RIMS Kôkyûroku Bessatsu
B56 (2016), 107–186

Born–Jordan time-frequency analysis

Acknowledgement: *I want to thank ESI (Erwin Schrödinger Institute) and NuHAG (Numerical Harmonic Analysis Group) for the relaxed and inspiring atmosphere during my 2012 and 2014 visits to Vienna. Moreover, my 2014 pleasant visits to the RIMS workshop in Kyoto, and University of Hokkaido in Sapporo influenced shaping this exposition. Also, thanks to University of Torino for the hospitality in 2014, during writing these notes. The research was partially funded by the grants from the Finnish Academy of Science and Letters (Vilho, Yrjö and Kalle Väisälä Foundation) and Magnus Ehrnrooth Foundation.*

By

Ville Turunen*

Contents

Received August 1, 2015. Revised December 11, 2015.
2010 Mathematics Subject Classification(s): 42A99, 42B10, 47A07, 47G30, 44A15, 81S30, 94A12
Key Words: *time-frequency distributions, pseudo-differential operators, Born–Jordan transform.*
Supported by JAPAN SUPPORT
*Aalto University, Department of Mathematics and Systems Analysis, P.O. Box 11100, FI-00076 Aalto, Finland.
e-mail: ville.turunen@aalto.fi

Abstract

Born–Jordan quantization originates from the early quantum mechanics, leading to sharp time-frequency localization of signals. The related Born–Jordan transform provides an attractive alternative to short-time Fourier transform. We review the essential time-frequency analysis, characterizing the Born–Jordan transform within Cohen's class, and show how all this works in audio signal processing. Computationally, our Born–Jordan approach is as complex as using spectrograms (which suffer from arbitrariness of chosen analysis window, resulting in inferior localization). We relate this to singular integral operators, and compare the Weyl quantization to the Born–Jordan case.

§ 1. Introduction

In this treatise, we present sharp stable operations on "signals" of finite energy. This is achieved by so-called Born–Jordan transform in time-frequency analysis. Time-frequency analysis is an essential part Fourier analysis, having applications in e.g. audio signal processing (acoustics, phonetics, sound synthesis), visualization and diagnostics of medical data (ECG, EEG), analysis of radar signals, seismology, quantum physics etc. A signal could be "digital" or "analog" (discrete time series or continuous time signal). We shall start by quickly reviewing the notation and important results for Fourier integral transform, and we hope that the text will be useful not only to mathematicians but also to a wide audience in engineering and applied sciences.

The origins of time-frequency analysis are in both quantum mechanics and signal processing, and these subjects are closely related to each other, as explained e.g. in [9].

In the sequel, we shall mostly appeal to the signal processing metaphors. Simplifying a bit, we may think that *Fourier analysis* answers to the question "how often" something happens in a signal. In this vein, *time-frequency analysis* is a subfield of Fourier analysis, where we try to answer simultaneously "when and how often" something happens in a signal.

For the sake of argument, suppose our signal $u : \mathbb{R} \to \mathbb{C}$ is a piece of music: this is a function of *time variable*, and from u it is easy to read "**when** something happens in u". The dual problem to "**when**" is "**how often** something happens in u": this can be solved by moving to the Fourier transform $\widehat{u} : \mathbb{R} \to \mathbb{C}$, which is a function of *frequency variable*. In other words, the Fourier transform $\widehat{u}$ may reveal rhythms and notes in music u. These separate descriptions u and $\widehat{u}$ in time and frequency are not enough for demanding operations on signals. Even though the Fourier transform $u \mapsto \widehat{u}$ is invertible, it is not easy to simultaneously see "**when and how often** something happens in u" — this is the fundamental problem in time-frequency analysis.

One of the first problems in time-frequency analysis is to decide which transform and the corresponding energy density to use: there are literally infinitely many candidates available. Leon Cohen's time-frequency distributions [7, 8, 9] provide infinitely many alternatives for "energy density" of signal u in the phase-space (time-frequency plane). Once some preferred density is chosen, we try to understand the corresponding quantization rule, in order to be able to manipulate our functions (or signals) in a desired way. In these notes, we shall show that the **Born–Jordan distribution** has a simple well-motivated characterization within Cohen's class, and we shall study its properties. Moreover, we discretize and periodize the related transforms, applying them to real-life signal processing.

§ 2. Fourier analysis background

Now we review the main properties of Fourier integral transform, where the real line $\mathbb{R} =]-\infty, \infty[$ is the model for time. We wish to keep the presentation informal for the benefit of a wide audience in science and engineering: the reader may find the details of the relevant mathematical analysis in [24, 23, 22]. We use rather straightforward discretizations to compute the pictures in this article. In practical applications, time might be measured in seconds $[s]$, and frequency correspondingly in Hertz $[Hz = 1/s]$ (occurences per second). To make the theory as transparent as possible, *time-like* variables will be noted by Latin letters $(x, y, \cdots \in \mathbb{R})$, and *frequency-like* variables by corresponding Greek letters $(\xi, \eta, \cdots \in \widehat{\mathbb{R}})$. Even though mathematically $\widehat{\mathbb{R}}$ is the same as $\mathbb{R}$, we want to use this distinct time-frequency notation to stress the physical difference of time $x \in \mathbb{R}$ and frequency $\eta \in \widehat{\mathbb{R}}$ ("**When** something happens?" versus

"**How often** something happens?"). In our notation, the Cartesian product

$$\mathbb{R} \times \widehat{\mathbb{R}} = \left\{ (x,\eta) : \; x \in \mathbb{R}, \; \eta \in \widehat{\mathbb{R}} \right\}$$

is called the *time-frequency plane* (or the *phase space*).

An *analog signal* u is a "nice enough" function $u : \mathbb{R} \to \mathbb{C}$ (later, we may allow also the Schwartz tempered distributions). For example, at time (or position) $x \in \mathbb{R}$, $u(x) \in \mathbb{C}$ could be pressure/temperature/luminosity/position/wave function etc. In Fourier analysis, we describe signals u as "infinite linear combinations" of complex-valued waves

$$x \mapsto \mathrm{e}^{\mathrm{i}2\pi x\cdot\eta} = \cos(2\pi x \cdot \eta) + \mathrm{i}\sin(2\pi x \cdot \eta).$$

Such a wave has frequency $\eta \in \widehat{\mathbb{R}}$ (which can also be negative), and this wave has "weight" $\widehat{u}(\eta) \in \mathbb{C}$ (defined by the Fourier transform (2.3)) within signal u, in view of the Fourier inverse formula (2.7). Signal u has *energy density* or *power* $|u|^2 : \mathbb{R} \to [0,\infty]$ (physical unit e.g. *J/s*, *Joule per second*), meaning that

$$\int_{[a,b]} |u(x)|^2 \,\mathrm{d}x \quad \in \quad [0,\infty] \tag{2.1}$$

is the energy during the time interval $[a,b] \subset \mathbb{R}$ (physical unit e.g. *J*, *Joule*). The *energy* of signal u is

$$\|u\|^2 := \int_{\mathbb{R}} |u(x)|^2 \,\mathrm{d}x. \tag{2.2}$$

Fourier (integral) transform of a "nice enough" analog signal $u : \mathbb{R} \to \mathbb{C}$ is another analog signal $\mathscr{F}u = \widehat{u} : \widehat{\mathbb{R}} \to \mathbb{C}$ of *frequency* variable $\eta \in \mathbb{R}$, where

$$\widehat{u}(\eta) := \int_{\mathbb{R}} \mathrm{e}^{-\mathrm{i}2\pi y\cdot\eta}\, u(y) \,\mathrm{d}y. \tag{2.3}$$

What did we mean by "nice enough" signals? At least *absolute integrability* $\|u\|_{L^1} := \int_{\mathbb{R}} |u(x)| \,\mathrm{d}x < \infty$ is "nice enough" here, because $|\widehat{u}(\eta)| \leq \int_{\mathbb{R}} |u(x)| \,\mathrm{d}x$. Write $u \in L^1(\mathbb{R})$ if $\|u\|_{L^1} < \infty$.

Example 2.1. Let $|c| = 1$ for $c \in \mathbb{C}$, and let $u : \mathbb{R} \to \mathbb{C}$, where

$$u(x) := \begin{cases} c\, 2\pi\varepsilon\, \mathrm{e}^{-2\pi\varepsilon\,(x-t_0)}\, \mathrm{e}^{\mathrm{i}2\pi(x-t_0)\cdot\alpha} & \text{when } x > t_0, \\ 0 & \text{when } x \leq t_0. \end{cases}$$

This is a complex-valued "vibration" at frequency $\alpha \in \widehat{\mathbb{R}}$, starting at time $t_0 \in \mathbb{R}$, decaying at rate $\varepsilon > 0$. Then

$$\widehat{u}(\eta) = \int_{\mathbb{R}} \mathrm{e}^{-\mathrm{i}2\pi x\cdot\eta}\, u(x) \,\mathrm{d}x = \int_{t_0}^{\infty} \cdots \,\mathrm{d}x = \ldots = c\, \frac{\mathrm{e}^{-\mathrm{i}2\pi t_0\cdot\eta}}{1 + \mathrm{i}(\eta-\alpha)/\varepsilon}.$$

The energy densities $|u|^2$ in time and $|\widehat{u}|^2$ in frequency are thereby

$$|u(x)|^2 = \begin{cases} (2\pi\varepsilon)^2\, \mathrm{e}^{-4\pi\varepsilon\,(x-t_0)} & \text{when } x > t_0, \\ 0 & \text{when } x \leq t_0, \end{cases}$$

$$|\widehat{u}(\eta)|^2 = \frac{1}{1+(\eta-\alpha)^2/\varepsilon^2} \qquad \text{for all } \eta \in \widehat{\mathbb{R}}.$$

Obviously, u cannot be retrieved back from $|u|^2$ and $|\widehat{u}|^2$, but yet $\widehat{u}$ contains essentially all the information about u: this enables useful time-invariant operations (i.e. convolutions) on signals. Also, the energy is conserved in the Fourier transform: $\|\widehat{u}\|^2 = \|u\|^2$.

Let us now deal with a class of particularly well-behaving signals: Schwartz test functions $u : \mathbb{R} \to \mathbb{C}$ are "*smooth and rapidly decaying*". More precisely:

Definition 2.2. *Schwartz test function space* $\mathscr{S}(\mathbb{R})$ consists of those infinitely smooth functions $u : \mathbb{R} \to \mathbb{C}$ for which

$$\lim_{|x|\to\infty} x^n\, u^{(m)}(x) = 0 \tag{2.4}$$

for all $m, n \in \mathbb{N} = \{0, 1, 2, 3, 4, 5, \cdots\}$.

There are many test signals:

Example 2.3. If $u \in C^\infty(\mathbb{R})$ and $u(x) = 0$ whenever $|x| \geq 1$ then $u \in \mathscr{S}(\mathbb{R})$; e.g. define $u(x) := \exp(1/(x^2-1))$ for $|x| < 1$. Also Gaussian signals $x \mapsto \mathrm{e}^{ax^2+bx+c}$ are examples of Schwartz test functions (when $\mathrm{Re}(a) < 0$, $a, b, c \in \mathbb{C}$). Clearly, $\mathscr{S}(\mathbb{R}) \subset L^1(\mathbb{R}) \neq \mathscr{S}(\mathbb{R})$.

Example 2.4. Let $k \in \mathbb{N}$, $\lambda \in \mathbb{C}$, $u, v \in \mathscr{S}(\mathbb{R})$, and let $q : \mathbb{R} \to \mathbb{C}$ be a polynomial. Then $\lambda u, u+v, u^{(k)}, qu, uv \in \mathscr{S}(\mathbb{R})$.

The Fourier transform treats polynomial multiplication and differentiation in a symmetric fashion: If $u \in \mathscr{S}(\mathbb{R})$ then $\widehat{u} \in \mathscr{S}(\widehat{\mathbb{R}})$, because

$$\widehat{u}\,'(\eta) = -\mathrm{i}2\pi\, \widehat{v}(\eta), \qquad \widehat{u'}\,(\eta) = +\mathrm{i}2\pi\, \eta\, \widehat{u}(\eta), \tag{2.5}$$

where $v(x) := x\, u(x)$. These formulas motivate the definition of Schwartz test signals. Hence the Fourier transform gives a linear mapping

$$(u \mapsto \widehat{u}) : \mathscr{S}(\mathbb{R}) \to \mathscr{S}(\widehat{\mathbb{R}}). \tag{2.6}$$

Example 2.5. Let $u_\varepsilon(x) = \mathrm{e}^{-\varepsilon\pi x^2}$ (Gaussian), where $\varepsilon > 0$. First,

$$u_\varepsilon'(x) = -2\varepsilon\pi\, x\, u_\varepsilon(x) \overset{(2.5)}{\Longrightarrow} +\mathrm{i}2\pi\,\eta\,\widehat{u_\varepsilon}(\eta) = (\varepsilon/\mathrm{i})\,\widehat{u_\varepsilon}\,'(\eta) \iff \widehat{u_\varepsilon}(\eta) = \widehat{u_\varepsilon}(0)\,\mathrm{e}^{-\pi\eta^2/\varepsilon},$$

and integrating in the plane in polar coordinates, here

$$\widehat{u_\varepsilon}(0) = \int_{\mathbb{R}} u_\varepsilon(x)\,\mathrm{d}x = \left[\int_{\mathbb{R}}\int_{\mathbb{R}} u_\varepsilon(x)\,u_\varepsilon(y)\,\mathrm{d}x\,\mathrm{d}y\right]^{1/2} = \left[\int_0^\infty\int_0^{2\pi} \mathrm{e}^{-\varepsilon r^2}\, r\mathrm{d}\theta\,\mathrm{d}r\right]^{1/2} = \frac{1}{\sqrt{\varepsilon}}.$$

(Especially, $\widehat{u_\varepsilon} = u_\varepsilon$, when $u(x) = \mathrm{e}^{-\pi x^2}$.) Applying this to for any $u \in \mathscr{S}(\mathbb{R})$, we find

$$u(x) = \lim_{0<\varepsilon\to 0}\int_{\mathbb{R}} u(y)\,\frac{1}{\sqrt{\varepsilon}}\,\mathrm{e}^{-\pi(y-x)^2/\varepsilon}\,\mathrm{d}y = \lim_{0<\varepsilon\to 0}\int_{\widehat{\mathbb{R}}} \mathrm{e}^{-\varepsilon\pi\eta^2}\mathrm{e}^{+\mathrm{i}2\pi x\cdot\eta}\int_{\mathbb{R}} u(y)\,\mathrm{e}^{-\mathrm{i}2\pi y\cdot\eta}\,\mathrm{d}y\,\mathrm{d}\eta,$$

which proves the *Fourier inverse formula*

$$u(x) = \int_{\widehat{\mathbb{R}}} \mathrm{e}^{+\mathrm{i}2\pi x\cdot\eta}\,\widehat{u}(\eta)\,\mathrm{d}\eta. \tag{2.7}$$

Thus the Fourier transform $\mathscr{F} = (u \mapsto \widehat{u}) : \mathscr{S}(\mathbb{R}) \to \mathscr{S}(\widehat{\mathbb{R}})$ is bijective.

Definition 2.6. *Inner product* between signals $u, v \in \mathscr{S}(\mathbb{R})$ is

$$\langle u, v\rangle := \int_{\mathbb{R}} u(x)\,\overline{v(x)}\,\mathrm{d}x \in \mathbb{C} \tag{2.8}$$

In the sequel, for clarity, we denote the complex conjugation by $\lambda^* := \overline{\lambda} \in \mathbb{C}$.

