3.3 creates artifacts at the same location but with shifted frequencies. The artifacts here however could move to different locations, as it happens for the Radon transform, for example.

Note that A^{-1} has canonical relation C^{-1} but the latter acts on the image of $T^*\Omega \setminus 0$ if a priori $\operatorname{supp} f \subset \Omega$ and we restrict the reconstruction there. Some of the shifted singularities of Af may fall outside that domain of C and they will create no singularities in the reconstruction. Therefore, we may have shifts in the reconstructed f, full or partial cancellation (and interference patterns as a result), and removal of singularities even if their images are still present in Af.

4.4. **(iv) Locally averaged sampling.** We study now what happens when the measurements Af are locally smoothened either to avoid aliasing or for some practical reasons. One way to model this is to assume that we are given samples of $\phi_h * Af$, where $\phi_h = h^{-m}\phi(\cdot/h)$, where m is the dimension of the data space, and ϕ is smooth with $\int \phi = 1$. If $\hat\phi(\eta)$ is approximately supported in the ball $|\eta| \leq B'$, how to sample $\phi_h * Af$ and what does this tell us for f? The answer to the first question is given by the sampling theorems — we need to sample at rate smaller than $\pi h/B'$ assuming that Af a priori has even higher frequencies than B'. The second question is more interesting. As explained in the Introduction, we have that $\phi_h *$ is a h-ΨDO and if A is associated

to a canonical map, then and by Egorov's theorem, $\phi_h * Af = AP_hf + O(h^\infty)f$, where P_h is a h-ΨDO of order zero with principal symbol $\hat{\phi}\circ C$. If A is well posed, we can recover P_hf up to small error which is a certain regularized version of f.

We can make this more general. Assume that the convolution kernel can depend on the sampling point. We model that by assuming that we are sampling Q_hAf, where Q_h is an h-ΨDO of order zero with essential support of its symbol contained in (y,η) for which $|\eta|\le B'$. This hides the implicit assumption that the convolution kernels cannot change rapidly when we make h smaller and smaller because the symbol $q(y,\eta)$ of Q_h must satisfy the symbol estimate (12), and in particular, $\partial_x p$ cannot increase with h^{-1}. Then it is not hard to see that at every sampling point $y=y_j$, $Q_hAf(y_j)$ is the Fourier multiplier with $q(y_j,\eta)$ restricted to the point $y=y_j$. Therefore, each measurement is really a convolution. An application of the semiclassical Egorov theorem [14] combined with the remark following Theorem 2.2 yields the following.

Proposition 4.1. *Let f be semiclassically band limited. For $\varepsilon>0$, let $f=f_1+f_2$ with $\Sigma_h(f_1)\subset\{|\xi|\le\varepsilon\}$ and $\Sigma_h(f_2)\subset\{|\xi|\ge 2\varepsilon\}$. Let F be a classical FIO associated with a diffeomorphism C and let P_h be a h-ΨDO. Then $Q_hFf_1=FP_hf_1+O(h^\infty)f_1$ with P_h a h-ΨDO with principal symbol $q_0=p_0\circ C$, where q_0 is the principal symbol of Q_h. Moreover, the full symbol p of P_h is supported in $C^{-1}(supp(q))$, where q is the full symbol of Q_h.*

Therefore, having locally averaged data Q_hAf instead of Af, with A elliptic, allows us to reconstruct the smoothened P_hf (plus a function with much lower frequencies) instead of f. If we want to choose P_h first, for example to be a specific convolution, we can find the operator Q_h which applied to the data Af results in the desired regularization. Finally, in applications, $Q_hA(y_j)$ may not be realized as convolutions with compact supports of their Fourier transforms but if those supports are approximately compact with some error estimates, we can apply the asymptotic isometry property (26) to estimate the resulting error.

5. Non-uniform sampling: A lower bound of the sampling rate

Assume we want to sample a semiclassically limited f_h on a non-uniform grid, see also Section 3.2. One reason to do that would be to reduce the number of the sampling points if the shape of $\mathrm{WF}_h(f_h)$ allows for this, as in the Radon transform examples (where we sample $\mathcal{R}f$). We prove a theorem similar to one of the results of Landau [18] in the classical case. In case of non-uniform sampling, we establish a lower bound of the sampling rate of f with $\mathrm{WF}_h(f)$ contained in a fixed compact set, equal to the phase volume of the interior of the latter. In most applications, $\partial\mathcal{K}$ would have measure zero, and then $\mathrm{Vol}(\mathcal{K}^{\mathrm{int}})=\mathrm{Vol}(\mathcal{K})$. The theorem below is different that the corresponding results in [18] in the following way. Aside from being semiclassical, our bound is in terms of the volume of $\mathrm{WF}_h(f)$, i.e., it is microlocalized, rather than being $(2\pi)^{-n}\mathrm{Vol}(\Omega)\mathrm{Vol}(\mathcal{B})$ for the number of points in every Ω when $\mathrm{supp}\,\hat{f}\subset\mathcal{B}$. In other words, we express the bound in terms of the volume of $\mathrm{WF}_h(f_h)$ instead of the volume of the minimal bounding product $\Omega\times\mathcal{B}$.

Theorem 5.1. *Let $\{x_j(h)\}_{j=0}^{N(h)}$ be a set of points in $\mathbf{R}^n$. Let $\mathcal{K}\subset T^*\mathbf{R}^n$ be a compact set. If*

$$\|f_h\|^2\le\frac{C}{N(h)}\sum_{j=1}^{N(h)}|f_h(x_j(h))|^2+O(h^\infty) \tag{35}$$

for every semiclassically band limited f_h with $\mathrm{WF}_h(f)\subset\mathcal{K}$, then

$$N(h)\ge(2\pi h)^{-n}\mathrm{Vol}(\mathcal{K}^{\mathrm{int}})(1-o(h)), \tag{36}$$

where $\mathrm{Vol}(\mathcal{K}^{\mathrm{int}})=\int_{\mathcal{K}^{\mathrm{int}}}\mathrm{d}x\,\mathrm{d}\xi$ is the measure of the interior $\mathcal{K}^{\mathrm{int}}$ of $\mathcal{K}\subset T^*\mathbf{R}^n$.

Proof. By the properties of the Lebesgue measure, given $\delta > 0$, we can find a closed (and necessarily compact in this case) set $F_\delta \subset \mathcal{K}^{\text{int}}$ so that $\text{Vol}(\mathcal{K}^{\text{int}} \setminus F_\delta) < \delta$. Let $0 \le p(x,\xi) \le 1$ be a real valued smooth function supported in $\mathcal{K}^{\text{int}}$ and equal to 1 on F_δ. Let $P_h = p^{\text{w}}(x, hD)$ be the Weyl quantization of p. Then P_h is self-adjoint and compact.

Let E_h be the eigenspace of P_h spanned by the eigenfunctions corresponding to the eigenvalues in $[c, 2]$ with some $c \in (0, 1)$. The upper bound 2 can be replaced by any number greater than 1 since the eigenvalues of P_h cannot exceed $1 + O(h)$. By Sard's theorem, we can choose c so that it is a non-critical value for p. Every $f_h \in E_h$ satisfies $\text{WF}_h(f_h) \subset \mathcal{K}$. Indeed, for every unit eigenfunction ϕ_h of P_h with an eigenvalue $\lambda_h \in [c, 2]$, and for every $q \in S^{0,0}$ with $\operatorname{supp} q \cap \mathcal{K} = \emptyset$ we have $q(x, hD)\phi_h = (1/\lambda_h)q(x, hD)p(x, hD)\phi_h = O(h^\infty)$, and this yields the same conclusion for every tempered $f_h \in E_h$.

By [8], we have the following Weyl asymptotic

$$\dim E_h = (2\pi h)^{-n} \text{Vol}\{c \le p \le 2\}(1 + O(h)). \tag{37}$$

We will show that $N(h) \ge \dim E_h$ for $0 < h \ll 1$. If this is not true, we would have $N(h) < \dim E_h$ for a sequence $h = h_j \to 0$. Then for every $h \in \{h_j\}$, we can find a unit $f_h \in E_h$ so that all the samples of f_h vanish for such h. As we showed above, $\text{WF}_h(f_h) \subset \mathcal{K}$. This contradicts (35) however.

Therefore, by (37),

$$\text{Vol}(F_\delta) < \text{Vol}\{c \le p \le 2\} \ \le \ \liminf_{h\to 0} (2\pi h)^n N(h).$$

Take the limit $\delta \to 0$ to complete the proof. □

Remark 5.1. The proof holds if we replace the error term in (35) by $o(h)\|f\|$.

Remark 5.2. Existence and a characterization of optimal sampling sets where (36) would be an equality is a harder problem which we do not study here. Estimate (36) is sharp in case of uniform sampling described in Corollary 3.1 at least, see Remark 3.3 which generalizes easily to different band limits B_j for each component of x as in Corollary 3.1. It is straightforward to show that (36) is also sharp in the case of more general uniform sampling described in Theorem 3.2.

Remark 5.3. The statement of the theorem is preserved under (non-linear) diffeomorphic transformations because WF_h and its phase volume are invariant. If for some $\mathcal{K}$, we can choose a non-linear transformation which would fit the problem in the situation handled by Theorem 3.2, we can construct a sampling set with the optimal number of sampling points by transforming the periodic lattice by that transformation. Doing this piecewise, as suggested in Section 3.2, would provide a smaller sampling set. We show how this can be done in our Radon transform examples.

Corollary 5.1. *Let A be a classical FIO of order m associated with a diffeomorphic canonical relation C. Then the minimal asymptotic number of points (up to an $o(h)$ relative error) to sample Af_h guaranteed by Theorem 5.1 does not exceed the number of points needed to sample f_h; and if A is elliptic, it is the same.*

The proof follows from the fact that C is symplectic and in particular it preserves the phase volume; and from Theorem 2.2.

In the examples we consider, A happened to be an 1-to-2 diffeomorphism, and each branch is elliptic. Then we can apply the corollary to each "half" of C. Then the number of points to sample Af stably would be twice that for f; but the number of points to recover f stably up to a function with WF_h in a small neighborhood of the zero section is half of that.

In the remainder of the paper, we present a few applications.

6. The X-ray/Radon transform in the plane in the parallel geometry

We present the first example: the X-ray/Radon transform $\mathcal{R}$ in the plane in the so-called parallel geometry parameterization. The analysis of this transform in this and in the fan-beam representation can and has been studied with traditional tools [2,6,25,28], also [24, Ch. III], when the weight is constant. The sampling of $\mathcal{R}f$ requires estimates of the Fourier transform of $\mathcal{R}f$, which is done in those papers in a non-rigorous and approximate way by Bessel functions expansions and their asymptotics. An actual proof would require the use of two-parameter asymptotics of Bessel functions as in [26]. The statements are actually asymptotic which is not made clear. An alternative approach, not based on sampling theory, was proposed recently by Katsevich in [16, 17, 17]. The goal in those paper is to analyze the reconstruction of jump type singularities.

We go a bit deeper than that even when the weight is constant and we treat variable weights as well. The main purpose of this section is to demonstrate the general theory on a well studied transform, where one can write explicit formulas; and some sampling analysis has been done (for constant weights), so we can compare the results.

The numerical simulations in this and in the next section have been done in Matlab. The phantoms f are defined by formulas and sampled first on a very fine grid. Then we compute their Radon transforms $\mathcal{R}f$ numerically. To simulate coarser sampling of $\mathcal{R}f$, we sample the so computed $\mathcal{R}f$. To simulate inversion with coarsely sampled $\mathcal{R}f$, we upsample the downsampled data to the original grid to simulate a function of continuous variables. Instead of using the Whittaker-Shannon interpolation formula (25), we use the `lanczos3` interpolation which is a truncated version of the latter. Then we perform the inversion on that finer grid. Note that our goal is not to reduce the computational grid at this point, it rather is to show the amount of data and the artifacts contained in data sampled in a certain way. We compute the Fourier transforms of f and $\mathcal{R}f$ using the discrete Fourier transform command in Matlab. Since we work with f vanishing near the boundary of the square $[-1,1]^2$, and $\mathcal{R}f$ is vanishing in the p variable near $p=\pm 1$ (for such f's) and is periodic in its angular variables, the discrete Fourier transform, which in fact is a transform on a torus, creates no artifacts.

