SAMPLING LINEAR INVERSE PROBLEMS WITH NOISE

PLAMEN STEFANOV AND SAMY TINDEL

Abstract. We study the effect of additive noise to the inversion of FIOs associated to a diffeomorphic canonical relation. We use the microlocal defect measures to measure the power spectrum of the noise in the phase space and analyze how that power spectrum is transformed under the inversion. In general, white noise, for example, is mapped to noise depending on the position and on the direction. In particular, we compute the standard deviation, locally, of the noise added to the inversion as a function of the standard deviation of the noise added to the data. As an example, we study the Radon transform in the plane in parallel and fan-beam coordinates, and present numerical examples.

1. Introduction

The purpose of this work is to study how noise in discrete measurements affects the reconstruction in linear inverse problems

$$Af = g, \tag{1.1}$$

where A is a Fourier Integral Operator (FIO). Examples are the Radon transform and the geodesic X-ray transforms in two dimensions, at least, thermoacoustic tomography, and the linearization of some non-linear inverse problems like boundary and lens rigidity. We assume that A is associated with a local diffeomorphism (which condition can be relaxed to the clean intersection condition in principle), and elliptic. Then a parametrix exists, which we will denote by A^{-1}, also an FIO of the same type. One can regard the problem as mapping noise by FIOs, rather than under by their inverses but we keep the former point of view.

We want to emphasize that we are not trying to remove noise. That would be only possible with a priori, say statistical information about f, but this is not the goal of this work. On the other hand, understanding well the structure of the noise under the action of the inverse would allow for better understanding of what part of f (in phase space) is most affected by noise and would hopefully allow for more efficient noise reduction.

We study additive noise first. Such noise is typically created by noisy detectors which add certain constant (but usually low) noise to the signal or by background noise. In section 7 we study examples of non-additive noise: multiplicative noise, Poisson noise as an example of modulation noise, and noise appearing in CT scan. In case of additive noise, we are given the noisy data $g + g_{\text{noise}}$, where g_{noise} (a function) is the noise. Then we are trying to solve

$$Af = g + g_{\text{noise}} \tag{1.2}$$

instead. The right-hand side (r.h.s.) may not be in the range of A so a solution may not even exist. What is often done is to apply the adjoint (assuming some Hilbert structure)

$$A^*Af = A^*g + A^*g_{\text{noise}},$$

Date: May 18, 2023.

P.S. partially supported by the National Science Foundation under grant DMS-1900475.

S.T. partially supported by the National Science Foundation under grant DMS-1952966.

which automatically cuts the part of g_{noise} perpendicular to the range of A, and then invert A^*A, assuming that A is injective in the first place. If not, we invert A^*A on the range of A^*. This can be viewed also as the least squares approximation, and it is what the Landweber iteration does, for example. So the inversion is

$$f_{\text{recovered}} = (A^*A)^{-1}A^*g + (A^*A)^{-1}A^*g_{\text{noise}} = f_0 + (A^*A)^{-1}A^*g_{\text{noise}}, \tag{1.3}$$

where f_0 is the so described least-squares reconstruction of f without noise. Of course, we could do a different "inversion". One way to do it to choose a different Hilbert structure. What is described above is very common however and it is known as the Moore-Penrose inverse. We do not have to assume that the inversion is the Moore-Penrose inverse; it could be any parametrix of A, and the Moore-Penrose inverse is such a parametrix under the assumptions we made on A.

With the above considerations in mind, we can think of the added noise as

$$f_{\text{noise}} := A^{-1}g_{\text{noise}}, \tag{1.4}$$

where, as above, A^{-1} is a parametrix, and $A^{-1}g_{\text{noise}}$ is well-defined even if g_{noise} is not in the range of A. This is also the so described solution of

$$Af_{\text{noise}} = g_{\text{noise}}.$$

We can drop the f_{noise} and the g_{noise} notation now and just study (1.1) with g not necessarily in the range of A, i.e., g is the noise now.

Example 1. The example we will use in this paper is the Radon transform $\mathcal{R}$ in $\mathbf{R}^2$

$$\mathcal{R}f(\omega, p) = \int_{x\cdot\omega=p} f(x)\,\mathrm{d}\ell, \qquad p\in\mathbf{R},\ \omega\in S^1, \tag{1.5}$$

where $\mathrm{d}\ell$ is the Euclidean line measure. It is written in "parallel geometry" coordinates. We study this example in more detail in section 5; and in section 6, we will study the same problem for the Radon transform in fan-beam coordinates. It is known that $\mathcal{R}$ is an FIO of order 1/2 with a canonical relation a graph of a local diffeomorphism (1-to-2). The most popular inversion formula is the "filtered back projection"

$$f = \frac{1}{4\pi}\mathcal{R}'|D_p|g, \quad g = \mathcal{R}f, \tag{1.6}$$

where $\mathcal{R}'$ is the transpose in distribution sense; and its versions with adding an additional filter. We view (1.6) as a unfiltered inversion and that with an additional filter, see (5.10), as a filtered one. Now, one can define a norm in the g space by $\||D_p|^{1/2}g\|_{L^2(\mathbf{R}\times S^1)}$. Then $\mathcal{R}^* = \mathcal{R}'|D_p|$ and (1.6) takes the form $f = (4\pi)^{-1}\mathcal{R}^*g$. Formula (1.6) is used all the time with noisy data not in the range of $\mathcal{R}$. In addition, we have $\mathcal{R}^*\mathcal{R} = 4\pi\,\mathrm{Id}$. Therefore, the relation $f = (4\pi)^{-1}\mathcal{R}^*g$ can be recast as $f = (\mathcal{R}^*\mathcal{R})^{-1}\mathcal{R}^*g$, which is exactly (1.3).

On the other hand, we may assume that the natural space for g is $L^2(\mathbf{R}\times S^1)$. Then $\mathcal{R}^*\mathcal{R} = 4\pi|D|^{-1}$; then the inversion is

$$f = \frac{1}{4\pi}|D|\mathcal{R}'g, \quad g = \mathcal{R}f.$$

This inversion formula is equivalent to (1.6). Note that the inverse is a version of (1.3) again.

Assume that in discrete measurements, the added noise consists of random variables with a known autocorrelation. The simplest case is independent identically distributed (i.i.d.) random variables at each "pixel" (white noise). The distribution could be Gaussian, uniform, etc. We convert the discrete measurements to a function on a "continuous", space, i.e., locally a function on $\mathbf{R}^n$. Then we invert the data by applying a parametrix, as in (1.4). The discretization rate is

assumed to be proportional to a small parameter $h > 0$ and we are interested in the asymptotic properties as $h \to 0$. Our main goal is a characterization of the induced noise f_{noise} after the inversion.

The novelties of our approach are the following. First, we view discretization and the inverse process — interpolation from a given discretization, as the step size tends to 0, in the semiclassical setting, where the small parameter $h > 0$ is proportional to the step size. This point of view was proposed in the first author's paper [17]. This allows us to use tools from semiclassical analysis to estimate the sharp sampling rate of Af, knowing the band limit of f, characterize aliasing artifacts during inversion if Af is undersampled, give a sharp limit of the resolution, etc. In this paper, we assume that we do not undersample Af.

The second novelty is moving the analysis of the spectral character of the noise to the phase space; roughly speaking, instead of localizing in the dual variable ξ only, to localize in both the spatial one x, and ξ. In the applied literature, there are two main ways to characterize noise: through its standard deviation (which assigns just one number) and through its *power spectral density* (or power spectrum). The latter is $|\hat{f}(\xi)|^2$, where f is the noise, as a function of the frequency ξ. Knowing that, we can recover the standard deviation as well, by Parseval's identity. Even though not always explicitly stated, when the noise is not expected to be homogeneous (translation invariant), one can localize in the base variable x by taking the modulus squared of the windowed Fourier transform $|\widehat{\phi f}(\xi)|^2$ with some $\phi \in C_0^\infty$. We propose going one step further: consider the power spectrum in the phase space of points x and (co)directions ξ. With the presence of the small parameter h, the natural framework is the semiclassical analysis again. The semiclassical version of localizing both in space and momentum is to localize near some x_0 in the x space with a smooth cutoff of size $h^{1/2}$ and then take the Fourier transform with ξ replaced by ξ/h, see [19] for a discussion. The natural candidate of the power spectrum in the phase space then would be the so-called semiclassical defect measure $\mathrm{d}\mu(x,\xi)$ which, roughly speaking, measures the spectral content of $f = f_h(x)$ in the phase space. We call that measure power spectrum as well.

The third novelty is looking at the noise in ergodic sense, which we also call "spatial", i.e., the noise in one measurement. There are two ways to look at the statistical properties of the noise. First, one might be interested in the expected value of the noise pointwise as we keep repeating the same experiment over and over again (in our context, if we have a series of noisy data sets and do an inversion for each one of them). We call this "temporal" view, and the analysis of the temporal properties is easier. In applications, we have one such experiment however. Our goal is to analyze the statistics of the noise in the inversion for a single experiment, as the sampling rate gets smaller and smaller, hence the term "spatial". In statistics, an estimate with a single experiment is possible when the variables are i.i.d., and we rely on the ergodic properties of the sequence.

We start with analysis of discrete white noise. The flatness of its spectrum in temporal sense, see (8.6), is well-known, which justifies its name. In spatial (ergodic) sense, this is true only in a certain averaged sense, see Theorem 8.1. For the white noise interpolated to a "continuous" function, we show that the defect measure $\mathrm{d}\mu$ is flat as well in Theorem 4.1. In Theorem 4.2 we study the spectrum of more general, correlated noise.

Next, we study propagation of noise under FIOs A^{-1} (or simply A) of the mentioned type. With the semiclassical view of noise and is power spectrum, the analysis of the power spectrum of the result $A^{-1}g$ is reduced to the mapping property of a (semiclassical) defect measure under a (classical) FIO. The answer is given by the Egorov's theorem with some extra care of the zero section. Then the tools described above would allow us the characterize the spectrum of the resulted noise in the reconstruction. We want to emphasize that even if we start with white noise g, which has a flat spectrum, the noise $A^{-1}g$ is not homogeneous in general — its power spectrum depends

on the position x and the codirection ξ. In particular, its standard deviation may change from a neighborhood of one point to another.

As we mentioned already, our analysis is not restricted to (additive) white noise only, see also section 7 for non-additive noise. We can have data with added non-white noise as well, as long as its power density in the general sense we consider it, is well defined. It could be pink, blue noise, if can be anisotropic noise, varying from point to point, or even noise corresponding to a non-absolutely continuous defect measure. For example, we may have the Radon transform $\mathcal{R}f(p,\omega)$ with added noise depending on one of those two variables only, then the associated measure would be singular. Theorems 4.1 and (4.3) still apply and describe the power density of the noise in the reconstruction.

Instead of developing the general abstract theory further, we present its application to the inversion of the Radon transform in the plane. In "parallel geometry", we show that the spectral density of the added noise is independent of the position x and proportional to $|\xi|^{1/2}$ up to the Nyquist limit (and the spectral power, which is the square of the density, is proportional to $|\xi|$). In "fan-beam coordinates", the noise depends on the position x, on $|\xi|$ proportional to $|\xi|^{1/2}$ again but depends on the direction of ξ (relative to x) as well. We present many numerical simulations.

Noise is a major concern in the applied inverse problems and has been considered in the literature; nevertheless, we are not aware of directly related works. We will mention only a few more theoretical works about noise and inverse problems. Reconstruction of Riemannian manifolds with noisy data has been studied in [5]. Using noise a source for a reconstruction has been studied in [1, 2, 7, 8].

The structure of the paper is as follows. In section 2, we recall some basic facts about semiclassical analysis, needed for our exposition. We also study the relation between classical and semiclassical FIOs. In section 3, we summarize and develop further some of the results in [17] about sampling in the semiclassical limit. In Theorem 4.1 in section 4, we prove that the power spectral density of white noise is uniform, by computing its microlocal defect measure. We also show that more general noise satisfying some assumptions, has a well defined microlocal defect measure as well. Then we apply Egorov's theorem to describe how that measure transforms under FIOs associated with a canonical diffeomorphism. Sections 5 and 6 are devoted to an application of the theory to the Radon transform on the plane in parallel and to fan-beam coordinates. We present many numerical examples as well. Multiplicative noise and other type of noise are analyzed in section 7. Finally, in section 8, we analyze discrete white noise without converting it to noise of a continuous variable. We show that it has flat spectrum *on average*.

Acknowledgments. The authors would like to thank Kiril Datchev for his advice and Magda Peligrad for making us aware of the reference [10].

2. Preliminaries on semiclassical analysis

We recall some basic facts from semiclassical analysis. For more details, we refer to [3, 12, 19]. Before that, a few words about the notation. All norms $\|\cdot\|$ are in L^2 unless indicated otherwise; also $\langle\xi\rangle := (1+|\xi|^2)^{1/2}$. We denote by $\mathcal{S}$ the Schwartz class; and $\mathcal{E}'$ is the space of the compactly supported distributions. For a linear operator A, A' is the transpose in distribution sense, while A^* is the L^2-adjoint.

2.1. Semiclassical wave front set.

The semiclassical Fourier transform $\mathcal{F}_h f$ in $\mathbf{R}^n$ of a function depending also on $h>0$ is given by

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

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. The Sobolev spaces are the semiclassical ones defined by the norm

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

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 $\phi(x_0) \neq 0$, and

$$\mathcal{F}_h(\phi f_h) = O(h^\infty) \quad \text{for } \xi \text{ in a neighborhood of } \xi_0$$

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. Elements of the zero section can be in $\mathrm{WF}_h(f)$.

Sjöstrand proposed essentially 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

$$\mathcal{F}_h(\phi f_h) = O(h^\infty \langle \xi \rangle^{-\infty}) \quad \text{for } \xi \text{ in a conical neighborhood of } \xi_0 \tag{2.1}$$

with ϕ as above.

2.2. **Semiclassical pseudo-differential operators (h-ΨDOs).** We define the symbol class $S^{m.k}$ of symbols in $\mathbf{R}^n$ as the smooth functions $p(x, \xi)$ on $\mathbf{R}^{2n}$, depending also on h, satisfying the symbol estimates

$$|\partial^\alpha_x \partial^\beta_\xi p(x, \xi)| \leq C_{\alpha,\beta,K} h^k \langle \xi \rangle^{m-|\beta|}, \tag{2.2}$$

for x in any compact set K, see, e.g., [6]. In fact, we are going to work with symbols supported in a fixed compact set in the ξ variable, so the behavior in ξ above does not matter; one may also work with the symbol class $h^k S^m(1)$, see [12, 19] where $S^m(1)$ is defined as (2.2) with $k = m = 0$. Given $p \in S^m$, we write $P = P_h = p(x, hD)$ with

$$Pf(x) = (2\pi h)^{-n} \iint e^{\mathrm{i}(x-y)\cdot\xi/h} p(x, \xi) f(y) \, \mathrm{d}y \, \mathrm{d}\xi, \tag{2.3}$$

where the integral has to be understood as an oscillatory one. This is the standard quantization; sometimes it is convenient to work with the Weyl one $p^{\mathrm{w}}(x, hD)$, where $p(x, \xi)$ is replaced by $p((x + y)/2, \xi)$ in (2.3). Then real symbols correspond to symmetric operators, in particular. Negligible operators are those with $O(h^\infty)$ norms in any pair of Sobolev spaces.

2.3. **Semiclassically band limited functions.** In [19], it is said that a tempered f_h is localized in phase space, if there exists $p \in C^\infty_0(\mathbf{R}^{2n})$ so that

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

All functions in this paper will be of this type.

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

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

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.

In [17], we gave the following definition.

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*

$$|\mathcal{F}_h f(\xi)| \le C_N h^N \langle \xi \rangle^{-N} \quad \text{for } \xi \notin U \tag{2.4}$$

for every $N > 0$.

We showed in [17] that f_h is semiclassical band limited if and only if it is localized in phase space and if and only of $\mathrm{WF}_h(f)$ is finite (no points of the type (2.1)) and compact.

In applications, we take $\mathcal{B}$ to be $[-B,B]^n$ with some $B > 0$ or the ball $|\xi| \le B$.

An 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 $\hat\phi \in C_0^\infty$. Then $\phi_h * f$ is semiclassically band limited with $\mathcal{B} = \operatorname{supp}\hat\phi$.

2.4. **Classical ΨDOs as semiclassical ΨDOs.** In the applications we have in mind, we deal with classical ΨDOs and FIOs and want to treat them as semiclassical ones. The negligible operators in the classical calculus are the smoothing ones. We showed in [17] that for every $f \in \mathcal{E}'(\mathbf{R}^n)$ and for every smoothing K, we have $\mathrm{WF}_h(Kf) \subset \mathbf{R}^n \times \{0\}$. Next, every classical ΨDO of order m can be written as an oscillatory integral of the kind (2.3) with $h = 1$ and a symbol $a(x,\xi)$ vanishing for $|\xi| \le 1$, plus a smoothing operator. Then formally, that oscillatory integral is an h-ΨDO with symbol $a(x,\xi/h)$. Then we can replace $\langle\xi\rangle$ in (2.2) by $|\xi|$ to obtain an equivalent estimate, and

$$|\partial_x^\alpha \partial_\xi^\beta a(x,\xi/h)| \le C_{\alpha,\beta,K} h^{-|\beta|} |\xi/h|^{m-|\beta|} = C_{\alpha,\beta,K} h^{-m} |\xi|^{m-|\beta|}.$$

On the support of the symbol, we have $|\xi| \ge h$, therefore the factor $|\xi|^{m-|\beta|}$ is not uniformly bounded near $\xi = 0$ when $m < |\beta|$. On the other hand, it is uniformly bounded when $|\xi| \ge \varepsilon$ with $\varepsilon > h$. This allowed us in [17], for every $\varepsilon > 0$, to split $a(x,D)$ into an h-ΨDO with symbol $a(x,\xi/h)(1-\chi(\xi/\varepsilon))$ with some cut-off function $\chi \in C_0^\infty$ plus an operator mapping semiclassically band limited functions into functions with semiclassical wave front set in an $O(\varepsilon)$ neighborhood of the zero section. We show below that we can do the same thing for FIOs associated with canonical diffeomorphisms.

Let A be a properly supported FIO with a canonical relation which is a graph of a homogeneous canonical transformation. Then up to a smoothing operator, A is of the form

$$Af(x) = (2\pi)^{-n} \iint e^{\mathrm{i}(\phi(x,\eta) - y\cdot\eta)} a(x,\eta) f(y)\, \mathrm{d}y\, \mathrm{d}\eta, \tag{2.5}$$

see [9, sec.25.3], with a a classical symbol and a phase $\phi(x,\eta)$ homogeneous in η of odder 1, satisfying $\det \phi_{x\eta} \ne 0$, $\phi_x \ne 0$ for $\eta \ne 0$. The smoothing "error" can still be written in this form with $\phi(x,\eta) = x\cdot\eta$ (a ΨDO) and an amplitude of order $-\infty$, so the arguments below apply to it as well. Let $\psi \in C_0^\infty$ have support in $B(0,2)$, $\psi = 1$ on $B(0,1)$, and fix $\varepsilon > 0$. Then $A = A_{h,\varepsilon} + R_{h,\varepsilon}$,

where

$$
\begin{aligned}
A_{h,\varepsilon}f(x) &= \left[A(\mathrm{Id}-\psi(hD/\varepsilon))f\right](x)\\
&= (2\pi h)^{-n}\iint e^{\mathrm{i}(\phi(x,\eta)-y\cdot\eta)/h}a(x,\eta/h)(1-\psi(\eta/\varepsilon))f(y)\,\mathrm{d}y\,\mathrm{d}\eta,\\
R_{h,\varepsilon}f(x) &= [A\psi(hD/\varepsilon)f](x) = (2\pi h)^{-n}\iint e^{\mathrm{i}(\phi(x,\eta)-y\cdot\eta)/h}a(x,\eta/h)\psi(\eta/\varepsilon)f(y)\,\mathrm{d}y\,\mathrm{d}\eta.
\end{aligned}
\tag{2.6}
$$

Theorem 2.1. *Under the assumptions above,*

(a) The operator $A_{h,\varepsilon}$ is an h-FIO with a (semiclassical) canonical relation the same as the (classical) one of A. Moreover, for every semiclassically band limited f with $\mathrm{WF}_h(f)\cap(\mathbf{R}^n\times B(0,\varepsilon))=\emptyset$, we have $Af=A_{h,\varepsilon}f+O(h^\infty)$.

(b) For every $f_h\in\mathcal{E}'(\mathbf{R}^n)$ with support in some compact set independent of h, satisfying $|\mathcal{F}_hf_h|\le Ch^{-N}$ for some N, we have $\mathrm{WF}_h(R_{h,\varepsilon}f)\subset\mathbf{R}^n\times B(0,C\varepsilon)$ with some $C>0$.

Proof. Since A is properly supported, and f is either $O(h^\infty)$ outside some fixed compact set in (a), or vanishes in (b), we can assume that the x–support of a is compact and independent of h as well. It follows from (2.2) that if a is a classical symbol of order m, then $a(x,\eta/h)$ is a semiclassical one of order $(m,-m)$ for $|\eta|/h>1$. Therefore our claim (a) is true for $|\eta|>\varepsilon/2$ and $0<h<\varepsilon$, which is true on the support of the symbol $a(x,\eta/h)(1-\psi(\eta/\varepsilon))$ of $A_{h,\varepsilon}$. Hence $\tilde a$ is a semiclassical symbol of order $(-m,m)$. The second part of (a) is immediate.