The inner product is preserved by the Fourier transform: $\langle \widehat{u}, \widehat{v}\rangle = \langle u, v\rangle$, because

$$\langle \widehat{u}, \widehat{v}\rangle = \int_{\widehat{\mathbb{R}}} \widehat{u}(\eta)\left[\int_{\mathbb{R}} \mathrm{e}^{-\mathrm{i}2\pi x\cdot\eta}\,v(x)\,\mathrm{d}x\right]^*\mathrm{d}\eta = \int_{\mathbb{R}}\int_{\widehat{\mathbb{R}}} \mathrm{e}^{+\mathrm{i}2\pi x\cdot\eta}\,\widehat{u}(\eta)\,\mathrm{d}\eta\,v(x)^*\,\mathrm{d}x = \langle u, v\rangle.$$

Especially, Fourier transform preserves *energy* $\|u\|^2 := \langle u, u\rangle$ of signal $u \in \mathscr{S}(\mathbb{R})$:

$$\|\widehat{u}\|^2 = \|u\|^2. \tag{2.9}$$

Definition 2.7. The *convolution* of absolutely integrable $u, v \in L^1(\mathbb{R})$ is the signal $u * v : \mathbb{R} \to \mathbb{C}$ such that

$$u * v(x) = (u * v)(x) := \int_{\mathbb{R}} u(x-y)\,v(y)\,\mathrm{d}y. \tag{2.10}$$

The reader may then verify the absolute integrability of $u * v$:

$$\|u * v\|_{L^1} \leq \|u\|_{L^1}\,\|v\|_{L^1} < \infty.$$

"Convolution in time" is "multiplication in frequency", that is

$$\widehat{u * v}(\eta) = \widehat{u}(\eta)\,\widehat{v}(\eta). \tag{2.11}$$

This is a useful property in signal processing. Moreover, for differentiation we have

$$(u * v)' = u' * v, \tag{2.12}$$

if also u' is absolutely integrable: hence convolution makes signal v smoother. Furthermore, $u * v \in \mathscr{S}(\mathbb{R})$ when $u, v \in \mathscr{S}(\mathbb{R})$.

Definition 2.8. *Translation* of $u \in \mathscr{S}(\mathbb{R})$ by time-lag $b \in \mathbb{R}$ is $T(b)u \in \mathscr{S}(\mathbb{R})$, where

$$T(b)u(x) := u(x-b). \tag{2.13}$$

Modulation of $u \in \mathscr{S}(\mathbb{R})$ by frequency-lag $\alpha \in \widehat{\mathbb{R}}$ is $M(\alpha)u \in \mathscr{S}(\mathbb{R})$, where

$$M(\alpha)u(x) := \mathrm{e}^{+\mathrm{i}2\pi x\cdot\alpha}u(x). \tag{2.14}$$

The Fourier transform intertwines between the translations and the modulations: $\widehat{M(\alpha)s} = T(\alpha)\widehat{u}$ and $\widehat{T(b)s} = M(-b)\widehat{u}$, that is

$$\begin{aligned}\widehat{M(\alpha)s}(\eta) &= T(\alpha)\widehat{u}(\eta),\\ \widehat{T(b)s}(\eta) &= M(-b)\widehat{u}(\eta).\end{aligned}$$

We want to transform an input signal $u \in \mathscr{S}(\mathbb{R})$ to the output signal $Au = A(u) \in \mathscr{S}(\mathbb{R})$. Suppose $A : \mathscr{S}(\mathbb{R}) \to \mathscr{S}(\mathbb{R})$ is *linear*, i.e.

$$\begin{aligned}A(u+v) &= A(u) + A(v) \quad \text{and}\\ A(\lambda u) &= \lambda\, A(u)\end{aligned}$$

for all signals $u, v \in \mathscr{S}(\mathbb{R})$ and constants $\lambda \in \mathbb{C}$. Linear transform A is formally presented as an *integral operator*:

$$Au(x) = \int_{\mathbb{R}} K_A(x,y)\, u(y)\,\mathrm{d}y, \tag{2.15}$$

where K_A is the *Schwartz distribution kernel* of A. Notice that integral operator $A : \mathscr{S}(\mathbb{R}) \to \mathscr{S}(\mathbb{R})$ has "essentially unique" distributional kernel K_A (provided that $u \mapsto Au$ is "naturally continuous" — precise statement in so-called *Schwartz kernels theorem*).

Let operator A be *time-invariant*, meaning $T(b)A = AT(b)$ for all $b \in \mathbb{R}$, i.e.

$$T(b)Au(x) = AT(b)u(x) \tag{2.16}$$

for all signals u and for all $x, b \in \mathbb{R}$; in other words, $A = T(-b)AT(b)$, which means

$$\begin{aligned}\int_{\mathbb{R}} K_A(x,y)\, u(y)\,\mathrm{d}y = Au(x) \;&=\; T(-b)AT(b)u(x) \;=\; AT(b)u(x+b)\\ &= \int_{\mathbb{R}} K_A(x+b,y)\, T(b)u(y)\,\mathrm{d}y\\ &= \int_{\mathbb{R}} K_A(x+b,y)\, u(y-b)\,\mathrm{d}y\\ &= \int_{\mathbb{R}} K_A(x+b,y+b)\, u(y)\,\mathrm{d}y.\end{aligned}$$

Thus $K_A(x,y) = K_A(x+b, y+b)$ for all $b, x, y \in \mathbb{R}$, especially $K_A(x,y) = K_A(x-y,0) = v(x-y)$ for some signal $v : \mathbb{R} \to \mathbb{C}$: hence, $Au = u * v$ is a convolution.

Test signals $u \in \mathscr{S}(\mathbb{R})$ are "tame"; we shall extend Fourier analysis to "wilder" signals. "Size" of signal $u : \mathbb{R} \to \mathbb{C}$ is measured by the Lebesgue norms

$$\|u\|_{L^p} \overset{1 \le p < \infty}{:=} \left[\int_{\mathbb{R}} |u(x)|^p \, \mathrm{d}x \right]^{1/p},$$

$$\|u\|_{L^\infty} := \operatorname{ess\,sup}_{x \in \mathbb{R}} |u(x)| \overset{\text{If } u \text{ continuous}}{=} \sup_{x \in \mathbb{R}} |u(x)| = \lim_{p \to \infty} \|u\|_{L^p}.$$

We denote $u \in L^p(\mathbb{R})$, if $\|u\|_{L^p} < \infty$, where spaces $L^p(\mathbb{R})$ are so-called *Lebesgue spaces*, see details in e.g. [24, 23]. Then there is the following terminology:

- $u \in L^1(\mathbb{R})$ is *absolutely integrable*: $\int_{\mathbb{R}} |u(x)| \, \mathrm{d}x = \|u\|_{L^1}$.
- $u \in L^2(\mathbb{R})$ has *finite energy*: $\|u\|^2 = \int_{\mathbb{R}} |u(x)|^2 \, \mathrm{d}x = [\|u\|_{L^2}]^2$.
- $u \in L^\infty(\mathbb{R})$ is *essentially bounded*: $|u(x)| \le \|u\|_{L^\infty}$ for almost all x.

Write $u = v$ if $\|u - v\|_{L^p} = 0$ for $u, v \in L^p(\mathbb{R})$ (which happens whenever $u(x) = v(x)$ for almost every $x \in \mathbb{R}$). Here $\mathscr{S}(\mathbb{R}) \subset L^p(\mathbb{R})$ for all $p \in [1, \infty]$. Functions $u \in L^p(\mathbb{R})$ certainly can be discontinuous. Nevertheless, if $u \in L^1(\mathbb{R})$ and

$$v(x) := \int_0^x u(y) \, \mathrm{d}y$$

then $v \in L^\infty(\mathbb{R})$ (satisfying $\|v\|_{L^\infty} \le \|u\|_{L^1}$ clearly), and $v' = u$ in sense that $v'(x) = u(x)$ for almost all $x \in \mathbb{R}$ (this is so-called *Lebesgue differentiation theorem*). If $1 < p < \infty$ and $u \in L^p(\mathbb{R})$ then $u = u_1 + u_\infty$, where $u_1 \in L^1(\mathbb{R})$ and $u_\infty \in L^\infty(\mathbb{R})$. Why? Simply define

$$u_\infty(x) := \begin{cases} u(x) & \text{when } |u(x)| \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

Thus, if we want to find Fourier transform $\widehat{u}$ for signal $u \in L^p(\mathbb{R})$, we need to understand the special cases $p = 1$ and $p = \infty$: For $p = 1$, we already have the nice Fourier integrals; Case $p = \infty$ leads naturally to so-called *distributions* (which are a generalization of ordinary functions). Let us try to approximate $u \in L^p(\mathbb{R})$ by test functions $u_k \in \mathscr{S}(\mathbb{R})$. If $g, v \in \mathscr{S}(\mathbb{R})$, then $g * (uv) \in \mathscr{S}(\mathbb{R})$: this is smoothing by convolution. For $k \in \mathbb{Z}^+$, define $g_k, v_k \in \mathscr{S}(\mathbb{R})$ by

$$v_k(\eta) = \widehat{g_k}(\eta) := \mathrm{e}^{-\pi(\eta/k)^2},$$

so that $u_k := g_k * (v_k \, u) \in \mathscr{S}(\mathbb{R})$. Now, if $1 \le p < \infty$ then

$$\lim_{k \to \infty} \|u - u_k\|_{L^p} = 0, \quad \text{in other words} \quad u = \lim_{k \to \infty} u_k \quad \text{in} \quad L^p(\mathbb{R}).$$

This means that $\mathscr{S}(\mathbb{R})$ is *dense* in $L^p(\mathbb{R})$, when $1 \leq p < \infty$. But $\mathscr{S}(\mathbb{R})$ is not dense in $L^\infty(\mathbb{R})$: for instance, think of the constant function $\mathbf{1} \in L^\infty(\mathbb{R})$, for which $\|u-\mathbf{1}\|_{L^\infty} \geq 1$ for every $u \in \mathscr{S}(\mathbb{R})$. Thereby we cannot define Fourier transform for $u \in L^\infty(\mathbb{R})$ by a bounded linear extension of $(u \mapsto \widehat{u}) : \mathscr{S}(\mathbb{R}) \to \mathscr{S}(\widehat{\mathbb{R}})$. However, there is another method, which discuss soon.

We have the linear energy-preserving Fourier integral transform

$$(u \mapsto \widehat{u}) : \mathscr{S}(\mathbb{R}) \to \mathscr{S}(\widehat{\mathbb{R}}), \quad \|\widehat{u}\|^2 = \|u\|^2.$$

If $u \in L^2(\mathbb{R})$, by density of $\mathscr{S}(\mathbb{R}) \subset L^2(\mathbb{R})$, take $u_k \in \mathscr{S}(\mathbb{R})$ so that

$$\lim_{k\to\infty} \|u - u_k\| = 0, \quad \text{i.e.} \quad u = \lim_{k\to\infty} u_k \quad \text{in} \quad L^2(\mathbb{R}).$$

No matter which approximations u_k we choose, the energy-preservation guarantees the uniqueness of the limit

$$\widehat{u} := \lim_{k\to\infty} \widehat{u_k} \quad \in \quad L^2(\widehat{\mathbb{R}}) \cong L^2(\mathbb{R}).$$

This defines the linear energy-preserving Fourier transform

$$(u \mapsto \widehat{u}) : L^2(\mathbb{R}) \to L^2(\widehat{\mathbb{R}}), \quad \|\widehat{u}\|^2 = \|u\|^2. \tag{2.17}$$

This is automatically a bijection, and also *unitary*, which means $\langle \widehat{u}, \widehat{v} \rangle = \langle u, v \rangle$ for all $u, v \in L^2(\mathbb{R})$, where the inner product is

$$\langle u, v \rangle = \langle u, v \rangle_{L^2(\mathbb{R})} := \int_{\mathbb{R}} u(x)\, v(x)^*\, \mathrm{d}x.$$

Integrals $\widehat{u}(\eta) := \int_{\mathbb{R}} \mathrm{e}^{-\mathrm{i}2\pi x\cdot\eta}\, u(x)\, \mathrm{d}x$ define the Fourier transform for $u \in L^1(\mathbb{R})$. However, such integrals do not converge absolutely if $u \notin L^1(\mathbb{R})$. For $u \in L^2(\mathbb{R})$ and $\psi \in \mathscr{S}(\mathbb{R})$ we have

$$\begin{aligned} \langle \widehat{\widehat{u}}, \widehat{\widehat{\psi}} \rangle = \langle \widehat{u}, \widehat{\psi} \rangle &= \langle u, \psi \rangle \\ = \int_{\mathbb{R}} u(x)\psi(x)^* \mathrm{d}x &= \int_{\mathbb{R}} u(x)\widehat{\widehat{\psi}}(-x)^* \mathrm{d}x = \int_{\mathbb{R}} u(-x)\widehat{\widehat{\psi}}(x)^* \mathrm{d}x : \end{aligned}$$

thus $\widehat{\widehat{u}}(x) = u(-x)$ for almost every $x \in \mathbb{R}$.

Example 2.9. Let $u(x) = 1$ for $|x| < 1/2$, and $u(x) = 0$ otherwise. Then $u \in L^1(\mathbb{R})$, and $\widehat{u} = \mathrm{sinc} \in L^\infty(\widehat{\mathbb{R}})$ is the *cardinal sine*, where

$$\mathrm{sinc}(\eta) := \begin{cases} \frac{\sin(\pi\eta)}{\pi\eta} & \text{for } \eta \neq 0, \\ 1 & \text{for } \eta = 0. \end{cases}$$

Now $\mathrm{sinc} \in L^2(\widehat{\mathbb{R}})$ but $\mathrm{sinc} \notin L^1(\widehat{\mathbb{R}})$. However, $\widehat{\mathrm{sinc}}(x) = \widehat{\widehat{u}}(x) = u(-x) = u(+x)$ for almost every $x \in \mathbb{R}$, so that $\widehat{\mathrm{sinc}} = u \in L^2(\mathbb{R})$.

For $u \in L^\infty(\mathbb{R})$, $m \in \mathbb{N}$, polynomial $r : \mathbb{R} \to \mathbb{C}$ and $\psi \in \mathscr{S}(\mathbb{R})$, let

$$\langle r\, u^{(m)}, \psi \rangle := (-1)^m \int_{\mathbb{R}} u(x)\, (\psi^*\, r)^{(m)}(x)\, \mathrm{d}x \tag{2.18}$$

(where the m^{th} derivative $u^{(m)}$ makes classically sense if $u \in \mathscr{S}(\mathbb{R})$ — formula (2.18) is just inspired from formal integration by parts). Here $u^{(m)}$ is called the m^{th} *distribution derivative* of $u \in L^\infty(\mathbb{R})$. If $r_1, \cdots, r_n : \mathbb{R} \to \mathbb{C}$ are polynomials and $u_1, \cdots, u_n \in L^\infty(\mathbb{R})$, then

$$u = \sum_{m=1}^{n} r_m\, u_m^{(m)} \tag{2.19}$$

is called a *Schwartz tempered distribution* $u \in \mathscr{S}'(\mathbb{R})$. The Fourier transform $\widehat{u} \in \mathscr{S}'(\widehat{\mathbb{R}}) \cong \mathscr{S}'(\mathbb{R})$ is then defined by

$$\langle \widehat{u}, \widehat{\psi} \rangle := \langle u, \psi \rangle = \sum_{m=1}^{n} \langle r_m\, u_m^{(m)}, \psi \rangle \tag{2.20}$$

(which is again classically justified if $u \in \mathscr{S}(\mathbb{R})$). Now we have obtained the bijective Fourier transform

$$(u \mapsto \widehat{u}) : \mathscr{S}'(\mathbb{R}) \to \mathscr{S}'(\widehat{\mathbb{R}}). \tag{2.21}$$

The space $\mathscr{S}'(\mathbb{R})$ of tempered distributions is rather large:

Example 2.10. $L^p(\mathbb{R}) \subset \mathscr{S}'(\mathbb{R})$ for every $p \in [1, \infty]$: remember that $L^p(\mathbb{R}) \subset L^1(\mathbb{R}) + L^\infty(\mathbb{R})$, and that L^1-functions are derivatives of L^∞-functions. If $u \in \mathscr{S}'(\mathbb{R})$ then the distribution derivatives $u^{(m)} \in \mathscr{S}'(\mathbb{R})$ for every $m \in \mathbb{N}$.