6.1. $\mathcal{R}_\kappa$ as an FIO.

Let $\mathcal{R}_\kappa$ be the weighted Radon transform in the plane

$$\mathcal{R}_\kappa f(\omega,p)=\int_{x\cdot\omega=p}\kappa(x,\omega)f(x)\,\mathrm{d}\ell, \tag{38}$$

where κ is a smooth weight function, $p\in\mathbf{R}$, $\omega\in S^1$, and $\mathrm{d}\ell$ is the Euclidean line measure on each line in the integral above. If $\kappa=1$, we write $\mathcal{R}=\mathcal{R}_\kappa$. Each line is represented twice: as (ω,p) and as $(-\omega,-p)$ but it is represented only once as a directed one. In general, the weight does not need to be even in the ω variable, so it is natural to think of the lines as directed ones. Let $\omega^\perp$ be ω rotated by $\pi/2$. We parameterize ω by its polar angle φ and write

$$\omega(\varphi)=(\cos\varphi,\sin\varphi). \tag{39}$$

The Schwartz kernel of $\mathcal{R}_\kappa$ is $\kappa\delta(x\cdot\omega(\varphi)-p)$. Then it is straightforward to show that $\mathcal{R}_\kappa$ is an FIO of order $-1/2$ with canonical relation

$$C=\left\{\Big(\varphi,\underbrace{x\cdot\omega(\varphi)}_{p},\underbrace{-\lambda(x\cdot\omega^\perp(\varphi))}_{\hat\varphi},\underbrace{\lambda}_{\hat p},x,\underbrace{\lambda\omega(\varphi)}_{\hat x=\xi}\Big),\ \lambda\neq 0\right\}, \tag{40}$$

where we used the non-conventional notation of denoting the dual variable of p by $\hat p$, etc. Set $\xi=\lambda\omega$. Given $\xi\neq 0$, there are two solutions for λ, ω: either $\lambda_+=|\xi|$, $\omega=\xi/|\xi|$ or $\lambda_-=-|\xi|$,

$\omega = -\xi/|\xi|$. Therefore, $C = C_+ \cup C_-$, where

$$C_\pm : (x,\xi) \longmapsto \Big(\underbrace{\arg(\pm\xi)}_{\varphi}, \underbrace{\pm x\cdot\xi/|\xi|}_{p}, \underbrace{-x\cdot\xi^\perp}_{\hat\varphi}, \underbrace{\pm|\xi|}_{\hat p} \Big). \tag{41}$$

The Schwartz kernel of $\mathcal{R}_\kappa$ is a delta function on $Z := \{x\cdot\omega(\varphi) = p\}$ which is invariant under the symmetry $\tau(\varphi,p) = (\varphi+\pi, -p)$ (the latter modulo 2π). Then C is invariant under τ lifted to the cotangent bundle:

$$(\varphi, p, \hat\varphi, \hat p) \quad\longmapsto\quad (\varphi+\pi, -p, \hat\varphi, -\hat p), \tag{42}$$

and in fact this is an isomorphism between C_+ and C_-.

Take C_+ first. We see that a singularity (x,ξ) of f can create a singularity of $\mathcal{R}_\kappa f(p,\omega)$ at $p = x\cdot\xi/|\xi|$ and $\omega = \xi/|\xi|$, i.e., at $\varphi = \arg(\xi)$ in the codirections $|\xi|(-(x\cdot\omega^\perp)\omega^\perp, 1)$. Note first that such (p,ω) determines the oriented line through x normal to ξ and the normal is consistent with the orientation. Taking C_- next, we see that (x,ξ) may affect the wave front set of $\mathcal{R}_\kappa$ at $p = -x\cdot\xi/|\xi|$ and $\omega = -\xi/|\xi|$, and at the corresponding codirections. That is the same line as before but with the opposite orientation and the weight κ on it might be different.

So we see that $\mathcal{R}_\kappa$ is an FIO associated with $C_-\cup C_+$ and each one of them is a local diffeomorphism. Indeed, $(x,\xi) = C_\pm^{-1}(\varphi,p,\hat\varphi,\hat p)$ is given by

$$x = p\omega(\varphi) - (\hat\varphi/\hat p)\omega^\perp(\varphi), \quad \xi = \hat p\omega(\varphi). \tag{43}$$

It is well defined for $\hat p \neq 0$ but if we want x in the image to be in $|x| < R$, we need to require $p^2 + (\hat\varphi/\hat p)^2 < R^2$, see also (46) and Figures 2 and 3.

6.2. (i) Sampling $\mathcal{R}_\kappa f$.

6.2.1. *Sampling $\mathcal{R}_\kappa f$ on periodic grids.* Assume that $f = f_h$ satisfies

$$\mathrm{WF}_h(f) \subset \{(x,\xi);\ |x| < R,\ |\xi| < B\} \tag{44}$$

with some $R > 0$, $B > 0$, i.e., up to $O(h^\infty)$, f is essentially supported in $B(0,R)$ and $\mathcal{F}_h f$ is essentially supported in $|\xi| < B$. The number N_F of points to sample f stably has a lower bound

$$N_f \sim (2\pi h)^{-2}\pi R^2 \times \pi B^2 = h^{-2}R^2B^2/4, \tag{45}$$

where $\sim$ means equality up to an $(1+o(h))$ relative error, see Theorem 5.1 and Remark 5.2.

A sharp upper interval containing $\hat p$ is $[-|\xi|, |\xi|]$ and a sharp interval for $\hat\varphi$ is $[-R|\xi|, R|\xi|]$, therefore, the smallest rectangle containing $\mathcal{F}_h\mathcal{R}_\kappa f$ for all such possible f's is

$$(\hat\varphi,\hat p) \in RB[-1,1]\times B[-1,1].$$

Note that this rectangle does not describe $\Sigma_h(\mathcal{R}_\kappa f)$ when $\Sigma_h(f)$ is included in $|\xi| \le B$. The latter is the cone $|\hat\varphi| \le R|\hat p|$ in that rectangle, i.e.,

$$\{(\hat\varphi,\hat p);\ |\hat\varphi| \le R|\hat p|,\ |\hat p| \le B\}, \tag{46}$$

see Figure 2. Suppose we sample on a rectangular grid with sampling rates $s_\varphi h$ in the φ variable in $[0,2\pi]$; and with a sampling rate $s_p h$ in the p variable in $[-R,R]$. Then the Nyquist condition is equivalent to

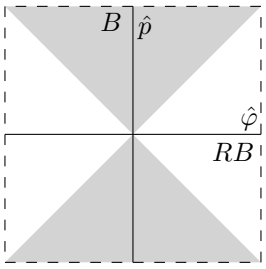


Figure 2. The frequency set of $\mathcal{R}_\kappa f$.

$$s_\varphi \le \frac{\pi}{RB}, \quad s_p \le \frac{\pi}{B}. \tag{47}$$

This means taking $2RB/h \times 2RB/(\pi h) = h^{-2}4R^2B^2/\pi$ samples to recover $\mathcal{R}_\kappa$ approximately, see also [24]. Note that this is $8/\pi$ times the estimate in (45) for $2N_f$, which by Corollary 5.1 is the theoretical asymptotic minimum since the canonical relation C is 1-to-2. For the recovery of $\mathrm{WF}_h(f) \setminus 0$ we need half of those samples, i.e., $\sim h^{-2}2R^2B^2/\pi$. This is again $8/\pi$ times the asymptotic minimum in Corollary 5.1.

On the other hand, the conic shape of the range of $\mathrm{WF}_h(f)$ in (46) suggests a more efficient sampling, see also [24,28]. One can tile the plane with that set by using the elementary translations by (RB, B) and by $(0, 2B)$. If $2\pi(W^*)^{-1}$ has those columns, then

$$W = \frac{\pi}{RB}\begin{pmatrix} 2 & -1 \\ 0 & R \end{pmatrix}. \tag{48}$$

Then one should sample on a grid $shW\mathbf{Z}^n$ with $s < 1$, by Theorem 3.2. Since $\det W = 2\pi^2/RB^2$, and the area of the region we sample is $2\pi \times 2R = 4\pi R$, we see that we need $\sim h^{-2}2R^2B^2/\pi$ points which is $4/\pi$ larger than the phase volume of $\mathcal{R}_\kappa f$ for all such possible f; and for proper recovery of $\mathrm{WF}_h(f)$ we need a half of that, i.e., $\sim 4/\pi$ times N_f. The coefficient $4/\pi$ is close to 1 but not equal to 1 as it is clear from the next section and Figure 3.

6.2.2. *Microlocalization and non-uniform sampling.* The sampling requirements above were based on the following. To determine the sampling rate for $\mathcal{R}_\kappa f$, we find $\Sigma_h(R_\chi f)$ as the projection of $\mathrm{WF}_h(\mathcal{R}_\kappa f) \subset C \circ \mathrm{WF}_h(f)$ to its phase variable. Since we are interested in sampling f with $\mathrm{WF}_h(f) \subset \{|x| \le R,\ |\xi| \le B\}$ and to find a sharp sampling rate, we projected $C \circ \{|x| \le R,\ |\xi| \le B\}$ onto its phase variable to get the smallest closed set (46) containing $\Sigma_h(\mathcal{R}_\kappa f)$ for every such f. This answers the question if we are interested in sampling on a periodic grid for all such possible f's. The analysis allows us to localize or microlocalize some of those arguments.

The dependence of the sampling requirements on $\operatorname{supp} f$**.** The sampling frequency in the angular variable on a rectangular grid should be smaller than $\pi h/(RB)$, and the dependence on R may look strange since the Radon transform has a certain translation invariance. The reason for it is that we assume that we know that f is supported in a disk and we reconstruct it there only. Numerical experiments confirm that: when the sampling rate in the angular variable decreases, artifacts do appear and they move closer and closer to the original when the rate decreases.

Non-uniform sampling. We are interested first in the optimal sampling rate for $\mathcal{R}_\kappa f$ locally, near some (φ, p). The latter is determined by the frequency set $C \circ \{|x| \le R,\ |\xi| \le B\}$ projected to its phase variables $(\hat\varphi, \hat p)$ with (φ, p) fixed. It is straightforward to see that on the range of C,

$$|\hat\varphi| = \pm|\hat p|\sqrt{|x|^2 - p^2}, \quad |\hat p| \le B, \quad \text{where } |p| \le |x|. \tag{49}$$

Since x ranges in $|p| \le |x| \le R$, we get

$$|\hat\varphi| \le |\hat p|\sqrt{R^2 - p^2}, \quad |\hat p| \le B. \tag{50}$$

We plot those double triangles in Figure 2 at a few points in the rectangle $[0, 2\pi] \times [-1, 1]$ (where $\varphi = 0$ and $\varphi = 2\pi$ should be identified) in the (φ, p) plane. The phantom f consists of six small Gaussians in the unit disk ($R = 1$). We also superimpose a density plot of $\mathcal{R}f$. One can see from this figure that the set of the conormals to the curves in the plot of $\mathcal{R}f$ fall inside those triangles but the semiclassical wave front set also captures the range of the magnitudes of the frequencies. When the stripes are horizontal, the magnitude drops a bit and the stripes a bit more blurred. Since $R = 1$ in this example, the Nyquist sampling limits of the sampling rates $s_\varphi h$ and $s_p h$ given in (47) are equal, i.e., the optimal grid would be a square one.

This analysis suggests the following non-uniform sampling strategy. For a fixed k, we divide the interval $[-1, 1] \ni p$ into $2k$ sub-intervals $I_j^\pm = \pm[(j-1)/k, j/k]$, $j = 1, 2, \dots, k$, each one of length

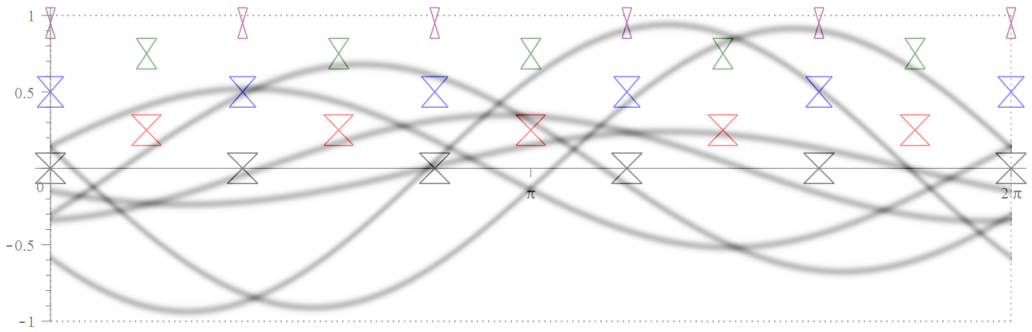


Figure 3. $\mathcal{R}f(\varphi, p)$ for f consisting of six randomly placed small Gaussians in the unit ball. The double triangles, shown in the upper half only, represent $\mathrm{WF}_h(\mathcal{R}f)$ localized at their vertices (φ, p).