To prove (b), multiply $R_{h,\varepsilon}f$ by $\rho\in C_0^\infty$ and apply $\mathcal{F}_h$:

$$
\mathcal{F}_h\rho R_{h,\varepsilon}f(\xi) = (2\pi h)^{-n}\iint e^{\mathrm{i}(\phi(x,\eta)-x\cdot\xi)/h}\rho(x)a(x,\eta/h)\psi(\eta/\varepsilon)\mathcal{F}_hf(\eta)\,\mathrm{d}\eta\,\mathrm{d}x.
$$

For the phase $\Phi:=\phi(x,\eta)-x\cdot\xi$ we have $\Phi_x=\phi_x(x,\eta)-\xi$. By the homogeneity of ϕ, for $|\eta|\le 2\varepsilon$, and $|\xi|>C\varepsilon$, we have $\Phi_x\neq 0$. Then a stationary phase argument implies $\mathcal{F}_h\rho R_{h,\varepsilon}f(\xi)=O(h^\infty)$ for such ξ. This proves (b). □

2.5. **Semiclassical defect measures.** Given f_h with $\|f_h\|\le C$, one can show that there exists a sequence $h_j\to 0$ so that the limit

$$
\lim_{h=h_j\to 0+}(p(x,hD)f_h,f_h)_{L^2}=\int p(x,\xi)\,\mathrm{d}\mu_f(x,\xi)
\tag{2.7}
$$

exists for every symbol $p\in C_0^\infty$, see [12, 19], and defines a Borel measure $\mathrm{d}\mu_f(x,\xi)\ge 0$ called a semiclassical defect measure associated to f. That measure may not be unique. Note that $\mathrm{d}\mu_f$ is invariantly defined on $T^*\mathbf{R}^n$. On the other hand, its definition (2.7) depends on the choice of the measure (respectively the coordinates) used to define the L^2 space there. We can use every quantization of p in (2.7), for example the Weyl one $p^{\mathrm{w}}(x,hD)$ which guarantees that (2.7) is real when p is real-valued.

When f is semiclassically band limited, $\mathrm{WF}_h(f)$ is compact, hence $\mathrm{d}\mu_f$ has compact support as well, and

$$
\|f_{h_j}\|_{L^2}^2=\int \mathrm{d}\mu_f+o(1).
\tag{2.8}
$$

This in particular implies that our assumption guarantees that $\|f_{h_j}\|_{L^2}$ is asymptotically constant as $h_j\to 0$. In fact, some authors require $\|f_h\|=1$, see [12].

3. Sampling in the semiclassical limit

3.1. **Sampling semiclassically band limited functions.** We recall some results in [17] first. 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

$$f(x) = \sum_{k\in\mathbf{Z}^n} f(sk)\chi_k(x), \quad \chi_k(x) := \prod_{j=1}^{n} \operatorname{sinc}\Big(\frac{1}{s}(x_j - sk_j)\Big), \tag{3.1}$$

where we adopt the "engineering" definition of the sinc function

$$\operatorname{sinc}(x) = \sin(\pi x)/\pi x.$$

Moreover,

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

where $\|\cdot\|$ is the L^2 norm, see, e.g., [14] or [4].

The proof is based on viewing the samples $f(sk)$ as the (inverse) Fourier coefficients of $\hat{f}$, extended as $2\pi/s$-periodic function. We reproduce the proof below in the semiclassical case.

In [17], we formulated this, and related results in the semiclassical setting. One of those theorems is the following. Recall that $\Sigma_h(f)$ is defined in Definition 2.1.

Theorem 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 $[-\pi, \pi]$, 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*

$$f_h(x) = \sum_{k\in\mathbf{Z}^n} f_h(s_1hk_1, \dots, s_nhk_n) \prod_j \chi_j\left(\frac{1}{s_jh}(x_j - s_jhk_j)\right) + O_{\mathcal{S}}(h^\infty)\|f\|, \tag{3.2}$$

and

$$\|f_h\|^2 = s_1 \dots s_n h^n \sum_{k\in\mathbf{Z}^n} |f_h(s_1hk_1, \dots, s_nhk_n)|^2 + O(h^\infty)\|f\|^2. \tag{3.3}$$

One could think of χ_j as somewhat better versions of the sinc function: they decay faster if we choose $\hat{\chi}_j$ to be smooth. We can do this because $\Sigma_h(f)$ (which is compact) is assumed to be included in the interior of the closed $\prod[-B_j, B_j]$. In case of an equality, we must take $\chi_j(x) = \operatorname{sinc}(x)$.

Assume now for simplicity that all B_j and s_j are equal to some B and s, respectively. We can always choose a linear transformation $y = Wx$ to get back to (3.2) or even more general sampling grids, and the dual one $\xi = W^*\eta$ for the dual variables. Set $\chi(x) = \chi_1(x_1)\cdots\chi_n(x_n)$. Then (3.2) and (3.3) take the form

$$f_h(x) = \sum_{k\in\mathbf{Z}^n} f_h(shk)\chi_k(x) + O_{\mathcal{S}}(h^\infty)\|f\|, \quad \chi_k(x) := \chi\left(\frac{1}{sh}(x - shk)\right) \tag{3.4}$$

and

$$\|f_h\|^2 = (sh)^n \sum_{k\in\mathbf{Z}^n} |f_h(shk)|^2 + O(h^\infty)\|f\|^2. \tag{3.5}$$

The proof of Theorem 3.1 is based on the following observation. Since $\mathcal{F}_h f$ is supported in $\prod(-B_j, B_j)$ up to an $O(h^\infty|\xi|^{-\infty})$ error, and $\pi/s > B_j$, we have

$$(\mathcal{F}_h f)_{\text{ext}}(\xi) = (sh)^n \sum_k f(shk) e^{-\mathrm{i}sx\cdot\xi} + O_{\mathcal{S}}(h^\infty), \tag{3.6}$$

where $(\mathcal{F}_h f)_{\text{ext}}(\xi)$ is the periodic extension of $\mathcal{F}_h f(\xi)$ with period $2\pi/s$ in each 1D variable. Multiply this by $\hat{\chi}(s\xi)$ to get

$$\mathcal{F}_h f(\xi) = (sh)^n \hat{\chi}(s\xi) \sum_k f(shk) e^{-\mathrm{i}s\xi\cdot k} + O_{\mathcal{S}}(h^\infty).$$

If χ_k is the interpolating function in (3.4), then

$$\mathcal{F}_h \chi_k(\xi) = (sh)^n \hat{\chi}(s\xi) e^{-\mathrm{i}s\xi\cdot k}. \tag{3.7}$$

Take $\mathcal{F}_h^{-1}$ to complete the proof. The full details can be found in [17]. Also, χ does not need to be of product type, as shown there.

Remark 3.1. In the limit case $\Sigma_h(f) = (-B, B)^n$, which is not allowed by the theorem since $\Sigma_h(f)$ is compact, we have $\hat{\chi} = \mathbf{1}_{[-\pi,\pi]^n}$, where $\mathbf{1}_{[-\pi,\pi]^n}$ stands for the characteristic function of $[-\pi, \pi]^n$. Then χ is a product of sinc functions, see (3.1). We will use the notation

$$\operatorname{sinc}_k(x) := \prod_{j=1}^n \operatorname{sinc}\Big(\frac{1}{sh}(x_j - shk_j)\Big). \tag{3.8}$$

Then (3.7) takes the form

$$\mathcal{F}_h \operatorname{sinc}_k(\xi) = (sh)^n \mathbf{1}_{[-\pi/s,\pi/s]^n}(\xi) e^{-\mathrm{i}sk\cdot\xi}. \tag{3.9}$$

The functions sinc_k form an orthogonal system, and

$$\phi_k := (sh)^{-n/2} \operatorname{sinc}_k \tag{3.10}$$

is an orthonormal basis in the subspace $\mathbf{1}_{[-\pi/s,\pi/s]^n}(hD)L^2(\mathbf{R}^n)$. For future reference, we want to mention that for every $m \geq 2$ integer, if $\operatorname{sinc}_k^{(m)}(\xi)$ is defined by

$$\mathcal{F}_h \operatorname{sinc}_k^{(m)}(\xi) = (sh/m)^n \mathbf{1}_{[-m\pi/s,m\pi/s]^n}(\xi) e^{-\mathrm{i}sk\cdot\xi} \tag{3.11}$$

then

$$\operatorname{sinc}_k^{(m)}(x) = \prod_j \operatorname{sinc}\Big(\frac{m}{sh}(x_j - shk_j)\Big) \tag{3.12}$$

instead. Then

$$\phi_k^{(m)} := (sh/m)^{-n/2} \operatorname{sinc}_k^{(m)}$$

is an orthonormal system in $\mathbf{1}_{[-m\pi/s,m\pi/s]^n}(hD)L^2(\mathbf{R}^n)$ but not a basis. To make it a basis, notice that s was replaced by s/m and k was replaced by mk there. Allowing the original k to run over all integer points, i.e., replacing k by k/m there would complete (3.12) to a basis.

3.2. **Constructing a semiclassically band limited function from a discrete sequence.** The next question is how to associate a semiclassically band limited function to a set of numbers $\boldsymbol{f}_k$, $k \in \mathbf{Z}^n$, which we view as its samples. Without the band limited requirement, this can be done in infinitely many ways, of course, by various ways to interpolate between the samples. On the other hand, if we fix the band limit B, then for $B < \pi/s$, such a function, if exists, would be oversampled, and those samples can be shown to be dependent. We can only hope that this problem always has a solution when $B \geq \pi/s$. The next proposition, proven in [17], shows that it can be done when $B = \pi/s$ with a sinc interpolation.

Proposition 3.1. *Let $\Omega \Subset \Omega_1$ be both open. Fix $s > 0$. For $0 < h \ll 1$, let $K(h) \subset \mathbf{Z}^n$ be the set of those k for which $shk \in \Omega$. Then for every collection of complex numbers $\{\boldsymbol{f}_{k,h}\}$, $k \in K(h)$ with $\sum_k |\boldsymbol{f}_{k,h}|^2$ tempered, there exists a semiclassically band limited f_h with $\mathrm{WF}_h(f) \subset \Omega_1 \times [-\pi/s, \pi/s]^n$ so that $f(shk) = \boldsymbol{f}_{k,h}$.*

One such choice is given by

$$\tilde{f}_h = \psi f_h, \quad \text{with} \quad f_h(x) = \sum_{k \in K(h)} \boldsymbol{f}_{k,h} \operatorname{sinc}_k(x), \tag{3.13}$$

where $\psi \in C_0^\infty(\Omega_1)$ is equal to 1 near $\bar{\Omega}$. Moreover, (3.5) holds.

Proof. With f_h as above, we have $f_h(shk) = \boldsymbol{f}_h(k)$ and

$$\mathcal{F}_h f_h(\xi) = \mathbf{1}_{[-\pi/s,\pi/s]^n}(\xi)(sh)^n \sum_{k \in K(h)} \boldsymbol{f}_{k,h} e^{-\mathrm{i}sk\cdot\xi}, \tag{3.14}$$

compare to (3.6) and (3.9). Then $\mathrm{WF}_h(f_h) \subset \mathbf{R}^n \times [-\pi/s, \pi/s]^n$. Let ψ be as in the theorem. Then $\tilde{f}_h = \psi f_h$ has the required properties. □

By Theorem 3.1, (3.13) is the only such representation when f_h is restricted to Ω, up to an $O(h^\infty)$ error, if we want to keep the band limit B to be the sharp one $B = \pi/s$.

The expansion (3.13) has the usual downsides associated with the presence of the sinc functions there — they decay too slowly at infinity allowing the influence of each term to extend too far. When $B > \pi/s$ (strictly), we can have the localized interpolation functions χ_k of Theorem 3.1 in principle. The situation is different than that in Theorem 3.1 though. The functions χ_j in (3.2) do not necessarily satisfy $\chi_{k_1}(shk_2) = \delta_{k_1,k_2}$, where δ_{k_1,k_2} stands for the Kronecker symbol. In the case under consideration, they have to (up to an $O(h^\infty)$ error). Also, when $B > \pi/s$, the corresponding function f_h would be undersampled rather than oversampled. Next, in interpolations like these, the desire is to make it as smooth as possible.

One way to enforce $f_h(shk) = \boldsymbol{f}_{k,h}$ is to replace the sinc function in (3.12) with itself, multiplied by some $\phi \in \mathcal{S}$ with $\phi(0) = 1$, $\hat{\phi} \in C_0^\infty$, i.e., to put a product of $\operatorname{sinc}(x)\phi(x)$ for each point x_j. Then $\mathcal{F}(\phi \operatorname{sinc}) = (2\pi)^{-1}\mathbf{1}_{[-\pi,\pi]} * \hat{\phi}$ has support larger than $[-\pi, \pi]$ which corresponds to a band limit greater than π/s, compare with (3.14). One can also have a rapidly decreasing $\hat{\phi}$ instead of a C_0^∞ one, and the resulting error by replacing it with a suitable a C_0^∞ one can be estimated easily.

3.3. **Lanczos-3 interpolation and other convolution based interpolations.** One practical and approximate realization of the idea above is the Lanczos-3 interpolation. It is part of the family of the Lanczos-k interpolations with the number 3 below replaced by an integer k. In it, the functions χ_j in (3.2) are taken to be

$$\mathrm{Lan3}(x) := H(3 - |x|) \operatorname{sinc}(x) \operatorname{sinc}(x/3),$$

where H is the Heaviside function, and x stands for each coordinate function x_i. Its Fourier transform is not of compact support but decays like $O(|\xi|^{-2})$ with a small leading term; and it is very small outside $|\xi| \le 2\pi$, see Figure 1; as opposed to $\operatorname{sinc}(x)$ which Fourier transform is supported on $|\xi| \le \pi$. The kernel $\mathrm{Lan3}(x)$ is easy to compute numerically, has a small support, and preserves the property $f(shk) = \boldsymbol{f}_{k,h}$ because $\mathrm{Lan3}(k) = \delta_{k,0}$ for k an integer. So for all practical purposes, choosing χ in (3.13) to be Lan3, provides an interpolation with a band limit no greater than $B = 2\pi/s$ and even $B = 1.5\pi/s$; 1.5 to 2 times that of (3.13), see Figure 1.

If the samples in (3.15) with χ a Lanczos-3 kernel, are those of a function with a band limit $B = \pi/s$ (the Nyquist limit for that step size), then the reconstruction will leave frequencies below $B/2$ mostly unchanged, and will attenuate and alias those between $B/2$ and B as in Figure 1. The

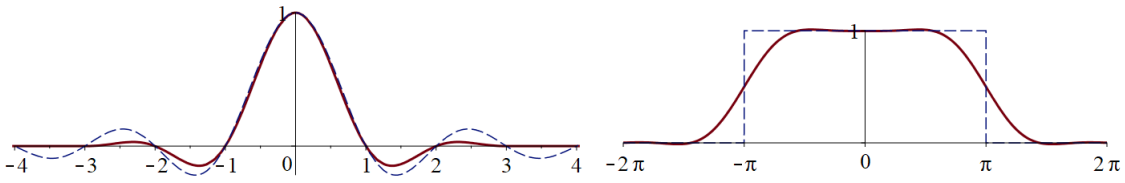


Figure 1. The Lanczos-3 kernel Lan3 and its Fourier transform. The sinc kernel and its Fourier transform are shown as dashed lines.

resulting aliasing will be "small" because the amplitude is "small" away from $|\xi| \le B$ (in Figure 1, $B = \pi$). The created f_h will have an essential band limit larger than B, as explained above, and this is true even if the samples are arbitrary.

The property of the Lanczos-3 kernel to be almost 1 in $[-\pi/2, \pi/2]$ can be used to practical interpolations with an explicit kernel with small support approximating well enough χ_j in (3.2). For this, it is enough to oversample twice or even 1.5 times only in each coordinate and use the Lanczos-3 kernel. We use this technique in the numerical computations later. This way, we work with a very well localized kernel rather than with the sinc one.

The Lanczos-3 interpolation belongs to the family of the convolution based interpolations of the type

$$f_h(x) = \sum_{k \in K(h)} \boldsymbol{f}_{k,h} \chi_k(x), \quad \chi_k(x) := \chi\Big(\frac{1}{sh}(x - shk)\Big), \tag{3.15}$$

with various compactly supported kernels χ. It is easy to see that this is the case when the interpolation is translation invariant, has a finite domain of influence, and is a linear operator. The simplest examples are the nearest neighbor (χ_k are characteristic functions of boxes in $\mathbf{R}^n$) and the linear interpolation. Some of the higher order ones are the third order cubic Catmull-Rom spline and a fourth order cubic spline proposed by Keyes, see [11, 13]. Without going into detail, we will mention that those two are very similar to Lanczos-2 and Lanczos-3, respectively with the Keyes one being a bit more smoothing that Lanczos-3. The Fourier transforms $\hat{\chi}$ related to the cubic interpolations and the Lanczos-3 ones decay fast enough to be well approximated with compactly supported ones. Then we have the following.

Proposition 3.2. *Let $\hat{\chi} \in L^\infty_{\text{comp}}$. Then for f_h given by (3.15) we have*

$$\|f_h\|^2 \le C(sh)^n \sum_{k \in K(h)} |\boldsymbol{f}_{k,h}|^2, \quad C := \|\hat{\chi}\|^2_{L^\infty}.$$

Proof. Let $1 \le m \in \mathbf{Z}$ be such that $\operatorname{supp} \hat{\chi} \in [-m\pi, m\pi]^n$. Let $\operatorname{sinc}_k^{(m)}$ be defined by (3.12). Then $\phi_k = (sh/m)^{-n/2} \operatorname{sinc}_k^{(m)}$ form an orthonormal system, see Remark 3.1. For

$$g_h(x) := \sum_{k \in K(h)} \boldsymbol{f}_{k,h} \operatorname{sinc}_k^{(m)}(x) \tag{3.16}$$

we have

$$\|g_h\|^2 = (sh/m)^n \sum_{k \in K(h)} |\boldsymbol{f}_{k,h}|^2. \tag{3.17}$$

Multiply (3.11) by $\hat{\chi}(s\xi)$, we get

$$\hat{\chi}(shD)\,\mathrm{sinc}_k^{(m)} = m^{-n}\hat{\chi}_k, \tag{3.18}$$

where we used (3.7), valid for χ_k as in (3.15) with every χ as in the proposition. Therefore, $f_h = m^n\hat{\chi}(shD)g_h$. Hence it is easily seen that $\|f_h\| \le m^n\|\hat{\chi}\|_{L^\infty}\|g_h\|$, where we recall that g_h is defined by (3.16). Combining this with (3.17), we complete the proof. □

3.4. **Noisy samples.** Let us say we restore a semiclassically band limited function from noisy samples. Assume oversampling, i.e., $B < \pi/s$ (strictly). Without noise, we would use the formula (3.4) where $f_h(\cdot)$ are the samples which we call here $\boldsymbol{f}_{k,h}$. In other words, we would take f_h as in (3.15) with χ so that

$$\hat{\chi} = 1 \text{ on } [-sB, sB]^n\,, \qquad \mathrm{supp}\,\hat{\chi} \subset (-\pi,\pi)^n, \tag{3.19}$$

which we take to be in the Schwartz class; and we can do this since we can choose $\hat{\chi}$ to be in C_0^∞.

If we do the same thing with the noisy samples, the added noise will be given by (3.15) again. Then $f(shk) = \boldsymbol{f}_{k,h}$ would be true for the noise free samples since a priori, f_h has a band limit B. This would not be true for the noisy samples, in general, because they are not necessary samples of such a function. In fact, one of the goals of the current contribution is to tackle this issue.

As in the proof of Proposition 3.2, note that (3.15) and the sinc reconstruction (3.13) are closely connected: one can get the former from the latter by applying the convolution operator $\hat{\chi}(shD)$ to it.

3.5. **Delta type of expansion.** We can view the convolution based interpolation (3.15) as a convolved delta type of expansion in which χ is formally replaced by the Dirac delta. Indeed, start with

$$f_h^\delta(x) := (sh)^n \sum_{k\in K(h)} \boldsymbol{f}_{k,h}\delta(x - shk), \tag{3.20}$$

then $f_h = \chi_h * f_h^\delta$ with $\chi_h(x) = (sh)^{-n}\chi(x/(sh))$. On the Fourier side, we have (3.14) without the cutoff function **1** there.

4. Noise and defect measures

4.1. **Microlocal defect measures as a generalization of power density.** We start this section by specifying the kind of *white noise* considered in the sequel, see also section 6.

Hypothesis 4.1. *For every $h > 0$, the noise is modeled by a family $\{\boldsymbol{f}_{k,h};\, k \in \mathbf{Z}^n\}$ of independent and identically distributed (i.i.d.) real valued random variables defined on the same probability space $(\mathbb{X}, \mathcal{F}, \mathbb{P})$. The random variables $\boldsymbol{f}_{k,h}$ have zero expected values and a common finite variance σ^2. For our computations we also make the following technical assumption on the common higher moments: there exists a constant $\delta > 0$ such that*

$$\mathbb{E}\left(\boldsymbol{f}_{k,h}^4\,(\log(1 + |\boldsymbol{f}_{k,h}|))^{1+\delta}\right) < \infty. \tag{4.1}$$

The variables $\boldsymbol{f}_{k,h}$ model the noise at each cell/pixel $x_k = shk$, with the relative step $s > 0$ fixed, and $h > 0$ a small parameter. In Hypothesis 4.1 we allow $\boldsymbol{f}_k$ to depend on h, but h will often be omitted for notational sake. In the numerical examples later, we use either normally distributed $\boldsymbol{f}_k$ or uniformly distributed ones. For a fixed bounded domain Ω, the number of sampling points $x_k = shk$ in it (we called that index set $K(h)$ in Proposition 3.1) is $|\Omega|(sh)^{-n}(1 + o(h))$. For each $h > 0$, only that many $\boldsymbol{f}_{k,h}$'s will be used eventually; therefore, we have a triangular array of random variables $\boldsymbol{f}_{k,h}$, $h > 0$, $k \in K(h)$.

As explained in the introduction, there are two types of statistical properties we are interested in. First, what we call "temporal" mean, variance, etc., are the moments of each $\boldsymbol{f}_k$ as a random variable. They are determined by the process which creates them and in practical applications correspond to repeated experiments, hence the term "temporal". We use the notation $\mathbb{E}(\boldsymbol{f}_k)$, $\mathbb{E}(\boldsymbol{f}_k^2)$, etc., for the expectation. The second, and the more interesting kind of properties are for a single experiment as $h \to 0$, i.e., when the number $N \sim h^{-n}$ of $\boldsymbol{f}_k$ grows. The mean is just the mean of those finitely many numbers, and the variance is the mean of their squares. We call them empirical spatial mean and variance, using the notation VAR for the latter and STD for the spatial standard deviation. Limit theorems for averaged random quantities with certain invariances are called sometimes ergodic properties; we view them as "spatial" ones, interpreting $\boldsymbol{f}_k$ as samples of some function in space. By the strong law of large numbers, the mean of $\sim h^{-n}$ of $\boldsymbol{f}_k$'s converges to zero almost surely, and its spatial variance converges to σ^2 almost surely, as $N \to \infty$. Below, we define similar quantities for continuous function-valued random variables.