Example 2.11. Let $u(x) = e_\beta(x) := \mathrm{e}^{\mathrm{i}2\pi x \cdot \beta}$. Then $e_\beta \in L^\infty(\mathbb{R})$, and

$$\langle \widehat{e_\beta}, \widehat{\psi} \rangle := \langle e_\beta, \psi \rangle = \int_{\mathbb{R}} \mathrm{e}^{+\mathrm{i}2\pi x \cdot \beta}\, \psi(x)^*\, \mathrm{d}x = \widehat{\psi}(\beta)^* =: \int_{\widehat{\mathbb{R}}} \delta_\beta(\eta)\, \widehat{\psi}(\eta)^*\, \mathrm{d}\eta,$$

where $\delta_\beta := \widehat{e_\beta} \notin L^p(\widehat{\mathbb{R}})$ is the *Dirac delta distribution* at $\beta \in \widehat{\mathbb{R}}$.

Think $\delta_b \in \mathscr{S}'(\mathbb{R})$ as a unit mass (or a unit impulse) at $x = b$. Roughly, $\delta_b(x) = 0$ if $x \neq b$, but beware: δ_b is not a function, because if u is a function such that $u(x) = 0$ for almost every $x \in \mathbb{R}$ then $\int_{\mathbb{R}} u(x)\, \psi(x)^*\, \mathrm{d}x = 0$ for all $\psi \in \mathscr{S}(\mathbb{R})$. No function $u : \mathbb{R} \to \mathbb{C}$ satisfies $\int_{\mathbb{R}} u(x)\, \psi(x)^*\, \mathrm{d}x = \psi(b)^*$ for all $\psi \in \mathscr{S}(\mathbb{R})$.

Example 2.12. Dirac delta $\delta_b \notin L^p(\mathbb{R})$ for any $p \in [1, \infty]$. Yet here

$$\langle \widehat{\delta_b}, \widehat{\psi} \rangle := \langle \delta_b, \psi \rangle = \psi(b)^* = \int_{\widehat{\mathbb{R}}} \mathrm{e}^{-\mathrm{i}2\pi b \cdot \eta}\, \widehat{\psi}(\eta)^*\, \mathrm{d}\eta = \langle e_{-b}, \widehat{\psi} \rangle,$$

giving $\widehat{\delta_b} = e_{-b} \in L^\infty(\mathbb{R})$. An alternative, informal computation is

$$\widehat{\delta_b}(\eta) = \int_{\mathbb{R}} \delta_b(x)\, \mathrm{e}^{-\mathrm{i}2\pi x\cdot\eta}\, \mathrm{d}x = \mathrm{e}^{-\mathrm{i}2\pi b\cdot\eta} = e_{-b}(\eta).$$

Example 2.13. *Signum* function $\mathrm{sgn} \in L^\infty(\mathbb{R})$ is defined by $\mathrm{sgn}(x) := x/|x|$ for $x \neq 0$. Notice that the derivative

$$\mathrm{sgn}'(x) := \lim_{h\to 0} \frac{\mathrm{sgn}(x+h) - \mathrm{sgn}(x)}{h} \in \mathbb{R}$$

exists if and only if $x \neq 0$. For distribution derivative $\mathrm{sgn}' = \mathrm{sgn}^{(1)}$,

$$\begin{aligned} \langle \mathrm{sgn}', \psi\rangle := -\langle \mathrm{sgn}, \psi'\rangle &= -\int_{\mathbb{R}} \mathrm{sgn}(x)\, \psi'(x)^*\, \mathrm{d}x \\ = \int_{-\infty}^{0} \psi'(x)^*\, \mathrm{d}x - \int_{0}^{\infty} \psi'(x)^*\, \mathrm{d}x &= \psi(0)^* + \psi(0)^*. \end{aligned}$$

Hence the distribution derivative is $\mathrm{sgn}' = 2\,\delta_0 \in \mathscr{S}'(\mathbb{R})$.

Example 2.14. For $\varepsilon > 0$, let us define $s_\varepsilon \in L^1(\mathbb{R})$ by $s_\varepsilon(x) := \mathrm{e}^{-\varepsilon|x|}\,\mathrm{sgn}(x)$. Then $\|s_\varepsilon - \mathrm{sgn}\|_{L^\infty} \not\to 0$ as $\varepsilon \to 0^+$, yet

$$\lim_{\varepsilon\to 0^+} s_\varepsilon(x) = \mathrm{sgn}(x),$$

and for $\eta \neq 0$ we have

$$\widehat{\mathrm{sgn}}(\eta) = \lim_{\varepsilon\to 0^+} \widehat{s_\varepsilon}(\eta) = \ldots = \frac{1}{\mathrm{i}\pi\eta}.$$

Thus if $r(y) := \frac{1}{\pi y}$ for $y \neq 0$, then $\widehat{r}(\eta) = +\mathrm{i}\,\mathrm{sgn}(-\eta) = -\mathrm{i}\,\mathrm{sgn}(\eta)$ formally. This suggests that the *Hilbert transform* $H = (u \mapsto Hu) : L^2(\mathbb{R}) \to L^2(\mathbb{R})$, for which

$$\widehat{Hu}(\eta) = -\mathrm{i}\,\mathrm{sgn}(\eta)\,\widehat{u}(\eta),$$

should satisfy convolution-type singular integral formula

$$Hu(x) = \int_{\mathbb{R}} \frac{u(x-y)}{\pi y} \mathrm{d}y := \lim_{\varepsilon\to 0^+} \left(\int_{-\infty}^{-\varepsilon} \frac{u(x-y)}{\pi y} \mathrm{d}y + \int_{\varepsilon}^{\infty} \frac{u(x-y)}{\pi y} \mathrm{d}y \right).$$

For the absolute convergence of Fourier integrals

$$u(x) = \int_{\widehat{\mathbb{R}}} \mathrm{e}^{+\mathrm{i}2\pi x\cdot\eta}\, \widehat{u}(\eta)\, \mathrm{d}\eta = \int_{\widehat{\mathbb{R}}} \mathrm{e}^{+\mathrm{i}2\pi x\cdot\eta} \int_{\mathbb{R}} \mathrm{e}^{-\mathrm{i}2\pi y\cdot\eta}\, u(y)\, \mathrm{d}y\, \mathrm{d}\eta,$$

we must have $u, \widehat{u} \in L^1(\mathbb{R})$, and then $u, \widehat{u}$ are also continuous and belong to all L^p-spaces: this is true certainly if $u, u', u'' \in L^1(\mathbb{R})$ (or more generally if $u, u' \in L^1(\mathbb{R})$ and $u' \in L^2(\mathbb{R})$). However, we can extend Fourier interpretations beyound L^p-spaces to tempered distributions. Thus, it is not harmful to write such Fourier integral formulas for signals outside $L^1(\mathbb{R})$, too. For $u \in \mathscr{S}'(\mathbb{R})$, in sense of distributions,

$$\begin{aligned}\int_{\widehat{\mathbb{R}}} \mathrm{e}^{+\mathrm{i}2\pi x\cdot\eta}\, \widehat{u}(\eta)\,\mathrm{d}\eta \;&=\; \int_{\widehat{\mathbb{R}}} \mathrm{e}^{+\mathrm{i}2\pi x\cdot\eta} \int_{\mathbb{R}} \mathrm{e}^{-\mathrm{i}2\pi y\cdot\eta}\, u(y)\,\mathrm{d}y\,\mathrm{d}\eta \\ = \int_{\mathbb{R}} \int_{\widehat{\mathbb{R}}} \mathrm{e}^{\mathrm{i}2\pi(x-y)\cdot\eta}\,\mathrm{d}\eta\, u(y)\,\mathrm{d}y \;&=\; \int_{\mathbb{R}} \delta_0(x-y)\,u(y)\,\mathrm{d}y \;=\; u(x).\end{aligned}$$

So, we have the bijective time-to-frequency Fourier transforms

$$\begin{array}{ccccc} \mathscr{S}(\mathbb{R}) & \subset & L^2(\mathbb{R}) & \subset & \mathscr{S}'(\mathbb{R}) \\ \downarrow & & \downarrow & & \downarrow \\ \mathscr{S}(\widehat{\mathbb{R}}) & \subset & L^2(\widehat{\mathbb{R}}) & \subset & \mathscr{S}'(\widehat{\mathbb{R}}) \end{array}$$

where

- $\mathscr{S}(\mathbb{R}) \cong \mathscr{S}(\widehat{\mathbb{R}})$ contains all the smooth rapidly decaying signals.
- $L^2(\mathbb{R}) \cong L^2(\widehat{\mathbb{R}})$ contains all the finite energy signals.
- $\mathscr{S}'(\mathbb{R}) \cong \mathscr{S}'(\widehat{\mathbb{R}})$ contains "nearly all the signals we ever meet".

With these Fourier bijections, we may present the signal

either *in time* **or** *in frequency*,

whatever is convenient for manipulation. But we would like to operate **simultaneously**

both *in time* **and** *in frequency*

— this is what we shall do in combined time-frequency analysis!

§ 3. Overview of time-frequency analysis

The main object for us to study here is the infinite family of Cohen class time-frequency transforms, which are natural sesquilinear translation-modulation invariant forms, trying to capture the idea of an energy density for a signal. We are especially interested in the Born–Jordan transform, which was deduced by Leon Cohen in [7] starting from the Born–Jordan quantization rule derived in [4].

Time-frequency analysis for signal processing is closely connected to the quantum mechanics. In 1925, the matrix mechanics was created by Werner Heisenberg, Max Born and Pascual Jordan, see [16], [4], and [5] (discussed e.g. in [26] and [10]). Heisenberg's postulates lead to a unique quantization or correspondence rule, so-called *Born–Jordan quantization.*

Other quantization rules were proposed, but all of them violate some of Heisenberg's original postulates in matrix mechanics. Of these, perhaps the most famous one is Hermann Weyl's quantization from 1927, in [28]. See also Weyl's monograph [29].

In 1932 in [30], Eugene Wigner introduced an idealized phase space energy distribution for quantum statistical mechanics: this is nowadays called the *Wigner distribution* (or the *Wigner–Ville distribution*, referring to Jean-André Ville who introduced it for signal analysis in [27]). Though the Wigner distribution has many desirable properties, it is of little use in practical applications due to its sensitivity to noise.

From the 1930s to the 1960s, many fundamental ideas of time-frequency analysis were introduced, most notably the *short-time Fourier transforms* and the related spectrograms, pioneered by researchers at the Bell Labs during the 2nd World War, and soon independently Dennis Gabor in [12]. Description of the early history of time-frequency methods can be found e.g. from [8] and [9]. Much of these developments became special instances of a wide class of quadratic time-frequency distributions, introduced in 1966 by Leon Cohen in [7]: roughly speaking, any such distribution is a convolution smoothing of the Wigner distribution. Each Cohen class time-frequency distribution corresponds to a quantization rule, and vice versa. Especially, the Wigner distribution corresponds to the Weyl quantization. One of Cohen's original examples was the deduction of the Born–Jordan time-frequency distribution out of the Born–Jordan quantization rule from [4]. For more information on the Cohen class time-frequency analysis, see [8], [9] and [14]. In recent years, there has been increasing interest to the Born–Jordan distribution: see e.g. the results by Paolo Boggiatto, Alessandro Oliaro et al in [2, 3]. Actually, the author of this article had his first encounter with the Born–Jordan distribution in a 2008 talk by Alessandro Oliaro.

Frequency-like variables will be denoted by the Greek letters $\xi, \eta, \cdots \in \widehat{\mathbb{R}} := \mathbb{R}$, corresponding Fourier dually to the *time-like variables* $x, y, \cdots \in \mathbb{R}$. To connect this analysis to signal processing applications, usually in the sequel we shall call variable x the *time*, and variable y is called the *lag* (which is actually a shift in time, to be precise: then $x + y$ is a new time instant). The corresponding Fourier dual variable $\eta = \widehat{y}$ is called the *frequency*, and $\xi = \widehat{x}$ is called the *doppler* variable (which is a shift in frequency, analogous to the lag in time — the name comes from the Doppler effect, introduced in 1842 by Christian Andreas Doppler).

As before, a *signal* u is a nice enough complex-valued function of real variable. The Fourier transform $\mathscr{F}u = \widehat{u} \in \mathscr{S}(\widehat{\mathbb{R}})$ of Schwartz test function $u \in \mathscr{S}(\mathbb{R})$ is defined by

$$\widehat{u}(\eta) := \int_{\mathbb{R}} \mathrm{e}^{-\mathrm{i}2\pi y\cdot\eta}\, u(y)\, \mathrm{d}y,$$

with the inverse Fourier transform $\mathscr{F}^{-1} = (\widehat{u} \mapsto u)$ by

$$u(x) = \int_{\widehat{\mathbb{R}}} \mathrm{e}^{+\mathrm{i}2\pi x\cdot\xi}\, \widehat{u}(\xi)\, \mathrm{d}\xi.$$

So, $\widehat{u}$ is another signal. The *symplectic Fourier transform* is $F = \mathscr{F} \otimes \mathscr{F}^{-1}$,

$$F = \mathscr{F} \otimes \mathscr{F}^{-1} : \mathscr{S}(\mathbb{R} \times \widehat{\mathbb{R}}) \to \mathscr{S}(\widehat{\mathbb{R}} \times \mathbb{R}). \tag{3.1}$$

The inverse of the symplectic Fourier transform is $F^{-1} = \mathscr{F}^{-1} \otimes \mathscr{F}$.

Fourier transform preserves energy, invertibly taking signals $x \mapsto u(x)$ of time variable $x \in \mathbb{R}$ to signals $\eta \mapsto \widehat{u}(\eta)$ of frequency variable $\eta \in \widehat{\mathbb{R}}$. Then $x \mapsto |u(x)|^2$ is the energy density of the signal in time x, and $\eta \mapsto |\widehat{u}(\eta)|^2$ the corresponding energy density in frequency η. How about finding an energy density $(x, \eta) \mapsto \delta(x, \eta)$ in combined time-frequency (x, η)?

Given a tempered distribution $\psi \in \mathscr{S}'(\mathbb{R} \times \widehat{\mathbb{R}})$ in the time-frequency plane, the corresponding Cohen class time-frequency transform of nice enough signals $u, v : \mathbb{R} \to \mathbb{C}$ is

$$W_\psi(u, v) = \psi * W(u, v) : \mathbb{R} \times \widehat{\mathbb{R}} \to \mathbb{C}, \tag{3.2}$$

where the Wigner transform $W(u, v) : \mathbb{R} \times \widehat{\mathbb{R}} \to \mathbb{C}$ is defined by

$$W(u, v)(x, \eta) := \int_{\mathbb{R}} \mathrm{e}^{-\mathrm{i}2\pi y \cdot \eta}\, u(x + y/2)\, v(x - y/2)^*\, \mathrm{d}y. \tag{3.3}$$

Let us define equivalence $u \sim v$ of measurable functions u, v if $u(x) = \lambda v(x)$ for almost all $x \in \mathbb{R}$, where $\lambda \in \mathbb{C}$ is a constant with $|\lambda| = 1$. Then let $[u] := \{v : \ u \sim v\}$ denote the corresponding equivalence class. The time-frequency distribution of $[u]$ is then

$$W_\psi[u] = \psi * W(u, u) : \mathbb{R} \times \widehat{\mathbb{R}} \to \mathbb{C}. \tag{3.4}$$

Here, $W_\psi[u](x, \eta)$ can be thought to describe "time-frequency energy density of $[u]$ at (x, η)" (with respect to convolution kernel ψ).