$1/k$. For each k, we take $|\hat{\varphi}| \le |\hat{p}|\sqrt{R^2 - j^2/k^2}$, $|\hat{p}| \le B$ as a sharp cone where $(\hat{\varphi}, \hat{p})$ lies, see (49). Then in $[0, 2\pi] \times I_j$, we sample on the grid $\delta h W_j \mathbf{Z}^n$, where $\delta < 1$ is fixed and W_j is as in (48) with $R = 1/j$. Then we can get closer and closer to the sharp number of the sampling point for $\mathcal{R}_\chi$ stated in Corollary 5.1 and Theorem 5.1, which should be $\sim 2N_f$ for a stable recovery of $R_\kappa f$ and N_f for a stable recovery of f itself, see (45).

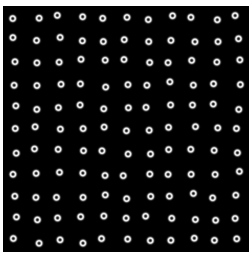
6.3. **(ii) Resolution limit on f posed by the sampling rate of $\mathcal{R}_\kappa f$.** Let s_φ and s_p be the relative sampling rates for φ and p, respectively. Lack of aliasing is equivalent to $\hat{\varphi} \le \pi/s_\varphi$, $\hat{p} < \pi/s_p$, see (33), (34). By (41), this is equivalent to

$$|x \cdot \xi^\perp| \le \pi/s_\varphi, \quad |\xi| \le \pi/s_p. \tag{51}$$

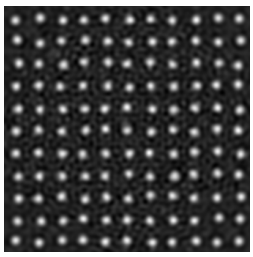
If the sampling rates satisfy the sharp Nyquist condition (47), the latter condition above implies the former. Actually, the first condition in (51) is most critical for (x, ξ) with x close to the boundary $|x| = R$ and $\xi \parallel x$, which are represented by radial lines close to $|x| = R$. In Figure 4b, which is undersampled in φ, we see evidence of that; another evidence is Figure 7c, where $\mathcal{R}f$ is blurred in φ.

When the sampling rates do not necessarily satisfy the Nyquist condition (47) in $|x| < R$, we illustrate the significance of (51) in Figure 4. The relative sampling rate s_p imposes a universal limit on the resolution, independent on x and the direction of ξ. In Figure 4b, this is demonstrated a function f in the square $[-1, 1]^2$ with $s_p = 1/30$ (and $h = 1$), i.e., the square is fully covered by 60 horizontal or vertical lines. On the other hand, the second inequality imposes a locally non-uniform and a non-isotropic resolution limit. Assuming $s_p \ll 1$ (which is true in practical applications), in optics terms, the resolution of saggital (radial) lines deteriorates gradually away from the center; there, $|x \cdot \xi/|\xi|| \ll 1$, so $|x \cdot \xi^\perp/|\xi^\perp||$ is close to its maximum for that x, which restricts $|\xi|$ by (51). Resolution of meridional (circular) lines is the greatest; there $|x \cdot \xi^\perp/|\xi^\perp|| \ll 1$, so for a given x, $|\xi|$ could be large by (51). In Figure 4c, we choose an angular step $s_\phi = 4$ degrees, i.e., 90 angles of view. In both cases, there are also aliasing artifacts explained below.

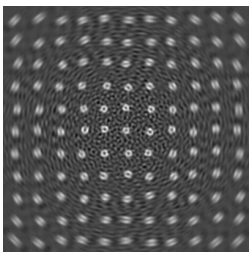
6.4. **(iii) Aliasing.** We study now what happens if $\mathcal{R}_\kappa f$ is undersampled. It might be undersampled in the φ or the p variable or in both.



(a) f on $[-1,1]^2$



(b) Reconstructed f undersampled in p.



(c) Reconstructed f, angular step $s_\varphi = 4^\circ$

Figure 4. (a) f on $[-1,1]^2$, (b) f reconstructed with $s_\varphi \ll 1$ and undersampled in the p variable; (c) f reconstructed with $s_p \ll 1$ and an angular step $s_\varphi = 4$ degrees.

6.4.1. *Angularly undersampled $\mathcal{R}_\kappa f$.* Assume (44) as above and assume that $s_\varphi > \pi h/(RB)$, i.e., the first Nyquist condition in (47) is violated. Then the aliasing of $\mathcal{R}_\kappa f$ can be described as a sum of h-FIOs, see (30), (31), with canonical relations

$$S_k : (\hat p, \hat\varphi) \longmapsto (\hat p, \hat\varphi + 2\pi k/s_\varphi), \quad k = 0, \pm 1, \dots . \tag{52}$$

In typical cases with not very severe undersampling, k is restricted to $k = \pm 1$ plus $k = 0$ which is the original image but blurred by $\hat\chi$ in (30). A direct computation shows that the aliasing artifacts are described by an h-FIO with canonical relations

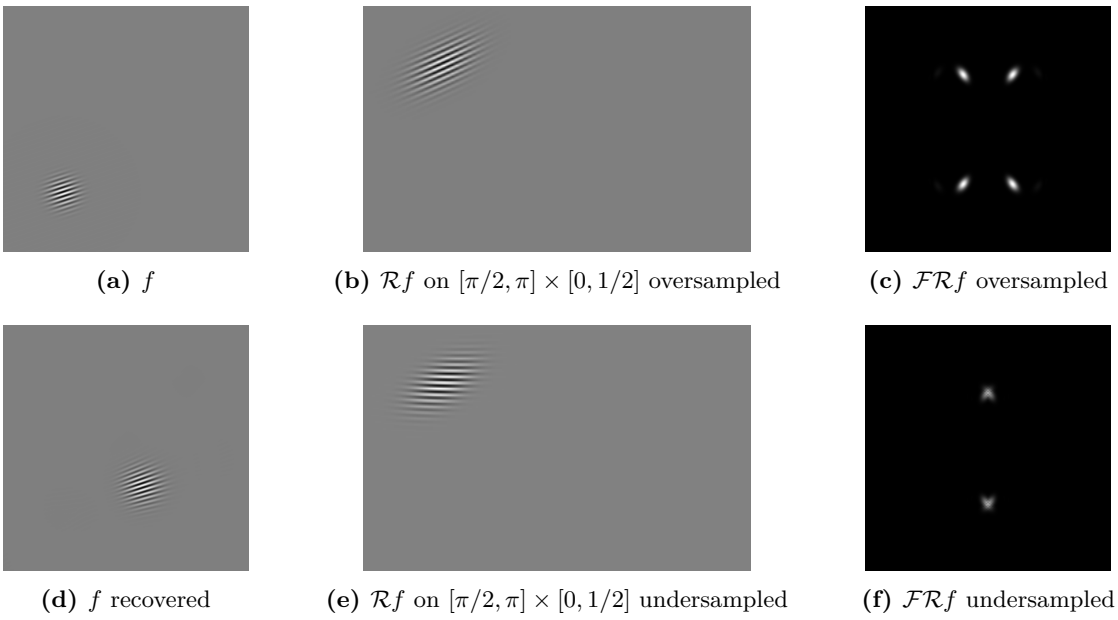
$$(x,\xi) \quad \longmapsto \quad C_\pm^{-1} \circ S_k \circ C_\pm(x,\xi) = \Big(x - \frac{2\pi k}{s_\varphi}\frac{\xi^\perp}{|\xi|^2}, \xi\Big) \tag{53}$$

when $\hat\varphi + 2k\pi/s_\varphi \in [-\pi/s_\varphi, \pi/s_\varphi]$, i.e., when

$$-x\cdot\xi^\perp + 2k\pi/s_\varphi \in [-\pi/s_\varphi, \pi/s_\varphi]. \tag{54}$$

Those are shifts of (x,ξ) in the x variable, in the direction of $\xi^\perp$, at distance $2\pi k/(s_\varphi|\xi|)$. By (54), k depends on (x,ξ) and in particular for $|x\cdot\xi^\perp| \ll 1$, we have $k = 0$ only and then there is no aliasing. In general, the reconstructed f will have the singularities of f shifted by (53) for various $k = 0, \pm 1, \dots$, as long as they satisfy (54). The value $k = 0$ corresponds to $\mathrm{WF}_h(f)$ (not shifted). Note that only finitely many of them would stay in the ball $|x| < R$. It is even possible all of them to be outside that ball and $\mathcal{R}_\kappa f$ to be undersampled and therefore aliased. Then the reconstructed image in $|x| < R$ will not have a singularity corresponding to that one.

We illustrate this with a numerical example in Figure 5d. We choose f to be a coherent state as in Figure 1. In Figure 5, we plot f, a crop of its Radon transform $\mathcal{R}f$, oversampled, on $[\pi/2,\pi]\times[0,1/2]$ (the only other significant part is symmetric to it and we do not show it), and the Fourier transform of $\mathcal{F}\mathcal{R}f$. Since $\mathcal{R}f$ is even and real valued, $\mathcal{F}\mathcal{R}f$ has two symmetries. A reconstruction of f with oversampled data, not shown, looks almost identical to f. Next we undersample $\mathcal{R}f$ using a $s_\varphi = 3$ degree step in φ. In Figure 5d, we show the reconstructed f which looks like f shifted along the direction of the pattern. The undersampled $\mathcal{R}f$ used to get reconstruction is shown in (b). Compared to (e), the pattern changed its orientation (and the magnitude of its frequency), similarly to the classical aliasing effect illustrated in Figure 1. The effect on the reconstructed f, see (d), however is very different and in an agreement with (53). In (f), we plot $\mathcal{F}\mathcal{R}f$ where $\mathcal{R}f$ now is the aliased version of the Radon transform of f. We see that the bright spots where $\mathcal{R}f$ is essentially supported have shifted compared to (c): the ones to the left have shifted to the right and vice versa, as explained earlier.



(a) f **(b)** $\mathcal{R}f$ on $[\pi/2,\pi]\times[0,1/2]$ oversampled **(c)** $\mathcal{FR}f$ oversampled

(d) f recovered **(e)** $\mathcal{R}f$ on $[\pi/2,\pi]\times[0,1/2]$ undersampled **(f)** $\mathcal{FR}f$ undersampled

Figure 5. Top: f, $\mathcal{R}f$ and the Fourier transform $\mathcal{FR}f$ of $\mathcal{R}f$ when $\mathcal{R}f$ is angularly undersampled. Bottom: f reconstructed with a back-projection, $\mathcal{R}f$ and $\mathcal{FR}f$ when $\mathcal{R}f$ is angularly undersampled with a 3 degrees step: the pattern has shifted.

In this case, only the values $k=\pm1$ in (55) contribute to singularities because the singularity of f does not satisfy (56) with $k=0$, i.e., the original singularity is not within the resolution range.

A similar example, not shown, with the pattern moved close to the center is reconstructed well (see also Figure 4c) is reconstructed well without an artifact even though the artifact computed by (53) would still fit in the square shown. The reason for it is condition (54) which for small $|x|$ and the other parameters unchanged is valid for $k=0$ only.