Our terminology could be confusing since for random processes, that is families $\{f(t)\}_{t\in\mathbf{R}}$ of real-valued random variables, t is naturally interpreted as a time parameter. However, in our case the parameter (denoted as x) is a spatial variable and $\boldsymbol{f}$ has to be considered as a random field.

With Hypothesis 4.1 in hand, we think of each discrete noise as identified with a function f_h as in (3.15) with some $\hat{\chi} \in C_0^\infty$ without necessarily assuming (3.19) for now. Clearly, $\mathbb{E}(f_h) = 0$, which is a temporal characteristic. We now state a lemma for the spatial mean and variance of f_h.

Lemma 4.1. *Let $\{\boldsymbol{f}_{k,h};\, k \in \mathbf{Z}^n\}$ be a noise satisfying Hypothesis 4.1, and define the function f_h according to (3.15). Then $\mathbb{P}$-almost surely we have*

$$\mathrm{MEAN}(f_h) := \frac{1}{|\Omega|}\int_\Omega f_h \,\mathrm{d}x \to 0, \quad \text{as } h \to 0. \tag{4.2}$$

As far as the spatial variance of f_h is concerned, we get the following $\mathbb{P}$-almost sure limit,

$$\mathrm{VAR}_\Omega(f_h) := \frac{1}{|\Omega|}\int_\Omega f_h^2 \,\mathrm{d}x \to \sigma^2, \quad \text{as } h \to 0. \tag{4.3}$$

Proof. We will only prove (4.3), the proof of (4.2) being similar. To this aim, starting from (3.15) and using the fact that $\{\chi_k;\, k \in K(h)\}$ is an orthogonal system we get

$$\|f_h\|^2 = \sum_{k\in K(h)} \boldsymbol{f}_{k,h}^2 \|\chi_k\|^2 = c_\chi (sh)^n \sum_{k\in K(h)} \boldsymbol{f}_{k,h}^2, \tag{4.4}$$

where $c_\chi = \|\chi\|^2$. Plugging (4.4) into the definition (4.3) of $\mathrm{VAR}_\Omega(f_h)$, we obtain

$$\mathrm{VAR}_\Omega(f_h) = c_\chi \frac{(sh)^n}{|\Omega|} \sum_{k\in K(h)} \boldsymbol{f}_{k,h}^2. \tag{4.5}$$

Taking limits in (4.5) now amounts to applying an almost sure limit theorem for the triangular array $\{\boldsymbol{f}_{k,h}^2;\, k \in K(h), h > 0\}$. This is ensured by the relation $\mathrm{Card}(K(h)) = |\Omega|(sh)^{-n}(1 + o(h))$ and classical theorems on strong law of large numbers for triangular arrays (see e.g [10, Corollary on p. 378]), as soon as the random variables $\boldsymbol{f}_{k,h}^2$ satisfy Hypothesis 4.1. The proof of our claim (4.3) is now easily achieved. □

By (4.3), f_h is L^2 bounded almost surely, therefore it almost surely has a microlocal defect measure (possibly not unique) associated to it. In this paper, we consider every such semiclassical defect measures $\mathrm{d}\mu_f(x,\xi) \geq 0$, defined in Section 2.5, as a spectral density of f_h. In Theorem 4.1 below however, we show that the limit is unique and it holds for every sequence $h \to 0$ in the case we consider.

One can see that $\mathrm{d}\mu_f$ makes sense as the variance density in the phase space. In fact for a domain Ω, the quantity

$$\mathrm{VAR}^0_\Omega(f) := \frac{1}{|\Omega|}\iint_{T^*\Omega} \mathrm{d}\mu_f \tag{4.6}$$

corresponds formally to p being the characteristic function of Ω, divided by $|\Omega|$, which would correspond to the usual variance definition if $\lim_{h\to 0} f_h$ existed. We are not claiming that the latter limit exists however but when f is white noise, the defect measure exists as a limit in mean square sense, as we prove in Theorem 4.1 below. The superscript 0 in (4.6) is a reminder that this is a quantity in the limit $h \to 0$. We want to emphasize that VAR^0 is just defined by (4.6) and (4.8) below for any f for which $\mathrm{d}\mu_f$ exists and it is not necessarily connected to any random f. When f is random (noise), $\mathrm{VAR}^0(f)$ is related to it as in Theorem 4.1 below. We define the standard deviation $\mathrm{STD}(f)$ as the square root of the variance $\mathrm{VAR}(f)$ (with or without the superscript 0).

Assume now

$$\mathrm{d}\mu_f = \gamma_f \,\mathrm{d}x\,\mathrm{d}\xi \tag{4.7}$$

with some continuous $\gamma_f \geq 0$. Then taking the limit as Ω converges to a point, we set

$$\mathrm{VAR}^0_x(f) := \int \gamma_f(x,\xi)\,\mathrm{d}\xi. \tag{4.8}$$

Hence $\mathrm{VAR}^0_x(f)$ can be viewed as the asymptotic variance density of the noise at x.

4.2. **A remark about the Wigner function.** In this section, we will relate the Wigner function to the defect measures at a heuristic level. For a noise $\boldsymbol{f}$ satisfying Hypothesis 4.1, we set

$$(p^{\mathrm{w}}(x,hD)f_h, f_h) = \int p(x,\xi) W_f(x,\xi)\,\mathrm{d}x\,\mathrm{d}\xi, \tag{4.9}$$

where W_f is the Wigner function, see [2],

$$W^h_f(x,\xi) = (2\pi h)^{-n}\int e^{-\mathrm{i}z\cdot\xi/h} f_h(x+z/2)\bar{f}_h(x-z/2)\,\mathrm{d}z.$$

Note that $W^h_f\,\mathrm{d}x\,\mathrm{d}\xi$ is h-dependent and not a measure in general since it may take negative values. However, the existence theorem of defect measures says that there exits at least one sequence $h_j \to 0$ for which W^h_f converges to some $\mathrm{d}\mu$. Moreover, we have

$$\int W^h_f(x,\xi)\,\mathrm{d}\xi = |f(x)|^2, \qquad \int W^h_f(x,\xi)\,\mathrm{d}x = (2\pi h)^{-n}|\mathcal{F}_h f(\xi)|^2. \tag{4.10}$$

In [2], de Verdière considers random vector fields $f(x)$, $x \in \mathbf{R}^n$, and defines their auto-correlation by

$$\mathrm{ACor}_f(x,y) = \mathbb{E}(f(x)\bar{f}(y))$$

Then he defines the power spectrum of f by

$$P_h(x,\xi) = \mathbb{E}(W^h_f(x,\xi)).$$

This lifts the notion of power spectrum to the phase space but the limit $h \to 0$ is not taken.

Following the steps of the forthcoming Theorem 4.1 and using crucially the fact that $\mathbb{E}(\boldsymbol{f}_k\boldsymbol{f}_l) = \sigma^2\delta_{k,l}$, we let the patient reader check that

$$\begin{aligned}\mathbb{E}(p^{\mathrm{w}}(x,hD)f_h, f_h) &= (sh)^n\sigma^2\,\mathrm{tr}(Q(h)) \\ &= \frac{s^n\sigma^2}{(2\pi)^n}\Big(\iint |\hat{\chi}(s\xi)|^2 p(x,\xi)\,\mathrm{d}x\,\mathrm{d}\xi + O(h)\Big),\end{aligned} \tag{4.11}$$

where Q is defined by (4.19). Thanks to (4.9), this leads to the expected value of the Wigner function W_f^h up to an $O(h)$ error in a weak sense; and eventually, it could lead to the expected value of the defect measure, if we can take limits as $h \to 0$ in any reasonable probabilistic sense. There are several difficulties with this approach. We have to treat and estimate the remainder as a measure applied to p; different subsequences h_j could converge to different defect measures for a fixed $\boldsymbol{f}_k$ while the expected value applies to all such sequences, etc. The latter is the important reason we do not pursue this approach. In addition, the Wigner function method characterizes the power spectrum of the noise after repeated experiments (in temporal sense), while we want to study a single one (in ergodic sense).

4.3. **The defect measure of white noise.** Let $\boldsymbol{f}_k$, $k \in \mathbf{Z}^n$ have values in $\mathbf{R}$. As before, $\Omega \subset \mathbf{R}^n$ is a bounded domain. In the theorem below, given $h > 0$, we associate a semiclassically band limited function f_h to $\{\boldsymbol{f}_k\}$ by (3.15). This uses $|\Omega|(sh)^{-n}(1 + o(1))$ terms of the sequence $\boldsymbol{f}_k$. We allow $\{\boldsymbol{f}_k\}$ to depend on h. Then we get a triangular array of random variables.

The following theorem is the main technical result of this paper.

Theorem 4.1. *Assume that $\{\boldsymbol{f}_{k,h};\, k \in \mathbf{Z}^n\}$ is a noise satisfying Hypothesis 4.1, with L^4 moments only. Namely the random variables $\boldsymbol{f}_k$, $k \in \mathbf{Z}^n$ take values in $\mathbf{R}$ and are created by a white noise process with variance $\sigma^2 > 0$ and a bounded fourth moment.*

(a) Let f_h^δ be the associated distribution given by (3.20) with some fixed $s > 0$. Then for every $p \in C_0^\infty(T^*\Omega)$,

$$\Big(p^{\mathrm{w}}(x, hD) f_h^\delta, f_h^\delta\Big)_{L^2} \longrightarrow \int p(x,\xi)\, \mathrm{d}\mu_{f^\delta}(x,\xi), \quad \text{as } h \to 0+ \text{ in mean square sense,} \tag{4.12}$$

where

$$\mathrm{d}\mu_{f^\delta}(x,\xi) = \sigma^2 s^n \frac{\mathrm{d}x\, \mathrm{d}\xi}{(2\pi)^n}. \tag{4.13}$$

(b) Let f_h be the associated function given by (3.15) with some fixed $s > 0$ and with $\hat\chi \in C_0^\infty$ not necessarily satisfying (3.19). Then for every $p \in C_0^\infty(T^*\Omega)$,

$$(p^{\mathrm{w}}(x, hD) f_h, f_h)_{L^2} \longrightarrow \int p(x,\xi)\, \mathrm{d}\mu_f(x,\xi), \quad \text{as } h \to 0+ \text{ in mean square sense,} \tag{4.14}$$

where

$$\mathrm{d}\mu_f(x,\xi) = \sigma^2 s^n |\hat\chi(s\xi)|^2 \frac{\mathrm{d}x\, \mathrm{d}\xi}{(2\pi)^n}. \tag{4.15}$$

Proof. Notice first that the l.h.s. of (4.12) is well-defined in distribution sense since the Schwartz kernel of $p^{\mathrm{w}}(x, hD)$, see (4.33), is Schwartz class. Let $\hat\chi \in C_0^\infty$ be such that $p(x,\xi)\hat\chi(\xi) = p(x,\xi)$. Then f_h^δ can be replaced by $\chi_h * f_h^\delta$ as in section 3.5; which is (3.15). Therefore, we need to prove (b) only.

We start with the easier case when (3.19) is satisfied (with $B < \pi/s$). This corresponds to the practical situation of restoring an oversampled function with white noise added, and the theorem studies how the noise is added to the result.

Recall that the functions sinc_k were defined in (3.8) and that $\phi_k = (sh)^{-n/2}\,\mathrm{sinc}_k$ form an orthonormal basis in the space $\mathbf{1}_{[-\pi/s,\pi/s]^n}(hD)L^2(\mathbf{R}^n)$, as mentioned earlier. The interpolation function χ satisfies $\hat\chi\, \mathbf{1}_{[-\pi,\pi]^n} = \hat\chi$ by (3.19), therefore,

$$\chi_k = \hat\chi(shD)\,\mathrm{sinc}_k = (sh)^{n/2}\hat\chi(shD)\phi_k. \tag{4.16}$$

Since $\Omega \supset \mathrm{supp}_x\, p$, we have

$$(p^{\mathrm{w}}(x,hD)f_h, f_h) = \sum_{k,l\in K(h)^2} \boldsymbol{f}_k \boldsymbol{f}_l (p^{\mathrm{w}}(x,hD)\chi_k, \chi_l) = \sum_{k,l\in K(h)^2} p_{kl}\boldsymbol{f}_k\boldsymbol{f}_l, \tag{4.17}$$

where, as before, $K(h) = \{k \in \mathbf{Z}^n,\, shk \in \Omega\}$, $K^2 = K \times K$, and

$$p_{kl} := (p^{\mathrm{w}}(x,hD)\chi_k, \chi_l) = (sh)^n (Q\phi_k, \phi_l), \tag{4.18}$$

with

$$Q(h) := \bar{\hat{\chi}}(shD) p^{\mathrm{w}}(x,hD) \hat{\chi}(shD). \tag{4.19}$$

We shall prove in Lemma 4.2 that $|p_{kl}| \le C(sh)^n$. Our aim in (4.29) is to prove that in the $L^2(\mathbb{X})$ sense we have

$$\lim_{h\to 0} (p^{\mathrm{w}}(x,hD)f_h, f_h)_{L^2} = \frac{s^n}{(2\pi)^n}\sigma^2 \int p(x,\xi)\, \mathrm{d}\mu_f(x,\xi). \tag{4.20}$$

We now split the proof of (4.20) in several steps.

Step 1: A decomposition: Split the summation in (4.17) over elements (k,l) on the diagonal $\Delta := \{k = l\}$ and away from it:

$$(p^{\mathrm{w}}(x,hD)f_h, f_h)_{L^2} = W_1 + W_2, \tag{4.21}$$

where

$$W_1 := \sum_{k\in K(h)} p_{kk}\boldsymbol{f}_k^2, \quad W_2 := \sum_{k,l\in K(h)^2\setminus\Delta} p_{kl}\boldsymbol{f}_k\boldsymbol{f}_l. \tag{4.22}$$

Furthermore, according to (4.29) below we have

$$\sum_{k\in K(h)} p_{kk} = \frac{s^n}{(2\pi)^n}\int q(x,\xi)\, \mathrm{d}x\, \mathrm{d}\xi. \tag{4.23}$$

Thus owing to the fact that $q = p + O(h)$, we can recast (4.21) as

$$(p^{\mathrm{w}}(x,hD)f_h, f_h)_{L^2} - \frac{s^n}{(2\pi)^n}\sigma^2 \int p(x,\xi)\, \mathrm{d}\mu_f(x,\xi) = W_{1,0} + W_2, \tag{4.24}$$

where the term $W_{1,0}$ is defined by

$$W_{1,0} = \sum_{k\in K(h)} (\boldsymbol{f}_k^2 - \sigma^2)\, p_{kk}.$$

We are now reduced to prove that both $W_{1,0}$ and W_2 in (4.24) converge to 0 in $L^2(\mathbb{X})$.

Step 2: Analysis of $W_{1,0}$: Observe that the random variables $\boldsymbol{f}_k^2 - \sigma^2$ are independent, have zero expectation and a finite variance $\tilde{\sigma}^2 = \mathbb{E}(f_k^4) - \sigma^4$ under our fourth moment assumptions. Then $\mathbb{E}(W_{1,0}) = 0$. Moreover, invoking the forthcoming inequality (4.28) and the fact that $\mathrm{Card}(K(h)) \le c|\Omega|(sh)^{-n}$, we get

$$\mathbb{E}(W_{1,0}^2) = \sum_{k\in K(h)} \tilde{\sigma}^2 p_{kk}^2 \le (sh)^{2n}\tilde{\sigma}^2|\Omega|(sh)^{-n} \le Ch^n. \tag{4.25}$$

Therefore, $W_{1,0}$ converges to 0 as $h \to 0$, in the $L^2(\mathbb{X})$ sense.

Step 3: Analysis of W_2: The random variables $\boldsymbol{f}_k\boldsymbol{f}_l$, $k \ne l$, have expected values zero and variance σ^4. Next, $\boldsymbol{f}_k\boldsymbol{f}_l$ and $\boldsymbol{f}_{k'}\boldsymbol{f}_{l'}$ are not independent unless neither k' nor l' are equal to k or l but they are uncorrelated. Indeed, we only need to check that when, say $k = k'$ and even then,

$\mathbb{E}\left((\boldsymbol{f}_k\boldsymbol{f}_l)(\boldsymbol{f}_k\boldsymbol{f}_{l'})\right) = \mathbb{E}(\boldsymbol{f}_k^2)\mathbb{E}(\boldsymbol{f}_l)\mathbb{E}(\boldsymbol{f}_{l'}) = 0$ because all $\boldsymbol{f}_k$ have expectation zero. Therefore some elementary $L^2(\mathbb{X})$ considerations, together with (4.28), reveal that

$$\mathbb{E}(W_2^2) = \sigma^4 \sum_{k,l\in K(h)^2\setminus\Delta} p_{kl}^2 \le Ch^n. \tag{4.26}$$

Therefore, $W_2 \to 0$ in mean square sense.

Summarizing our considerations so far, the proof of the case when (3.19) holds is easily achieved by plugging (4.25) and (4.26) into (4.24).

Step 4: Dropping the assumption (3.19). Let m be such that $\operatorname{supp}\hat{\chi} \subset (-m\pi, m\pi)$. Let $\operatorname{sinc}_k^{(m)}$ be as in (3.12). Then (4.16) takes the form, see also (3.18),

$$\chi_k = m^n\hat{\chi}(shD)\operatorname{sinc}_k^{(m)} = m^n(sh/m)^{n/2}\hat{\chi}(shD)\phi_k^{(m)} = (shm)^{n/2}\hat{\chi}(shD)\phi_k^{(m)}. \tag{4.27}$$

The necessary modifications of the proof above in this case are as follows. For the deterministic term featuring in (4.22) we have the same formula but now,

$$p_{kl} := (p^{\mathrm{w}}(x,hD)\chi_k,\chi_l) = (shm)^n(Q\phi_k^{(m)},\phi_l^{(m)}).$$

The set $\{\phi_k^{(m)}\}$ is an orthonormal system in $\mathbf{1}_{[-m\pi/s,m\pi/s]^n}(hD)L^2(\mathbf{R}^n)$ but not a basis, see Remark 3.1. The missing elements are those with fractional indices in $\mathbf{Z}^n/m$. Then there are many "gaps" in the sum $W_{1,0}$ compared to the one with a basis, giving us a trace as in Lemma 4.2. On the other hand, the extra factor m^n in (4.27) allow us to think of each term m^np_{kk} as an approximation of all m^n terms in a box around k of size one, which would add the missing terms. The error is $O(h^{n+1})$ (multiplied by the constant m^n), by (4.31). Since $K(h)/m$ has $O((m/h)^{-n})$ points, this introduces an $O(h)$ error, thus (4.14) is preserved. □

The following lemma was used in the proof above. Below, $\|\cdot\|_{\mathrm{HS}}$ stands for the Hilbert-Schmidt norm.

Lemma 4.2. *For p_{kl} defined by (4.18), we have*

$$|p_{kl}| \le C(sh)^n, \tag{4.28}$$

$$\sum_k p_{kk} = (sh)^n \operatorname{tr} Q = \frac{s^n}{(2\pi)^n}\int q(x,\xi)\,\mathrm{d}x\,\mathrm{d}\xi, \tag{4.29}$$

$$\sum_{k,l}|p_{kl}|^2 = (sh)^{2n}\|Q\|_{\mathrm{HS}}^2 = \frac{s^{2n}h^n}{(2\pi)^n}\int |q(x,\xi)|^2\,\mathrm{d}x\,\mathrm{d}\xi, \tag{4.30}$$

where q is the complete symbol of the h-ΨDO Q in (4.19). Next,

$$\begin{aligned} p_{kl} &= (p^{\mathrm{w}}(x,hD)\chi_k,\chi_l) \\ &= s^{2n}h^n\iint \check{p}\Big(\frac{sh}{2}(x+y+k+l), s(x-y+k-l)\Big)\chi(x)\chi(y)\,\mathrm{d}x\,\mathrm{d}y, \end{aligned} \tag{4.31}$$

where $\check{p}$ is the inverse Fourier transform of p w.r.t. ξ

Proof. Inequality (4.28) follows directly from the fact that $\|P(h)\|$ is bounded uniformly in h, see, e.g., [19, Theorem 4.21]. If we add the basis elements of $(\mathrm{Id} - \mathbf{1}_{[-\pi/s,\pi/s]^n}(hD))L^2(\mathbf{R}^n)$ to the ϕ_k terms in (4.18), we will get zero contribution, so we consider it done. Then the first equality in (4.29) follows by the definition of a trace. The second part follows from [3, Ch. 9].

To prove (4.30), write

$$\|Q\|_{\rm HS}^2 = \operatorname{tr}(Q^*Q) = \sum_k \|Q\phi_k\|^2 = \sum_{k,l} |(Q\phi_k, \phi_l)|^2 = (sh)^{-2n} \sum_{k,l} |p_{kl}|^2, \tag{4.32}$$

see also the proof of [16, Theorem VI.23]. This proves the first part of (4.30). For the second part, notice that by [3, Ch. 9] again, the Hilbert-Schmidt norm of a classical ΨDO $R := r(x, D)$ is given by

$$\|R\|_{\rm HS}^2 = \frac{1}{(2\pi)^n} \int |r(x,\xi)|^2 \,\mathrm{d}x \,\mathrm{d}\xi.$$

We can turn R into a classical ΨDO by setting formally $r(x,\xi) = q(x, h\xi)$ to get

$$\|Q(h)\|_{\rm HS}^2 = \frac{1}{(2\pi h)^n} \int |q(x,\xi)|^2 \,\mathrm{d}x \,\mathrm{d}\xi.$$

Combining this with (4.32), we complete the proof of (4.30) as well.

Finally, the Schwartz kernel of $p^{\rm w}(x, hD)$ is given by

$$h^{-n}\check{p}((x+y)/2, (x-y)/h), \tag{4.33}$$

and $\check{p}$ is in the Schwartz class. Then

$$(p^{\rm w}(x, hD)\chi_k, \chi_l) = h^{-n} \iint \check{p}\Big(\frac{x+y}{2}, \frac{x-y}{h}\Big)\chi\Big(\frac{1}{sh}(x - shk)\Big)\chi\Big(\frac{1}{sh}(y - shl)\Big) \,\mathrm{d}x \,\mathrm{d}y.$$

Make the change of variables $\tilde{x} = (x - shk)/(sh)$, $\tilde{y} = (x - shl)/(sh)$; then $x = sh(\tilde{x} + k)$, $y = sh(\tilde{y} + l)$ to get (4.31). □

Remark 4.1.