A rule of thumb is that the Wigner distribution $W[u] = W(u, u)$ is too sensitive to noise, so that some ψ-convolution smoothing is needed. On the other hand, e.g. invertibility would be desirable: if $u \in L^2(\mathbb{R})$, we would still like to recover $[u]$ from $W_\psi[u]$. But invertibility is destroyed when smoothing too much: for instance, information is numerically lost for all the spectrograms (which are related to the short-time Fourier transform). However, it will turn out that e.g. the Born–Jordan transform is invertible while it tolerates lots of noise; moreover, the Born–Jordan transform has many other pleasant properties, and it provides also an attractive computational tool for real-life applications.

§ 4. Unitary transformations and symmetry groups

Vaguely speaking, time-frequency analysis is about investigating families of operators on the Hilbert space $H = L^2(\mathbb{R})$ so that the natural time-frequency symmetries are taken into account. Let us consider *unitary operators* $A \in \mathscr{U}(L^2(\mathbb{R}))$, that is bijective linear mappings $A : L^2(\mathbb{R}) \to L^2(\mathbb{R})$ that are *isometric*, i.e.

$$\|Au\| = \|u\|. \tag{4.1}$$

For instance, the Fourier transform $\mathscr{F} : L^2(\mathbb{R}) \to L^2(\mathbb{R})$ is unitary. Let $y, \xi \in \mathbb{R}$, $p \in \mathbb{R}$ and $\lambda \in \mathbb{R} \setminus \{0\}$. Let us define $T(y), M(\xi), U(p), D(\lambda) \in \mathscr{U}(L^2(\mathbb{R}))$ (*translations, modulations, units, dilations*) by

$$T(y)u(x) := u(x-y), \tag{4.2}$$
$$M(\xi)u(x) := \mathrm{e}^{\mathrm{i}2\pi x\cdot\xi}u(x), \tag{4.3}$$
$$U(p)u(x) := \mathrm{e}^{\mathrm{i}2\pi p}u(x), \tag{4.4}$$
$$D(\lambda)u(x) := |\lambda|^{1/2}u(\lambda x). \tag{4.5}$$

Notice that we have group homomorphisms

$$\begin{aligned} &T, M : \mathbb{R} \to \mathscr{U}(L^2(\mathbb{R})),\\ &U : \mathbb{R} \to \mathscr{U}(L^2(\mathbb{R})),\\ &D : \mathbb{R}\setminus\{0\} \to \mathscr{U}(L^2(\mathbb{R})). \end{aligned}$$

Clearly, $U(p)$ commutes with all $\mathscr{F}$, $T(x)$, $M(\xi)$ and $D(t)$. Notice also that

$$\mathscr{F}\,T(y) = M(-y)\,\mathscr{F}, \tag{4.6}$$
$$\mathscr{F}\,M(\xi) = T(\xi)\,\mathscr{F}, \tag{4.7}$$
$$\mathscr{F}\,D(\lambda) = D(1/\lambda)\,\mathscr{F}. \tag{4.8}$$

Moreover,

$$T(y)\,M(\xi) = M(\xi)\,T(y)\,U(-\xi\cdot y), \tag{4.9}$$
$$D(\lambda)\,T(y) = T(y/\lambda)\,D(\lambda), \tag{4.10}$$
$$D(\lambda)\,M(\xi) = M(\lambda\xi)\,D(\lambda). \tag{4.11}$$

Thereby, for instance,

$$\begin{aligned} &(M(\xi_0)\,T(y_0)\,D(\lambda_0))\,(M(\xi_1)\,T(y_1)\,D(\lambda_1))\\ &= M(\xi_0+\lambda_0\xi_1)\,T(y_0+\lambda_0^{-1}y_1)\,D(\lambda_0\lambda_1)\,U(-\lambda_0 y_1\cdot\xi_0). \end{aligned}$$

Heisenberg group. There are several slightly different, yet essentially similar definitions for Heisenberg groups in the literature, see e.g. [11]. One may begin with e.g. quantum mechanical considerations or study matrix groups. We shall approach the subject via commutator relations of unitary operators. Recall that the Fourier transform intertwines modulations with translations: $\mathscr{F}M(\xi) = T(\xi)\mathscr{F}$. The *Heisenberg group* $\mathbb{H}$ is the minimal subgroup of $\mathscr{U}(L^2(\mathbb{R}))$ containing translations $T(y)$ and modulations $M(\xi)$ for all $y \in \mathbb{R}$ and $\xi \in \widehat{\mathbb{R}}$. How to parametrize the Heisenberg group? Noticing

$$\begin{aligned} &M(q)\,T(p)\,U(t-p\cdot q/2)\\ &= T(p)\,M(q)\,U(t+p\cdot q/2), \end{aligned}$$

we define $\sigma : \mathbb{R} \times \mathbb{R} \times \mathbb{R} \to \mathscr{U}(L^2(\mathbb{R}))$ by

$$\sigma(p, q, t) := M(q)\, T(p)\, U(t - p \cdot q/2), \tag{4.12}$$

for which

$$\begin{aligned} &\sigma(p_0, q_0, t_0)\, \sigma(p_1, q_1, t_1) \\ &= \sigma\left(p_0 + p_1, q_0 + q_1, t_0 + t_1 - \frac{p_0 \cdot q_1 - q_0 \cdot p_1}{2}\right). \end{aligned}$$

Thus we may identify $\mathbb{H}$ with $\mathbb{R} \times \mathbb{R} \times \mathbb{R}$, endowed with the group operation

$$(p_0, q_0, t_0)\, (p_1, q_1, t_1) := \left(p_0 + p_1, q_0 + q_1, t_0 + t_1 - \frac{p_0 \cdot q_1 - q_0 \cdot p_1}{2}\right), \tag{4.13}$$

so $\sigma : \mathbb{H} \to \mathscr{U}(L^2(\mathbb{R}))$ is a group homomorphism. The reader should be warned that the group operation in [11] is given by

$$\left(p_0 + p_1, q_0 + q_1, t_0 + t_1 + \frac{p_0 \cdot q_1 - q_0 \cdot p_1}{2}\right);$$

this group operation comes from studying $\rho(p, q, t) := M(q)\, T(-p)\, U(t + p \cdot q/2)$ instead of $\sigma(p, q, t)$. Homomorphism $\sigma : \mathbb{H} \to \mathscr{U}(L^2(\mathbb{R}))$ could be called the *Schrödinger representation* of the Heisenberg group. Explicitly,

$$\sigma(p, q, t)u(x) = \mathrm{e}^{\mathrm{i}2\pi x \cdot q}\, \mathrm{e}^{\mathrm{i}2\pi(t - p \cdot q/2)}\, u(x - p), \tag{4.14}$$

$$\rho(p, q, t)u(x) = \mathrm{e}^{\mathrm{i}2\pi x \cdot q}\, \mathrm{e}^{\mathrm{i}2\pi(t + p \cdot q/2)}\, u(x + p). \tag{4.15}$$

Nevertheless,

$$(p, q, t)^{-1} = (-p, -q, -t) \tag{4.16}$$

is the inversion in $\mathbb{H}$, and the center is $Z(\mathbb{H}) = \{(0, 0, t) \in \mathbb{R} \times \mathbb{R} \times \mathbb{R} : \ t \in \mathbb{R}\}$.

Translation-dilation group. Let the *translation–dilation group* $\mathbb{D}$ be the minimal subgroup of $\mathscr{U}(L^2(\mathbb{R}))$ containing translations $T(y)$ and dilations (or scalings) $D(\lambda)$ for every $y \in \mathbb{R}$ and $\lambda \in \mathbb{R} \setminus \{0\}$. Recall that

$$D(1/s)\, T(y) = T(sy)\, D(s),$$

so

$$\begin{aligned} &(T(y_0)\, D(1/s_0))\, (T(y_1)\, D(1/s_1)) \\ &= T(y_0 + s_0 y_1)\, D(1/(s_0 s_1)). \end{aligned}$$

Consequently, $\mathbb{D}$ can be identified with $\mathbb{R} \times (\mathbb{R} \setminus \{0\})$, endowed with the group operation

$$(y_0, s_0)(y_1, s_1) := (y_0 + s_0 y_1, s_0 s_1). \tag{4.17}$$

The translation-dilation group would lead to *time-scale analysis* (e.g. continuous wavelet transforms) — however, in this text we shall deal with the translation-modulation group, leading to *time-frequency analysis* (e.g. short-time Fourier transforms). It must be emphasized that there are time-frequency transforms that are also dilation-invariant! Examples of these are the Wigner and the Born–Jordan transforms.

§ 5. STFT (Short-Time Fourier Transform) and spectrogram

Spectrograms are currently the most commonly used time-frequency distributions. Let us briefly describe them. Let us now view an analog signal $u : \mathbb{R} \to \mathbb{C}$ through a window, which is another signal $w : \mathbb{R} \to \mathbb{C}$. The *w-windowed Fourier transform* (or **STFT**, the *Short-Time Fourier Transform*) for signal u is function $\mathcal{G}(u, w) : \mathbb{R} \times \widehat{\mathbb{R}} \to \mathbb{C}$, defined by the formula

$$\mathcal{G}(u, w)(x, \eta) := \widehat{u\, w_x^*}(\eta), \tag{5.1}$$

where $w_x(y) = w(y - x)$. In other words,

$$\mathcal{G}(u, w)(x, \eta) = \int_{\mathbb{R}} u(y)\, w(y - x)^* \, \mathrm{e}^{-\mathrm{i}2\pi y \cdot \eta} \, \mathrm{d}y.$$

Here $\mathcal{G}$ refers to Gabor [12]. The natural idea here is that the Fourier transform $\widehat{u}(\eta)$ measures the "content" of the signal u at frequency $\eta \in \mathbb{R}$ over all time instants, whereas $\mathcal{G}(u, w)(x, \eta)$ measures the "content" of the signal u at time-frequency point $(x, \eta) \in \mathbb{R} \times \widehat{\mathbb{R}}$ (when u is viewed through the window w). It is obviously appropriate to choose w so that most of its time-frequency content is nearby $(x, \eta) = (0, 0)$, especially

$$w(x) \approx 0 \approx \widehat{w}(\eta) \quad \text{for large} \quad |x|, |\eta|.$$

To normalize energy, it is natural to require $\|w\|^2 = 1$.

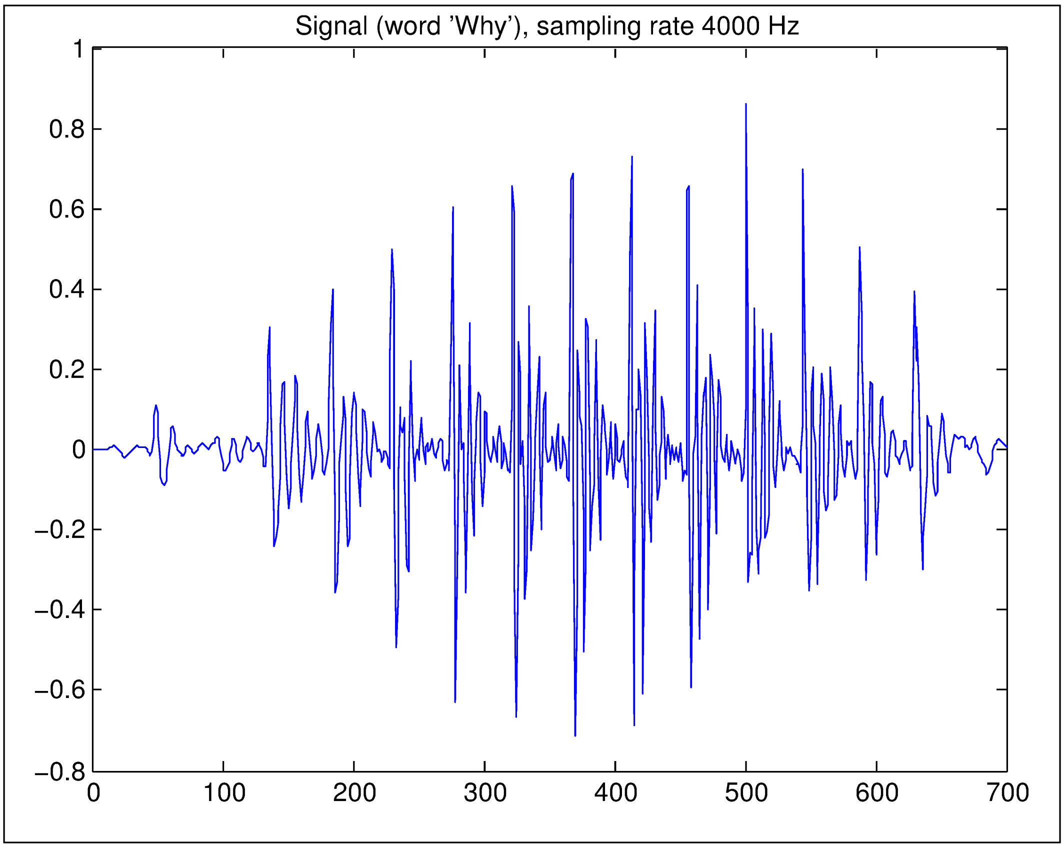
Spectrogram (Sonogram). The *w-spectrogram* (related to the w-windowed Fourier transform) for signal $u : \mathbb{R} \to \mathbb{C}$ is

$$|\mathcal{G}(u, w)|^2 : \mathbb{R} \times \widehat{\mathbb{R}} \to \mathbb{R}^+. \tag{5.2}$$

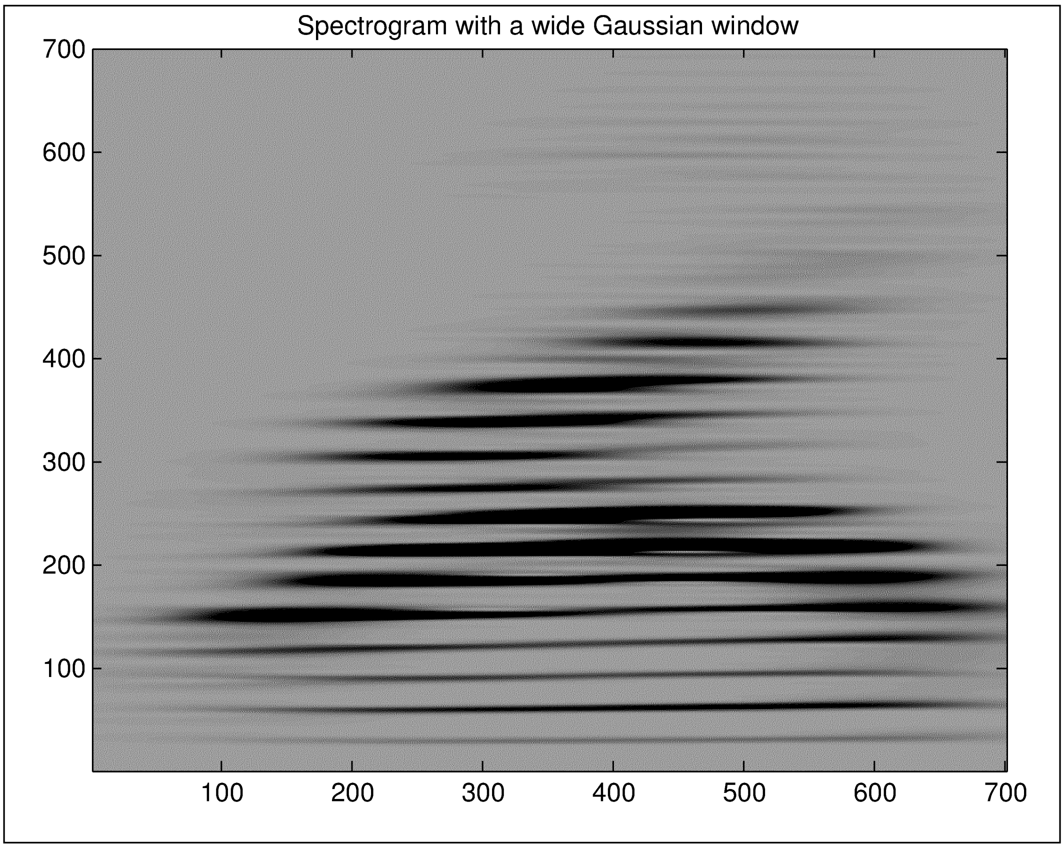
Idea here is that $|\mathcal{G}(u, w)(x, \eta)|^2 \geq 0$ is the "energy density" of signal $u : \mathbb{R} \to \mathbb{C}$ at the time-frequency point $(x, \eta) \in \mathbb{R} \times \widehat{\mathbb{R}}$ (when signal u is viewed through window w).