6.4.2. *$\mathcal{R}f$ undersampled in the p variable.* Assume that s_p is not small enough to satisfy the sampling conditions but s_φ is. The aliasing of $\mathcal{R}_\kappa f$ then can be computed, using (43) and (31), to be

$$(x,\xi) \quad\longmapsto\quad \Big(x - x\cdot\xi^\perp\left(\frac{1}{|\xi|\pm 2\pi k/s_p}-\frac{1}{|\xi|}\right)\frac{\xi^\perp}{|\xi|},\xi+\frac{2\pi k}{s_p}\frac{\xi}{|\xi|}\Big), \tag{55}$$

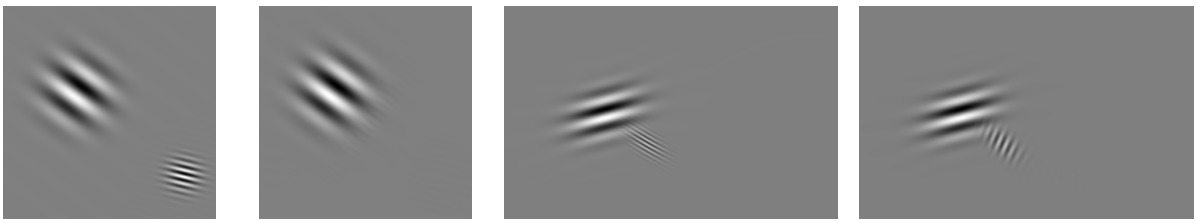
when $\hat p+(2\pi k/s_p)\in[-\pi/s_p,\pi/s_p]$, i.e., when

$$\pm|\xi|+2k\pi/s_p\in[-\pi/s_p,\pi/s_p]. \tag{56}$$

Those are still shifts along $\xi^\perp$ but they are not equally spaced (with k). Also, the magnitude of the frequency changes but the direction does not. In case of mild aliasing, we have $k=\pm1$ (when we are recovering f in $|x|<R$) and they generate shifts of different sizes. In general, there are infinitely many artifacts outside the ball $|x|<R$ regardless of the sampling rate and the band limit of f (our criterion whether Rf is aliased or not depends on R).

In Figure 6, we present an example where one of the patterns disappears from the computational domain $[-1,1]^2$ due to undersampling in the p variable. The other one remains.

6.5. **(iv) Locally averaged measurements.** Assume now that we measure $Q_h\mathcal{R}_\kappa f$ with Q_h an h-ΨDO of order $(0,0)$ (or simply a convolution) limiting the frequency set $\Sigma_h(\mathcal{R}_\kappa f)$ of the data. If



(a) f **(b)** f reconstructed **(c)** $\mathcal{R}f$ on $[0,\pi]\times[-1,1]$ oversampled **(d)** $\mathcal{R}f$ on $[0,\pi]\times[-1,1]$ undersampled

Figure 6. (a) f and (b) reconstructed f with $\mathcal{R}f$ undersampled in p. (c) $\mathcal{R}f$ and (d) $\mathcal{R}f$ aliased.

$q_0(\varphi,p,\hat{p},\hat{\varphi})$ is the principal symbol of Q_h, then by Proposition 4.1, a backprojection reconstructs P_hf where P_h has a principal symbol

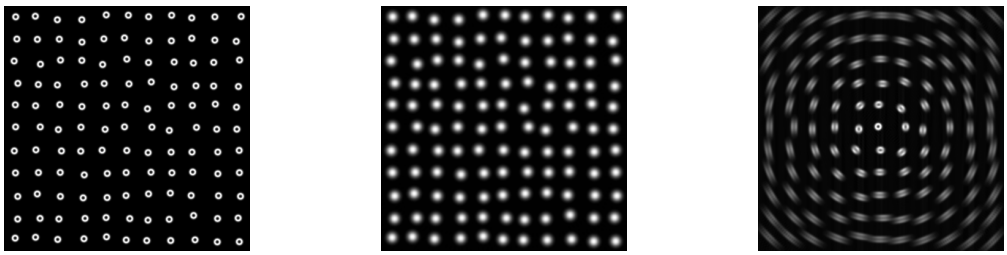
$$p_0(x,\xi)=\frac{1}{2}q_0\circ C_+ + \frac{1}{2}q_0\circ C_-. \tag{57}$$

If, in particular, Q_h is a convolution with a kernel of the type $q_0=\psi(a\hat{\varphi}^2+b\hat{p}^2)$ with $\hat{\psi}\in C_0^\infty$, $a>0$, $b>0$, then

$$p_0(x,\xi)=\psi\big(a|\xi|^2+b|x\cdot\xi^\perp|^2\big). \tag{58}$$

If $\hat{\psi}$ is decreasing, for example, then this symbol takes its smallest values for x near the boundary and $\xi\perp x$, and those are the covectors with the lowest resolution as well. The effect of Q_h is then non-uniform, it blurs f the most at those covectors. If we want a uniform blur, then we choose $p_0=\psi(a|\xi|^2)$ and compute $q_0=\psi(a\hat{p}^2)$. This is not surprising in view of the classical intertwining property $\mathrm{d}_p^2\mathcal{R}_\kappa=\mathcal{R}_\kappa\Delta$ when $\kappa=1$ (true modulo lower order terms for general κ). In other words, only convolving w.r.t. the p variable is needed. This means integrating over "blurred lines". If ψ limits $\mathrm{WF}_h(Q_h\mathcal{R}_\kappa f)$ to, say, $|\hat{p}|\le B'$, then this limits $\hat{\varphi}$ as well by the first inequality on (46), to $|\varphi|\le|\hat{p}|$. Therefore, $(\hat{\varphi},\hat{p})$ are restricted to a smaller cone of the type (46) which imposes sampling requirements as above. Then we can recover stably P_hf.

In Figure 7, we show a reconstructed image with data averaged in the p variable (then $b=0$ in (58)) and the angular variables (then $a=0$ in (58)) . Note that in Figure 7c the image is blurred angularly but in contrast to Figure 4b, there are no aliasing artifacts.



(a) f on $[-1,1]^2$ **(b)** reconstruction with $\mathcal{R}f$ averaged in p **(c)** reconstruction with $\mathcal{R}f$ averaged in φ

Figure 7. f and a reconstructed f on $[-1,1]^2$ with data averaged in the p and the φ variable.

6.6. **Example: Resolution limit of `radon` and `iradon` in MATLAB.** As a simple application, we will analyze the resolution limit of the Radon transform in MATLAB R2019a if one uses the built in routines `radon` for computing $\mathcal{R}f$ and `iradon` for inversion. The angular step is fixed in its default use: $s_\varphi = \pi/180$, i.e., one degree (it can be made smaller with the right parameters). The function f is given by an $N \times N$ square array $\{f_{ij}\}$. Without loss of generality we may assume that $f(x)$ is originally defined on $[-1,1]^2$ (and supported there) and $\{f_{ij}\}$ are samples of f with a step $s_{x^1} = s_{x^2} = 2/N$. Then `radon` returns an $M \times 360$ array with $M \approx \sqrt{2}N$ approximating $\mathcal{R}(\phi, p)$ for $|p| \le \sqrt{2}$ with a step $s_p = 2\sqrt{2}/M \approx 2/N$ and for $\varphi \in [0, 2\pi]$. In fact, it only computes $\mathcal{R}(p, \phi)$ for 180 angles but it can be extended if needed to all angles using the known symmetry, see (42). That choice of M is dictated by the requirement that the equally spaced lines passing through $[-N/2, N/2]^2$, parallel to a diagonal (the worst scenario case) have step equal to $s_{x^1} = s_{y^1}$ which is the resolution limit of f but along horizontal and vertical directions. We will analyze the sharpness of that condition below.

Since $s_{x^1} = 2/N$, this corresponds to a band limit $B_{x^1} = B_{x^2}$ with $\pi/B_{x^1} = 2/N$ for each of the two variables x^1 and x^2, i.e., $B_{x^1} = \pi N/2$, see Corollary 3.1. We take $h = 1$ formally but since we are interested in the behavior as $N \gg 1$, we may think of N as $N = N(h) = C/h$ with a fixed C, say $C = 2$, see below.

We have f supported in $\Omega = [-1,1]^2$ with

$$\Sigma_h(f) \subset [-1,1]^2 \times [-B_{x^1}, B_{x^1}]^2, \tag{59}$$

and this is the smallest set with that property for all such f. In the analysis above, that set was a product of balls. We could compute the image of $[-1,1]^2 \times [-B_{x^1}, B_{x^1}]^2$ under the canonical relation of $\mathcal{R}$, which is routine but still takes some efforts. Instead of that, we include it first in the set $B(0,\sqrt{2}) \times B(0, \sqrt{2}B_{x^1})$. By (47), since $R = \sqrt{2}$, $B = \sqrt{2}N\pi/2$, we must have

$$\frac{\pi}{180} = s_\varphi \le \frac{1}{N}, \quad s_p \le \frac{\sqrt{2}}{N}. \tag{60}$$

The first inequality says that we need $N \le N_0 := 180/\pi \approx 57.3$ in order to avoid aliasing in the angular variable, i.e., $B_{x^1} = B_{x^2} \le 90$. The $N \times N$ mesh might be with a larger N but the band limit of f should not be exceeded; i.e., f would be oversampled then; which is beneficial for the accuracy of the computations. In fact, the Radon transform in `radon` is computed with interpolation different than the sinc one (we chose "spline" in our numerical simulations) which has a smoothing effect and `iradon` applies some smoothing even with no additional filter applied (Ram-Lac); so for full resolution, some oversampling is desired. The numerical experiments in Figure 8 below however show that as long as aliaisng is concerned, the bounds we drived are pretty sharp.

The second inequality in (60) poses a stricter condition on the step s_p compared to what `radon` does: $s_p \approx 2/N$. On the other hand, the first condition is the most limiting one. If we take $N = N_0$ to be the highest N guaranteeing no angular aliasing (more precisely, the band limit B_{x^1} associated with that), then we need $s_p \le \sqrt{2}/N_0 = \sqrt{2}\pi/180 \approx 0.025$. This will only affect the resolution if N is close to its critical value $N_0 \approx 57$ and it will not if $N > \sqrt{2}N_0 \approx 81$, which is often the case.

Therefore, the angular sampling rate of one degree imposes a hard limit on the resolution of f even with the best inversion; it is $B_{x^1} = B_{x^2} = 90$ band limit regardless of how large N is (and we want $N > 81$). The worst resolved singularities (x, ξ) are when x is close to $|x| = \sqrt{2}$ and ξ is tangential to that circle, i.e., $\xi \perp x$ (like edges parallel to x). Now we can see that this worst scenario case can still happen when $\Sigma_h(f)$ is in the smaller set (59) instead of the product of the

disks $B(0,\sqrt{2})\times B(0,\sqrt{2}B_{x^1})$ as we assume for convenience above — when x is close to a corner and ξ is normal to it.

In Figure 8, we illustrate this effect. The phantom is a pattern close to a critical singularity as explained above, with an angle of the pattern close to $\pi/4$. The frequency $|\xi|$ is chosen to be $72\sqrt{2}$, $90\sqrt{2}$ and $180\sqrt{2}$. Then the projections of ξ to the horizontal and to the vertical axes are 72, 90, and 180 by absolute values. As we saw above, 90 is the critical value. In the first case, there are no artifacts. In the second one, we see a part of a larger round artifact with a center outside of the computational region, along the diagonal, staying at distance close to the length of the diagonal, as predicted in section 6.4.1. In the third case we see two artifacts at approximate distances from the phantom half of the length of the diagonal. They correspond to $k=\pm1,\pm2$ in (53). In Figure 5d, the original phantom is not visible when there is angular undersampling; and here, it is. The reason is that we do not do the inversion on a finer mesh upsampling the data; we use `iradon` directly. There are artifacts coming from computing $\mathcal{R}f$ with `radon` and from that inversion as well.

(a) $|\xi|=72\sqrt{2}$, no artifacts **(b)** $|\xi|=90\sqrt{2}$ **(c)** $|\xi|=180\sqrt{2}$

Figure 8. A pattern in the lower left corner and its artifacts in the reconstruction along the diagonal with `radon` and `iradon`. The frequencies $|\xi|/\sqrt{2}$ of the pattern are compared to the critical value 90.

Finally, we connect the discussion above with the asymptotic nature of our analysis. We make the problem more general: assume that the number of the uniformly spaced grid points on the unit circle (in the angular variable φ) is N_φ, and the number of the points on $[-\sqrt{2},\sqrt{2}]\ni p$ is N_p, how should N, N_φ and N_p be related? Above we had $N_\varphi=360$. Now, we think of N, N_φ, N_p as constant multiples of $1/h$ and the conclusions holds asymptotically when $h\gg 1$.