(a) The presence of the parameter s in (4.15) is to be expected. The random sequence $\boldsymbol{f}_k$ is not related to any distance scale, while sh is the distance between two adjacent points on the sampling grid after we associate $\boldsymbol{f}_k$ to f_h. Then s reflects the choice of that scale.

(b) For every x, we have, see (4.6),

$$\mathrm{VAR}_x^0(f) = \int \gamma_f(x,\xi) \,\mathrm{d}\xi = \frac{\sigma^2}{(2\pi)^n} \|\hat{\chi}\|^2 = \sigma^2 \|\chi\|^2, \tag{4.34}$$

in mean square sense, see also (4.10). In particular, if χ is a product of sinc functions, we get σ^2, i.e., f_h has the same variance as that of $\boldsymbol{f}_k$, in a limit. If $\chi = \mathrm{LAN3}$, then $\|\chi\|^2 \approx 0.888$ in one dimension. In dimension n, we have a product of such $\chi(x_j)$'s, then the factor would be $\|\chi\|^{2n}$ instead, therefore, $\mathrm{STD}_x^0(f) \approx 0.94^n\sigma$. Note that there is no dependence on s here. For the linear interpolation, $\|\chi\|^2 = 2/3$, therefore, $\mathrm{STD}_x^0(f) = (2/3)^{n/2}\sigma \approx 0.816^n\sigma$. All those equalities are mean square limits in the sense of the theorem.

(c) If we are interested in the expected value of the variance in repeated experiments, the equivalent of (4.34) is easy to get. We can think of f_h as a linear operator, say Ψ, applied to $\boldsymbol{f} = \{\boldsymbol{f}_k\}$, i.e., $f_h = \Psi\boldsymbol{f}$. Then

$$\mathbf{E}(\|f_h\|^2) = \mathbf{E}(\Psi^*\Psi\boldsymbol{f}, \boldsymbol{f}) = \sigma^2 \operatorname{tr}(\Psi^*\Psi) = \sigma^2 \|\Psi\|_{\rm HS}^2,$$

where the latter norm is the Hilbert-Schmidt one. Then the equivalent of (4.34) can be derived from this formula. That requires repeated experiments however.

(d) The variance (4.6) is like the l.h.s. of (4.14) with p being the characteristic function of Ω divided by its volume. The theorem requires p to be smooth though, so we may think of (4.6) as an approximation of (pf_h, f_h) with $p \in C_0^\infty(\Omega)$ (independent of ξ) approximating that normalized characteristic function.

(e) Theorem 4.1 says that the noisy $|\hat{f}_h|^2$ in (3.15) converges in weak sense to $\frac{s^{2n}h^n}{(2\pi)^n}\sigma^2|\hat{\chi}(s\xi)|^2$.

(f) We can assume that the noise is not homogeneous, for example that $\boldsymbol{f}_{k,h}$ are replaced by $\zeta(shk)\boldsymbol{f}_{k,h}$ with some smooth ζ. This case can be handled as explained in section 7.1, where $g = \zeta$ and the problem with ∇g described there does not exist in this case. This would introduce the extra factor $|\zeta(x)|^2$ in (4.15). In principle, one can consider noise inhomogeneous in phase space, i.e., ζ being a suitably sampled ΨDO or an h-ΨDO.

In Figure 2, we present an one dimensional numerical example. In sections 5 and 6 we show two-dimensional ones. We take a discrete $\boldsymbol{f}$ with $N = 100$ components, upsize it to a 200 point grid with the Lanczos3 algorithm, and plot $|\hat{\boldsymbol{f}}|$, where the hat stands for the Discrete Fourier Transform, then the same quantity computed as a square root of $|\hat{\boldsymbol{f}}|^2$ averaged over 10^2 and 10^5 experiments, for frequencies in $[0, 100]$. This illustrates (4.11). The limiting profile looks very close to the profile in Figure 1, right, as expected from our Remark (e) above. At the right hand side of the plot, it is not as close to zero as the profile in Figure 3 because of the $O(1/N)$ error in (4.11); here $N = 100$ only. The plot on the right is essentially the expected value of the Wigner function W_f^h.

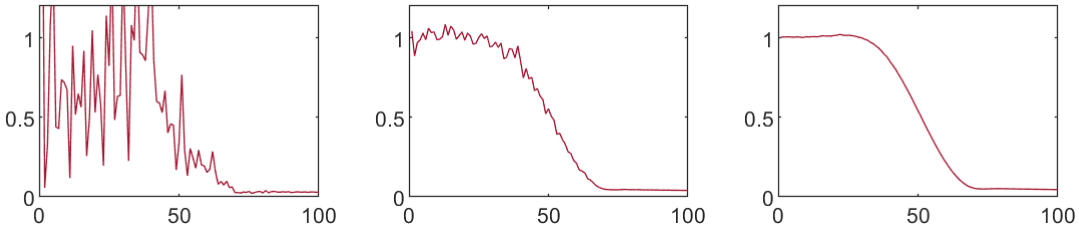


Figure 2. Plot of $|\hat{f}|$ for $N = 100$, with $|\hat{f}|^2$ averaged over $1, 10^2$, and 10^5 experiments.

In Figure 3, the setup is as above but we show the smoothing effect of averaging the power spectrum within a single experiment, illustrating relation (4.14). To this aim we consider $\boldsymbol{f}$ with $N = 10^2$, 10^4, and 10^6 components. The frequency interval is divided into 25 subintervals and averaged there, similarly to Figure 16. The plot on the left is very close to the plot of the modulus of the Fourier transform of the Lanczos3 filter in Figure 1.

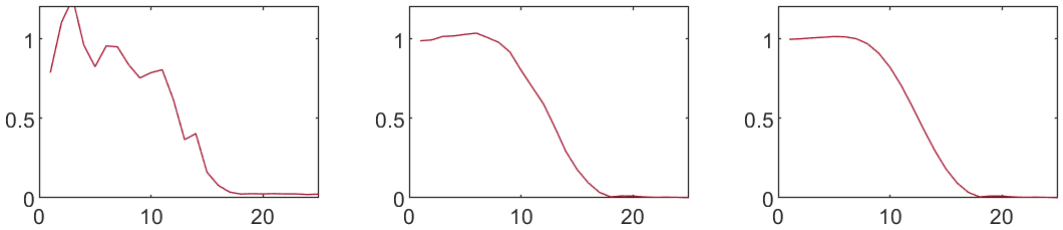


Figure 3. Plot of $|\hat{f}|$ with a single experiment, for $N = 10^2, 10^4, 10^6$, with averaging over 25 subintervals.

4.4. **Microlocal defect measure of more general noise.** We consider more general noise now. First, we assume that the random variables $\boldsymbol{f}_{k,h}$ might be correlated with the neighboring ones; and second, we assume that this correlation might be position dependent. Since the position of $\boldsymbol{f}_{k,h}$ would be at $x_k = shk$, this more general noise would be assumed to satisfy the following.

Hypothesis 4.2. *For every $h > 0$, the noise is modeled by a family $\{\boldsymbol{f}_{k,h};\, k \in \mathbf{Z}^n\}$ of real valued random variables defined on the same probability space $(\mathbb{X}, \mathcal{F}, \mathbb{P})$ with zero expected values. They are all assumed to satisfy (4.1) with a uniform bound. For the autocorrelation* $\mathrm{ACor}(\boldsymbol{f}_{k,h}, \boldsymbol{f}_{k+m,h})$ *we assume*

$$\mathrm{ACor}(\boldsymbol{f}_{k,h}, \boldsymbol{f}_{k+m,h}) = \beta(skh, m), \tag{4.35}$$

where $\beta(x,k)$, $x \in \mathbf{R}^n$, $k \in \mathbf{Z}^n$, is smooth in x, and supported in a bounded set w.r.t. both variables.

Note that we are no longer requiring, in particular, $\boldsymbol{f}_{k,h}$ to have the same variance. They are not identically distributed, in general.

Let

$$\check{\beta}(x,\xi) = \sum_m e^{\mathrm{i}sm\cdot\xi}\beta(x,m) \tag{4.36}$$

be the inverse Fourier series of β with respect to the m variable. This is essentially the Wigner distribution related to the auto-correlation, in the limit $h \to 0$. Since $\beta(s(k+m)h, -m) = \beta(skh, m)$, we must have $\beta(x,m) = \beta(x,-m)$ for all (x,m). Then (4.36) is just a cosine series, and in particular real. The theorem above shows that it is in fact non-negative.

The generalization of Theorem 4.1 to this case is the following.

Theorem 4.2. *Assume that $\{\boldsymbol{f}_{k,h};\, k \in \mathbf{Z}^n\}$ is a noise satisfying Hypothesis 4.2, with L^4 moments only. Let f_h be the associated function given by (3.15) with some fixed $s > 0$ and with $\hat{\chi} \in C_0^\infty$ not necessarily satisfying (3.19). Then (4.14) remains true with*

$$\mathrm{d}\mu_f(x,\xi) = \frac{s^n}{(2\pi)^n}\check{\beta}(x, s\xi)|\hat{\chi}(s\xi)|^2\,\mathrm{d}x\,\mathrm{d}\xi. \tag{4.37}$$

Proof. We follow the proof of Theorem 4.1. We replace the diagonal Δ in it by $\Delta = \{(k,l);\ |k-l| \le M\}$, where M is so that $\beta(\cdot, m) = 0$ for $|m| > M$. The off–Δ terms do not contribute to the limit (4.14) as above. For the rest, we estimate their contribution for every fixed m, and then sum up the results. The analog of W_1 now, depending on m, is

$$W_1 = \sum_{k\in K(h)} p_{k\,k+m}\boldsymbol{f}_k\boldsymbol{f}_{k+m} = \sum_{k\in K(h)} \beta(skh, m)p_{k\,k+m} + W_{1,0}, \tag{4.38}$$

where

$$W_{1,0} = \sum_{k\in K(h)} \left(\boldsymbol{f}_k\boldsymbol{f}_{k+m} - \beta(skh,m)\right) p_{k\,k+m}.$$

The analysis of $W_{1,0}$ is similar: the random variables $\boldsymbol{f}_k\boldsymbol{f}_{k+m} - \beta(skh,m)$ have zero expectation, thus $\mathbb{E}(W_{1,0}) = 0$. They have a uniformly bounded variance. To estimate $\mathbb{E}(W_{1,0}^2)$, notice that only $O(m^2h^{-n})$ terms in the expansion would have a non-zero expectation; and by (4.28), $\mathbb{E}(W_{1,0}^2) = O(h^n)$ again. It remains to compute the β term in (4.38).

Recall the definition (4.18) of p_{kl}. With $l = k + m$ there, an easy calculation shows that $Q\phi_{k+m} = q_m(x, hD)\phi_k$ for any h-ΨDO $Q = q(x, hD)$, with $q_m(x,\xi) = e^{\mathrm{i}sm\cdot\xi}q(x + shm, \xi)$ (which is a symbol as well, notice that there is no h in the phase). The principal symbol of that is just $e^{\mathrm{i}sm\cdot\xi}q(x,\xi)$. Then the β term in (4.38) takes the form

$$(sh)^n \sum_{k\in K(h)} (\beta(skh,m)\phi_k, q_m(x,hD)\phi_k).$$

By the properties of ϕ_k, recall (3.8) and (3.9), replacing $\beta(skh, m)$ above with $\beta(x, m)$ would result in an $O(sh)$ error in each term, and a total error $O(h)$. Considering this done, and moving the β factor to the right, we get a quadratic form with q multiplied by $\beta(x, m)$:

$$(sh)^n \sum_{k\in K(h)} (\phi_k, \tilde{q}_m(x, hD)\phi_k),$$

where $\tilde{q}_m(x,\xi) = e^{\mathrm{i}sm\cdot\xi}\beta(x, m)q(x,\xi)$.

So far, m was fixed. Summing over m (the number of those terms is $2M + 1$), we get to the situation of the proof of Theorem 4.1 with q replaced by

$$\sum_m e^{\mathrm{i}sm\cdot\xi}\beta(x, m)q(x,\xi) = \check{\beta}(x, s\xi)q(x,\xi).$$

The theorem then follows as in the proof of Theorem 4.1. □

4.5. **Spectral density under an FIO.** We want to find out how a spectral density transforms under an action of a classical FIO of order m. It is easier to answer this question for semiclassical FIOs since the defect measures are a semiclassical object, and we will reduce the classical case to the semiclassical one.

Theorem 4.3. *Let A be a classical FIO of order m on $\mathbf{R}^n$ with a homogeneous principal symbol associated with a canonical relation which is a graph of a local diffeomorphism κ (called the canonical transformation of A). Let $f = f_h$ be semiclassically band limited and uniformly bounded in L^2. Then for every defect measure $\mathrm{d}\mu_f$ given as the limit (2.7) for some $h = h_j \to 0$, the defect measure $\mathrm{d}\mu_{h^m Af}$ associated to the same sequence h_j exists as well and it satisfies*

$$\mathrm{d}\mu_{h^m Af} = \kappa^{-1^*}(b\,\mathrm{d}\mu_f) \quad \text{on } T^*\mathbf{R}^n \setminus 0,$$

where b is the (classical) principal symbol of A^*A.

Proof. By (2.7),

$$\begin{aligned}\int p(x,\xi)\,\mathrm{d}\mu_{h^m Af} &= \lim_{h=h_j\to 0} (p(x, hD)h^m Af_h, h^m Af_h)_{L^2} \\ &= \lim_{h=h_j\to 0} (h^m A^* p(x, hD)h^m Af_h, f_h)_{L^2}\,.\end{aligned} \tag{4.39}$$

Since we need to find $\mathrm{d}\mu_{h^m Af}$ away from the zero section, it is enough to assume that $p = 0$ near $\xi = 0$.

If for a moment we ignore the need to cut near $\xi = 0$, then we can think of A as in (2.5) as an h-FIO with symbol $a(x, \xi/h) = h^{-m}a(x,\xi)$ for $|\xi| \gg 1$. Then by the semiclassical Egorov's theorem [12, Theorem 5.5.5], which an analog of the classical one, (Theorem 25.3.5 in [9]), we would get

$$(h^m A^* p(x, hD)h^m Af_h, f_h)_{L^2} = (Qf_h, f_h)_{L^2}\,, \tag{4.40}$$

where Q is an h-ΨDO with a principal symbol $b(p \circ \kappa)$, with b the (classical) principal symbol of A^*A and κ is the canonical transformation of A. Note that the canonical relations of A and its semiclassical version after the change $\xi \mapsto \xi/h$ are the same.

To deal with the fact that we have a classical FIO and a semiclassical ΨDO, we apply Theorem 2.1. Let $A = A_{h,\varepsilon} + R_{h,\varepsilon}$ be as in (2.6). For $\varepsilon \ll 1$, the remainder $R_{h,\varepsilon}$ would contribute an $O(h^\infty)$ error to (4.39) if we replace A there by $A_{h,\varepsilon}$ because $p = 0$ near $\xi = 0$. Therefore, we can consider this done. Then $A_{h,\varepsilon}$ is an h-FIO, see (2.6) with symbol $\tilde{a} := a(x, \eta/h)(1 - \psi(\eta/\varepsilon)) \in h^{-m}S^0$

supported where $\eta \geq \varepsilon$. On the support, $|\eta|/h \geq \varepsilon/h$, and there, $a(x, \eta/h)$ is homogeneous for $h \ll 1$; therefore $\tilde{a} = h^{-m} a(x, \eta)(1 - \psi(\eta/\varepsilon))$

Then we can apply the semiclassical version of Egorov's theorem [12, Theorem 5.5.5]. For that, we need to compare the principal symbol of the h-ΨDO $A^*_{h,\varepsilon} A_{h,\varepsilon}$ to that of the classical ΨDO A^*A and see how the cutoff $(1 - \psi(\eta/\varepsilon))$ near the zero section affects that.

The principal symbol of $A^*_{h,\varepsilon} A_{h,\varepsilon}$ is given by

$$c(x, \xi, h) = |\tilde{a}(\pi_1 \circ \kappa(x, \xi)), \xi, h)|^2 \, J(x, \xi),$$

where π_1 is the projection on the fist variable, and $J > 0$ is a smooth Jacobian, homogeneous of order zero w.r.t. ξ, depending on the phase function only. For $|\xi| > 2\varepsilon$ we have $\tilde{a} = h^{-m} a(x, \eta)$; therefore

$$c(x, \xi, h) = h^{-2m} \, |a(\pi_1 \circ \kappa(x, \xi)), \xi)|^2 \, J(x, \xi), \quad |\xi| \geq 2\varepsilon.$$

This is the principal symbol of A^*A as a classical ΨDO as well without the factor h^{-2m}. Therefore, the limit of (4.40), as $h = h_j \to 0$, would be

$$\int b(p \circ \kappa)(x, \xi) \, \mathrm{d}\mu_f$$

as long as $p = 0$ for $|\xi| \leq 2\varepsilon$. Make the change of variables $\kappa(x, \xi) = (y, \eta)$, and using the fact that κ is symplectic, in particular an isometry, we would get

$$\int p(x, \xi) \, \mathrm{d}\mu_{h^m Af} = \int p(x, \xi) \kappa^{-1^*}(b \mathrm{d}\mu_f),$$

when $p = 0$ for $|\xi| \leq 2\varepsilon$, where κ^{-1^*} is the pull-back under κ^{-1}. Since $\varepsilon > 0$ is arbitrary, this holds when $0 \notin \mathrm{supp}_\xi \, p$. Then

$$\mathrm{d}\mu_{h^m Af} = \kappa^{-1^*}(b \mathrm{d}\mu_f). \tag{4.41}$$

So far A was microlocalized near pair of points, where κ is a (global) diffeomorphism. Since it is only a local one, we can do the same for each branch, and add the results. Then b would be the principal symbol of A^*A with all branches combined, as stated. □

Note that in particular, if (4.7) holds, then $\kappa^{-1^*}(\gamma_f \, \mathrm{d}x \, \mathrm{d}\xi) = \gamma_g \circ \kappa^{-1} \, \mathrm{d}y \, \mathrm{d}\eta$.

Remark 4.2. The proof also implies that if $Q = q(x, hD)$ and $R = r(x, hD)$ are h-ΨDOs, then

$$\mathrm{d}\mu_{h^m QARf} = |q|^2 \kappa^{-1^*}(b|r|^2 \, \mathrm{d}\mu_f) \quad \text{on } T^*\mathbf{R}^n \setminus 0, \tag{4.42}$$

where b still denotes the principal symbol of A^*A.

Example 2. Take $R = r(x)$ (i.e., a multiplication) with r smooth. Then, up to $O(h)$, equality (3.15) for rf_h takes a similar form but now $\boldsymbol{f}_{k,h}$ are replaced by $r(shk)\boldsymbol{f}_{k,h}$. This is an example of non-homogeneous noise, depending on the position, for which Theorem 4.1 applies but then the measure is as in (4.42).

Example 3. Let R be a convolution with $h^{-n}\psi(x/h)$ with some $\psi \in C_0^\infty$. This is an h-ΨDO with symbol $\check{\psi}(\xi)$, therefore we get the factor $|r|^2 = |\check{\psi}(\xi)|^2$ in (4.42). An elementary computation shows that, up to $O(h)$, Rf_h is obtained from $\tilde{\boldsymbol{f}}_{k,h} = \sum_m \psi(s(k-m))\boldsymbol{f}_{m,h}$. Those are correlated (in general) random variables. They model sensors with cross-talk. Then Theorem 4.1 applies with the measure is as in (4.42).

Both examples are covered by Theorem 4.2 as well if you think of Rf as f but generated by correlated noise $\boldsymbol{f}_{k,h}$.

4.6. **Back to the inverse problem.** We return to the inverse problem (1.1) now. Let A be an FIO as in Theorem 4.3, and elliptic. More precisely, let $\Omega \subset \mathbf{R}^n$ be a bounded domain, and let Ω' be another such domain so that the canonical transformation κ of A maps $T^*\Omega$ into $T^*\Omega'$. By a compactness argument, if A is defined first as $A : \mathcal{E}'(\Omega) \to \mathcal{E}'(\mathbf{R}^n)$, then the range of κ projected to its base variable is a bounded set, thus such an Ω' exists. Outside $\bar{\Omega}'$, the image of A is smooth. The measurement g, supposedly equal to Af for some $f \in \mathcal{E}'(\Omega)$ but corrupted by noise, is a function defined in Ω'. Then (1.1) is microlocally solvable: $f = A^{-1}g$ (we do not have problems with g not being in the range because A^{-1} is a parametrix) and we are in the situation above with A replaced by A^{-1}. The added noise is given by (1.4). Dropping the subscript "noise" as we already did, we assume that g is given first as discrete noise $\{\boldsymbol{g}_k\}$ and then converted to a semiclassically band limited function g as in (3.15). Then

$$A^{-1}g = \sum_{shk \in \Omega'} \boldsymbol{g}_k A^{-1}\chi_k.$$

We have not defined what noise is but we can think of this as noise because it is a linear combination of $\{A^{-1}\chi_k\}$ with random coefficients. It has zero mean in the sense of (4.2). Then

$$\mathrm{d}\mu_{h^{-m}A^{-1}g} = \kappa^*(b^{-1}\mathrm{d}\mu_g) \quad \text{on } T^*\mathbf{R}^n \setminus 0, \tag{4.43}$$

where κ is the canonical transformation of A and b is the principal symbol of AA^*. By Egorov's theorem again applied to the operator $A^*(AA^*)A = (A^*A)^2$, the principal symbol of it is that of A^*A multiplied by $b \circ \kappa$. Therefore, $b \circ \kappa$ is the principal symbol of A^*A.

The defect measure (4.43) then describes the power spectrum of the noise in the reconstruction *away from the zero section* $\xi = 0$. We cannot expect to get an estimate near the zero section in this case since A may not be even injective. For example, the interior region of interest problem for the Radon transform in the plane has no unique solution and the practical solution is a parametrix. Then every element in the kernel would be smooth and could be considered as noise with zero frequency.

Next theorem is a direct consequence of (4.43). The operator Q is needed to cut the zero section, and R is a filter which we may want to apply to the data, see also next section. Below, $\sigma_p(Q)$ stands for the principal symbol of Q.