Note on time-frequency visualizations. We shall later explain how to discretize the time-frequency analysis. In the time-frequency pictures in the sequel computed with `Matlab`, the time always runs on the horizontal axis from left to right, and the frequency in the vertical axis from bottom to top; the time and frequency units do not matter at the moment. All our visualized signals will be real-valued, so that the symmetric time-frequency distributions for negative frequencies would be mirror images of the positive frequencies: thus we cut away the negative frequency part of the picture. In the spectrograms, the energy density is presented by gray-scale: the light gray corresponds to zero density, and darker gray to higher positive density (and we reserve the whiter shades of gray for the *negative* (sic!) energy densities).

Examples of spectrograms. Notice that the windowed Fourier transform and the corresponding spectrogram heavily depend on the choice of window w! The following spectrograms depict the same signal (male voice saying the word "*Why*", extracted from the signal from P. C. Hansen's website [15]). First, here is the waveform of the signal:

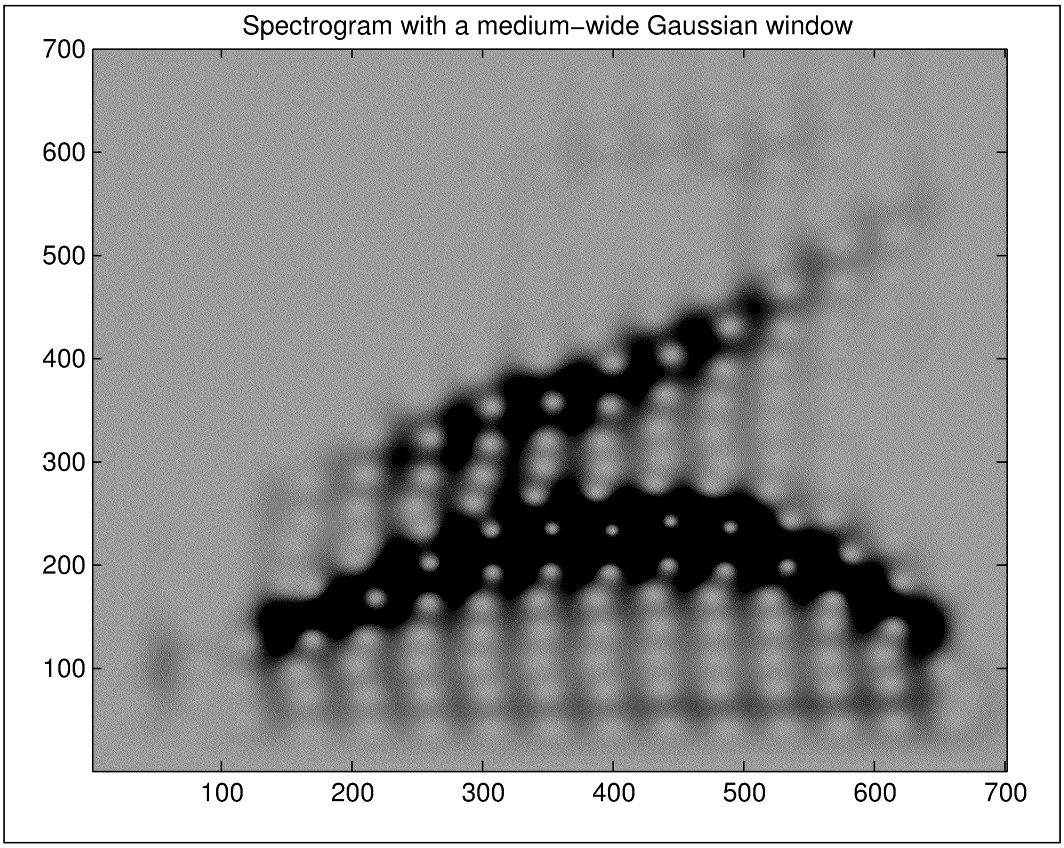


Here is the first of the three spectrograms for the signal. It is able to hazily locate frequencies, but not time-instants:



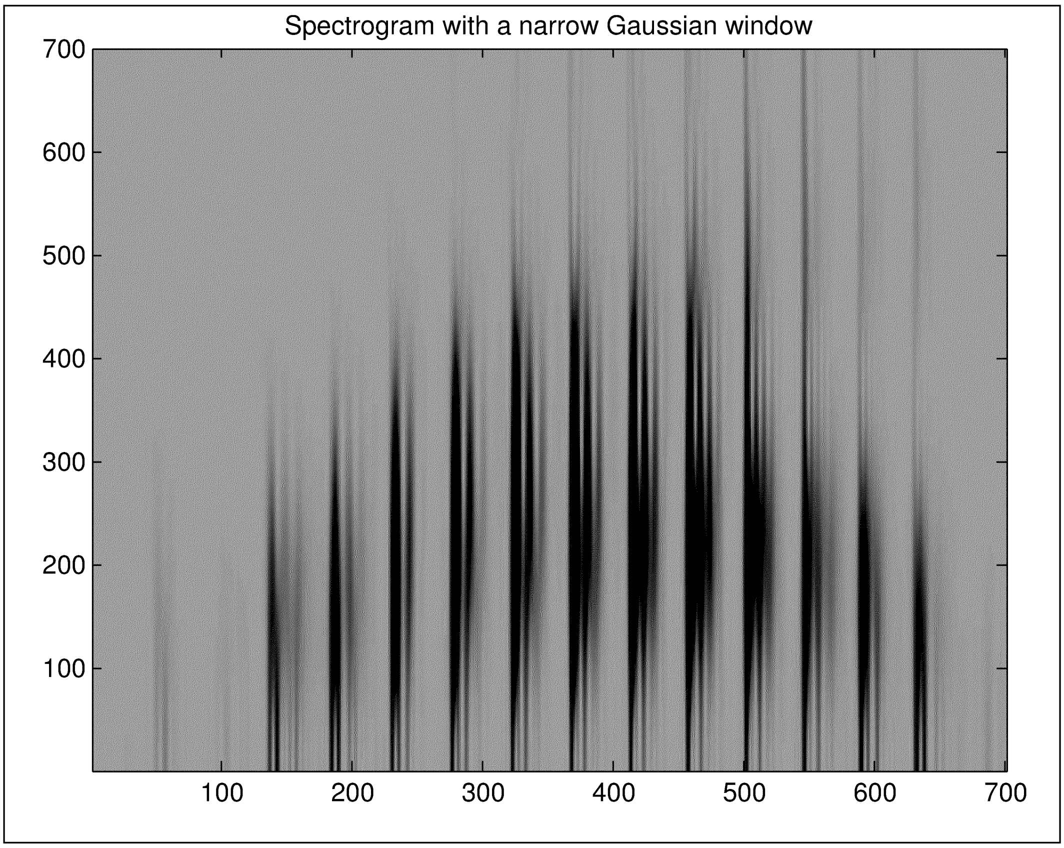
Later we will see that the Born–Jordan distribution can perfectly localize both frequencies and time-instants, also in practical cases like here.

Here is the second of the three spectrograms for the signal:



In the second spectrogram picture, in the case of the medium-wide Gaussian window, we have slightly increased intensity to enhance the visibility of the underlying fuzzy grid-like pattern that will become sharp and clear with the Born–Jordan distribution. Nevertheless, we emphasize that in any spectrogram, the information about the signal is lost for good, due to too much of diffusion. Also, spectrograms fail to be time-local, frequency-local, they do not have correct time nor frequency marginals, they are not scale invariant. They are pointwise positive, though, should that comfort someone.

Here is the third of the three spectrograms for the signal:



In the picture above, we can see hazy localizations of sudden snaps. The second spectrogram was a sort of an obscure mixture of the first and the third spectrograms.

In these spectrograms for the same signal, the sampling rate is 4000 Hz, and we took 700 samples (attaching enough zeros to the both ends of the signal). On the vertical axis, each numerical unit corresponds to $2000/700 = 20/7$ Hz. In each of these spectrograms, the time analysis window w is a Gaussian function $x \mapsto c_1 \exp(-c_2 x^2)$: in the first picture, this window is widest, and in the third picture the window is most narrow. A wide window locates frequencies quite well, whereas a narrow window is more capable of locating sudden transitions in the signal (like gnarly feature in the human voice, caused by the glottis pulse).

Roughly speaking, the obscurity of spectrograms is due to **both** "arbitrary" choice of the time-analysis window **and** "suffering" *twice* from the Heisenberg uncertainty principle in Fourier analysis. For Born–Jordan time-frequency distribution, we do not have to choose any time-analysis window, and we shall suffer only once from the uncertainty principle.

Example 5.1. Let δ_p be the Dirac delta at the time instant $p \in \mathbb{R}$, and let $e_\alpha : \mathbb{R} \to \mathbb{C}$, where $e_\alpha(x) = \mathrm{e}^{\mathrm{i}2\pi x\cdot\alpha}$. How does the choice of the window w show up in the spectrograms for signals δ_p, e_α? Simply, $|\mathcal{G}(\delta_p, w)(x,\eta)|^2 = |w(p-x)|^2$ and $|\mathcal{G}(e_\alpha, w)(x,\eta)|^2 = |\widehat{w}(\alpha-\eta)|^2$, hinting that the spectrograms might look quite hazy.

Example 5.2. Suppose we know the windowed Fourier transform $\mathcal{G}(u,w)$ together with w having no zeroes. How do we find signal u? Just take the inverse Fourier transform of $\eta \mapsto \mathcal{G}(u,w)(x,\eta)$, and do the easy arithmetic. However, the corresponding spectrogram $(x,\eta) \mapsto |\mathcal{G}(u,w)(x,\eta)|^2$ is unfortunately numerically not invertible.

§ 6. Ambiguity transform

The *ambiguity transform* (or *Woodward's radar ambiguity transform*) of signals $u, v \in \mathscr{S}(\mathbb{R})$ is $\chi(u,v) : \widehat{\mathbb{R}} \times \mathbb{R} \to \mathbb{C}$, where

$$\chi(u,v)(\xi,y) := \int_{\mathbb{R}} \mathrm{e}^{-\mathrm{i}2\pi x\cdot\xi}\, u(x+y/2)\, v(x-y/2)^*\, \mathrm{d}x. \tag{6.1}$$

The *ambiguity function* (or the *characteristic function*) of signal u is then $\chi[u] := \chi(u,u) : \widehat{\mathbb{R}} \times \mathbb{R} \to \mathbb{C}$, satisfying $\chi[u](\xi,y)^* = \chi[u](-\xi,-y)$. An application of the Cauchy–Schwarz inequality gives

$$|\chi[u](\xi,y)| \leq \chi[u](0,0) = \|u\|^2, \tag{6.2}$$

and it is easy to see that

$$|\chi[u](\xi,y)| < \chi[u](0,0) \tag{6.3}$$

for all $(\xi,y) \neq (0,0)$; this extends from the Schwartz test functions $u \in \mathscr{S}(\mathbb{R})$ to hold for all finite-energy signals $u \in L^2(\mathbb{R})$. For instance, if $u(t) = \mathrm{e}^{-\pi t^2}$ (so that $\widehat{u} = u$), we have

$$\chi[u](\xi,y) = \frac{1}{\sqrt{2}}\, \mathrm{e}^{-\pi(\xi^2+y^2)/2}.$$

The ambiguity function was first used in radar detection by Philip Woodward, [31]. For instance, suppose we fire a short effectively narrow bandwidth signal u_0 at time 0 at an object, which moves with radial velocity v_0 (small compared to the speed c of light) at relatively short distance d_0 from us: for simplicity, think of a complex Gaussian $u_0(x) = \mathrm{e}^{-\mu x^2}\mathrm{e}^{-\mathrm{i}2\pi x\cdot\eta_0}$, when $\widehat{u_0}$ is effectively concentrated nearby frequency η_0. The radar signal u_0 is reflected from the object back to us as signal u_1, with approximate time lag $y_0 := 2d_0/c$. At frequency η, the corresponding Doppler shift is $\xi = -2\eta v_0/c$. With $\xi_0 := -2\eta_0 v_0/c$, find (ξ_0, y_0) (and thus (v_0, d_0) that we actually want) approximately by maximizing the correlation

$$(\xi,y) \mapsto |\langle M(\xi)T(y)u_0, u_1\rangle|.$$

Here the relation to the ambiguity function is the following: forgetting other distortions assume that $u_1(x) \approx \lambda\, \mathrm{e}^{\mathrm{i}2\pi x\cdot\xi_0} u_0(x-y_0) = \lambda\, M(\xi_0)T(y_0)u_0(x)$ (for a constant λ), so

$$\begin{aligned}|\langle M(\xi)T(y)u_0, u_1\rangle| &\approx |\langle M(\xi)T(y)u_0, \lambda\, M(\xi_0)T(y_0)u_0\rangle| \\ &= |\lambda|\, |\chi[u_0](\xi-\xi_0, y-y_0)| \\ &\leq |\lambda|\, \chi[u_0](0,0) = |\lambda|\, \|u_0\|^2.\end{aligned}$$

§ 7. Wigner transform

Main properties of the Wigner distribution can be found e.g. in [8], [9], [11] and [14]. These properties are important, as the other Cohen class time-frequency distributions can be regarded as convolution-smoothings of the Wigner distribution.