With $\Omega=[-1,1]^2$ as above, we can set $h=2/N$. Then f is sampled at $x=hk$, with k running over a square $N\times N$ lattice. In other words, $s=1$ in Corollary 3.1. This requires, in the critically sampled case, $B_{x^1}=B_{x^2}=\pi$. Compared to $B_{x^1}=\pi N/2$ we had above, the new B_{x^1} is just the old one multiplied by h because it is a semiclassical band limit. If we are sampling a fixed function f independent of h, we can take the convolution $f_h=\phi_h * f$ with $\operatorname{supp}\mathcal{F}_h\phi\subset B(-\pi,\pi)$ and sample f_h instead; a routine anti-aliasing procedure. Then we can replace N by $2/h$ above, and replace all sampling steps s by sh (which is the notational convention we use in Theorem 3.2 and Corollary 3.1). In particular, then the resolution band limit for f imposed by N_φ alone, assuming s_p satisfies (60), would be $B_{x^1}=B_{x^2}=N_\varphi/4$ modulo an $(1+O(N_\varphi^{-\infty}))$ factor.

On the other hand, if f is given on an $N\times N$ grid, without any additional restrictions on its frequency band other that the Nyquist one implied by the size of that grid, then we need $2N$ points for $p\in[-\sqrt{2},\sqrt{2}]$ (vs. $\sqrt{2}N$ ones `radon does`) and $2\pi N$ points for φ on $[0,2\pi]$. As noted above,

we actually need about a half of them due to the symmetry of $\mathcal{R}$ but if we allow a variable weight, we need all of them.

Another example is shown in Figure 9. The original Barbara image is 250×250. We compute its Radon transform with `radon` and invert it with `iradon`. We see non-local aliasing artifacts as the theory predicts. The most visible ones are on forehead and on the hair which are shifted from the pattern on scarf below the chin.

(a) original **(b)** reconstruction

Figure 9. The Barbara image on a 250×250 grid, reconstructed with `iradon` from its Radon transform computed with `radon`. The one degree angular resolution creates non-local aliasing artifacts.

We want to emphasize again that `radon` and `iradon` do not use sampling based interpolation to compute $\mathcal{R}f$ and to invert it. The requirements on the sampling rates of $\mathcal{R}$ are therefore higher for a close to perfect resolution.

7. The X-ray/Radon transform $\mathcal{R}$ in the plane in fan-beam coordinates

7.1. **$\mathcal{R}_\kappa$ as an FIO.** We parametrize $\mathcal{R}_\kappa$ now by the so-called fan-beam coordinates. Each line is represented by an initial point $R\omega(\alpha)$ on the boundary of $B(0,R)$, where f is supported, and by an initial direction making angle β with the radial line through the same point, see Figure 10. It is straightforward to see that this direction is given by $\omega(\alpha+\beta)$. Then the lines through $B(0,R)$ are given by

$$x \cdot \omega(\alpha+\beta-\pi/2) = R\sin\beta, \quad \alpha \in [0,2\pi),\ \beta \in [-\pi/2,\pi/2]. \tag{61}$$

This allows us to conclude that in this representation $\mathcal{R}_\kappa$ is an FIO again (being FIO is invariant under diffeomorphic changes) and to compute its canonical relation using the rules of transforming covectors. We will do it directly however. We regard α as belonging to $\mathbf{R}$ modulo $2\pi\mathbf{Z}$. The relationship between this and the parallel geometry parameterization (φ,p) is given by

$$\varphi = \alpha+\beta-\pi/2, \quad p = R\sin\beta. \tag{62}$$

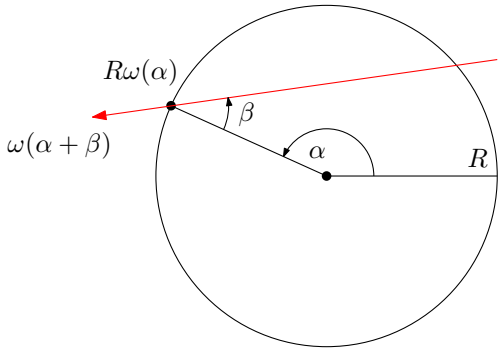


Figure 10. The fan-beam coordinates.

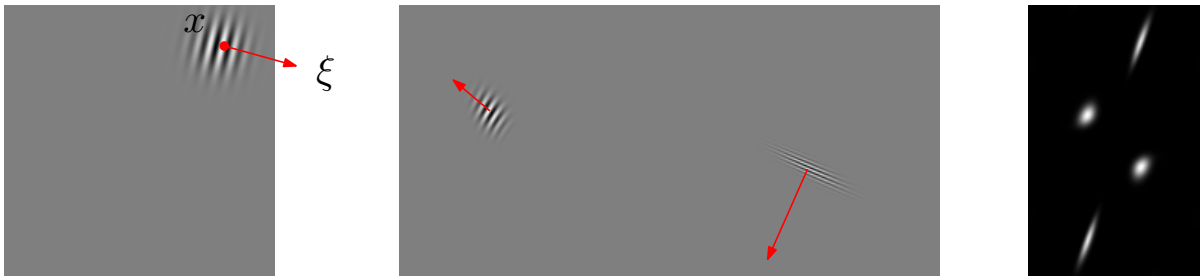
Each undirected line is given by a pair (φ, p) and $(\varphi+\pi, -p)$; which in the parallel beam coordinates corresponds to (α, β) and $(\alpha + 2\beta - \pi, -\beta)$. The Schwartz kernel of $\mathcal{R}_\kappa$ in this parameterization is a smooth factor times a delta function on the manifold (61). As above, when $\kappa = 1$, we write $\mathcal{R} = \mathcal{R}_\kappa$. Then

$$\mathcal{R}f(\alpha,\beta) = \mathcal{R}f(\alpha+2\beta-\pi,-\beta). \tag{63}$$

In general, that change of (α, β) is a symmetry of (61). The canonical relation is given by

$$C = \left\{ \Big(\alpha, \beta, \underbrace{\lambda(-x\cdot\omega(\alpha+\beta))}_{\hat\alpha}, \underbrace{\lambda R\cos\beta - \lambda x\cdot\omega(\alpha+\beta)}_{\hat\beta}, x, \underbrace{\lambda\omega(\alpha+\beta-\pi/2)}_{\hat x=\xi}\Big),\ \lambda\neq 0\right\}. \tag{64}$$

Therefore, with $\omega = \omega(\alpha+\beta)$, we have $\xi = -\lambda\omega^\perp$, $\hat\alpha = -\lambda\omega\cdot x = x\cdot\xi^\perp$, $\hat\beta = \lambda R\cos\beta + \hat\alpha$. If $\lambda > 0$, then $\lambda = |\xi|$ and then $\lambda R\cos\beta = |\xi|\sqrt{R^2 - (x\cdot\xi/|\xi|)^2}$. Also, by (61), $x\cdot\xi = R|\xi|\sin\beta$. For the dual variables, we have $\hat\beta = |\xi|\sqrt{R^2-(x\cdot\xi/|\xi|)^2} + \hat\alpha$.



(a) f on $[-1,1]^2$ **(b)** $\mathcal{R}f$ **(c)** $|\widehat{\mathcal{R}f}|$

Figure 11. The canonical relation of $\mathcal{R}$ in fan-beam coordinates. (a) a coherent state f. (b) the image of $(x,\xi)\in \mathrm{WF}_h(f)$ under C_+ and C_-. (c) $|\widehat{\mathcal{R}f}|$.

If $\lambda < 0$, we get another solution by formally replacing $|\xi|$ by $-|\xi|$. Therefore, the canonical relations $C_\pm$ are given by

$$\beta = \pm\sin^{-1}\frac{x\cdot\xi}{R|\xi|},\quad \alpha = \arg\xi - \beta \pm \frac{\pi}{2}\quad \hat\alpha = x\cdot\xi^\perp,\quad \hat\beta = \pm|\xi|\sqrt{R^2-(x\cdot\xi/|\xi|)^2} + \hat\alpha. \tag{65}$$

Then $C_\pm$ are isomorphic under the symmetry mentioned above lifted to the tangent bundle

$$(\alpha,\beta,\hat\alpha,\hat\beta) \quad\longmapsto\quad (\alpha+2\beta-\pi,-\beta,\hat\alpha,2\hat\alpha-\hat\beta). \tag{66}$$

We illustrate the canonical relations on Figure 11. On Figure 12, $\pi_2\circ C_\pm(x,\xi)$ are marked by crosses.

The inverses $C_\pm^{-1}$ are given by

$$x = R\sin\beta\,\omega(\alpha+\beta-\pi/2) - \frac{\hat\alpha}{\hat\beta-\hat\alpha}R\cos\beta\,\omega(\alpha+\beta),\quad \xi = \frac{\hat\beta-\hat\alpha}{R\cos\beta}\omega(\alpha+\beta-\pi/2). \tag{67}$$

In particular, we recover the well known fact that C is 1-to-2, as in the previous case.

7.2. **(i) Sampling.** We assume (44) again.

7.2.1. *Sampling on a rectangular lattice.* The smallest closed rectangle including the range of $\hat\alpha$ and $\hat\beta$ if $|x|\le R$, $|\xi|\le B$ is

$$RB[-1,1]\times 2RB[-1,1]. \tag{68}$$

Therefore, for the relative sampling rates s_α and s_β in the α and in the β variables, respectively, in $[0,2\pi)\times[-\pi/2,\pi/2]$, we have the Nyquist limits

$$s_\alpha \le \frac{\pi}{RB},\quad s_\beta \le \frac{\pi}{2RB}, \tag{69}$$

compare with (47). This means taking more than $2RB/h\times 2RB/h$ samples. This is π times more than in the parallel geometry case, see (47), and $\pi\times 8/\pi = 8$ times more than the expected optimal rate. For a recovery of $\mathrm{WF}_h(f)\setminus 0$, we need a half of that.

To analyze the actual range, it is enough to analyze the range of $(\hat\alpha,\hat\beta') = (\hat\alpha,\hat\beta-\hat\alpha)$, i.e., the l.h.s. of (33). Notice first that on C, one can parameterize the line corresponding to (α,β) as

$$x = R\omega(\alpha) - t\omega(\alpha+\beta),\quad 0\le t\le 2R\cos\beta. \tag{70}$$

Then

$$\hat\alpha = -|\xi|(R\cos\beta - t),\quad \hat\beta' = \pm|\xi|R\cos\beta,\quad 0\le t\le 2R\cos\beta. \tag{71}$$

Therefore, for a fixed (α,β), the range of $(\hat\alpha,\hat\beta')$ is independent of α and when ξ varies over $|\xi|\le B$, that range fills the double triangle $|\hat\alpha|\le|\hat\beta'|\le RB\cos\beta$. Over the whole range of β, see (61), this fills $|\hat\alpha|\le|\hat\beta'|\le RB$. Then we can get the range of $(\hat\alpha,\hat\beta)$, we take the inverse linear transformation.

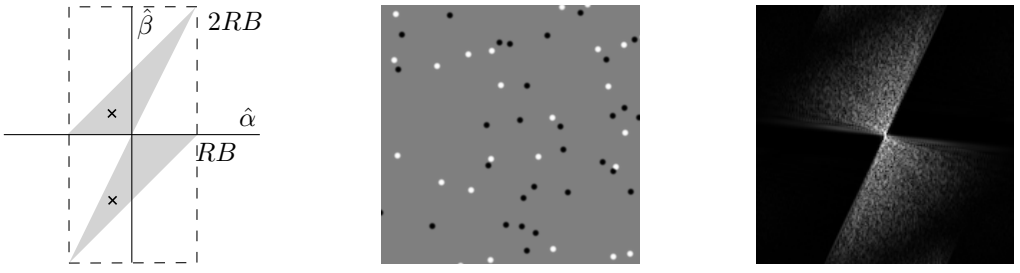


Figure 12. Left: the range of $\mathcal{F}_hR$ for $|\xi|\le B$ in fan-beam coordinates. Center: f as a sum of randomly placed small Gaussians with mean value zero. Right: $|\mathcal{F}_hRf|$ in fan-beam coordinates on a log scale.