Theorem 4.4. *Let A be as above, and elliptic, and let $g = g_h$ be semiclassically band limited with $\mathrm{WF}_h(g) \subset T^*\Omega'$, uniformly bounded in $L^2(\Omega')$. If $R = r(x, hD)$ is any h-ΨDO in Ω' with an h-independent symbol, and if $Q = q(x, hD)$ is a similar h-ΨDO in Ω with $q = 0$ near the zero section, then*

$$\begin{aligned} \mathrm{VAR}^0_\Omega(Qh^{-m}A^{-1}Rg) &= \frac{1}{|\Omega|}\int_{T^*\Omega} |q|^2 \sigma_p(A^*A)^{-1}\kappa^*\left(|r|^2\,\mathrm{d}\mu_g\right) \\ &= \frac{1}{|\Omega|}\int_{T^*\Omega'} |q\circ\kappa^{-1}|^2 \sigma_p(AA^*)^{-1}|r|^2\,\mathrm{d}\mu_g \end{aligned} \tag{4.44}$$

for every g (called there f) as in Theorem 4.3.

Proof. By Remark 4.2 about Theorem 4.3 and (4.6),

$$\mathrm{VAR}^0_\Omega(Qh^{-m}A^{-1}Rg) = \frac{1}{|\Omega|}\int_{T^*\Omega} |q|^2 \kappa^*\left(b^{-1}|r|^2\mathrm{d}\mu_g\right).$$

Make the change of variables $(y, \eta) = \kappa(x, \xi)$, where (y, η) are the variables in the phase space of g, using the fact that κ is symplectic, and therefore an isometry, to get the second equality of the theorem. □

A typical use of this theorem is to take q to cut off smoothly a small neighborhood of the zero section. Then, for g being white noise, for example, the effect of that on the r.h.s. would be small. Then if we formally take $q = 1$, hence $Q = \mathrm{Id}$, we get a good approximation of the variance of the noise in the reconstruction away from the zero frequency noise, by Theorem 4.1. The operator R plays a role of a filter before the inversion.

We want to emphasize that g in Theorem 4.4 does not need to be white noise; we just need a well-defined $\mathrm{d}\mu_g$, which is the case for noise satisfying Hypothesis 4.2, by Theorem 4.2.

Remark 4.3. In some situations, like in the next two sections, the requirement $q = 0$ near the zero section can be removed, and the whole operator Q can be removed (replaced by Id). Assume that the filter r is compactly supported in the dual variable. Since we deal with semiclassically band limited g, we can always assume that. Assume that $\sigma_p(A^*A)^{-1}\kappa^*\mathrm{d}\mu_g$ is absolutely continuous near the zero section. In the case of the Radon transform in parallel geometry in the next section, for example, with g being white noise, that measure is $C|\xi|\mathrm{d}x\,\mathrm{d}\xi$, so this assumption is satisfied. Then the first integral in (4.44) has a limit when q (a priori vanishing near $\xi = 0$) tends to 1, and that limit is given by the same formula with $q = 1$. Then the l.h.s. has the same limit, too,, because we just defined it by that equality, see (4.6). A similar remark applies to the second integral.

5. The Radon transform in "parallel geometry"

We apply the theory to the Radon transform now. We study the parallel geometry parameterization first, where each (directed) line is parameterized by its signed distance p to the origin, and its normal ω, see (1.5). For

$$\omega(\varphi) = (\cos\varphi, \sin\varphi), \tag{5.1}$$

we choose the natural measures $\mathrm{d}\varphi$; and the standard measure $\mathrm{d}p$ for p. Based on that, we a define the microlocal defect measure $\mathrm{d}\mu_g(\varphi, p, \hat\varphi, \hat p)$ of $g = g_h(\varphi, p)$. If we restrict p to $|p| \le R$, corresponding to Radon transforms of functions supported in $B(0, R)$, since φ naturally belongs to $|\varphi| \le \pi$ (modulo 2π) (call that Ω), then

$$\mathrm{VAR}^0_\Omega(g) = \frac{1}{4\pi R}\int\int_\Omega \mathrm{d}\mu_g(\varphi, p, \hat\varphi, \hat p). \tag{5.2}$$

The Radon transform is an FIO of order $-1/2$ with a canonical relation given by the union of canonical relations corresponding to the canonical transformations

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

The ranges of $\kappa_\pm$ intersect in the zero section only, and in particular, $\pm\hat p \ge 0$ on the range of $\kappa_\pm$. Next, each branch is a local diffeomorphism. Indeed, $(x, \xi) = \kappa_\pm^{-1}(\varphi, p, \hat\varphi, \hat p)$ is given by

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

It is well defined for $\hat p \ne 0$ but if we want x in the image to be in $|x| < R$, we need to require $p^2 + (\hat\varphi/\hat p)^2 < R^2$; therefore $\kappa_\pm^{-1}$ are well defined away from the zero section. Then $\mathcal{R}^{-1}$ is associated with κ^{-1}, which is a local diffeomorphism as well. What prevents it from being global is that it is 2-to-1, i.e., and in particular, it is not injective.

5.1. **The unfiltered inversion.** The symbol of $\mathcal{R}\mathcal{R}^*$ is $b = 4\pi|\hat{p}|^{-1}$, where $\hat{p}$ is the dual of p. Applying the canonical transformation, we get $b \circ \kappa = 4\pi/|\xi|$. We could have obtained this as the principal (and full) symbol $4\pi/|\xi|$ of $\mathcal{R}^*\mathcal{R}$. Therefore, by (4.43),

$$\mathrm{d}\mu_{h^{1/2}\mathcal{R}^{-1}g}(x,\xi) = \frac{|\xi|}{4\pi}\kappa^*\mathrm{d}\mu_g \quad \text{on } T^*\mathbf{R}^2 \setminus 0. \tag{5.3}$$

The fact that κ is 1-to-2 presents some subtlety here, already accounted for in the proof of Theorem 4.3. Microlocally, one can express $\mathcal{R}$ as $\mathcal{R} = \mathcal{R}_+ + \mathcal{R}_-$; then each $\mathcal{R}_\pm$ has normal operator $\mathcal{R}_\pm^*\mathcal{R}_\pm$ with principal symbols one half of that $\mathcal{R}^*\mathcal{R}$; then we apply (4.43), and the combined result would be still the principal symbol of $\mathcal{R}^*\mathcal{R}$.

Let us say that we have f supported in $B(0,R)$ with a certain semiclassical band limit $B \geq |\xi|$. We take its Radon transform $\mathcal{R}f$. Here, f is not discretized, we can think of $\mathcal{R}f$ as the physical X-ray transform. The assumption on the band limit will be satisfied if the X-rays are not really ideal lines but have some thickness. Then we sample $\mathcal{R}f$ densely enough to satisfy the Nyquist requirements and add noise to it. The noise will have higher frequencies than those coming from f if $\mathcal{R}f$ is oversampled. When we invert $\mathcal{R}f$, we will get higher frequencies for f as well that do not originally belong to the set where the frequency set of f lies. We can apply a filter, cutting them to $|\xi| \leq B$. Note that this is a filter not affecting f, that is why we think of those as a unfiltered inversion. One way to do this is to restrict $\hat{p}$ to $|\hat{p}| \leq B$ before applying $\mathcal{R}'$ in (1.6).

More precisely, let $\operatorname{supp} f \subset B(0,R)$ and

$$\mathrm{WF}_h(f) \subset \{(x,\xi);\ |x| \leq R,\ |\xi| \leq B\}. \tag{5.4}$$

Then the range of the frequency sets $\Sigma(\mathcal{R}f)$ of all such f's (the projection of the semiclassical wave front set on the fiber variable) of $\mathcal{R}f$ is the double cone

$$\big\{(\hat{\varphi},\hat{p});\ |\hat{\varphi}| \leq R|\hat{p}|,\ |\hat{p}| \leq B\big\}, \tag{5.5}$$

included in the box $\mathcal{B} := \{|\hat{\varphi}| \leq RB,\ |\hat{p}| \leq B\}$, see Figure 4 and [17] for more details. The set (5.5) is the "worst scenario case" over all points (φ,p). For $|p| \gg 0$, the opening of the cone is much smaller: $|\hat{\varphi}| \leq |\hat{p}|\sqrt{R^2-p^2}$. We refer to [17] and Figure 3 there. This describes the range of κ. Therefore, some portion of the noise will not propagate back to the reconstructed f.

We assume that we sample $g = \mathcal{R}f$ at a rate smaller than the Nyquist requirement for the box $\mathcal{B}$. Moreover, we assume an interpolation kernel χ in (3.15) (with f replaced by g) is chosen so that $\hat{\chi} = 1$ in a neighborhood of $\mathcal{B}$. As we explained in the introduction, we assume that the data is (white) noise, since the problem is linear. Then the power spectrum of the noise (more precisely, the Wigner function) converges in mean sense to a defect measure $\mathrm{d}\mu_g$ that is absolutely continuous by Theorem 4.1, i.e., it has the form $\mathrm{d}\mu_g = \gamma_g\,\mathrm{d}x\,\mathrm{d}\xi$ of the kind (4.7) on $\mathcal{B}$, with γ_g as in (4.15). Then on $\mathcal{B}$, we have $\gamma_g = s^n\sigma^2/(2\pi)^n =: \gamma^\sharp$, and

$$\gamma_{h^{1/2}\mathcal{R}^{-1}g}(x,\xi) = \frac{|\xi|}{4\pi}\gamma^\sharp \quad \text{for } |\xi| \leq B. \tag{5.6}$$

This is "blue noise". Here and below, all equalities about the statistics of f are in the limit sense of Theorem 4.1, see (4.6) and (4.8). An important observation is that there is no x dependence in this case. The dependence on ξ is rotationally invariant. This is not the case with the Radon transform in fan-bean coordinates as we will see below.

By (5.2),

$$\mathrm{VAR}^0_\Omega(g) = \mathrm{VAR}^0_{p,\varphi}(g) = 4B_\varphi B_p\gamma^\sharp, \quad \forall(\varphi,p) \in S^1 \times [-R,R].$$

The two variances are equal because $\gamma_{\mathcal{R}^{-1}g}$ is independent of the position.

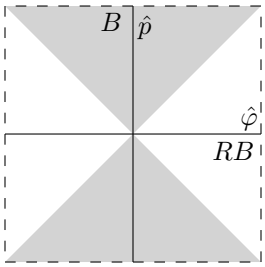


Figure 4. The frequency set of $\mathcal{R}f$.

Assume that the sampling rates of g are based on B_φ and B_p which take their sharp values not to allow undersampling: $B_p = B$, $B_\varphi = RB$, where B is the band limit of f as in (5.4). Then

$$\mathrm{VAR}^0_{p,\varphi}(g) = 4RB^2\gamma^\sharp.$$

Note that this is actually the sharp lower bound of the variation when the oversampling becomes asymptotically sharp sampling but it is not achievable in our theory; this would require a sinc interpolation while we need a rapidly decreasing kernel.

For the variance of $f = \mathcal{R}^{-1}g$, we have, see Theorem 4.4 and Remark 4.3,

$$\begin{aligned}\mathrm{VAR}^0_x(h^{1/2}\mathcal{R}^{-1}g) &= \int_{|\xi|\le B} \gamma_{\mathcal{R}^{-1}g}(x,\xi)\,\mathrm{d}\xi \\ &= \frac{1}{4\pi}\gamma^\sharp \int_{|\xi|<B} |\xi|\,\mathrm{d}\xi = \frac{1}{4\pi}\gamma^\sharp 2\pi\int_0^B \rho^2\,\mathrm{d}\rho \\ &= \frac{B^3\gamma^\sharp}{6}.\end{aligned} \tag{5.7}$$

We get the following theorem.

Theorem 5.1 (unfiltered inversion)**.** *Under the assumptions above, in particular assuming that g is white noise, and no undersampling, we have*

$$\mathrm{STD}^0(\mathcal{R}^{-1}g) = \frac{B^{3/2}}{\sqrt{24B_\varphi B_p h}}\,\mathrm{STD}^0(g). \tag{5.8}$$

If $g = \mathcal{R}f$ is sampled sharply, then

$$\mathrm{STD}^0(\mathcal{R}^{-1}g) = \left(\frac{B}{24Rh}\right)^{1/2}\mathrm{STD}^0(g). \tag{5.9}$$

Recall that we defined VAR^0, see (4.6), and similarly, STD^0, as integral of the defect measure. The implication of this theorem is that when we have g created by a white noise process, then for every $Q = q(hD)$ with $q = 0$ near the origin, $\mathrm{STD}^0(h^{1/2}Q\mathcal{R}^{-1}g)$ converges in mean square sense to a quantity (see (5.13)), which itself converges to the r.h.s. of (5.8), respectively (5.9), when $q \to 1$. In other words, the cutoff near $\xi = 0$ is removable at the expense of taking a double limit: first $h \to 0$, then $q \to 1$ (in L^1 sense).

5.2. **The filtered inversion.** The Radon transform is inverted often with a low-pass filter before applying $\mathcal{R}'$ in (1.6), i.e.,

$$f = \frac{1}{4\pi}\mathcal{R}'\nu(D_p)|D_p|g, \tag{5.10}$$

where ν is an even function decaying away from the origin. Assuming a band limit B_p for the p variable, determined by the sampling rate s_p, for example, one popular filter is the Hann filter:

$$\nu_{\text{Hann}}(\hat{p}) = \frac{1}{2}\left(1 + \cos\frac{\pi\hat{p}}{B_p}\right) = \cos^2\frac{\pi\hat{p}}{2B_p}, \quad |\hat{p}| \le B_p, \tag{5.11}$$

and $\nu_{\text{Hann}}(\hat{p}) = 0$ otherwise. Another commonly used filter is the cosine one

$$\nu_{\text{cosine}}(\hat{p}) = \cos\frac{\pi\hat{p}}{2B_p}, \quad |\hat{p}| \le B_p.$$

They are plotted in Figure 5.

Figure 5. The Hann and the cosine filters with $B = 1$.

There are many other filters (windows) used in signal processing and imaging. We assume that ν is continuous and supported in $|\hat{p}| \le B_p$. If the shape of the filter is fixed, say Hann, then $\nu(t) = \nu_0(t/B_p)$ with some fixed ν_0 supported in $[0,1]$, see, e.g., (5.11). Then (5.3) takes the form

$$\gamma_{h^{1/2}\mathcal{R}_\nu^{-1}g}(x,\xi) = \frac{|\xi|\nu_0^2(|\xi|/B_p)}{4\pi}\gamma_g \circ \kappa(x,\xi), \tag{5.12}$$

where $\mathcal{R}_\nu^{-1} = \mathcal{R}\nu(D_p)$ is the filtered inversion, defined as the operator applied to g in (5.10). Then the equivalent to (5.6) is

$$\gamma_{h^{1/2}{\mathcal{R}_\nu}^{-1}g}(x,\xi) = \frac{|\xi|\nu_0^2(|\xi|/B_p)}{4\pi}\gamma^\sharp.$$

Taking $B_p = B$ as before, similarly to (5.7) we get the following analog of (5.7)

$$\begin{aligned}
\operatorname{VAR}_x^0(h^{1/2}{\mathcal{R}_\nu}^{-1}g) &= \int_{|\xi|\le B} \gamma_{\mathcal{R}_\nu^{-1}g}(x,\xi)\,\mathrm{d}\xi \\
&= \frac{1}{4\pi}\gamma^\sharp \int_{|\xi|<B} |\xi|\nu_0^2(|\xi|/B)\,\mathrm{d}\xi = \frac{1}{4\pi}\gamma^\sharp 2\pi\int_0^B \rho^2\nu_0^2(\rho/B)\,\mathrm{d}\rho \\
&= \frac{1}{6}B^3\gamma^\sharp c_\nu,
\end{aligned} \tag{5.13}$$

where

$$c_\nu := 3\int_0^1 \rho^2\nu_0^2(\rho)\,\mathrm{d}\rho. \tag{5.14}$$

We proved the following.

Theorem 5.2 (filtered inversion)**.** *Under the assumptions above, in particular assuming white noise and no undersampling, with a filter $\nu_0(|D_p|/B)$, we have*

$$\operatorname{STD}^0(\mathcal{R}^{-1}g) = \frac{B^{3/2}\sqrt{c_\nu}}{\sqrt{24B_\varphi B_p h}}\operatorname{STD}^0(g). \tag{5.15}$$

If $\mathcal{R}f$ is sampled sharply, then

$$\mathrm{STD}^0(\mathcal{R}^{-1}g) = \left(\frac{Bc_\gamma}{24Rh}\right)^{1/2} \mathrm{STD}^0(g).$$

If there is no filter ($\nu_0 = 1$), we have $c_\nu = 1$, which explains the appearance of the factor $1/3$ in the definition of c_ν. For the Hann filter, $c_\nu = 3/8 - 45/(16\pi^2) \approx 0.0900$, then $\sqrt{c_\nu} \approx 0.3000$. For the cosine filter, $\sqrt{c_\nu} \approx 0.4427$. In (5.19) below, the constant would be approximately 0.07676 for the Hann filter and 0.11327 for the cosine one.

5.3. **Numerical experiments.** We use MATLAB and the built in `radon` and `iradon` routines to compute and invert numerically the Radon transform in the plane. The default angular step is one degree but it can be changed. Assume that f is given on an $N \times N$ lattice. Then by default, `radon` computes $\mathcal{R}f(\varphi, p)$ on a $360 \times N\sqrt{2}$ lattice, with $N\sqrt{2}$ rounded; the actual formula is $2\,\mathrm{ceil}\big(\sqrt{2}(N - \mathrm{floor}((N-1)/2) - 1)\big) + 3$. Then `iradon` inverts the data to the original grid (with N replaced by $N+1$ or $N+2$ which does not matter in view of our asymptotic setup).

As we showed in [17], this choice of the discretization of $\mathcal{R}f$ is suboptimal for $N \gg 1$; we need to compute $\mathcal{R}f$ on an $N_\varphi \times N_p$ lattice with $N_p = 2N$, $N_\varphi = 2\pi N$ at least, and some oversampling would be beneficial, see Figure 6. With most test images, the (dominating) frequencies are well below the Nyquist limit, that is why most of the time the inversion is satisfactory. When we add, say white noise, the Nyquist limit is reached, and the inversion with `iradon` will alias some of those frequencies.

5.4. **Discretization.**

Let us say we have f on an $M \times N$ grid. We think of that as discrete samples of f originally defined on, say, $[-a, a] \times [-b, b]$. This we have the steps $s_{x_1} = 2a/M$, $s_{x_2} = 2b/N$. Assume for a moment that we apply the classical sampling theory (no small parameter h) in a formal way at this point. Then those steps have to be π/B_{x_1}, respectively π/B_{x_2} at most, where B_{x_j} are the band limits in the x_j variable. Then we get $B_{x_1} = M\pi/(2a)$, $B_{x_2} = N\pi/(2b)$ as the least upper bounds of the band limits of f. For the band limit of $|\xi|$, we have $B = \big(B_{x_1}^2 + B_{x_2}^2\big)^{1/2}$, and the maximum is achieved at the vertices of the box $[-B_{x_1}, B_{x_1}] \times [-B_{x_2}, B_{x_2}]$. Note that the disk $|\xi| \le B$ contains more frequencies than can be properly sampled on the $M \times N$ grid; the extra ones lie outside that inscribed box.

We can connect the classical sampling theory to the semiclassical one as follows. Denote for a moment the semiclassical quantities with tildes over them. Let $M = \tilde{M}/h$, $N = \tilde{N}/h$, with $\tilde{M}$, $\tilde{N}$ fixed. The steps s (s_{x_1}, etc.) are equal to the semiclassical *relative* steps $\tilde{s}$ but since in our sampling theorems the absolute steps are sh, this means that the absolute steps are multiplied by h. Then our analysis holds as $h \to 0$, i.e., as $M \to \infty$, $N \to \infty$ (keeping the ratio constant) and the steps going to zero at a rate $\sim h$. This is the usual setup in numerical analysis where $\tilde{s} = 1$, i.e., the step is h.

For each such f we define the L^2 norm as

$$\|f\|^2 = \frac{4ab}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} |f_{ij}|^2.$$

This is consistent with formula (16) in [17] and approximates the L^2 norm of a continuous function on that box with samples f_{ij}. Then

$$\mathrm{STD}(f) = \Big(\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}|f_{ij}|^2\Big)^{\frac{1}{2}} = \frac{\|f\|}{2\sqrt{ab}}$$

is the standard deviation $\mathrm{STD}(f)$ of f when the mean of f is zero.

We will apply this to both f defined on $[-a,a]^2$ for some $a>0$, and to $\mathcal{R}f$ on $[-\pi,\pi]\times[-R,R]$.

Assume that g is a discrete representation of a function on $[-\pi,\pi]\times[-R,R]$ sampled on an $N_\varphi \times N_p$ lattice. Assume g is obtained by a white noise process (with zero mean) and variance σ^2. Then a slight extension of Lemma 4.1 shows that $\mathrm{VAR}(g)\to\sigma^2$ almost surely.

The sampling steps are $s_\varphi = 2\pi/N_\varphi$, $s_p = 2R/N_p$; hence to avoid aliasing, we need $B_\varphi \le N_\varphi/2$, $B_p \le \pi N_p/(2R)$.

Let f, to which $\mathcal{R}$ will be applied, represent a discretization of a function on $[-a,a]^2$, and assume that it is sampled on an $N\times N$ lattice. Then, similarly, the sharp band limit in each variable is $B_{x_1} = B_{x_2} = \pi N/(2a)$.

As we showed in [17], and it follows easily from (5.5), to avoid aliasing, we need

$$N_p \ge 2N, \quad N_\varphi \ge 2\pi N. \tag{5.16}$$

This inequality, as well as the inequalities and the equalities below are meant in asymptotic sense, i.e., one should multiply, say the r.h.s. in this case by $(1+o(1))$, as $N\to\infty$. Note that (5.16) follows from viewing f as supported in $B(0,\sqrt{2}a)$, i.e., $R=\sqrt{2}R$, with frequency set in $|\xi|\le B := \sqrt{2}B_{x_1}$. As we mentioned above, that ball contains more frequencies than those in its inscribed square. For every $g=\mathcal{R}f$ in the range of $\mathcal{R}$ with f as above, after an inversion we get f, of course, and then the frequencies will fall inside the inscribed square $[-B_{x_1},B_{x_1}]^2$. If we take g to be "noise", not in the range of $\mathcal{R}$, then by the mapping property of κ^{-1}, see [17], formula (51), the frequency set of $\mathcal{R}^{-1}g$ will generically fill the disk $|\xi|\le B = B_{x_1}\sqrt{2}$. If we want to avoid aliasing (without applying a filter), we would need to reconstruct $f=\mathcal{R}^{-1}g$ on an $N\sqrt{2}\times N\sqrt{2}$ grid or better. On the other hand, for all practical purposes, we would want to apply a filter.