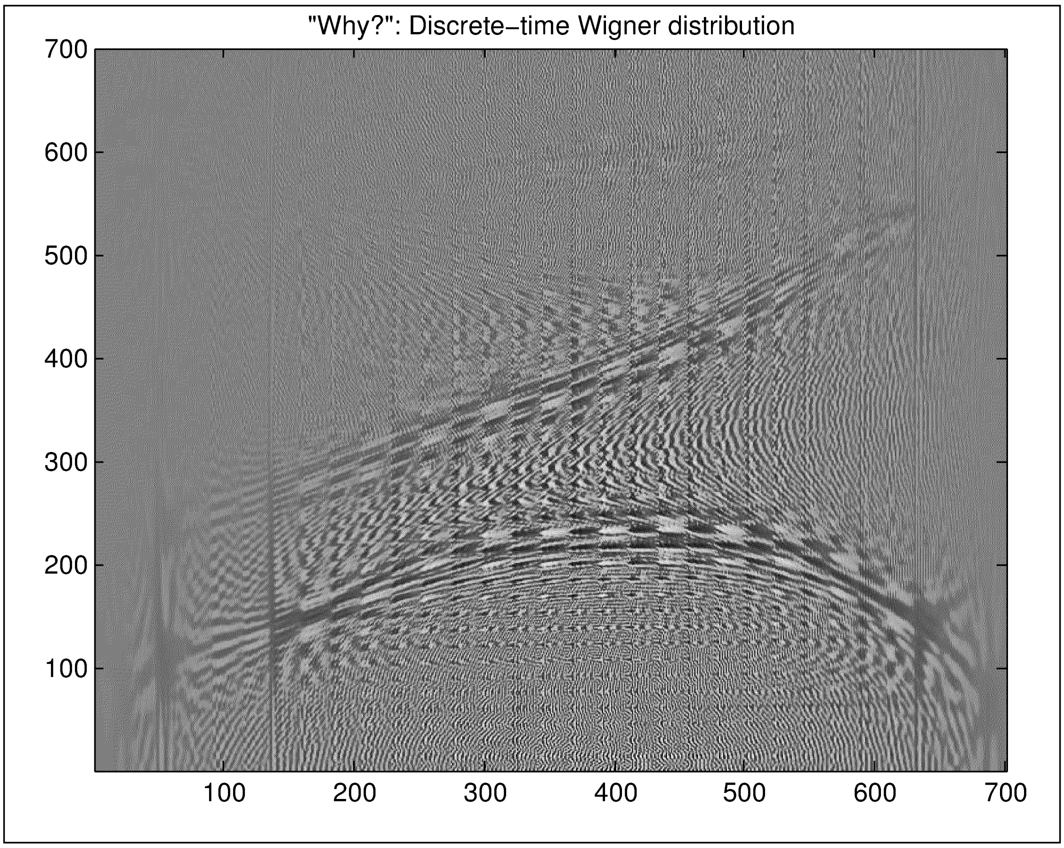
The *Wigner transform* $W(u,v)$ of signals u,v is given by

$$W(u,v)(x,\eta) := \int_{\mathbb{R}} \mathrm{e}^{-\mathrm{i}2\pi y\cdot\eta}\, u(x+y/2)\, v(x-y/2)^*\, \mathrm{d}y. \tag{7.1}$$

Clearly, this transform can be defined at least for signals $u,v \in \mathscr{S}(\mathbb{R})$, and it turns out that $W(u,v) : \mathbb{R}\times\widehat{\mathbb{R}} \to \mathbb{C}$ is continuous for also $u,v \in L^2(\mathbb{R})$. For tempered distributions, treat the integrals in a weak sense, as usual. Now $W[u] := W(u,u)$ is called *Wigner distribution* of signal u. For instance, if $u(t) = \mathrm{e}^{-\pi t^2}$ (so that $\widehat{u} = u$), we have

$$W[u](x,\eta) = \sqrt{2}\, \mathrm{e}^{-2\pi(x^2+\eta^2)}.$$

Here is the discrete-time Wigner distribution for the same signal that was depicted by the three different spectrograms in Section 5:



This illustrates how sensitive the Wigner distribution is to noise: the signal here was still rather well-behaving. To reduce the inherent interferences in the Wigner distribution, we are going to do smoothing by convolution. However, convolution of two Wigner distributions would lead back to the spectrograms, and we must eventually choose some other method:

Convolution of Wigner distributions. Spectrograms are actually obtained as time-frequency convolutions $W[v] * W[u] : \mathbb{R} \times \widehat{\mathbb{R}} \to \mathbb{R}$ of signals $v, u : \mathbb{R} \to \mathbb{C}$, because

$$\begin{aligned} W[v] * W[u](x,\eta) &= \int_{\widehat{\mathbb{R}}} \int_{\mathbb{R}} W[v](x-z,\eta-\omega)\, W[u](z,\omega)\, \mathrm{d}z\, \mathrm{d}\omega \\ &= \int_{\widehat{\mathbb{R}}} \int_{\mathbb{R}} \int_{\mathbb{R}} \int_{\mathbb{R}} \mathrm{e}^{-\mathrm{i}2\pi y\cdot(\eta-\omega)}\, v(x-z+y/2)\, v(x-z-y/2)^* \\ &\qquad \mathrm{e}^{-\mathrm{i}2\pi t\cdot\omega}\, u(z+t/2)\, u(z-t/2)^*\, \mathrm{d}t\, \mathrm{d}y\, \mathrm{d}z\, \mathrm{d}\omega \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} \mathrm{e}^{-\mathrm{i}2\pi y\cdot\eta}\, v(x-z+y/2)\, v(x-z-y/2)^*\, u(z+y/2)\, u(z-y/2)^*\, \mathrm{d}y\, \mathrm{d}z \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}2\pi(s-t)\cdot\eta}\, v(x-s)\, v(x-t)^*\, u(t)\, u(s)^*\, \mathrm{d}t\, \mathrm{d}s \\ &= \left| \int_{\mathbb{R}} \mathrm{e}^{-\mathrm{i}2\pi t\cdot\eta}\, v(x-t)^*\, u(t)\, \mathrm{d}t \right|^2 \\ &= |\mathscr{G}(u,w)(x,\eta)|^2, \end{aligned}$$

where the window is given by $w(y) := v(-y)$. Already this hints that there are serious problems with sharpness in spectrograms: as there is the Heisenberg uncertainty for each of $W[u]$ and $W[v]$, the time-frequency localization in spectrogram $|\mathscr{G}(u,w)|^2 = W[v] * W[u]$ suffers doubly from the uncertainty. For instance, consider the energy-normalized Gaussian window $w(y) = v(-y) = \mathrm{e}^{-\pi y^2/2}$: then we may think that $W[u]$ is the initial spatial temperature distribution for a heat flow, and after a while we see only the temperature distribution $W[v] * W[u]$ from which the initial data cannot be numerically recovered.

§ 8. Variants of Gabor transform

A *Gabor transform* of a signal $u \in L^2(\mathbb{R})$ with respect to a *window* $w \in L^2(\mathbb{R})$ could arguably be $\mathscr{G}_a(u,w) : \mathbb{R} \times \widehat{\mathbb{R}} \to \mathbb{C}$, where $a \in \mathbb{R}$ and

$$\begin{aligned} \mathscr{G}_a(u,w)(x,\eta) &:= \langle u, M(\eta)\, T(x)\, U(-ax\cdot\eta)\, w \rangle_{L^2(\mathbb{R})} \\ &= \int_{\mathbb{R}} u(y)\, w(y-x)^*\, \mathrm{e}^{-\mathrm{i}2\pi(y-ax)\cdot\eta}\, \mathrm{d}y. \end{aligned}$$

Notice here the close kinship to the short-time Fourier transform! Let $\mathscr{G} := \mathscr{G}_{1/2}$. Again, intuitively, here signal u is gazed through the window $M(\eta)T(x)w$ (or a variant), which is located around (x,η) in the phase space if the original signal w is located around $O = (0,0)$. For instance, when $n = 1$, $w(x) = \mathrm{e}^{-cx^2}$ would be a nice initial window function for some $c \in \mathbb{R}^+$.

Why should the case $a = 1/2$ be especially interesting among the continuum of the Gabor type transformations? As a first evidence, notice that

$$\mathscr{G}(u,w)(x,\eta) = \langle u, \sigma(x,\eta,0)w \rangle_{L^2(\mathbb{R})}\,. \tag{8.1}$$

Moreover,

$$\mathscr{G}_0(u,w)(x,\eta) = \langle u, M(\eta)\,T(x)\,w\rangle_{L^2(\mathbb{R})},$$
$$\mathscr{G}_1(f,g)(x,\eta) = \langle u, T(x)\,M(\eta)\,w\rangle_{L^2(\mathbb{R})}.$$

We shall learn more about the case $a = 1/2$ later. In any case, $|\mathscr{G}_a(u,w)| = |\mathscr{G}(u,w)|$, and a straight-forward calculation proves a Moyal-type equality

$$\langle \mathscr{G}_a(u_0,w_0), \mathscr{G}_a(u_1,w_1)\rangle_{L^2(\mathbb{R}\times\widehat{\mathbb{R}})} = \langle u_0,u_1\rangle_{L^2(\mathbb{R})}\,\langle w_0,w_1\rangle^*_{L^2(\mathbb{R})}. \tag{8.2}$$

Defining the *a-Gabor distribution*

$$\mathscr{G}_a[u] := \mathscr{G}_a(u,u), \tag{8.3}$$

we obtain

$$\langle \mathscr{G}_a[u], \mathscr{G}_a[w]\rangle_{L^2(\mathbb{R}\times\widehat{\mathbb{R}})} = \left|\langle u,w\rangle_{L^2(\mathbb{R})}\right|^2. \tag{8.4}$$

Some examples of Gabor distributions of tempered distributions:

$$\mathscr{G}_a[\delta_{x_0}](x,\eta) = \delta_0(x)\,\mathrm{e}^{-\mathrm{i}2\pi x_0\cdot\eta},$$
$$\mathscr{G}_a[\mathrm{e}_{\eta_0}](x,\eta) = \delta_0(\eta)\,\mathrm{e}^{+\mathrm{i}2\pi x\cdot\eta_0}.$$

Let us now deduce inversion formulas for the Gabor transforms. Since

$$\begin{aligned}\int_{\widehat{\mathbb{R}}} \mathrm{e}^{\mathrm{i}2\pi p\cdot\eta}\,\mathscr{G}_a(u,w)(z,\eta)\,\mathrm{d}\eta &= \int_{\widehat{\mathbb{R}}}\int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}2\pi(p-y+az)\cdot\eta}\,u(y)\,w(y-z)^*\,\mathrm{d}y\,\mathrm{d}\eta\\ &= u(p+az)\,w(p+(a-1)z)^*,\end{aligned}$$

by setting $x = p + az$ and $x_0 = p + (a-1)z$, we get

$$u(x) = (1/w(x_0)^*)\int_{\widehat{\mathbb{R}}} \mathrm{e}^{\mathrm{i}2\pi((1-a)x+ax_0)\cdot\eta}\,\mathscr{G}_a(u,w)(x-x_0,\eta)\,\mathrm{d}\eta. \tag{8.5}$$

One easily gets another inversion formula

$$u(x) = \|w\|^{-2}\int_{\widehat{\mathbb{R}}}\int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}2\pi(x-az)\cdot\eta}\,w(x-z)\,\mathscr{G}_a(u,w)(z,\eta)\,\mathrm{d}z\,\mathrm{d}\eta.$$

§ 9. Variants of Wigner transform

The Wigner distribution and the Wigner transform originated in quantum mechanics (see [30, 20]). Loosely speaking, the Wigner disribution provides idealized phase space pictures of signals. The a-Gabor transform $\mathscr{G}_a(u,w)(x,\eta)$ is a "*w-windowed Fourier transform of $u \in L^2(\mathbb{R})$*". Correspondingly, the a-Wigner transform $\mathscr{W}_a(u,w) =$

$(\mathscr{F}\otimes\mathscr{F}^{-1})\mathscr{G}_a(u,w):\widehat{\mathbb{R}}\times\mathbb{R}\to\mathbb{C}$ is a "*phase space picture of u with respect to w*". Let us describe this in detail: Recall the a-Gabor transform of $u,w\in L^2(\mathbb{R})$,

$$\mathscr{G}_a(u,w)(x,\eta)=\int_{\mathbb{R}} u(y)\,w(y-x)^*\,\mathrm{e}^{-\mathrm{i}2\pi(y-ax)\cdot\eta}\,\mathrm{d}y.$$

The *a-Wigner transform* of $u\in L^2(\mathbb{R})$ with respect to $w\in L^2(\mathbb{R})$ is the function $\mathscr{W}_a(u,w):\widehat{\mathbb{R}}\times\mathbb{R}\to\mathbb{C}$ defined by

$$\mathscr{W}_a(u,w)(\eta,x):=\left((\mathscr{F}\otimes\mathscr{F}^{-1})\mathscr{G}_a(u,w)\right)(\eta,x). \tag{9.1}$$

That is,

$$\begin{aligned}\mathscr{W}_a(u,w)(\eta,x)&=\int_{\widehat{\mathbb{R}}}\int_{\mathbb{R}}\mathrm{e}^{-\mathrm{i}2\pi y\cdot\eta}\,\mathrm{e}^{+\mathrm{i}2\pi x\cdot\xi}\,\mathscr{G}_a(u,w)(y,\xi)\,\mathrm{d}y\,\mathrm{d}\xi\\&=\int_{\widehat{\mathbb{R}}}\int_{\mathbb{R}}\int_{\mathbb{R}}\mathrm{e}^{-\mathrm{i}2\pi y\cdot\eta}\,\mathrm{e}^{+\mathrm{i}2\pi x\cdot\xi}\,\mathrm{e}^{-\mathrm{i}2\pi(t-ay)\cdot\eta}\,u(t)\,w(t-y)^*\,\mathrm{d}t\,\mathrm{d}y\,\mathrm{d}\xi\\&=\int_{\mathbb{R}}\mathrm{e}^{-\mathrm{i}2\pi y\cdot\eta}\,u(x+ay)\,w(x+(a-1)y)^*\,\mathrm{d}y.\end{aligned}$$

The $(1/2)$-Wigner transform gives the Wigner transform: let us denote $W(u,w)(x,\eta):=\mathscr{W}_{1/2}(u,w)(\eta,x)$. Now there is the obvious symmetry in

$$W(u,w)(x,\xi)=\int_{\mathbb{R}}\mathrm{e}^{-\mathrm{i}2\pi y\cdot\xi}\,u(x+y/2)\,w(x-y/2)^*\,\mathrm{d}y. \tag{9.2}$$

Let the *a-Wigner distribution* of $u\in L^2(\mathbb{R})$ be

$$\mathscr{W}_a[u]:=\mathscr{W}_a(u,u):\widehat{\mathbb{R}}\times\mathbb{R}\to\mathbb{C}. \tag{9.3}$$

Noticing that $(\mathscr{W}_a[u])^*=\mathscr{W}_{1-a}[u]$, the *Wigner distribution*

$$\mathscr{W}_{1/2}[u] \tag{9.4}$$

is a real-valued function on $\widehat{\mathbb{R}}\times\mathbb{R}$. Notice also that

$$\int_{\widehat{\mathbb{R}}}\mathscr{W}_a(u,w)(\eta,x)\,\mathrm{d}\eta=u(x)\,w(x)^*, \tag{9.5}$$

$$\int_{\mathbb{R}}\mathscr{W}_a(u,w)(\eta,x)\,\mathrm{d}x=\widehat{u}(\eta)\,\widehat{w}(\eta)^*. \tag{9.6}$$

One easily obtains a Moyal-type equality

$$\langle\mathscr{W}_a(u_0,w_0),\mathscr{W}_a(u_1,w_1)\rangle_{L^2(\widehat{\mathbb{R}}\times\mathbb{R})}=\langle u_0,u_1\rangle_{L^2(\mathbb{R})}\,\langle w_0,w_1\rangle^*_{L^2(\mathbb{R})}. \tag{9.7}$$