In Figure 12, we show the range for $|\xi|\le B$, and a numerically computed $|\mathcal{F}_hRf|$ of f representing a sum of several well concentrated randomly placed Gaussians. It has the symmetry (66). This

result can be obtained from the parallel geometry analysis, of course, see (46) and Figure 2, by the change of variables on the cotangent bundle induced by (61).

As in the previous case, we can tile the plane with the regions in Figure 12 on the left by taking translations by $(RB, 0)$ and $(0, 2RB)$, see also [24]. If $2\pi(W^*)^{-1}$ has those columns, then

$$W = \frac{\pi}{RB}\begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}. \tag{72}$$

Then by Theorem 3.2, the most efficient sampling would be on a grid $hW\mathbf{Z}^n$, see also [24, 25]. This is $\sim 4N_f$, see (45) and is twice as sparse in each dimension compared to the previous criterion. For a recovery of $\mathrm{WF}_h(f) \setminus 0$, we need a half of that, i.e., $\sim 2N_f$, and both numbers are twice as much as the sharp bound in Corollary 5.1. Note that this however requires a reconstruction formula of the type (15) with χ there having a Fourier transform supported in the gray region in Figure 12 on the left, and equal to one on $\mathrm{WF}_h(f)$ instead of the formula based on the sinc functions. The reason that the number of points is not $\sim N_f$ is clear from the analysis below and from Figure 13 as well: we based our estimate on the area of the largest of the frequency sets plotted in Figure 13 rather on their sum (integral) over the base points.

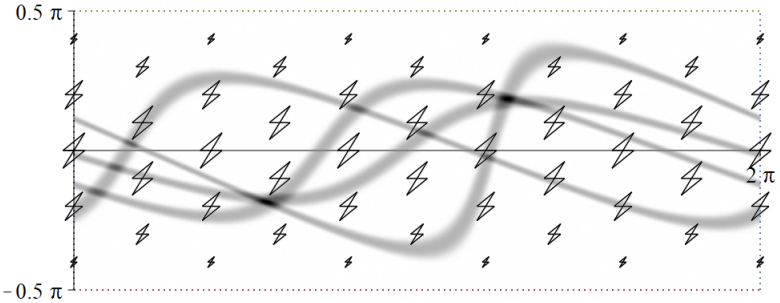


Figure 13. $\mathcal{R}f(\alpha, \beta)$ for f consisting of four randomly placed small Gaussians in the unit ball. The double triangles, shown in the upper half only, represent $\mathrm{WF}_h(\mathcal{R}f)$ localized at their vertices (α, β).

In Figure 13, we plot $\mathcal{R}f(\alpha, \beta)$ on $[0, 2\pi] \times [-\pi/2, \pi/2]$ for f consisting of four small Gaussians. We also plot the range of $\mathrm{WF}_h(f)$ for all possible f satisfying (44) at each (α, β), i.e, we plot (71). The double triangles represent the set of all possible conormals of singularities in $\mathcal{R}f(\alpha, \beta)$ with their lengths. As we can see (and prove), the highest oscillations can occur on the $\beta = 0$ line and they are along the direction $(1, 2)$. If we do non-uniform sampling, this is where we need the highest rate. This is also confirmed by the shape and the thickness of the stripes there.

The analysis above and Figure 13, suggests the following improvement: we need to sample denser when β is closer to 0. In fact, one can set $p = R \sin \beta$ (the R factor is not essential), as in (62) and sample uniformly in p. We will not explore this route here.

7.3. **(ii) Resolution limit given the sampling rate of $\mathcal{R}_\kappa f$.** Let s_α, s_β be the relative sampling rates in α and β, respectively. The Nyquist limit for $(\hat\alpha, \hat\beta)$ is given by $|\hat\alpha| \le \pi/s_\alpha$, $|\hat\beta| \le \pi/s_\beta$. By (65), this is equivalent to

$$|x \cdot \xi^\perp| \le \pi/s_\alpha, \quad \left|\pm\sqrt{R^2|\xi|^2 - (x\cdot\xi)^2} + x \cdot \xi^\perp\right| \le \pi/s_\beta. \tag{73}$$

Let θ be the angle which ξ makes with x when $x \neq 0$, more precisely, θ is such that $x \cdot \xi = |x| \cos\theta$, $x \cdot \xi^{\perp} = |x| \sin\theta$. Then

$$|x||\xi| |\sin\theta| \leq \pi/s_\alpha, \quad |\xi| \left| \pm\sqrt{R^2 - |x|^2 \cos^2\theta} + |x| \sin\theta \right| \leq \pi/s_\beta.$$

We plot the regions determined by the inequalities above with $s_\alpha = 2s_\beta$, see (69) to get the resolution diagram plotted on Figure 14, where $R = 1$. The horizontal lines represent the resolution

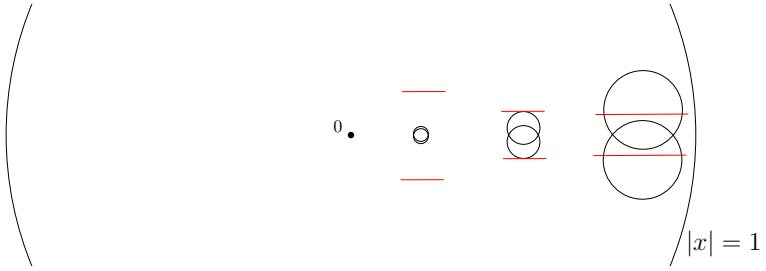


Figure 14. The resolution diagram of $\mathcal{R}_\kappa$ in fan-beam coordinates in the unit ball. For each x, the circles around it represent the frequency limit imposed by s_β as a function of the direction. A small radius means small frequency and therefore a smaller resolution. The horizontal lines mark the resolution limit imposed by s_α. The diagram is rotationally symmetric.

limit imposed by s_α. It is greatest near the origin and decreases (in vertical direction) away from the center. As it can be seen from (73), the second inequality in (73) (satisfied for both signs) implies the first one; so that actual resolution is controlled by the double circles there except at $|x| = 1$, where the lines are tangent to the lens shaped region. Next, the symmetry relation (66) has an interesting implication. Let $(\alpha_\pm, \beta_\pm, \hat{\alpha}_\pm, \hat{\beta}_\pm)$ be the image of (x, ξ) under $C_\pm$, related by (66), see Figure 11. Then the resolution limit on f at x in various directions posed by the sampling rate s_β near (α_+, β_+) *and* near (α_-, β_-) are given by (73) with both choices of the signs $\pm$. On Figure 14, they are represented by the intersection of the two disks at each x. We can see that near the origin, it is quite small and close to isotropic. Near $|x| = 1$, the resolution increases and it is better for ξ close to radial (for example, for circular lines). On the other hand, since C is 1-to-2, we need only one of $(\alpha_\pm, \beta_\pm, \hat{\alpha}_\pm, \hat{\beta}_\pm)$ to recover (x, ξ). Therefore, the data $\mathcal{R}_\chi f$ actually contains stable information about recovery the singularities of f in the union of those disks, instead of its intersection, if we can use that information. It follows from (65) that the better resolution is coming from that of the two lines through x normal to ξ with a source $R\omega(\alpha)$ which is closer to x. In the example in Figure 11, for example, if the sampling rate of $\mathcal{R}_\kappa f$ is not sufficient to sample the right-hand pattern, we can just cut it off smoothly and use the other one only.

Another approach is to note that the shape the double triangles in Figure 12 allow for under-sampling up to half of the rate, and when there is aliasing (overlapped shifted triangles), it affects both images of every (x, ξ) equally. To benefit from this however, instead of using a sinc type of interpolation, we need to use χ in (25) with $\hat{\chi}$ supported in the double triangle in Figure 12. Even better, we can sample on a parallelogram type of lattice as in (72). Unlike [24, 25] we could have a non-uniform sampling set as in Section 6.2.2 by dividing $[0, 2\pi] \times [-\pi/2, \pi/2]$ into horizontal strips and using parallelogram-like lattices in each one of varying densities using the fact that the wave front set size decreases when approaching $\beta = \pm\pi/2$.

In Figure 15 below, we present numerical evidence of this analysis. The phantom consists of two coherent states; each one a parallel transport of the other. Their wave front sets are localized in

the x and the ξ variables. For the state on the left, we have x almost parallel to ξ on the wave front set, while for the state on the top, x is almost perpendicular to ξ. As a result, the singularities of the first state are mapped to the lower frequency ones on the plot of $\mathcal{R}f$ closer to the corners. The state on the top creates the higher frequency oscillations of $\mathcal{R}f$ along the equatorial line of the plot of $\mathcal{R}f$. The Fourier transform on the right in Figure 15 confirms that — the four streaks closer to the borders correspond to the top phantom. Note that the horizontal axis in the Fourier transform plot is stretched twice compared to Figure 12 because the sampling requirement requires the same number of points on each axis, and then the discrete Fourier transform maps a square to a square.

(a) f on $[-1,1]^2$ **(b)** $\mathcal{R}f$ **(c)** $|\mathcal{F}\mathcal{R}f|$

Figure 15. (a): f having $\mathrm{WF}_h(f)$ at two points. (b): $\mathcal{R}f$ in fan-beam coordinates. (c): $|\mathcal{F}_h\mathcal{R}f|$. The two patterns on the central horizontal line correspond to the phantom on the top. They are close to being critically sampled. The two patterns on the diagonal correspond to the phantom on the left.

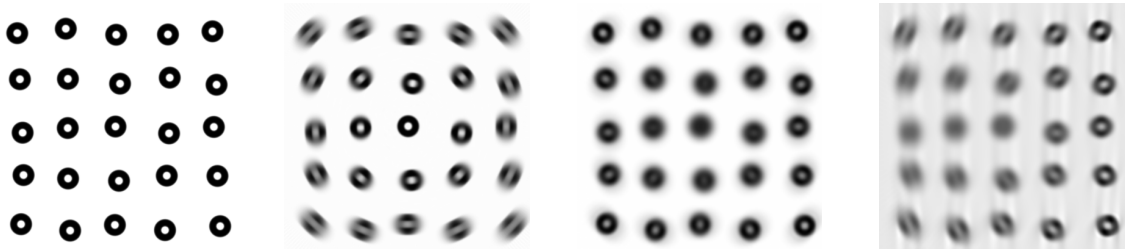
7.4. **(iii) Aliasing artifacts.** If $\mathcal{R}_\kappa f$ is undersampled in either variable, we would get aliasing artifacts as h-FIOs related to shifts of $\hat\alpha$ and $\hat\beta$, see (31), (32) Section 4.3. By (67) this would create shifts in the x variable along $\xi^\perp$ and a possible change of the magnitude of ξ but not its direction. We observed similar effects in the parallel parameterization case. In Figure 18b one can see that the aliasing artifacts are extended outside the location of the "doughnuts" there.

7.5. **(iv) Averaged measurements.** As in section 6.5, assume we measure $Q_h\mathcal{R}_\kappa f$ with Q_h an h-ΨDO of order $(0,0)$. If q_0 is the principal symbol of Q_h, then a backprojection reconstructs P_hf with P_h having principal symbol as in (57). In particular, if $q_0=\psi(a|\hat\alpha|^2+b|\hat\beta|^2)$, then

$$
\begin{aligned}
q_0 = \frac{1}{2}\psi\Big(a|x\cdot\xi^\perp|^2 + b\big|x\cdot\xi^\perp + \sqrt{R^2|\xi|^2-(x\cdot\xi)^2}\big|^2\Big)\\
+\frac{1}{2}\psi\Big(a|x\cdot\xi^\perp|^2 + b\big|x\cdot\xi^\perp - \sqrt{R^2|\xi|^2-(x\cdot\xi)^2}\big|^2\Big).
\end{aligned}
\tag{74}
$$

This formula reveals something interesting, similar to the observations above: the loss of resolution coming from each term is different. For each (x,ξ), the reconstructed f is $q_0(x,hD)f$ plus a lower order term, which is a sum of two with different (and direction dependent) losses of resolution. Let us say that ψ is radial and decreasing as $r=|x|$ increases. If $x\cdot\xi^\perp>0$, then the first term attenuates at that frequency more than the second one, and vice versa. Therefore, the reconstruction with full data in specific regions and directions would have less resolution that one with partial data. This is also illustrated in Figure 14: the intersection of the circles there reflects the resolution limit if we use full data and the union — the resolution limit with partial data chosen to maximize the resolution. To take advantage of that, we would need to take a h-ΨDO Q_h, not just a convolution.