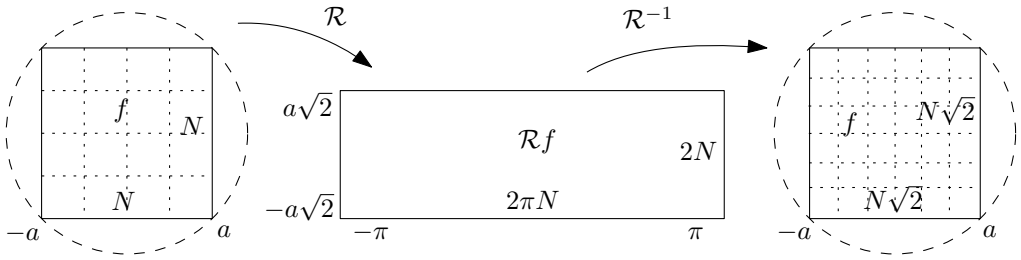


Figure 6. The sampling sets of f, $\mathcal{R}f$ and the reconstructed f with sharp sampling requirements.

Therefore, the discrete version of (5.15), including a filter now, is

$$\mathrm{VAR}(\mathcal{R}_\nu^{-1}g) = \frac{\pi^2 N^3 c_\nu}{12a^2 N_\varphi N_p}\,\mathrm{VAR}(g), \tag{5.17}$$

where the formula has the same asymptotic and probabilistic meaning as explained after Theorem 5.1.

Assume now that we sample sharply, i.e., we have equalities in (5.16). Then $N_p = 2N$, $N_\varphi = 2\pi N$ and we get

$$\mathrm{VAR}(\mathcal{R}_\nu^{-1} g) = \frac{\pi N c_\nu}{48 a^2} \mathrm{VAR}(g). \tag{5.18}$$

Therefore,

$$\mathrm{STD}(\mathcal{R}_\nu^{-1} g) \approx 0.2558 \sqrt{c_\nu} \frac{\sqrt{N}}{a} \mathrm{STD}(g). \tag{5.19}$$

We can make the following conclusions from (5.17), (5.18) and (5.19).

- With a sharp sampling rate, the noise ratio, measured as its standard deviation relative to that of g, increases as $\sqrt{N}$. This is understandable since we are allowing for higher frequencies, and $\mathcal{R}^{-1}$ is of order $1/2$. At the same time, we can handle f with higher frequencies because N is proportional to the Nyquist bound.
- The noise ratio, for a fixed N, is minimized when we sample sharply.
- In many applications, increasing N_ϕ and N_p decreases the size of the detectors, and then the discrete samples g_{ij} are scaled down by constants times N_ϕ and N_p. If the added noise is expressed in units relative to that, then the quotient in (5.17) would be proportional to $N^3 N_\varphi N_p$, i.e., the noise ratio increases with N_ϕ and N_p. This is known in engineering.

Default `iradon` inversion. First we present an inversion with the default one degree angular step. We choose $N = 601$, $N_\varphi = 360$ by default and $N_p = 853$ is chosen by `radon` as an approximation to $601\sqrt{2}$. We choose g to be normally distributed (Gaussian) noise with standard deviation one. Then we invert it with `iradon`. A plot of the modulus $|\hat{f}|$ of the Fourier transform $\hat{f}$ of the inversion f is shown in Figure 7. We chose to plot here and below $|\hat{f}|$ rather than $|\hat{f}|^2$ for clarity.

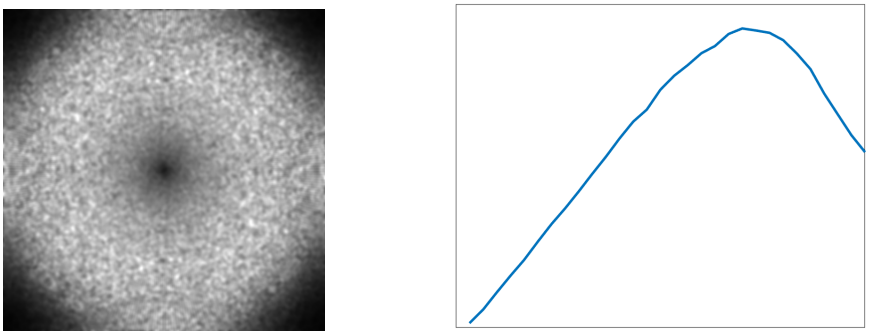


Figure 7. Left: $|\hat{f}|$ where $f = \mathcal{R}^{-1} g$ and g is white noise. Right: radial profile of $|\hat{f}|^2$ from the center to one of the sides (but not all the way along the diagonal to a vertex).

With an exact inversion, as $N \to \infty$, we should be seeing a density plot of square root of (5.6), i.e., $c|\xi|^{1/2}$, filling the whole square. We see is that the density increases in the radial variable $|\xi|$ from the center but at some point starts to decease until it visibly becomes zero when $|\xi|$ is slightly larger than a half of the side, and it is radially symmetric. This behavior can be explained by the following. The default choice $N_p = N\sqrt{2}$ (rounded) of N_p actually lowers the Nyquist limit of the reconstructed f to $1/\sqrt{2}$ of its original value. Without that, the boundary of the disk in Figure 7 would be the circumscribed circle of that square but with that choice, it is the inscribed one. The gradual decrease close the border can be explained by an effectively low pass filter when inverting $\mathcal{R}$. Our numerical experiments below at much higher resolutions for $\mathcal{R}f$ confirm that.

A similar experiment with a uniformly distributed noise g in a symmetric interval around the origin produces virtually the same plot of $|\hat{f}|$, not shown. In both cases, the values of f look normally distributed.

High precision inversion. We present numerical inversions with a proper discretization. We want to model adding noise to discrete measurements of the "continuous" $\mathcal{R}f$; inverted with high precision; i.e., by upsampling first the discrete data several times to mimic inversion in the "continuous domain". We do the following.

(i) The function f is assumed to be defined on $[-1,1]^2$ and sampled on an $N \times N$ lattice.
(ii) We compute a high accuracy $\mathcal{R}f$ on a $N_\varphi \times N_p$ lattice, where $N_\varphi \geq 2\pi N$, $N_p \geq 2N$. To do that, we perform the computations on a finer grid.
(iii) We add noise to the so-computed $\mathcal{R}f$.
(iv) We invert the noisy data by upsampling it first. The reconstructed f_{noisy} is either left sampled on a finer grid or downsampled to the original $N \times N$ one.

We give more details below. To do (ii), we upsample f on an $mN \times mN$ lattice with `Lanczos-3` with some $m \geq 1$. Typical m's we use are $m = 2$ and $m = 3$. Then we compute $\mathcal{R}f(\varphi, p)$ with `radon` on a $2\pi mN \times 2mN$ lattice which we view as $\mathcal{R}f(\varphi, p)$ on $[-\pi, \pi] \times [-\sqrt{2}, \sqrt{2}]$ sampled uniformly in each variable. The parameter m represents the degree of oversampling: $m = 1$ corresponds to the sharp lower bound for proper sampling. Since computing $\mathcal{R}f$ involves interpolation of f for computing the line integrals (we use the option 'spline' in `radon`), such an oversampling allows us to reduce the errors in such interpolation compared to the sinc inversion. Then we downsample the computed $\mathcal{R}f$ to a lower resolution $N_\varphi \times N_p$ (without interpolation; we take every m-th value in each row and column). This simulates a high precision $\mathcal{R}f$ computed on the $N_\varphi \times N_p$ grid. To do (iii), we add noise.

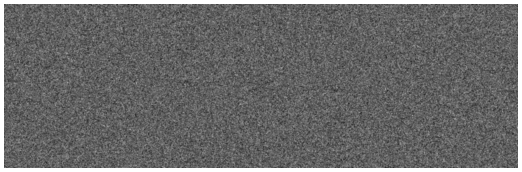
In (iv), we invert $\mathcal{R}$ on that lattice. We could resize to a different (but high enough resolution) before that but the results do not look much different. The resulting $f = \mathcal{R}^{-1}g$ is computed on an $mN \times mN$ lattice, which is viewed as f on $[-1,1]^2$ sampled uniformly. If needed, that f could be resampled to an $N \times N$ lattice but since it does not contain frequencies higher than the Nyquist limit $B = N\pi/2$ corresponding to N, this is not needed for computing the standard deviation, for example.

We want to emphasize that it is possible to do (close to) ideal upsampling, say from $N_x \times N_y$ to $\tilde{N}_x \times \tilde{N}_y$ with $\tilde{N}_x > N_x$ and $\tilde{N}_y > N_y$ which preserves the band limits $B_{x_1} = N_x\pi/2$ and $B_{x_2} = N_y\pi/2$ by using the Fourier transform. On the other hand, this is not what is usually done. When we use `Lanczos-3`, for example, the interpolation kernel is the inverse Fourier transform of a smoothened version of $\nu_{[-1,1]}$, see Figure 1, which is close to be equal to one in $[-0.5, 0.5]$ at least as explained in section 3.3. On the other hand, Theorem 3.1 in [17] requires some oversampling, and an interpolation kernel to be the Fourier transform of a function similar to that in Figure 1, equal to one on the (smaller) frequency band. Therefore if we choose $m \geq 2$ we are in this regime.

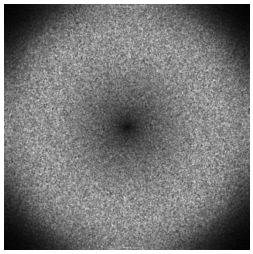
To do experiments with noise only, we take $f = 0$ in (1.2). Then steps (i) and (ii) are trivial, since $\mathcal{R}f = 0$. So our starting point is (iii), where we take g to be generated by either a normally or a uniformly distributed noise, on an $N_\varphi \times N_p$ grid. We upsample by a factor of m, i.e., to an $mN_\varphi \times mN_p$ grid an do the inversion there. We take $m = 2, 3, 4, 5$ in our experiments.

Non-filtered inversion. We test (5.8) now. To this end, we take g to be either Gaussian or uniformly distributed noise with zero mean on an $N_\varphi \times N_p$ grid as in (ii), with equalities there, i.e., $N_\varphi = 2\pi N$, $N_p = 2N$. Then we cut the Fourier transform of the result sharply to $1/m$-th of the frequency box corresponding to the original resolution $N_\varphi \times N_p$, $m = 2, 3, \ldots$; denote this by $\nu_m(D)g_m$, and apply $\mathcal{R}^{-1}$ to it without changing the grid size. This procedure provides more

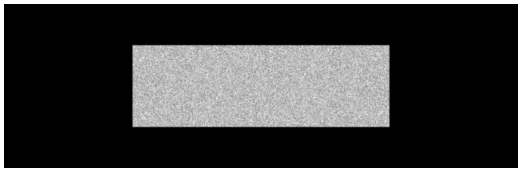
precise computation than just inverting the noise because it avoids the smoothing which happens in the part we cut off. If we had $\mathcal{R}f$ of a non-zero f polluted with noise, we would have upsized the data m times in each dimension first, and then would have performed that procedure.



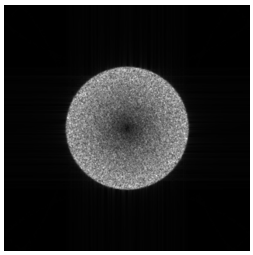
(a) $|\hat{g}_m|$ with $m = 2$: white noise



(b) reconstruction



(c) $|\hat{g}_m|$ with $m = 2$ cut to a half in each dimension



(d) reconstruction

Figure 8. Top: We choose g_m, $m = 2$ to be white noise; then $|\mathcal{R}^{-1}g_m|$ looks like in Figure (7). Bottom: $\nu_m\hat{g}_m$ and $\mathcal{R}^{-2}\nu_m g_m$, i.e., the Fourier transform of the reconstruction after the frequency cut-off of the noise.

Since we effectively multiply both N_φ and N_p by m, by (5.17), we see that (5.19) can be written in terms of the noise ratio as

$$\text{Noise ratio} := m\,\mathrm{STD}(\mathcal{R}^{-1}g)\Big/\frac{\sqrt{N}}{a}\,\mathrm{STD}(g) \approx 0.2558. \tag{5.20}$$

We take g first to be a Gaussian noise with several choices of N and m; doing five experiments for each choice. The results are in Table 1 below and in Figure 8, we illustrate the inversion with $m = 2$. Similar experiments with a uniformly distributed noise with mean zero generate similar

Noise ratio with Gaussian noise. Theoretical ratio: **0.2558**			
	$N = 100$	$N = 200$	$N = 300$
m=1	$0.2224 \pm 0.61\%$	$0.2223 \pm 0.32\%$	$0.2226 \pm 0.16\%$
m=2	$0.2552 \pm 0.38\%$	$0.2572 \pm 0.25\%$	$0.2578 \pm 0.17\%$
m=3	$0.2569 \pm 0.70\%$	$0.2584 \pm 0.41\%$	$0.2591 \pm 0.07\%$

Table 1. Noise experiments

numbers, not shown.

Filtered inversion. We perform similar experiments with the Hann and the cosine filter. Since the Hann filter is very small near the band limit B, see Figure!5, the smoothing effect of the

interpolation used by `iradon`, see Figure 7, plays a negligible role. Modeling that smoothening by the Lanczos-3 profile, for example, see Figure 1, by introducing an extra factor in (5.14) shows an error of less than 1% in $\sqrt{c_\nu}$. Then even with $m = 1$, we get a result close to the theoretical one, which is approximately 0.07676 for the Hann filter and 0.11327 for the cosine one, as we computed above. For $N = 300$, for example, we get $0.0767 \pm 0.11\%$ for Hann and $0.1105 \pm 0.21\%$ for cosine, where the smoothing effect of `iradon` is a bit less compensated for. The numbers for normally and uniformly distributed noise are very close.

For the cosine filter, we plot $|\hat{f}|$ (instead of $|\hat{f}|^2$ for clarity), the computed radial profile of $|\hat{f}|^2$, and its theoretical one $\rho\nu^2_{\text{cosine}}(\rho) = \rho\cos^2(\pi\rho/2)$ in Figure 9 below. The radial profile is computed as $|\hat{f}|^2$ averaged over 25 concentric rings. In this case, $|\hat{f}(\xi)|^2$ is proportional to the microlocal defect measure of f at any fixed x (it does not depend on x). The Hann filter behaves similarly,

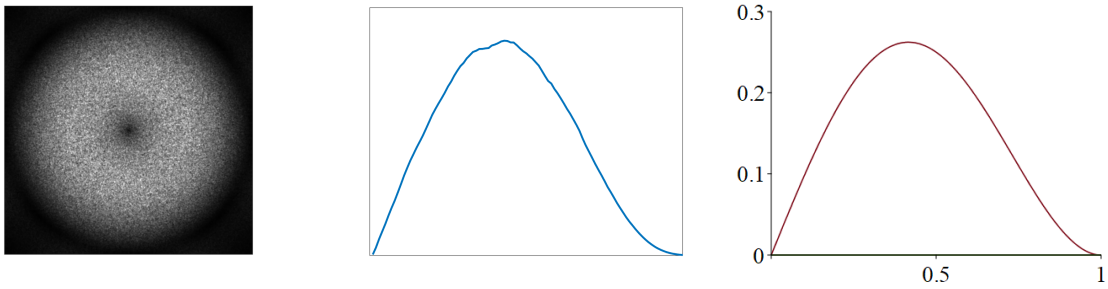


Figure 9. Cosine filter. Left: $|\hat{f}|$ where $f = \mathcal{R}^{-1}\nu_{\text{cosine}}g$ and g is white noise. Center: The computed radial profile of $|\hat{f}|^2$ from the center to one of the sides. Right: The theoretical profile $\rho\nu^2_{\text{cosine}}(\rho) = \rho\cos^2(\pi\rho/2)$.

with the computed radial profile of $|\hat{f}|^2$ very close to its theoretical one $\rho\cos^4(\pi\rho/2)$.

5.5. **Percentage of added noise.** In many numerical simulations, we add noise to the data, as a percentage of a certain norm of the data, and measure the percentage of the noise in the reconstruction. This is especially interesting in (mildly or not) ill-posed problems.

There is a lot of flexibility in choosing those norms. Let us say that we choose the $L^2(B(0,R))$ norm for f and the $L^2(S^1 \times (-R,R))$ norm for $\mathcal{R}$. Then the left inverse $\mathcal{R}^{-1}$ is not bounded in those spaces but on semiclassically bounded functions (which are smooth), it is; we refer to [17] for semiclassical estimates.

Let g_{noise} be the noise added to $g = \mathcal{R}f$, see (1.2). Its percentage is given my $\|g_{\text{noise}}\|/\|\mathcal{R}f\|$ (converted to percentage). We are interested in $\|f_{\text{noise}}\|/\|f\|$, where $f_{\text{noise}} = \mathcal{R}^{-1}g_{\text{noise}}$ is the noise in the reconstruction. We have

$$\frac{\|f_{\text{noise}}\|}{\|f\|} = K\frac{\|g_{\text{noise}}\|}{\|\mathcal{R}f\|}, \quad K := \frac{\|f_{\text{noise}}\|}{\|g_{\text{noise}}\|}\cdot\frac{\|\mathcal{R}f\|}{\|f\|}. \tag{5.21}$$

The coefficient K is the multiplier which relates the two percentages. Its first factor is proportional to the noise ratio we studied earlier since the L^2 norms are proportional to the standard deviations. The second one depends on f. To analyze it, write

$$\|\mathcal{R}f\|^2 \longrightarrow \int \gamma_{\mathcal{R}f}\,\mathrm{d}\varphi\,\mathrm{d}p\,\mathrm{d}\hat{\varphi}\,\mathrm{d}\hat{p},$$

where the convergence is in the sense of Theorem (4.1). Then we integrate over the semiclassical wave front. By (4.41),

$$\|\mathcal{R}f\|^2 \longrightarrow \int (b\gamma_f)\circ\kappa^{-1}\,\mathrm{d}\varphi\,\mathrm{d}p\,\mathrm{d}\hat{\varphi}\,\mathrm{d}\hat{p} = 4\pi\int \frac{\gamma_f\circ\kappa^{-1}}{|\hat{p}|}\,\mathrm{d}\varphi\,\mathrm{d}p\,\mathrm{d}\hat{\varphi}\,\mathrm{d}\hat{p}$$
$$= 4\pi\int \frac{\gamma_f(x,\xi)}{|\xi|}\,\mathrm{d}x\,\mathrm{d}\xi = 4\pi\big\||D|^{-1/2}f\big\|^2.$$

We used again the fact that κ is an isometry. This works for general operators but for $\mathcal{R}$ we actually know that $4\pi\|f\|^2 = \||D_p|^{1/2}\mathcal{R}f\|^2$. We can write $\tilde{f} = |D|^{1/2}f$, intertwine $|D|^{1/2}$ with $|D_p|^{1/2}$, to get the formula above as an exact one, not just a limit. Therefore, (5.21) yields

$$K = \frac{\|f_{\text{noise}}\|}{\|g_{\text{noise}}\|}\cdot\frac{4\pi\big\||D|^{-1/2}f\big\|}{\|f\|}. \tag{5.22}$$

Since the noise ratio is independent of f, we see that K would be large if, roughly speaking, f is low frequency. Most conventional images (with $f\geq 0$) have a very large zero frequency $\hat{f}(0)$ relative to the rest of the spectrum and the second quotient in (5.22) does not vary much. When $\int f(x)\mathrm{d}x = 0$, we have $\hat{f}(0) = 0$ and the variation of this quotient is higher. Then we do not need to isolate the zero section.

In Figure 10 we demonstrate this effect. We choose $N = 300$ and the dimensions of the grid for $\mathcal{R}f$ is chosen with equalities in (5.16), see also Figure 6. We add the same amount of normally distributed noise, 20% of $\|\mathcal{R}f\|$, to $\mathcal{R}f$. We measure different percentages of added noise to the reconstructed f depending on the frequency distribution of f, i.e., on the ratio in (5.22). Images with mostly lower frequencies suffer from noise more. On the other hand, given the a priori knowledge of their frequency band, that noise can be filtered out, unless we are looking for small high frequency detail in an overly lower frequency image. We chose non-negative f's in that figure only. Numerical experiments with f of mean value zero show lower added noise on a few examples. If in Figure 10(c) we allow random positive and negative amplitudes as well, for example (not shown), the added noise in (g) drops to 41%. It is worth mentioning that with many conventional images, the values we are getting are close. In fact, statistically, such images share similar power spectra distributions [18].

Therefore, measuring the sensitivity of a particular inversion to noise this way can be quite misleading. The added noise to the image depends on the noise ratio (5.20) which in turn depends on the grid chosen to discretize $\mathcal{R}f$; and also depends on the choice of the test image.