Especially,

$$\langle\mathscr{W}_a[u],\mathscr{W}_a[w]\rangle_{L^2(\widehat{\mathbb{R}}\times\mathbb{R})}=\left|\langle u,w\rangle_{L^2(\mathbb{R})}\right|^2. \tag{9.8}$$

Examples:

$$\mathscr{W}_0(u,w)(\eta,x) = u(x)\,\widehat{w}(\eta)^*\,\mathrm{e}^{-\mathrm{i}2\pi x\cdot\eta},$$
$$\mathscr{W}_1(u,w)(\eta,x) = \widehat{u}(\eta)\,w(x)^*\,\mathrm{e}^{+\mathrm{i}2\pi x\cdot\eta}.$$

In [11] it is shown that $\mathscr{W}_{1/2}[w] \geq 0$ if and only if either $w \equiv 0$ or

$$w(x) = \mathrm{e}^{-Ax^2+b\cdot x+c},$$

where $b, c \in \mathbb{C}$ and $A > 0$. Especially, if $w(x) = \mathrm{e}^{-\pi x^2}$ then

$$\mathscr{W}_{1/2}[w](\eta,x) = \sqrt{2}\,\mathrm{e}^{-2\pi(x^2+\eta^2)}. \tag{9.9}$$

Let us calculate

$$\begin{aligned}
&\mathscr{W}_a[M(\xi)T(y)w](\eta,x)\\
&= \int_{\mathbb{R}} \mathrm{e}^{-\mathrm{i}2\pi t\cdot\eta}\,M(\xi)T(y)w(x+at)\,(M(\xi)T(y)w(x+(a-1)t))^*\,\mathrm{d}t\\
&= \int_{\mathbb{R}} \mathrm{e}^{-\mathrm{i}2\pi t\cdot\eta}\,\mathrm{e}^{\mathrm{i}2\pi t\cdot\xi}\,w(x-y+at)\,w(x-y+(a-1)t)^*\,\mathrm{d}t\\
&= \mathscr{W}_a[w](\eta-\xi, x-y)\\
&= T(\xi,y)\,\mathscr{W}_a[w](\eta,x).
\end{aligned}$$

Thus

$$\langle \mathscr{W}_a[u], \mathscr{W}_a[M(\xi)T(y)w]\rangle_{L^2(\widehat{\mathbb{R}}\times\mathbb{R})} = \left|\langle u, M(\xi)T(y)w\rangle_{L^2(\mathbb{R})}\right|^2 \tag{9.10}$$
$$= |\mathscr{G}(u,w)(y,\xi)|^2\,. \tag{9.11}$$

Now

$$\begin{aligned}
\left((\mathscr{F}^{-1}\otimes I)\mathscr{W}_a(u,w)\right)(y,z) &= \int_{\widehat{\mathbb{R}}} \mathrm{e}^{\mathrm{i}2\pi y\cdot\eta}\,\mathscr{W}_a(u,w)(\eta,z)\,\mathrm{d}\eta\\
&= \int_{\widehat{\mathbb{R}}}\int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}2\pi(y-t)\cdot\eta}\,u(z+at)\,w(z+(a-1)t)^*\,\mathrm{d}t\,\mathrm{d}\eta\\
&= u(z+ay)\,w(z+ay-y)^*.
\end{aligned}$$

If we set here $x = z + ay$ and $x_0 = z + ay - y$, we get the inversion formula

$$u(x) = (1/w(x_0)^*)\int_{\widehat{\mathbb{R}}} \mathrm{e}^{\mathrm{i}2\pi(x-x_0)\cdot\xi}\,\mathscr{W}_a(u,w)(\xi,(1-a)x+ax_0)\,\mathrm{d}\xi. \tag{9.12}$$

Another inversion formula is obtained from the Gabor inversion:

$$\begin{aligned}
u(x) &= \|w\|^{-2}\int_{\widehat{\mathbb{R}}}\int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}2\pi(x-az)\cdot\eta}\,w(x-z)\,\mathscr{G}_a(u,w)(z,\eta)\,\mathrm{d}z\,\mathrm{d}\eta\\
&= \|w\|^{-2}\int_{\widehat{\mathbb{R}}}\int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}2\pi(x-az)\cdot\eta}\,w(x-z)\int_{\mathbb{R}}\int_{\widehat{\mathbb{R}}} \mathrm{e}^{-\mathrm{i}2\pi z\cdot\xi}\,\mathrm{e}^{\mathrm{i}2\pi y\cdot\eta}\,\mathscr{W}_a(u,w)(\xi,y)\,\mathrm{d}\xi\,\mathrm{d}y\,\mathrm{d}z\,\mathrm{d}\eta\\
&= \|w\|^{-2}\int_{\mathbb{R}}\int_{\widehat{\mathbb{R}}} w(x-z)\,\mathrm{e}^{-\mathrm{i}2\pi z\cdot\xi}\,\mathscr{W}_a(u,w)(\eta,x-az)\,\mathrm{d}\xi\,\mathrm{d}z.
\end{aligned}$$

Let us collect the intuition about the Wigner transform: for a signal $u \in L^2(\mathbb{R})$, the Wigner distribution $\mathscr{W}_{1/2}[u] : \widehat{\mathbb{R}}\times\mathbb{R} \to \mathbb{R}$ is a continuous function which is the frequency-time distribution of u, witnessed by e.g.

$$\left\langle \mathscr{W}_{1/2}[u], \mathscr{W}_{1/2}[M(\eta)T(x)w] \right\rangle_{L^2(\widehat{\mathbb{R}}\times\mathbb{R})} = \left| \int_{\mathbb{R}} u(y)\, w(y-x)^* \,\mathrm{e}^{-\mathrm{i}2\pi y\cdot\eta}\,\mathrm{d}y \right|^2. \tag{9.13}$$

Here, we may think that $M(\eta)T(x)w$ is a window located at (x,η) in the time-frequency space, through which the signal u is looked at. An important window is $w(x) = (2\pi c)^{1/4}\mathrm{e}^{-\pi c x^2}$, where $c \in \mathbb{R}^+$, for which

$$\mathscr{W}_{1/2}[M(\eta)T(x)w](\xi,y) = 2\,\mathrm{e}^{-2\pi(c(y-x)^2+(\xi-\eta)^2/c)}.$$

Here, larger c gives better time resolution at the expense of frequency, and vice versa. Pointwise values $\mathscr{W}_{1/2}[u](\eta,x) \in \mathbb{R}$ carry only limited meaning, the emphasis is on the averages as in equation (9.13).

§ 10. What is Cohen's class?

There is no commonly adopted definition for the Cohen class time-frequency distributions. Gröchenig defines in [14] that a Cohen's class time-frequency transform for signals u, v is of the form

$$W_\psi(u,v) := \psi * W(u,v), \tag{10.1}$$

for a tempered distribution $\psi \in \mathscr{S}'(\mathbb{R}\times\widehat{\mathbb{R}})$. The corresponding time-frequency distribution is then $W_\psi[u] = W_\psi(u,u)$. We already discussed such W_ψ in Section 3.

Now let $(u,v) \mapsto P(u,v) \in \mathbb{C}$ be sesquilinear, for u, v in a dense subspace of $L^2(\mathbb{R})$, with $P[u] := P(u,u)$ such that

$$P[T(x)M(\eta)u] = T(x,\eta)P[u] \quad \text{and} \quad |P(u,v)(0,0)| \le \textit{constant}\,\|u\|\,\|v\|. \tag{10.2}$$

Here, $P(u,v)(0,0) = \langle Au, v\rangle$ for a bounded linear operator $A : L^2(\mathbb{R}) \to L^2(\mathbb{R})$. Then applying the Schwartz kernel theorem, Gröchenig shows that $P(u,v) = W_\psi(u,v)$ for some $\psi \in \mathscr{S}'(\mathbb{R})$; also, there are ramifications of this result in [14]. It should be noted that in the applications we most often encounter *symmetric* time-frequency transforms, i.e. $P(u,v)(0,0) = \langle Au, v\rangle = \langle u, Av\rangle$, as then the energy density $P[u] = P(u,u)$ is real-valued.

However, while this generality of ψ would have advantages, in most of the real-life applications the distribution $\phi := F\psi \in \mathscr{S}'(\widehat{\mathbb{R}}\times\mathbb{R})$ is typically even a bounded function, and often smooth, though sometimes merely continuous: thus for this text, let us assume simply the boundedness

$$\phi = F\psi \in L^\infty(\widehat{\mathbb{R}}\times\mathbb{R}). \tag{10.3}$$

Actually, in practical examples often there are properties like $|\phi(\xi, y)| \leq \phi(0,0) = 1$, but we will come to the meaning of such extra properties later. We will use two notations $C^\phi = W_\psi$ for the *Cohen class time-frequency transforms*:

$$C^\phi(u,v) := W_\psi(u,v) = \psi * W(u,v), \tag{10.4}$$

where $\phi = F\psi \in L^\infty(\widehat{\mathbb{R}} \times \mathbb{R})$ is the symplectic Fourier transform of $\psi \in \mathscr{S}'(\mathbb{R} \times \widehat{\mathbb{R}})$. Notice that $\|C^\phi[u]\| \leq \|\phi\|_{L^\infty} \|u\|^2$, as

$$\|C^\phi(u,v)\| = \|\phi\, \chi(u,v)\| \leq \|\phi\|_{L^\infty} \, \|\chi(u,v)\| = \|\phi\|_{L^\infty} \, \|u\| \, \|v\|.$$

Let us briefly justify the generic convolution form $\psi * W(u,v)$ of the Cohen class: Nice enough $\psi : \mathbb{R} \times \widehat{\mathbb{R}} \to \mathbb{C}$ defines a time-frequency transform $(u,v) \mapsto \psi * W(u,v)$, where

$$\begin{aligned}
&\psi * W(u,v)(x,\eta) \\
&= \int_{\widehat{\mathbb{R}}} \int_{\mathbb{R}} \psi(x-t, \eta-\alpha)\, W(u,v)(t,\alpha)\, \mathrm{d}t\, \mathrm{d}\alpha \\
&= \int_{\widehat{\mathbb{R}}} \int_{\mathbb{R}} \int_{\mathbb{R}} \psi(x-t, \eta-\alpha)\, \mathrm{e}^{-\mathrm{i}2\pi a\cdot\alpha}\, u(t+a/2)\, v(t-a/2)^*\, \mathrm{d}a\, \mathrm{d}t\, \mathrm{d}\alpha \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} u(y)\, v(z)^* \left[\int \mathrm{e}^{\mathrm{i}2\pi(z-y)\cdot\alpha}\, \psi(x - \frac{y+z}{2}, \eta-\alpha)\, \mathrm{d}\alpha \right] \mathrm{d}y\, \mathrm{d}z.
\end{aligned}$$

Let us denote this previous innermost integral

$$L(x,y,z,\eta) := \int_{\widehat{\mathbb{R}}} \mathrm{e}^{\mathrm{i}2\pi(z-y)\cdot\alpha}\, \psi(x - \frac{y+z}{2}, \eta-\alpha)\, \mathrm{d}\alpha.$$

For a time-frequency transform P,

$$P(u,v)(0,0) = \langle Au, v\rangle = \int_{\mathbb{R}} \int_{\mathbb{R}} K_A(z,y)\, u(y)\, v(z)^*\, \mathrm{d}y\, \mathrm{d}z.$$

By the time-frequency invariance,

$$\begin{aligned}
P(u,v)(x,\eta) &= P(M_{-\eta}T_{-x}u, M_{-\eta}T_{-x}v)(0,0) \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} K_A(z,y)\, (M_{-\eta}T_{-x}u(y))\, (M_{-\eta}T_{-x}v(z))^*\, \mathrm{d}y\, \mathrm{d}z \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} K_A(z,y)\, u(y+x)\, v(z+x)^*\, \mathrm{e}^{\mathrm{i}2\pi(z-y)\cdot\eta}\, \mathrm{d}y\, \mathrm{d}z \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} u(y)\, v(z)^*\, \mathrm{e}^{\mathrm{i}2\pi(z-y)\cdot\eta}\, K_A(z-x, y-x)\, \mathrm{d}y\, \mathrm{d}z.
\end{aligned}$$

From this and from $L(x,y,z,\eta)$, we obtain

$$\mathrm{e}^{\mathrm{i}2\pi(z-y)\cdot\eta}\, K_A(z-x, y-x) = \int_{\widehat{\mathbb{R}}} \mathrm{e}^{\mathrm{i}2\pi(z-y)\cdot\alpha}\, \psi(x - \frac{y+z}{2}, \eta-\alpha)\, \mathrm{d}\alpha.$$

Hence

$$K_A(a,b) = \int_{\widehat{\mathbb{R}}} e^{i2\pi(a-b)\cdot(\alpha-\eta)}\, \psi(-\frac{a+b}{2}, \eta-\alpha)\, d\alpha =: \Psi(a+b, a-b).$$

This shows that kernels K_A, Ψ, ψ contain the same information!

§ 11. Symbols in ψ-quantization

Ideally, a time-frequency distribution $W_\psi[u]$ should act as an energy density of a signal u in the time-frequency plane (phase-space) $\mathbb{R} \times \widehat{\mathbb{R}}$. Any such energy density then dictates its own corresponding symbol-to-operator quantization, giving the natural linear signal processing operators for manipulating the signals in a desired manner: if we choose our energy density badly, we shall get bad outcomes in our signal processing.

Let us briefly consider how to define linear operators on a Hilbert space H. Clearly, linear $A : H \to H$ can be found by knowing the inner products $\langle Au, v\rangle \in \mathbb{C}$ for all $u, v \in H$. Actually, only $\langle Au, u\rangle \in \mathbb{C}$ (for all $u \in H$) is enough — this can be found from e.g. [22], but for the reader's convenience we present a short proof here:

Theorem. *Let $A, B : H \to H$ be linear. Then $A = B$ if for all $u \in H$*

$$\langle Au, u\rangle = \langle Bu, u\rangle.$$

Proof. Let $C = A - B$, $u, v \in H$, $\lambda \in \mathbb{C}$. Now

$$0 = \lambda 0 = \lambda\langle C(u+\lambda v), u+\lambda v\rangle = |\lambda|^2\langle Cu, v\rangle + \lambda^2\langle Cv, u\rangle.$$

Plug in $\lambda \in \{1, i\}$ to get $\langle Cu, v\rangle = 0$. Thus $A - B = C = 0$. □

Quantizations. The Wigner time-frequency transform corresponds to the *Weyl quantization* $\sigma \mapsto A_{W,\sigma}$, defined by the duality

$$\langle u, A_{W,\sigma} v\rangle_{L^2(\mathbb{R})} := \langle W(u,v), \sigma\rangle_{L^2(\mathbb{R}\times\widehat{\mathbb{R}})}, \tag{11.1}$$

leading to

$$A_{W,\sigma} v(x) = \int_{\mathbb{R}}\int_{\widehat{\mathbb{R}}} e^{i2\pi(x-y)\cdot\eta}\, v(y)\, \sigma(\frac{x+y}{2}, \eta)\, d\eta\, dy. \tag{11.2}$$

More generally, in the ψ-quantization, symbol $\sigma : \mathbb{R} \times \widehat{\mathbb{R}} \to \mathbb{C}$ gives rise to the *pseudo-differential operator* A^ψ_σ, where

$$\langle u, A^\psi_\sigma v\rangle = \langle \psi * W(u,v), \sigma\rangle. \tag{11.3}$$

In the Wigner–Weyl case above, we had $\psi = \delta_{(0,0)}$, i.e. $F\psi(\xi, y) = \phi(\xi, y) \equiv 1$. Thinking of the previous Theorem, the duality (11.3) in the special case $\langle u, A^\psi_\sigma u\rangle = \langle \psi * W[u], \sigma\rangle$ captures the idea of A^ψ_σ being the natural operator corresponding to the time-frequency weight function σ, as $\psi * W[u]$ acts as an energy density.