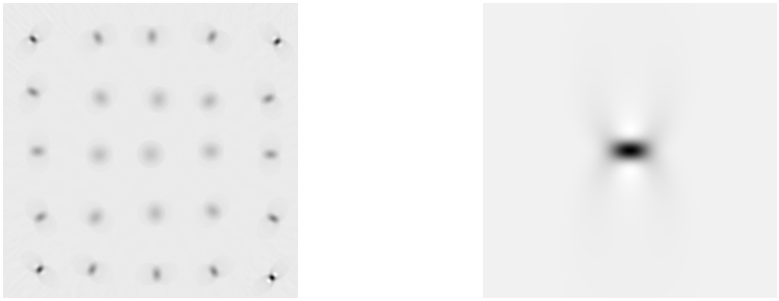
In Figure 16 we show an example. In (b), we show the reconstructed f with $\mathcal{R}f$ averaged in α. This corresponds to $b=0$ in (74). The worst resolution is where $|x\cdot\xi^\perp|$ is maximized, which



(a) f on $[-1,1]^2$ **(b)** f_{rec} with $\mathcal{R}f$ averaged in α **(c)** f_{rec} with $\mathcal{R}f$ averaged in β **(d)** f_{rec} with $\mathcal{R}f$ averaged in β and half-data

Figure 16. f and a reconstructed f_{rec} on $[-1,1]^2$ with data on the circumscribed circle and data averaged in the α and in the β variable. In (d), we use $\mathcal{R}f$ with sources on the right-hand part of the circle.

happens when $|x|$ is maximized and $x \perp \xi$, like for radial lines close to the boundary. We have the best resolution when $|x \cdot \xi^{\perp}|$ is small, and if we want that for all directions; this happens near the origin but circular lines away from the origin are resolved well, too. Averaging in β is represented by (c) and corresponds to $a = 0$ in (74). As explained above, we get a superposition of two images and the understand the plot better, one should look first at (d), where a reconstruction with α restricted to $[-\pi/2, \pi/2]$ (the r.h.s. of the circumscribed circle) is shown. There, for "doughnuts" closer to the right-hand side, radial lines (where $\xi \perp x$) are resolved better than circular ones, which corresponds to the union of the disks in Figure 14. On the left (far from the sources), it is the opposite: radial lines are very blurred, while circular ones are better resolved. This corresponds to the intersection of the disks in Figure 14 which predicts better resolution for $\xi \parallel x$. Then in (c), we have a superposition of two such images which have a combined resolution in which radial and circular blur are mixed: there is still better resolution of radial lines (but the effect is subtle in this example) and a larger radius blur in circular directions. The effect is stronger near the corners as compared to "doughnuts" near the edges but in the center of each side because the former are closer to the circumscribed circle.



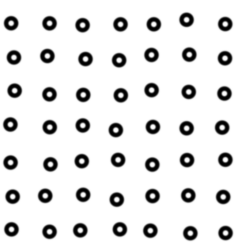
(a) f reconstructed on $[-1,1]^2$ **(b)** the predicted blur kernel at $R(0.88, 0)$

Figure 17. f is a 5×5 array of almost point-like Gaussians. (a) the reconstructed f on $[-1,1]^2$ with data on the circumscribed circle with $R = 1.45$ averaged in the β variable.; (b) the predicted blur kernel at $R(0.88, 0) \approx (1.28, 0)$, enlarged; then the radial line is horizontal.

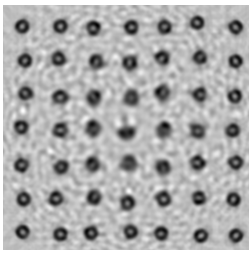
To illustrate this effect even better in Figure 17, we take f to be a slightly randomized 5×5 array of very well concentrated Gaussians and apply a Gaussian blur to $\mathcal{R}f$ in β. The reconstruction is

shown in Figure 17a. Then we compute numerically $\mathcal{F}^{-1}q_0$ see (74) with $a = 0$, which represents the convolution kernel of the reconstructed image at $x = R(0.88, 0)$, treating x as a constant. We plot it (enlarged) in (b). This is what the theory predicts to be the reconstructed image of a delta placed at that x. We can see a strong horizontal (i.e., radial) blur plus a fainter vertical (circular) one, spread over a larger area, with a negative sign. In this grayscale, black corresponds to the maximum and white corresponds to the minimum. In (a), one can see (smaller) similar images in the four corners, which are close to the circumscribed circle. Their orientations are along the radial lines, of course. As x moves closer to the center, the kernel looks more circularly symmetric and gets larger, which can be seen from Figure 17a and also from (74). At the origin, it is Gaussian as (74) predicts.

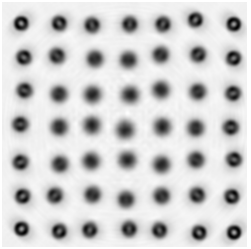
Anti-aliasing. In Figure 18, we present an example of f undersampled in the β variable and them blurred fist (in the same variable) and still undersampled at the same rate.



(a) original f on $[-1,1]^2$



(b) a reconstruction $\mathcal{R}f$ undersampled in β



(c) an anti-aliased reconstruction

Figure 18. f is a 7×7 array of "doughnuts". (b) The reconstructed f with data undersampled in the β variable, with 90 angles: a step $s_\beta = 2^o$; (b) f is blurred first and then sampled as in (b).

We see that the aliasing artifacts are mostly suppressed but some resolution is lost.

8. Thermo and Photo-Acoustic Tomography

Let Ω be a smooth bounded domain in $\mathbf{R}^n$. Let g_0 be a Riemannian metric in $\bar{\Omega}$, and let $c > 0$ be smooth. Assume that $c = 1$ and g_0 is Euclidean on $\partial\Omega$ (not an essential assumption). Fix $T > 0$. Let u solve the problem

$$\begin{cases} (\partial_t^2 - c^2\Delta_{g_0})u &= 0 \quad \text{in } (0,T)\times\mathbf{R}^n, \\ u|_{t=0} &= f, \\ \partial_t u|_{t=0} &= 0. \end{cases} \tag{75}$$

Here, $\partial_\nu = \nu^j \partial_{x^j}$, where ν is the unit outer normal vector field on $\partial\Omega$. The function f is the source which we eventually want to recover. The Neumann boundary conditions correspond to a "hard reflecting" boundary $\partial\Omega$. In applications, g_0 is Euclidean but the speed c is variable. The analysis applies to more general second order symmetric operator involving a magnetic field and an electric one, as in [29]. The metric determining the geometry is $g := c^{-2}g_0$. We assume that $\partial\Omega$ is convex.

Let $\Gamma \subset \partial\Omega$ be a relatively open subset of $\partial\Omega$, where the measurements are made. The observation operator is then modeled by

$$\Lambda f = u|_{[0,T]\times\Gamma}. \tag{76}$$

The inverse problem is to find f given Λf.

The natural space for f is the Dirichlet space $H_D(\Omega)$ defined as the completion of $C_0^\infty(\Omega)$ under the Dirichlet norm

$$\|f\|_{H_D}^2 = \int_\Omega |\nabla u|_g^2 \,\mathrm{d}\,\mathrm{Vol}\,. \tag{77}$$

The model above assumes acoustic waves propagating freely through $\partial\Omega$ where we make measurements. This means that the detectors have to be really small so that we can ignore their size. A different model studied in the literature is to assume that the waves are reflected from the boundary and measured there. To be specific, we may assume zero Neumann conditions on $\partial\Omega$ and then Λf would be the Dirichlet data but other combinations are possible. Then we solve first

$$\left\{\begin{array}{rcll} (\partial_t^2 - c^2\Delta_{g_0})u &=& 0 & \text{in } (0,T)\times\Omega,\\ \partial_\nu u|_{(0,T)\times\partial\Omega} &=& 0,\\ u|_{t=0} &=& f,\\ \partial_t u|_{t=0} &=& 0 \end{array}\right. \tag{78}$$

and define Λf as in (76) again but this time u is different.

As shown in [29] in the first case (75), Λ, restricted to f supported (strictly) in Ω, is an elliptic FIO of order zero with a canonical relation $C = C_- \cup C_+$, where

$$C_\pm : (x,\xi) \longmapsto \left(\underbrace{\pm s_\pm(x,\xi/|\xi|_g)}_{t}, \underbrace{\gamma_{x,\xi}(s_+(x,\xi))}_{y}, \underbrace{\mp|\xi|_g}_{\hat t=\tau}, \underbrace{\dot\gamma'_{x,\xi}(s_\pm(x,\xi))}_{\hat y=\eta} \right), \tag{79}$$

with $s_\pm(x,\xi)$ being the exit time of the geodesic starting from x in the direction $\pm g^{-1}\xi$ (this is ξ identified as a vector by the metric g) until it reaches $\partial\Omega$. We assume that $c^{-2}g_0$ is non-trapping; them those exit times are finite and positively homogeneous in ξ of degree -1. Also, $\dot\gamma'$ stands for the orthogonal (in the metric) projection of $\dot\gamma$ to $T\partial\Omega$. Clearly, the frequency range of C is the space-like cone $|\eta| < |\tau|$. The norm $|\xi|_g$ is the norm of η as a covector in the metric g, and similarly, $|\eta|$ is in the metric on $\partial\Omega$ induced by the Euclidean one on $\mathbf{R}^n$. We would have equality if ξ is tangent to $\partial\Omega$ but this cannot happen since $\operatorname{supp} f \subset \Omega$.

If we use (78) as a model instead (allowing for reflections) it was shown in [30] that the first singularities give rise to an FIO Λ with the same canonical relation, which is actually 2Λ modulo a lower order operator. After each reflection, we get an FIO with a canonical relation of the same type but reflected from the boundary. The sampling requirements are the same, and we will skip the details.

Assume now that $\Sigma_h(f) \subset \{|\xi| < B\}$. We have $|\xi|_g^2 = c^2 g_0^{ij}\xi_i\xi_j$. Let M^2 be the sharp lower bound of the metric form $c^{-2}g_0$ on the unit sphere over all x. Then $1/M^2$ is the sharp upper bound on $c^2 g^{-1}$ and $|\xi|_g \le B/M$ which is sharp. Then

$$\Sigma_h(\Lambda f) \le \{(\tau,\eta) \in \mathbf{R}\times T^*\partial\Omega;\ |\eta| < |\tau| \le B/M\}$$

and the r.h.s. is actually the range of $\Sigma_h(\Lambda f)$ for all f as above, see Figure 19. If we sample on a grid on $[0,T]\times\partial\Omega$, with the second variable in a fixed coordinate chart, we need to choose steps $\Delta t \le \pi M h/B$ and $|\Delta y^j| \le \pi M h/B$, where the latter norm is in the induced metric. Since Δy^j is constant (the superscript j refers to the j-th coordinate), for the Euclidean length Δy^j we must have $\Delta y^j \le \pi M' M h/B$, where $(M')^2$ is the sharp upper bound on the induced metric on the Euclidean sphere in that chart. In our numerical example below, the boundary is piecewise flat parameterized in an Euclidean way; then $M' = 1$ away from the corners.

Set $c_{\max} = \max c$. If g is Euclidean, $M = 1/c_{\max}$. The metric on $\mathbf{R}\times\partial\Omega$ is $\mathrm{d}t^2 + g'$, where g' is the Euclidean metric restricted to $T\partial\Omega$. The sampling requirements in any local coordinates on the

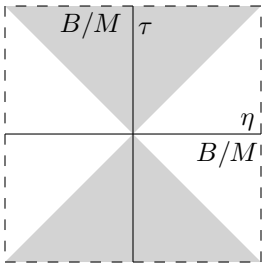


Figure 19. The frequency set of Λf.

boundary depend in those coordinates as explained above, with $M' = 1$. Therefore, the sampling rate in the (t, y) coordinates should be smaller than $\pi h/Bc_{\max}$.