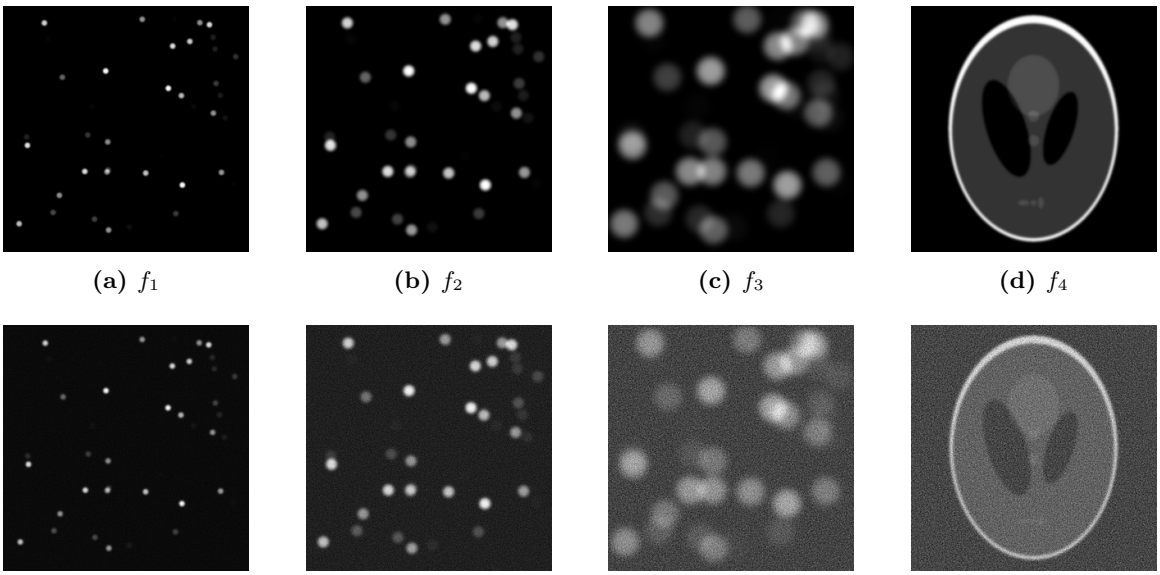
6. The Radon transform $\mathcal{R}$ in the plane in fan-beam coordinates

6.1. **$\mathcal{R}$ in fan-beam coordinates.** We parametrize $\mathcal{R}$ by the so-called fan-beam coordinates. Recall (5.1). 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 11. It is straightforward to see that this direction is given by $\omega(\alpha+\beta)$. Then the lines through $B(0,R)$ are given by

$$x\cdot\omega(\alpha+\beta-\pi/2) = R\sin\beta,\quad \alpha\in[-\pi,\pi],\ \beta\in[-\pi/2,\pi/2]. \tag{6.1}$$

The canonical relation is the union of the graphs of $\kappa_\pm$, see [17], given by

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



(a) f_1 **(b)** f_2 **(c)** f_3 **(d)** f_4

(e) 25%, 0.66 **(f)** 39.8%, 1.08 **(g)** 74.4%,2.06 **(h)** 79.8%, 2.18

Figure 10. Top: Four different choices of $f \geq 0$, $N = 300$. Bottom: f reconstructed with 20% noise added to $\mathcal{R}f$. The numbers show the added noise to f, and $\|\mathcal{R}f\|/\|f\|$.

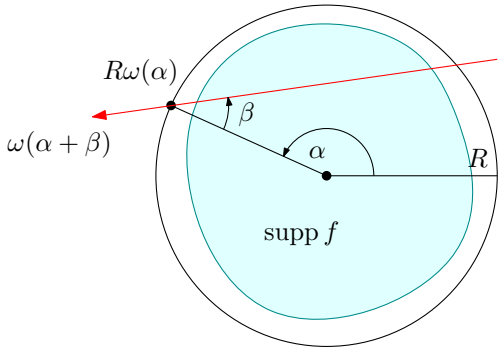


Figure 11. The fan-beam coordinates.

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

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

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

$$x = R\sin\beta\,\omega(\alpha+\beta-\pi/2) - \frac{\hat{\alpha}}{\hat{\beta}-\hat{\alpha}} R\cos\beta\,\omega(\alpha+\beta), \quad \xi = \frac{\hat{\beta}-\hat{\alpha}}{R\cos\beta}\omega(\alpha+\beta-\pi/2). \tag{6.2}$$

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

Set $(\varphi, p) = \Phi(\alpha, \beta)$, where

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

We have $\det \mathrm{d}\Phi = R\cos\beta$. Then $\mathcal{R}_{\mathrm{FB}} = \mathcal{R}\circ\Phi$. To compute $\mathcal{R}^*_{\mathrm{FB}}\mathcal{R}_{\mathrm{FB}}$, write

$$(\mathcal{R}^*_{\mathrm{FB}}\mathcal{R}_{\mathrm{FB}}f, f) = \int |\mathcal{R}_{\mathrm{FB}}f(\alpha,\beta)|^2\,\mathrm{d}\alpha\,\mathrm{d}\beta = \int |\mathcal{R}f(\varphi,p)|^2\frac{1}{\cos\beta}\,\mathrm{d}\varphi\,\mathrm{d}p.$$

Since $\sin\beta = p/R$, we have $\cos\beta = \sqrt{1-p^2/R^2}$. Therefore,

$$\mathcal{R}^*_{\mathrm{FB}}\mathcal{R}_{\mathrm{FB}} = \mathcal{R}^*(1-p^2/R^2)^{-1/2}\mathcal{R}.$$

The factor in the middle of the r.h.s. is a multiplication operator, and applying Egorov's theorem (one can actually do it even directly and without a remainder), one gets for the principal symbols, at least,

$$\sigma_p(\mathcal{R}^*_{\mathrm{FB}}\mathcal{R}_{\mathrm{FB}}) = \left(1-\frac{(x\cdot\xi)^2}{R^2|\xi|^2}\right)^{-1/2}\sigma_p(\mathcal{R}^*\mathcal{R}) = \frac{4\pi}{|\xi|}\left(1-\frac{(x\cdot\xi)^2}{R^2|\xi|^2}\right)^{-1/2}.$$

The equivalent to (5.3) then is

$$\gamma_{h^{1/2}\mathcal{R}^{-1}_{\mathrm{FB}}g}(x,\xi) = \frac{|\xi|}{4\pi}\left(1-\frac{(x\cdot\xi)^2}{R^2|\xi|^2}\right)^{1/2}\gamma_g\circ\kappa_{\mathrm{FB}}(x,\xi),\quad \xi\neq 0. \tag{6.3}$$

Therefore, the noise spectral distribution depends on x now, and it depends on the direction of ξ relative to x. For $x, |\xi|$ fixed, it is maximized when $\xi\perp x$, and minimized when $\xi\parallel x$.

6.2. **Sampling.** As above, if $\operatorname{supp} f\subset[-1,1]^2$ is sampled on an $N\times N$ grid, we have $B_{x_1} = B_{x_2} = N\pi/2$. As before, set $B = \sqrt{2}B_{x_1} = N\pi/\sqrt{2}$. Then we consider f having $\mathrm{WF}_h(f)$ in $B(0,R)\times B(0,\sqrt{2}B)$ with $R=\sqrt{2}$. The image of this product under the canonical map, projected to the dual variable $(\hat\alpha,\hat\beta)$ has the following smallest box containing it: $[-RB,RB]\times[-2RB,2RB]$, see [17]. The means taking at least $2RB\times 2RB$, i.e., $2N\pi\times 2N\pi$ samples over the intervals indicated in (6.1). Compared to (5.16), this requires π times the number of samples, which makes it a less efficient sampling geometry, as shown in [17].

In Figure 12, we present a numerical experiment to validate (6.3). We take g to be Gaussian noise and invert it with $\mathcal{R}_{\mathrm{FB}}$. Then we crop a small rectangle in the top left corner and take the modulus of its Fourier Transform. Then x is close to $x_0 = 0.8(-1,1)$ and the small black elongated oval in the center has a major axis along the same vector, formula (6.3) predicts.

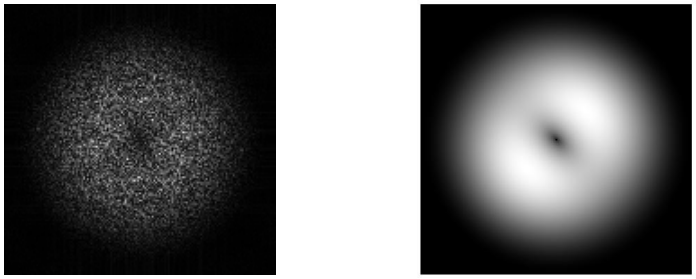


Figure 12. Spectral density of the noise in f with the Hann filter. Left: measured in the top left corner. Right: theoretical profile (6.3) at that corner.

6.3. **Noise Ratio.** We study the noise ratio with a filtered inversion. In `ifanbeam` in MATLAB, for example, $\mathcal{R}_{\rm FB}f$ is converted to parallel coordinates and the filter is applied after that. By (6.2), the filter $\nu(\hat{p})$, with ν even, takes the form $F := \nu\big(|\hat{\beta}-\hat{\alpha}|/(RB\cos\beta)\big)$, where B is the band limit of $|\xi|$. The inversion operator then is $\mathcal{R}^{-1}_{{\rm FB},\nu} = \mathcal{R}^{-1}_{\rm FB}F$ which equals $(\mathcal{R}^*_{{\rm FB},\nu}\mathcal{R}_{{\rm FB},\nu})^{-1}\nu(|D|)\mathcal{R}^*_{\rm FB}$ modulo lower order operators by Egorov's theorem. We get, similarly to (5.12) that (6.3) modifies as

$$\gamma_{h^{1/2}\mathcal{R}^{-1}_{{\rm FB},\nu}g}(x,\xi) = \frac{|\xi|}{4\pi}\left(1-\frac{(x\cdot\xi)^2}{R^2|\xi|^2}\right)^{1/2}\nu_0^2(|\xi|/B)\gamma_g\circ\kappa_{\rm FB}(x,\xi), \quad \xi\neq 0.$$

Assume that g is oversampled (related to B, see [17] for the sampling requirements), and it is white noise. Then the variance at a point, see (2.8) is given by

$$\begin{aligned}
{\rm VAR}_x(h^{1/2}f) &= \gamma^\sharp\int_{|\xi|\le B}\frac{|\xi|}{4\pi}\left(1-\frac{(x\cdot\xi)^2}{R^2|\xi|^2}\right)^{1/2}\nu_0^2(|\xi|/B_p)\,{\rm d}\xi,\\
&= \frac{\gamma^\sharp B^3}{4\pi}\int_0^{2\pi}\int_0^1\left(1-\frac{|x|^2}{R^2}\cos^2\theta\right)^{1/2}\rho^2\nu_0^2(\rho)\,{\rm d}\rho\,{\rm d}\theta\\
&= \frac{\gamma^\sharp B^3}{6}\frac{c_\nu}{2\pi}\int_0^{2\pi}\left(1-\frac{|x|^2}{R^2}\cos^2\theta\right)^{1/2}{\rm d}\theta,
\end{aligned}$$

compare with (5.13). The integral is of elliptic type and varies between 2π, when $|x|=0$, and 4 when $|x|=R$. To connect this to (5.13), the integrand in (5.13) there corresponds to $|x|=0$ formally; and then we get (5.13). Taking a square root, we see that the standard deviation would be higher in the center, the same as in the parallel geometry case, and will decrease slightly to about 80% at $|x|=R$, which corresponds to the four corners of the square in our numerical simulations.

Similarly to (5.7) and (5.13), we integrate over x in the inscribed disk $|x|\le 1$ in $[-1,1]^2$ and divide by its area π to get the variation in that disk. Then $R=\sqrt{2}$ and

$${\rm VAR}_{B(0,1)}(f) \approx \frac{2.93}{\pi}\frac{\gamma^\sharp B^3}{6h}c_\nu \approx 0.9328\frac{\gamma^\sharp B^3}{6h}c_\nu.$$

This is within 6-7% of the parallel geometry variance, and about 3% difference for the standard deviation.

7. Non-additive noise

In this section we discuss some types of non-additive noise. The exposition here will be more sketchy, we will point out how to fit those cases into the general framework we developed but will not go into detail.

7.1. **Multiplicative noise.** Assume the data $g=Af$ is subject to a multiplicative noise. This can happen if the detectors are not perfectly calibrated and each one reports a signal somewhat larger or smaller than it should be (non-uniform response). In imaging systems, photo response non-uniformity (PRNU) is an example of such noise. A generic way to model the kind of multiplicative noise we have in mind is the following: consider a sequence of discrete noise samples $\{\boldsymbol{w}_{k,h};\, h>0, k\in K(h)\}$, where

$$\boldsymbol{w}_{k,h} = 1+\boldsymbol{f}_{k,h}, \quad \text{or} \quad \boldsymbol{w}_{k,h} = \exp\left(\boldsymbol{f}_{k,h}\right), \tag{7.1}$$

and $\{\boldsymbol{f}_{k,h};\, h>0, k\in K(h)\}$ is the white noise considered in Hypothesis 4.1. Then we set

$$g_{\rm noise}(x) = \sum_{k\in K(h)}\boldsymbol{w}_{k,h}g(shk)\chi_k(x), \quad \chi_k(x)=\chi\Big(\frac{1}{sh}(x-shk)\Big), \tag{7.2}$$

where $\boldsymbol{w}_{k,h}$ are the discrete noise samples, playing the role of w_{noise} above, and g is the noise-free continuous signal. We will compare the noise $g_h(x)$ defined by (7.2), i.e., by the first formula in (7.1) (the second one can be treated similarly and one has to take into account that $\boldsymbol{w}_{k,h}$ is not necessarily centered), to a noise of the form

$$\tilde{g}_{\text{noise}}(x) := \sum_{k\in K(h)} \boldsymbol{w}_{k,h} g(x)\chi_k(x) = g(x) \sum_{k\in K(h)} \boldsymbol{w}_{k,h}\chi_k(x). \tag{7.3}$$

We have

$$\tilde{g}_{\text{noise}}(x) - g_{\text{noise}}(x) = \sum_{k\in K(h)} \boldsymbol{w}_{k,h} \left(g(x) - g(shk)\right) \chi_k(x).$$

Since

$$|g(x) - g(shk)| \le C|x - shk|,$$

with $C \le C_0\|\nabla g\|_{L^\infty}$, we get

$$|(g(x) - g(shk))\,\chi_k(x)| \le Csh \max(|x||\chi(x)|) \le C'h.$$

That factor h allows us to estimate, using Proposition 3.2, the error when replacing $g(shk)$ in (7.2) by $g(x)$. We would get an $O(h)$ error. The problem here is that we want to apply this to $g = Af$, all dependent on h, and in general, ∇Af grows like $h^{-1}|Af|$. This cancels the decay above. If we oversample a lot, the error will be "small". Also, in regions with $\mathrm{WF}_h(Af)$ far away from the Nyquist limit, that term will be small. If we ignore it for a moment, the noise added to Af is $\tilde{g}_h$ given by (7.3). It is white noise as above but multiplied by $g = Af$. The defect measure of the noise added to the data then is like in (4.15) with the additional factor $|Af(x)|^2$.

One important case which allows to overcome the difficulty above is when $g = \psi_h * g_0$, where $\psi_h(x) = h^{-n}\psi(x/h)$ with $\int \psi = 1$ (a Friedrichs mollifier) with $\hat{\psi} \in C_0^\infty$. This corresponds to averaged measurements of an h-independent function g_0. We refer to [17] for the sampling theory for such measurements. Then $\nabla g = \psi_h * \nabla g_0$, and assuming $g_0 \in C^1$ (either h-independent or uniformly bounded there in h), we have $|\nabla g| \le C$ (rather than C/h), which is the estimate we needed in the previous paragraph. Then the machinery we developed works and we need to multiply the noise measure by the additional factor $|g_0(x)|^2$ in (4.15), i.e., we get there

$$\mathrm{d}\mu_{g_{\text{noise}}}(x,\xi) = \frac{s^n}{(2\pi)^n}\sigma^2|g_0(x)|^2|\hat{\psi}(\xi)\hat{\chi}(s\xi)|^2\,\mathrm{d}x\,\mathrm{d}\xi$$

under the assumption $g = \psi_h * g_0$ in (7.2). Then (5.13) takes the form

$$\begin{aligned}\mathrm{VAR}_x^0(h^{1/2}{\mathcal{R}_\chi}^{-1} g_{\text{noise}}) &= \frac{1}{4\pi}\gamma^\sharp \int_{|\xi|<B} \left|\mathcal{R}f\big(x\cdot\xi/|\xi|, \arg(\xi)\big)\right|^2 |\xi||\nu_0(|\xi|/B)\hat{\psi}(\xi)|^2\,\mathrm{d}\xi \\ &= \frac{1}{4\pi}\gamma^\sharp \int_0^{2\pi}\int_0^B \left|\mathcal{R}f\big(x\cdot\theta,\theta\big)\right|^2 \rho^2|\hat{\psi}(\rho\theta)|^2\nu_0^2(\rho/B)\,\mathrm{d}\rho\,\mathrm{d}\theta.\end{aligned} \tag{7.4}$$

This shows that the standard deviation of the noise at x depends in particular on the line integrals of f along lines thorough x. Line integrals with large values would create stronger noise at x.

7.2. **Modeling noise in CT scan.** In CT scan tomography, what is measured is the attenuation along each ray. If I_0 is the initial intensity, and I is the one after the ray crosses the object, then the measurement is $I = \exp(-\mathcal{R}f)I_0$, by the Beer-Lambert law. Assuming an additive noise g_{noise}, we measure $I_{\text{noisy}} = \exp(-\mathcal{R}f)I_0 + g_{\text{noise}}$. If we invert this the same way as if there were no noise (which may not be the best strategy), we would get

$$f_{\text{noisy}} = -\mathcal{R}^{-1}\log(I_{\text{noisy}}/I_0) = -\mathcal{R}^{-1}\log(\exp(-\mathcal{R}f) + g_{\text{noise}}/I_0).$$

Obviously, increasing I_0 will decrease the effect of the added noise but in many applications this is not desirable and/or the noise level may depend on I_0. We take $I_0 = 1$, i.e., g_{noise} is the added noise relative to I_0. Then

$$\begin{aligned} -\log(\exp(-\mathcal{R}f) + g_{\text{noise}}) &= -\log\left(\exp(-\mathcal{R}f)(1 + \exp(\mathcal{R}f)g_{\text{noise}})\right) \\ &= \mathcal{R}f - \log\left(1 + \exp(\mathcal{R}f)g_{\text{noise}}\right). \end{aligned}$$

Therefore,

$$f_{\text{noisy}} = f - \mathcal{R}^{-1}\log\left(1 + \exp(\mathcal{R}f)g_{\text{noise}}\right) \tag{7.5}$$

If the noise is small enough, we can pass to a linearization to get

$$f_{\text{noisy}} \approx f - \mathcal{R}^{-1}\left(\exp(\mathcal{R}f)g_{\text{noise}}\right). \tag{7.6}$$

This is the multiplicative noise model above with gw_{noise} replaced by $e^g w_{\text{noise}}$. In (7.4), for example, the factor $|\mathcal{R}f|^2$ would be replaced by $\exp(2\mathcal{R}f)$.

7.3. **Modeling Poisson noise.** In SPECT, we measure the attenuated X-ray transform but the particle count at each detector is low. In this case, the predominant noise is of Poisson type: the number of particles at each detector is randomized by a Poisson distribution with probability of taking value k being $P(k, \lambda) = e^{-k}\lambda^k/k!$, where $\lambda \geq 0$ is the expected value at that detector, see [15, sec. 4.5]. Both the expected value and the variance of P equals λ. Then the particle count at each detector equals $\lambda + w\sqrt{\lambda}$, where w is a random variable with zero expected value and variance 1. Note that the probability distribution of w depends on λ and approximates a Gaussian one when $\lambda \gg 1$ and they are independent. Assuming locally averaged measurements, as above, we would get added noise $\boldsymbol{w}_{k,h}|\psi_h * \mathcal{R}f|^{1/2}$ when the units for $\mathcal{R}f$ are the number of particles; and α times that in general with some $\alpha > 0$. Note that $\boldsymbol{w}_{k,h}$ are not identically distributed (but Theorem 4.1 still applies) and are well approximated by Gaussian distributions when $\mathcal{R}f$ is not very small. The microlocal measure then would have the factor $\alpha\mathcal{R}f$ (we assume $f \geq 0$, thus $\mathcal{R}f \geq 0$). This is similar to multiplicative noise, where the factor was proportional to $|\mathcal{R}f|^2$.

7.4. **Numerical examples.** We present numerical simulations with the three types of non-additive noise in Figures 13 and 14. The phantoms are the Shepp-Logan one and three disks of different size and intensity, not shown there, both phantoms having ranges between 0 and 1. They are both rendered on a 300×300 grid discretizing the square $[-1, 1]^2$. Their Radon transforms are computed with 1,884 angular steps and 600 steps in the p variable covering the diagonal of the square. To simulate multiplicative noise, we choose the variance of $\boldsymbol{w}$ in (7.2) to be 0.2. To simulate CT noise, we use the non-linear model (7.5) (rather than the linearization (7.6)) with $\text{VAR}(g_{\text{noise}}) = 0.03$. In the Poisson noise case, each value of $\mathcal{R}f$ is randomized as follows: $\texttt{poissrnd}(80 * \mathcal{R}f)/80$; it is worth noticing that $\mathcal{R}f$ ranges from 0 to 0.51 in the Shepp-Logan case and to 0.56 in the disks case. We chose the noise parameters so that the noise would be of similar strength, visibly, in all three cases, and the distribution is Gaussian. Hann filter is applied to the inversion.

Note that the noise has different character compared to Figure 10(h), for example, and one can see individual lines (more precisely, line segments) in it. The multiplicative and the Poisson noise characters are somewhat similar; while the CT noise in the middle looks different. Our analysis shows that in the latter case, the standard deviation of the added noise in the linearization regime has range from $e^0 = 1$ to about $e^{0.5} \approx 1.65$ times $\text{STD}(g_{\text{noise}})$, see (7.6), while in the other two cases, the range is from 0 to a certain positive constant, which allows for almost zero noise locally before inversion. For this reason, individual lines are harder to distinguish in the CT case.