It is interesting that the sampling requirements do not depend on existence of conjugate points or not and are unaffected by possible presence of caustics. In fact, we can have caustics even if the geometry is Euclidean but we start from a concave wave front. In Figure 20 on the left, we plot

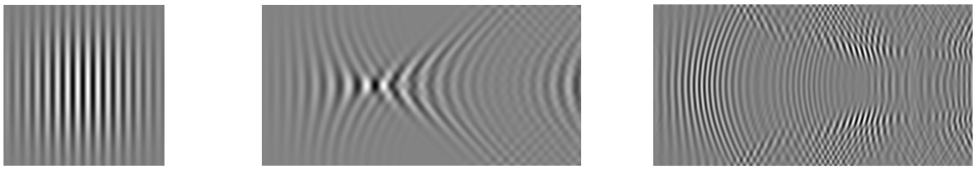


Figure 20. Left: f having $\mathrm{WF}_h(f)$ along horizontal directions, on the $[-1,1]^2$ square. Center: Λf on the right hand side cross time for $0 \le t \le 4$ with a speed with a slow region in the center. Despite the presence of caustics, Λf does not contain higher frequencies there. Right: Λf with a speed having a fast region in the center.

$f = e^{-2|x|^2} \sin((x - 0.3)/0.02)$ in the square $[-1, 1]^2$ computed with a high enough resolution. On the right, we plot Λf for the second model (78) on the right hand side of the square cross the time interval $[0, 4]$. The speed is $c = 1 - 0.5 \exp(-2|x|^2)$ having a slow region in the center and range $0.5 \le c \le 1-e^{-2}/2 \approx 0.93$. Then $M \sim 1$, and as noticed above, $M' = 1$. The sampling requirements of Λf on $[0, 4] \times \partial\Omega$ are therefore the same as those of f on $[-1, 1]^2$. Figure 20 demonstrates that fact by showing that the highest frequencies of Λf in the center are approximately the same as the highest ones on the left. Naturally, they occur where the rays hit $\partial\Omega$ at the largest angle with the normal which is represented by the slanted curves on the plot of Λf. Next, despite of presence of caustics a bit left of the center of the Λf plot, the oscillations are not of higher frequencies than elsewhere else. On the right, we plot Λf when $c = 1 + 0.5 \exp(-2|x|^2)$, i.e., there is a fast region in the middle. The speed range is approximately $[1.07, 1.5]$. There are higher frequencies than in the previous case and higher than in f. The sampling requirements are higher.

A more thorough analysis of this case in the context of this paper is presented in [22].

9. Discussion

9.1. **Summary.** We developed a framework allowing us to analyze the sampling requirements of Af given a priori information about the highest frequency content of f; or vice versa, we analyzed the resolution limit posed on f chosen by the level of discretization of Af. The linear operator A is assumed to be an FIO which happens very often in applications. We also analyzed the aliasing

artifacts when the Nyquist sampling conditions are not met and showed that they are non-local in general in contrast to classical sampling theory. We also analyzed the effect of averaging Af slightly modeling small but not point-like detectors.

9.2. **Other approaches under additional assumptions.** In classical sampling theory, it is known that one can go beyond the Nyquist limit if f is a priori known to be sparse in a certain basis, under additional conditions like random sampling or A satisfying some generic randomness conditions, see [5, 9]. We do not assume f to be sparse in some basis, the operators A we are interested in are very structured, and the grid chosen is usually periodic when the measurements are taken uniformly or when numerical simulations are done. In Figure 1, for example, and in all other aliasing examples in this paper, the measurements of the aliased image are virtually the same as those of the original, and no tricks can help. One needs to sample non-uniformly to resolve undersampled uniform patterns (under suitable sparsity assumptions on f). Compressed sensing ideas can be used here when those conditions are met and have been used in some particular cases but this is beyond the scope of this work.

Section 5 can be considered as a bridge between the theory we develop and compressed sensing. We are interested there in non-uniform sampling (but not from a probabilistic point of view). As explained in [32], the sampling requirements are (expected to be) dependent on the volume of the support of the signal in the frequency space, not just on the size of its bounding box. Thus sparser signals in the Fourier basis would be expected to be sampled with sparser sampling sets. Our Theorem 5.1 establishes a lower bound in terms of the volume *in the phase space*, not just in terms of the volume occupied by its projection on the second variable ξ. Non-uniform sampling has been studied extensively in the applied literature, see, e.g., [21], for example for multi-band signals, but the theory is not as complete as that for uniform sampling. It should be also noted that compressed sensing could require number of sampling points far exceeding the number of non almost-zero coefficients (but still below the Nyquist requirement) since their location is not a priori known, see [32] and the references there.

9.3. **Discretization of linear (inverse) problems.** Our approach sheds light on the proper way to discretize the linear problem (3), which we addressed already in section 6.6, for example. Let us say that f is approximated by a discrete function $\mathbf{f}$ defined on some grid. This can be done either by a numerical simulation or for a "real life" f without aliasing. We view $\mathbf{f}$ as the sampling coefficients of f. Following the classical sampling theory, those pointwise samples should be done after a slight blurring (anti-aliasing) of the original f if the latter has strong high-frequency components; in other words, they should be locally averaged. The popular Shepp-Logan phantom, for example, is poorly rendered in MATLAB with a direct call since the boundaries look like staircases and are a poor representation of the smooth curves they are supposed to represent. Instead, one can rendering on a higher resolution grid, apply a slight discrete blur and then downsample.

Then as we showed, one can estimate the minimal size of the grid needed to represent (the continuous) $g = Af$ properly, see also section 6.6. Assuming that this is done, let $\mathbf{g}$ be the discrete representation of g. Then we can define the discrete linear map (a matrix)

$$\mathbf{A} : \mathbf{f} \longmapsto f \stackrel{A}{\longmapsto} g \longmapsto \mathbf{g}.$$

Since passing from continuous to discrete representations and back has small (an $O(h^\infty)$) error in our setting, when we view $\mathbf{f}$, $\mathbf{g}$ just as the coefficients of f, g in a reconstruction formula of the type of (25), we get a discrete problem $\mathbf{Af} = \mathbf{g}$ equivalent to the original one $Af = g$ up to a small error. The original, and therefore the discretized problem might be stably solvable or not but the discretization does not lose information up to an $O(h^\infty)$ error.

In Inverse Problems, we often count the number of variables: if f depends on n variables, and $g = Af$ depends on m variables, the problem (3) is called under-determined if $m < n$. Then there cannot be a continuous left inverse if f and g belong to C^k or H^s spaces, by Functional analysis arguments; therefore the problem cannot be stable. Instead, we could count not dimensions but the number of numbers on which f and g depend, i.e., view them as belonging to finitely dimensional spaces. By Corollary 5.1, we need at least as many samples for g as we need for f, which also follows from linear algebra arguments once we associate f and g with $\mathbf{f}$ and $\mathbf{g}$. In some cases, as in the examples considered in this paper, we can actually take asymptotically the same number of measurements (samples) of g as those needed for f. In particular, since those numbers are proportional to h^n and h^m, respectively, we must have $m \geq n$; and then when the problem is well posed, it can be solved stably with those discretizations. If $m < n$, the discrete problem $\mathbf{Af} = \mathbf{g}$ has no unique (and stable) solution.

References

[1] F. Andersson, M. V. de Hoop, and H. Wendt. Multiscale discrete approximation of Fourier integral operators. *Multiscale Model. Simul.*, 10(1):111–145, 2012.

[2] R. N. Bracewell. Strip integration in radio astronomy. *Aust. J. Phys.*, 9:198–217, 1956.

[3] P. Caday. Computing Fourier integral operators with caustics. *Inverse Problems*, 32(12):125001, 33, 2016.

[4] E. Candès, L. Demanet, and L. Ying. A fast butterfly algorithm for the computation of Fourier integral operators. *Multiscale Model. Simul.*, 7(4):1727–1750, 2009.

[5] E. J. Candes, J. K. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, 59(8):1207–1223, 2006.

[6] A. M. Cormack. Sampling the Radon transform with beams of finite width. *Physics in Medicine and Biology*, 23(6):1141–1148, Nov 1978.

[7] M. V. de Hoop, G. Uhlmann, A. Vasy, and H. Wendt. Multiscale discrete approximations of Fourier integral operators associated with canonical transformations and caustics. *Multiscale Model. Simul.*, 11(2):566–585, 2013.

[8] M. Dimassi and J. Sjöstrand. *Spectral asymptotics in the semi-classical limit*, volume 268 of *London Mathematical Society Lecture Note Series*. Cambridge University Press, Cambridge, 1999.

[9] D. L. Donoho et al. Compressed sensing. *IEEE Transactions on information theory*, 52(4):1289–1306, 2006.

[10] S. Dyatlov and M. Zworski. *Mathematical theory of scattering resonances*. book in progress.

[11] C. L. Epstein. *Introduction to the mathematics of medical imaging*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, second edition, 2008.

[12] V. Guillemin. On some results of Gel'fand in integral geometry. In *Pseudodifferential operators and applications (Notre Dame, Ind., 1984)*, volume 43 of *Proc. Sympos. Pure Math.*, pages 149–155. Amer. Math. Soc., Providence, RI, 1985.

[13] V. Guillemin and S. Sternberg. *Geometric asymptotics*. American Mathematical Society, Providence, R.I., 1977. Mathematical Surveys, No. 14.

[14] V. Guillemin and S. Sternberg. *Semi-classical analysis*. International Press, Boston, MA, 2013.

[15] L. Hörmander. *The analysis of linear partial differential operators. IV*, volume 275. Springer-Verlag, Berlin, 1985. Fourier integral operators.

[16] A. Katsevich. A local approach to resolution analysis of image reconstruction in tomography. *SIAM J. Appl. Math.*, 77(5):1706–1732, 2017.

[17] A. Katsevich. Resolution analysis of inverting the generalized Radon transform from discrete data in $\mathbf{R}^3$. *SIAM Journal on Mathematical Analysis*, 52(4):3990–4021, 2020.

[18] H. J. Landau. Necessary density conditions for sampling and interpolation of certain entire functions. *Acta Math.*, 117:37–52, 1967.

[19] R. Marks. *Introduction to Shannon sampling and interpolation theory*. Springer Science & Business Media, 2012.

[20] A. Martinez. *An introduction to semiclassical and microlocal analysis*. Universitext. Springer-Verlag, New York, 2002.

[21] F. Marvasti. *Nonuniform sampling: theory and practice*. Springer Science & Business Media, 2012.

[22] C. Mathison. Sampling in thermoacoustic tomography. *J. Inverse Ill-Posed Probl.*, 28(6):881–897, 2020.

[23] F. Monard, P. Stefanov, and G. Uhlmann. The geodesic ray transform on Riemannian surfaces with conjugate points. *Comm. Math. Phys.*, 337(3):1491–1513, 2015.
[24] F. Natterer. *The mathematics of computerized tomography*. B. G. Teubner, Stuttgart, 1986.
[25] F. Natterer. Sampling in fan beam tomography. *SIAM J. Appl. Math.*, 53(2):358–380, 1993.
[26] F. W. J. Olver. *Asymptotics and special functions*. Academic Press, 1974. Computer Science and Applied Mathematics.
[27] D. P. Petersen and D. Middleton. Sampling and reconstruction of wave-number-limited functions in N-dimensional euclidean spaces. *Information and Control*, 5(4):279–323, 1962.
[28] P. Rattey and A. Lindgren. Sampling the 2-D Radon transform. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 29(5):994–1002, Oct 1981.
[29] P. Stefanov and G. Uhlmann. Thermoacoustic tomography with variable sound speed. *Inverse Problems*, 25(7):075011, 16, 2009.
[30] P. Stefanov and Y. Yang. Multiwave tomography in a closed domain: averaged sharp time reversal. *Inverse Problems*, 31(6):065007, 23, 2015.
[31] M. Unser. Sampling-50 years after Shannon. *Proceedings of the IEEE*, 88(4):569–587, Apr. 2000.
[32] L. P. Yaroslavsky. Can compressed sensing beat the Nyquist sampling rate? *Optical Engineering*, 54(7):1–4, 2015.
[33] M. Zworski. *Semiclassical analysis*, volume 138 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2012.

Department of Mathematics, Purdue University, West Lafayette, IN 47907