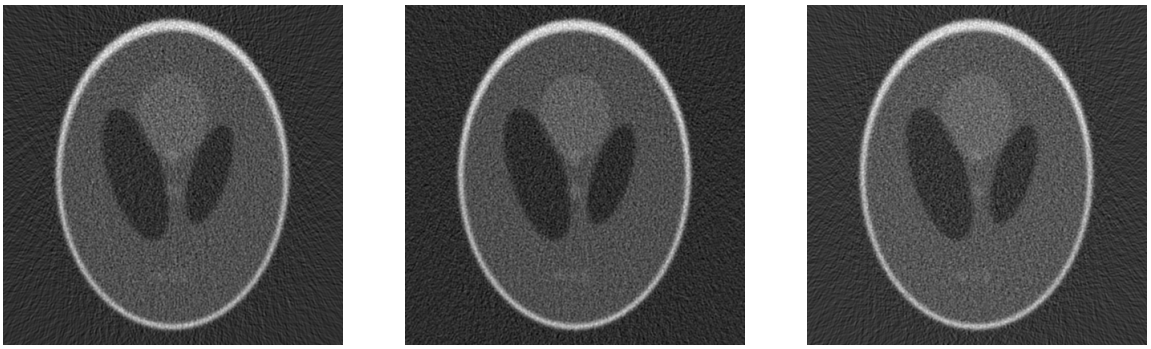


Figure 13. Shepp-Logan with the Hann filter and with (a) multiplicative noise; (b) CT type of noise; (c) Poisson noise

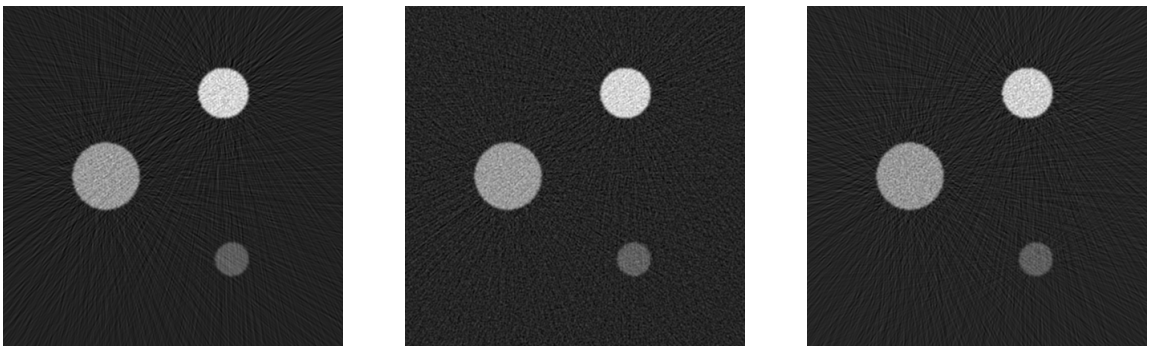


Figure 14. Three disks with the Hann filter and with (a) multiplicative noise; (b) CT type of noise; (c) Poisson noise

8. Discrete noise and its power spectrum

In this section, we analyze discrete white noise directly, without converting it to a continuous function. Here, $f(k)$ is a random vector on an $N \times \cdots \times N$ grid which we denoted by $\boldsymbol{f}_k$ before. We will denote by $\delta(k)$ the discrete delta function on $\mathbf{Z}^n$. In section 8.1, we follow mainly [15, Chapter 12], where f is a random variable depending on a (continuous) variable t; but most of it adapts to the discrete setting easily. We do a temporal analysis of the power spectrum for each fixed (discrete) frequency, with N fixed. We show that the spectrum of white noise is flat in the sense of expected value over repeated experiments, and we consider more general noise. On the other hand, for each experiment, the spectrum is quite, well, noisy and does not appear to smoothen as $N \to \infty$ numerically. In the second part, we study the ergodic properties of the power spectrum, with a single experiment, as $N \to \infty$. We show in Theorem 8.1 that the power spectrum is flat *on average*. That theorem is an analog of Theorem 4.1.

We want to emphasize that $f = \{f(k)\}$ depends on N, so we have a "triangular" array of random variables depending on the random outcome and increasing their size with N.

8.1. **Temporal analysis.** The discrete analog of the Fourier transform is the Discrete Fourier Transform (DFT) described below. It lives naturally on the discrete torus $\mathbf{T}_N^n = \mathbf{Z}^n / N\mathbf{Z}^n$ with period N. This shows that any time the DFT is used for spectral analysis, the original f is actually regarded as the restriction of a periodic function on a fundamental domain. We consider

$f : \mathbf{T}_N^n \to \mathbf{C}$, and we denote $f = \{f(k)\}$; with each element $f(k)$, $k = (k_1, \dots, k_n) \in \mathbf{T}_N^n$ a random variable in the same probability space. First, N will be fixed but eventually, we will take $N \gg 1$. We denote by fg the vector defined by $(fg)(k) = f(k)g(k)$, i.e., this is the multiplication of the functions of a discrete argument. Similarly, $|f|$ is the vector with components $|f(k)|$, while $\|f\|$ is the norm of f.

We define the (unitary) Discrete Fourier Transform (DFT) $\hat{f} = \mathcal{F}f$ by

$$\hat{f}(k^*) = \frac{1}{N^{n/2}} \sum_{k \in \mathbf{T}_N^n} f(k) e^{-2\pi \mathrm{i} k \cdot k^*/N}, \quad k^* \in \mathbf{T}_N^n.$$

Its inverse is the adjoint one

$$f(k) = \frac{1}{N^{n/2}} \sum_{k^* \in \mathbf{T}_N^n} \hat{f}(k^*) e^{2\pi \mathrm{i} k \cdot k^*/N}, \quad k \in \mathbf{T}_N^n.$$

Parseval's equality takes the form

$$f \cdot g = \hat{f} \cdot \hat{g}$$

for complex-valued f and g, where the dot-product is the natural one in $\mathbf{C}^N$. In particular, $\mathcal{F}$ is unitary. There is a natural (circular) convolution $f * g$ defined, and we have

$$\mathcal{F}(f * g) = N^{n/2} \hat{f}\hat{g}, \quad \mathcal{F}(fg) = N^{-n/2} \hat{f} * \hat{g}.$$

Next, we have

$$\mathcal{F}\delta = N^{-n/2}, \quad \mathcal{F}(1) = N^{n/2}\delta. \tag{8.1}$$

For each f with random entries, as above, define the *auto-correlation*

$$\mathrm{ACor}_f(m, k) = \mathbb{E}\{f(m)\bar{f}(k)\}. \tag{8.2}$$

The *auto-covariance* is defined as the auto-correlation of the centered f, i.e., of $f - \mathbb{E}(f)$, and it is easy to see that

$$\mathrm{ACov}_f(m, k) = \mathrm{ACor}(m, k) - \mathbb{E}\{f(m)\}\overline{\mathbb{E}\{f(k)\}}.$$

The process f is called *stationary*,[1] if $\mathrm{ACor}_f(m, k)$ is a function of $m - k$ only:

$$\mathrm{ACor}_f(m, k) = \mathrm{ACor}_f(m - k), \tag{8.3}$$

where, with some abuse of notation, we used the same notation ACor on the right. A process f is called *white noise* if

$$\mathrm{ACor}_f(m, k) = 0 \quad \text{for } m \neq k. \tag{8.4}$$

Then we must have

$$\mathrm{ACor}_f(m, k) = \sigma^2(m)\delta(m - k) \tag{8.5}$$

with $\sigma^2(m) = \mathrm{VAR}(f(m)) \geq 0$. We always assume that white noise has a zero mean. The process is wide-sense stationary (WSS) if it is stationary and its mean is constant. Then white noise is WSS if σ is constant. Note that WSS does not mean that $f(m)$ are independent from each other but if they are independent, they are uncorrelated, i.e., (8.4) holds.

Let $\Gamma(m^*, k^*)$ be the DFT of the auto-correlation of f, see (8.2), with respect to (m, k):

$$\Gamma(m^*, k^*) = \mathcal{F}(\mathrm{ACor}_f)(m^*, k^*).$$

[1] this terms comes from 1D processes, where x is the time

Then

$$\mathbb{E}\big\{\hat{f}(m^*)\bar{\hat{f}}(k^*)\big\} = \mathbb{E}\frac{1}{N^n}\sum_{m,k} f(m)\bar{f}(k)e^{-2\pi \mathrm{i}(m\cdot m^* - k\cdot k^*)/N} = \Gamma(m^*, -k^*).$$

In case of white noise satisfying (8.5), we have $\Gamma(m^*, k^*) = N^{-n/2}\widehat{\sigma^2}(m^* + k^*)$, thus we recover Theorem 11.2 in [15]:

$$\mathbb{E}\big\{\hat{f}(m^* + k^*)\bar{\hat{f}}(m^*)\big\} = N^{-n/2}\widehat{\sigma^2}(k^*), \quad \forall m^*.$$

This shows that even when f is not stationary, $\hat{f}$ is stationary with auto-correlation $\widehat{\sigma^2}$. If $\sigma =$ const., then each $f(m)$ has standard deviation σ, and $N^{-n/2}\widehat{\sigma^2} = \sigma^2\delta$, i.e.,

$$\mathbb{E}\big\{\hat{f}(k^*)\bar{\hat{f}}(m^*)\big\} = \sigma^2\delta(k^* - m^*). \tag{8.6}$$

In particular, $\mathbb{E}\{|\hat{f}(m^*)|^2\} = \sigma^2$ for all m, which shows a flat (expectation of a) spectrum. By Theorem 11.3 in [15], if f is real and Gaussian, then the covariance of $|\hat{f}(m^*)|^2$ and $|\hat{f}(k^*)|^2$ equals $N^{-n}(\widehat{\sigma^2})^2(m^* + k^*) + N^{-n}(\widehat{\sigma^2})^2(m^* - k^*)$, as we also show below. In particular, if $\sigma =$ const. in (8.5), we get covariance $\sigma^4(\delta(m^* + k^*) + \delta(m^* - k^*))$. Therefore, they are correlated when $k^* = m^*$ and $k^* = -m^*$ (because $\hat{f}$ is even) with standard deviation σ^4 for each Fourier coefficient except for the zeroth one when it is $2\sigma^4$. In fact, we do not need f to be Gaussian to have the same conclusion on asymptotic sense. We assume f real from now on.

Proposition 8.1. *Let f be real valued white noise with a finite fourth moment called μ_4. Then*

$$\mathrm{ACov}\left\{|\hat{f}(k^*)|^2, |\hat{f}(m^*)|^2\right\} = \sigma^4\delta(k^* - m^*) + \sigma^4\delta(k^* + m^*) + \frac{\mu_4 - 3\sigma^4}{N^n}. \tag{8.7}$$

Proof. We have

$$\begin{aligned}\mathrm{ACov}&\left\{|\hat{f}(k^*)|^2, |\hat{f}(m^*)|^2\right\} \\ &= \frac{1}{N^{2n}}\sum_{m_1,m_2,k_1,k_2} \mathbb{E}\{f(k_1)f(k_2)f(m_1)f(m_2)\}e^{-2\pi \mathrm{i}((k_1-k_2)\cdot k^* + (m_1-m_2)\cdot m^*)/N} - \sigma^4.\end{aligned}$$

The only non-zero expectation terms are those with two (equal or not) pairs of equal indices. Assume first that $k_1 = k_2$, $m_1 = m_2$. Then we have two cases for the expectation term above:

(1) If $m_1 \neq m_2$, the expectation term equals σ^4.
(2) Whenever $k_1 = k_2 = m_1 = m_2$, this expectation term is the fourth moment μ_4.

The latter number of terms is N^n, while the former is $N^{2n} - N^n$. Therefore, this set of indices contributes

$$\left(1 - \frac{1}{N^n}\right)\sigma^4 + \frac{1}{N^n}\mu_4$$

to the sum.

Consider the terms with $k_1 = m_1$, $k_2 = m_2$ but with $k_1 \neq k_2$ to exclude the previous case. Then the corresponding sum is

$$\begin{aligned}\frac{\sigma^4}{N^{2n}}\sum_{k_1=m_1\neq k_2=m_2} e^{-2\pi \mathrm{i}(k_1-k_2)\cdot(k^*+m^*)/N} &= \frac{\sigma^4}{N^{2n}}\sum_{k,m} e^{-2\pi \mathrm{i}k\cdot(k^*+m^*)/N} - \frac{\sigma^4}{N^n} \\ &= \sigma^4\delta(k^* + m^*) - \sigma^4/N^n.\end{aligned}$$

We performed the change $k = k_1 - k_2$, $m = k_2$ above, used (8.1) and compensated for the added terms corresponding to $k = 0$ in the second sum which are missing from the first one.

Finally, when $k_1 = m_2$, $k_2 = m_1$ but $k_1 \neq m_1$, the dot product in the phase function becomes $(k_1 - m_1) \cdot (k^* - m^*)$ and the same argument gives us

$$\frac{\sigma^4}{N^{2n}} \sum_{k_1 \neq m_1} e^{-2\pi \mathrm{i}(k_1 - m_1)\cdot(k^* - m^*)/N} = \frac{\sigma^4}{N^{2n}} \sum_{k \neq 0} \sum_{m} e^{-2\pi \mathrm{i} k \cdot (k^* - m^*)/N}$$
$$= \sigma^4 \delta(k^* - m^*) - \sigma^4 / N^n.$$

The analysis of those three cases completes the proof. □

Corollary 8.1. *If f in Proposition 8.1 is normal, then the last term in (8.7) vanishes.*

The proof follows from the well know fact that $\mu_4 = 3\sigma^4$ for normal distributions.

Remark 8.1. The results in Proposition 8.1 can be interpreted as follows. Up to an error $O(N^{-n})$, we get auto-covariance σ^4 if $k^* = m^* \neq 0$ and when $k^* = -m^* \neq 0$ (symmetry, because f is real), and $2\sigma^4$ if $k^* = m^* = 0$. If we stay in a fundamental domain of the type $k_j \in \{0, 1, \ldots, N-1\}$ then the symmetry becomes $|\hat{f}(k^*)|^2 = |\hat{f}(N - k_1^*, \ldots N - k_N)|^2$.

8.2. **Ergodic analysis. Flatness of the power spectrum on average.** Let α be a locally Riemann integrable function on $\mathbf{R}^n$, periodic of period 1 in each variable. Assume that f is real valued. We are interested in the following linear functional

$$\mu_N(\alpha) := \frac{1}{N^n} \sum_{k^* \in \mathbf{T}_N^n} \alpha(k^*/N) |\hat{f}(k^*)|^2.$$

This is a discrete analog of (4.14) with p there depending on the dual variable only. It is a weighted (not normalized) average of the power spectrum. What we do there is effectively rescale the spectrum from the integer points in $[0, N-1]^n$ (and then extended by periodicity) to the ones with fractional components of the kind k^*/N, forming a dense set in $[0,1]^n$ asymptotically. In statistics, this is done routinely in the study of peridograms, and k^*/N is replaced by a continuous variable.

Assume a white noise process (8.5) with $\sigma = \text{const}$. Then $\mathbf{E}(\mu_N(\alpha)) \to \sigma^2 \int \alpha(\xi)\, \mathrm{d}\xi$, as $N \to \infty$ by (8.6), where the integration is taken over the continuous torus in $\mathbf{R}^n$ with period one.

The random variables $|\hat{f}(k^*)|^2 - \sigma^2$ have zero expectation, correlation given by Proposition 8.1, and variance σ^4. Write

$$\mu_N(\alpha) = \sigma^2 \frac{1}{N^n} \sum_{k^* \in \mathbf{T}_N^n} \alpha(k^*/N) + \frac{1}{N^n} \sum_{k^* \in \mathbf{T}_N^n} \alpha(k^*/N) \left(|\hat{f}(k^*)|^2 - \sigma^2 \right).$$

The first term is a Riemannian sum. The second term has zero expectation and variance

$$\frac{\sigma^4}{N^{2n}} \sum_{k^* \in \mathbf{T}_N^n} \left(|\alpha(k^*/N)|^2 + \alpha(k^*/N) \bar{\alpha}(-k^*/N) \right) + O\Big(\frac{1}{N^n}\Big) \leq \frac{C}{N^n}.$$

The error terms come from the cross terms which are products of σ^4 and $O(N^{-n})$, by Proposition 8.1. There are N^{2n} of them. Therefore, we proved the following.

Theorem 8.1. *Let $f(k)$ be a white noise process on $\mathbf{T}_N^n$ (depending on N), with variance σ^2 and a finite fourth momentum. Then for every Riemann integrable function α on T_N^n we have*

$$\mu_N(\alpha) \longrightarrow \sigma^2 \int \alpha(\xi)\, \mathrm{d}\xi \quad \textit{in mean square sense,} \tag{8.8}$$

where the integral is taken over the torus in $\mathbf{R}^n$ with period one.

Therefore, the measure $N^{-n}\sum_{\mathbf{T}_N^n}|\hat{f}(m^*)|^2\delta(\xi-m^*/N)$ converges weakly to $\sigma^2\,\mathrm{d}\xi$ in mean square sense. In particular, if we take α to be the characteristic function of, say a box U in $\mathbf{T}_N^n$, then the average of the power spectrum on U tends to σ^2 in mean square sense.

8.3. **More general noise.** We assume now that the random variables $f_h(k)$ depend on h, have zero mean and have uniformly bounded fourth momenta but are not necessarily independent or equally distributed. If we assume (4.35), then the power spectrum is expressed in Theorem 4.2. One special but important case is when the auto-correlation is space independent (stationary, see (8.3)), then β in (4.35) is independent of x and we have

$$\mathrm{ACor}(f_h(k), f_h(k+m)) = \beta(m) \tag{8.9}$$

with some $\beta(m)$. Then (8.8) takes the form

$$\mu_N(\alpha) \longrightarrow \int \check{\beta}(\xi)\alpha(\xi)\,\mathrm{d}\xi \quad \text{in mean square sense,}$$

where $\check{\beta}$ is as in (4.36). In other words, the limit measure is $\check{\beta}(\xi)\,\mathrm{d}\xi$.

8.4. **Numerical examples.** We illustrate the temporal behavior of the spectrum first. In Figure 15, we take a random normally distributed vector f with $N = 200$ and variance $\sigma^2 = 1$. The power spectrum is plotted next to it. As we can see, it looks flat on average with mean value close to one but the variation is substantial. On the right we plot the histogram of the (spatial) standard deviation $\mathrm{STD}(|\hat{f}|^2)$ over $1,000$ experiments; it appears to have mean 1. We recall that $\mathrm{STD}(|\hat{f}|^2)$ is the square root of

$$\mathrm{VAR}\left(|\hat{f}|^2\right) = \frac{1}{N^n}\sum_{k^*\in\mathbf{T}_N^n}\left(|\hat{f}(k^*)|^2-\sigma^2\right)^2.$$

We have not proved a limit for it though. That would require estimating the auto-correlation of the summands above similarly to Proposition 8.1.

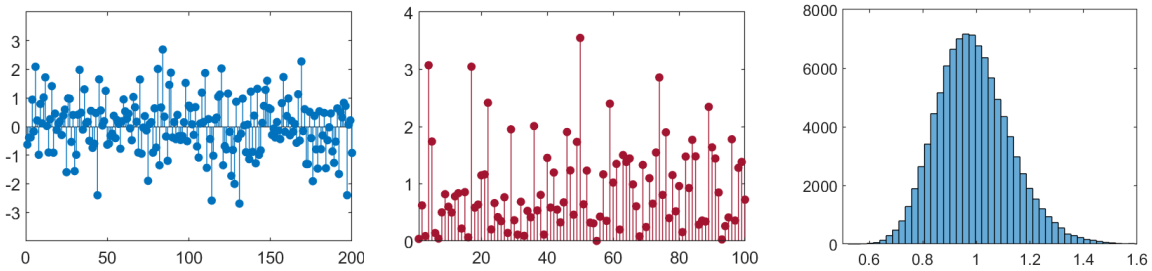


Figure 15. Left: a random normally distributed vector, $N = 200$, $\sigma^2 = 1$. Center: plot of $|\hat{f}|^2$ for indices from 0 to 100 ($|\hat{f}|^2$ is an even function with period 200). Right: the histogram of $\mathrm{STD}(|\hat{f}|^2)$ over $100,000$ experiments; it appears centered around 1.

Next, we illustrate the spatial (ergodic) behavior of the power spectrum. The averaged power spectrum for a normally distributed vector is shown in Figure 16. We divide the interval $[0, N/2]$ into 25 subintervals and average in each one of them. We take $N = 10^2, 10^3, 10^4$ and $N = 10^5$. As we can see, the averaged spectrum gets flatter and flatter. This illustrates Theorem 8.1.

If we keep N fixed but average over many experiments, the spectrum gets flatter as well numerically, as (8.6) suggests.

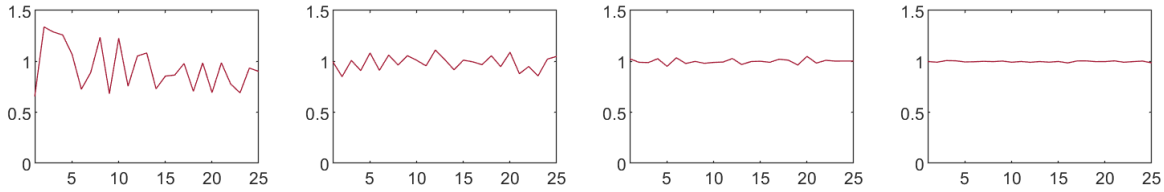


Figure 16. Plot of the averaged $|\hat{f}|^2$ for $N = 10^2, 10^3, 10^4$ and $N = 10^5$.

References

[1] Y. C. De Verdière. Semiclassical analysis and passive imaging. *Nonlinearity*, 22(6):R45, 2009.

[2] Y. C. De Verdière. A semi-classical calculus of correlations. *Comptes Rendus Geoscience*, 343(8-9):496–501, 2011.

[3] 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.

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

[5] C. Fefferman, S. Ivanov, M. Lassas, and H. Narayanan. Reconstruction of a Riemannian manifold from noisy intrinsic distances. *SIAM J. Math. Data Sci.*, 2(3):770–808, 2020.

[6] C. Gérard. Asymptotique des pôles de la matrice de scattering pour deux obstacles strictement convexes. *Mém. Soc. Math. France (N.S.)*, (31):1–146, 1988.

[7] T. Helin, M. Lassas, and L. Oksanen. Inverse problem for the wave equation with a white noise source. *Comm. Math. Phys.*, 332(3):933–953, 2014.

[8] T. Helin, M. Lassas, L. Oksanen, and T. Saksala. Correlation based passive imaging with a white noise source. *J. Math. Pures Appl. (9)*, 116:132–160, 2018.

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

[10] T.-C. Hu and R. L. Taylor. On the strong law for arrays and for the bootstrap mean and variance. *Internat. J. Math. Math. Sci.*, 20(2):375–382, 1997.

[11] R. Keys. Cubic convolution interpolation for digital image processing. *IEEE transactions on acoustics, speech, and signal processing*, 29(6):1153–1160, 1981.

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

[13] E. Meijering and M. Unser. A note on cubic convolution interpolation. *IEEE Transactions on Image processing*, 12(4):477–479, 2003.

[14] F. Natterer. *The mathematics of computerized tomography*. B. G. Teubner, Stuttgart, 1986.

[15] A. Papoulis and S. U. Pillai. *Probability, random variables, and stochastic processes*. Tata McGraw-Hill Education, 4th edition, 2002.

[16] M. Reed and B. Simon. *Methods of modern mathematical physics. III*. Academic Press [Harcourt Brace Jovanovich, Publishers], New York-London, 1979. Scattering theory.

[17] P. Stefanov. Semiclassical Sampling and Discretization of Certain Linear Inverse Problems. *SIAM J. Math. Anal.*, 52(6):5554–5597, 2020.

[18] A. Van der Schaaf and J. van Hateren. Modelling the power spectra of natural images: statistics and information. *Vision research*, 36(17):2759–2770, 1996.

[19] 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

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