ψ-quantization in terms of Weyl. How is the Weyl quantization connected to the ψ-quantization $\sigma \mapsto A^\psi_\sigma$? Now

$$\langle u, A^\psi_\sigma v\rangle = \langle \psi * W(u,v), \sigma\rangle = \langle (F\psi)\, \chi(u,v), F\sigma\rangle = \langle \chi(u,v), \phi^* F\sigma\rangle = \langle W(u,v), F^{-1}(\phi^* F\sigma)\rangle.$$

That is, $A^\psi_\sigma = A_{W,\tau}$, where the Weyl operator symbol is

$$\tau = F^{-1}(\phi^*\, F\sigma) : \mathbb{R} \times \widehat{\mathbb{R}} \to \mathbb{C}.$$

Let us now consider some examples of ψ-quantizations:

Born–Jordan quantization. The Born–Jordan time-frequency transform $(u, v) \mapsto Q(u, v) = C^\phi(u, v)$ corresponds to the Born–Jordan quantization, where

$$\phi(\xi, y) = \operatorname{sinc}(\xi \cdot y). \tag{11.4}$$

This will be discussed in more detail soon. An alternative way of thinking the Born–Jordan transform is finding the "arithmetic average" of the a-Wigner transforms. More precisely,

(11.5)
$$Q(u, v)(x, \eta) = \int_0^1 \mathscr{W}_a(u, v)(\eta, x)\, \mathrm{d}a = \int_0^1 \int_{\mathbb{R}} \mathrm{e}^{-\mathrm{i}2\pi y\cdot\eta}\, u(x + ay)\, v(x + (a-1)y)^*\, \mathrm{d}y\, \mathrm{d}a.$$

This is the approach e.g. in [2] and [3].

Kohn–Nirenberg quantization. The Rihaczek time-frequency transform corresponds to the *Kohn–Nirenberg quantization.* In this case, $F\psi(\xi, y) = \phi(\xi, y) = \mathrm{e}^{\mathrm{i}\pi\xi\cdot y}$, leading to

$$A^\psi_\sigma v(x) = \int_{\mathbb{R}} \int_{\widehat{\mathbb{R}}} \mathrm{e}^{\mathrm{i}2\pi(x-y)\cdot\eta}\, v(y)\, \sigma(x, \eta)\, \mathrm{d}\eta\, \mathrm{d}y = \int_{\widehat{\mathbb{R}}} \mathrm{e}^{\mathrm{i}2\pi x\cdot\eta}\, \widehat{v}(\eta)\, \sigma(x, \eta)\, \mathrm{d}\eta. \tag{11.6}$$

The Kohn–Nirenberg quantization on various groups is studied in [24].

Feynman quantization. In the so-called symmetric Feynman quantization,

$$A^\psi_\sigma v(x) = \int_{\mathbb{R}} \int_{\widehat{\mathbb{R}}} \mathrm{e}^{\mathrm{i}2\pi(x-y)\cdot\eta}\, v(y)\, \frac{\sigma(x, \eta) + \sigma(y, \eta)}{2}\, \mathrm{d}\eta\, \mathrm{d}y. \tag{11.7}$$

Here $\phi(\xi, y) = \cos(\pi\xi \cdot y)$.

Anti-Wick quantization. In the similar fashion, as above, taking a spectrogram for the time-frequency distribution, we obtain the corresponding ψ-quantization $\sigma \mapsto A^\psi_\sigma$. This is called the *anti-Wick quantization*, if the short-time Fourier transform for the spectrogram had the normalized Gaussian $x \mapsto e^{-\pi x^2/2}$ as the window function, yielding $\phi(\xi, y) = e^{-\pi(\xi^2+y^2)/2}$. We could talk about anti-Wick quantization for other window functions, too. Even though the trivial positivity of spectrograms guarantees the positivity of the corresponding quantization (i.e. positive symbols give positive operators), anti-Wick quantization has plenty of bad properties in practise, as the spectrograms fail to have many nice properties — this issue becomes clear in the sequel.

§ 12. Properties of different quantizations

Let us now consider properties of ψ-quantizations in terms of time-frequency and ambiguity kernels. The ambiguity kernel $\phi = (\mathscr{F} \otimes \mathscr{F}^{-1})\psi$ is the symplectic Fourier transform of the time-frequency kernel ψ.

We could think that the value $C^\phi[u] = W_\psi[u](x, \eta) \in \mathbb{C}$ is an "idealized energy density of signal u at time-frequency $(x, \eta) \in \mathbb{R} \times \widehat{\mathbb{R}}$" — however, single points (x, η) in the phase-space do not have a physical meaning (think of the uncertainty principle, to be discussed soon separately): only large-enough time-frequency areas, say greater than 1, are of interest. So, we should not expect an energy density to be pointwise positive (by which we strictly speaking mean *non-negative*) but hopefully still real-valued; the total energy should be positive.

Translation-modulation invariance. Within the Cohen class, *translation-modulation invariance* (or *time-frequency shift invariance*) is guaranteed automatically:

$$v = M(\xi)T(y)u \quad \Longrightarrow \quad W_\psi[v] = T(y, \xi)W_\psi[u], \tag{12.1}$$

i.e. if $v(x) = e^{i2\pi x\cdot\xi}u(x-y)$ then $W_\psi[v](x, \eta) = W_\psi[u](x-y, \eta-\xi)$. This corresponds to the intuition that if we translate the signal in time-frequency, the corresponding energy density should be likewise translated.

Scale invariance (dilation invariance). Time-frequency distribution $u \mapsto W_\psi[u]$ is *scale invariant* (or *dilation invariant*) if

$$W_\psi[v](x, \eta) = W_\psi[u](\lambda x, \eta/\lambda) \tag{12.2}$$

whenever $v(x) = D(\lambda)u(x) = \lambda^{1/2}u(\lambda x)$, where $\lambda > 0$ (notice the energy conservation $\|v\|^2 = \|u\|^2$). This means that $\phi(\xi, y) = f(\xi \cdot y)$ for some $f : \mathbb{R} \to \mathbb{C}$ (almost everywhere). Scale invariance means that λ-speeding up the time results in the natural symplectic stretching of the time-frequency distribution, with the same λ-factor. In this sense, the choice of physical time or frequency units does not matter in case of scale invariant time-frequency distributions. Physical dimensionless quantities, like the time-frequency areas, should be scale invariant. Let us justify the condition for scale invariance: Let $v(x) = \lambda^{1/2} u(\lambda x)$. Here

$$\begin{aligned}
W_\psi[v](x,\eta) &= \int_{\mathbb{R}} \int_{\widehat{\mathbb{R}}} \mathrm{e}^{\mathrm{i}2\pi x\cdot\xi}\, \mathrm{e}^{-\mathrm{i}2\pi y\cdot\eta}\, \phi(\xi,y)\, \chi[v](\xi,y)\, \mathrm{d}\xi\, \mathrm{d}y \\
&= \int_{\mathbb{R}} \int_{\widehat{\mathbb{R}}} \int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}2\pi(x-z)\cdot\xi}\, \mathrm{e}^{-\mathrm{i}2\pi y\cdot\eta}\, \phi(\xi,y)\, v(z+y/2)\, v(z-y/2)^*\, \mathrm{d}z\, \mathrm{d}\xi\, \mathrm{d}y \\
&= \int_{\mathbb{R}} \int_{\widehat{\mathbb{R}}} \int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}2\pi((x-z)\cdot\xi - y\cdot\eta)}\, \phi(\xi,y)\, \lambda\, u(\lambda(z+y/2))\, u(\lambda(z-y/2))^* \mathrm{d}z\, \mathrm{d}\xi\, \mathrm{d}y \\
&= \int_{\mathbb{R}} \int_{\widehat{\mathbb{R}}} \int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}2\pi((\lambda x-z)\cdot\xi - y\cdot\eta/\lambda)}\, \phi(\lambda\xi, y/\lambda)\, u(z+y/2) u(z-y/2)^*\, \mathrm{d}z\, \mathrm{d}\xi\, \mathrm{d}y \\
&= \int_{\widehat{\mathbb{R}}} \int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}2\pi(\lambda x\cdot\xi - y\cdot\eta/\lambda)}\, \phi(\lambda\xi, y/\lambda)\, \chi[u](\xi,y)\, \mathrm{d}\xi\, \mathrm{d}y.
\end{aligned}$$

For all $\lambda > 0$ and for all $u \in L^2(\mathbb{R})$, for this to be equal to

$$W_\psi[u](\lambda x, \eta/\lambda) = \int_{\mathbb{R}} \int_{\widehat{\mathbb{R}}} \mathrm{e}^{\mathrm{i}2\pi(\lambda x\cdot\xi - y\cdot\eta/\lambda)}\, \phi(\xi,y)\, \chi[u](\xi,y)\, \mathrm{d}\xi\, \mathrm{d}y,$$

we must have $\phi(\lambda\xi, y/\lambda) = \phi(\xi, y)$ for (almost) all $(\xi, y) \in \widehat{\mathbb{R}} \times \mathbb{R}$, so that $\phi(\xi, y) = f(\xi\cdot y)$ for some f.

Realness. The Wigner distribution is real-valued even for complex-valued signals, so that time-frequency distribution $W_\psi[u]$ is real-valued whenever the time-frequency kernel ψ is real-valued. That is, the ambiguity kernel satisfies

$$\phi(\xi, y)^* = \phi(-\xi, -y). \tag{12.3}$$

This realness means that real-valued symbols σ give symmetric operators, that is

$$\langle u, A^\psi_\sigma v\rangle_{L^2(\mathbb{R})} = \langle A^\psi_\sigma u, v\rangle_{L^2(\mathbb{R})}. \tag{12.4}$$

Positivity. In this text, we shall not require the "idealized energy density" $W_\psi[u]$ be pointwise positive; like said, time-frequency points $(x, \eta) \in \mathbb{R} \times \widehat{\mathbb{R}}$ are not physically meaningful. Of course, all spectrograms are positive, and so are their positively weighted integral averages: these are the only positive distribution examples within the Cohen class. However, such positive Cohen class distributions are not computationally stably invertible. And already Wigner observed that positive sesquilinear time-frequency transforms fail to satisfy the following marginal properties for time and frequency:

Time marginals. Time-frequency distribution $u \mapsto W_\psi[u]$ has correct *time marginals* if

$$\int_{\widehat{\mathbb{R}}} W_\psi[u](x,\eta)\,\mathrm{d}\eta = |u(x)|^2, \tag{12.5}$$

which for continuous ϕ means that $\phi(\xi,0) = 1$ for all $\xi \in \widehat{\mathbb{R}}$. In other words, the energy density in time is the natural $|u|^2$. On the level of the ψ-quantization $\sigma \mapsto A_{\psi,\sigma}$, this means that frequency-independent symbols correspond to multiplication operators: $\sigma(t,\eta) = (f \otimes \mathbf{1})(t,\eta) = f(t)$ gives the multiplication operator $u \mapsto fu$, i.e.

$$A_{\psi,f\otimes\mathbf{1}}u(x) = f(x)\,u(x). \tag{12.6}$$

Frequency marginals. Time-frequency distribution $u \mapsto W_\psi[u]$ has correct *frequency marginals* if

$$\int_{\mathbb{R}} W_\psi[u](x,\eta)\,\mathrm{d}x = |\widehat{u}(\eta)|^2, \tag{12.7}$$

which for continuous ϕ means that $\phi(0,y) = 1$ for all $y \in \mathbb{R}$. In other words, the energy density in frequency is the natural $|\widehat{u}|^2$. On the level of the ψ-quantization $\sigma \mapsto A_{\psi,\sigma}$, this means that time-independent symbols correspond to convolution operators: $\sigma(t,\eta) = \widehat{g}(\eta)$ gives the convolution operator $u \mapsto A_{\psi,\sigma}u = g * u$, i.e.

$$A_{\psi,\mathbf{1}\otimes\widehat{g}}u(x) = g * u(x) = \int_{\mathbb{R}} g(x-y)\,u(y)\,\mathrm{d}y. \tag{12.8}$$

Conservation of energy. Time-frequency distribution $u \mapsto W_\psi[u]$ *conserves energy* if

$$\int_{\mathbb{R}\times\widehat{\mathbb{R}}} W_\psi[u](x,\eta)\,\mathrm{d}(x,\eta) = \|u\|^2, \tag{12.9}$$

which for continuous ϕ means that $\phi(0,0) = 1$. Especially, this is guaranteed whenever having time marginals or frequency marginals.

Uncertainty principle. *Uncertainty principle in quantum mechanics* refers to the fundamental impossibility of measuring certain physical quantities simultaneously with arbitrary precision. *Uncertainty principle in Fourier analysis* refers to simultaneous non-localization of a signal and its Fourier transform, often expressed as an inequality like

$$\|u\|^2 \le 4\pi \left(\int_{\mathbb{R}} x^2\,|u(x)|^2\,\mathrm{d}x\right)^{1/2} \left(\int_{\widehat{\mathbb{R}}} \eta^2\,|\widehat{u}(\eta)|^2\,\mathrm{d}\eta\right)^{1/2}, \tag{12.10}$$

which can be shown by integrating by parts and applying the Cauchy–Schwarz inequality, noticing that the Fourier transform preserves the energy. In time-frequency analysis, if we want to have both time marginals and frequency marginals (recall that this means $\phi(\xi,0) = 1 = \phi(0,y)$ for all ξ, y), this amounts to that the time-frequency distributions of non-zero signals cannot essentially be concentrated in small time-frequency regions.

Time-locality. Time-frequency distribution $u \mapsto W_\psi[u]$ is *time-local* if the supporting intervals are respected as follows:

$$\text{supp}(W_\psi[u]) \subset [a,b] \times \widehat{\mathbb{R}} \quad \text{whenever} \quad \text{supp}(u) \subset [a,b] \tag{12.11}$$

for all signals u. This means that the time-lag kernel $(\mathscr{F}^{-1} \otimes I)\phi$ satisfies equation $(\mathscr{F}^{-1} \otimes I)\phi(x,y) = 0$ whenever $|y| \leq 2|x|$.

Let us justify the condition for time-locality. Let $\phi = F\psi : \widehat{\mathbb{R}} \times \mathbb{R} \to \mathbb{C}$ be the ambiguity kernel corresponding to the time-frequency kernel $\psi : \mathbb{R} \times \widehat{\mathbb{R}} \to \mathbb{C}$. Here

$$\begin{aligned} W_\psi[u](x,\eta) &= \int_{\mathbb{R}} \int_{\widehat{\mathbb{R}}} \mathrm{e}^{\mathrm{i}2\pi x\cdot\xi}\, \mathrm{e}^{-\mathrm{i}2\pi y\cdot\eta}\, \phi(\xi,y)\, \chi[u](\xi,y)\, \mathrm{d}\xi\, \mathrm{d}y \\ &= \int_{\mathbb{R}} \int_{\widehat{\mathbb{R}}} \int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}2\pi x\cdot\xi}\, \mathrm{e}^{-\mathrm{i}2\pi y\cdot\eta}\, \phi(\xi,y)\, \mathrm{e}^{-\mathrm{i}2\pi z\cdot\xi}\, u(z+y/2)\, u(z-y/2)^*\, \mathrm{d}z\, \mathrm{d}\xi\, \mathrm{d}y \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} \mathrm{e}^{-\mathrm{i}2\pi y\cdot\eta}\, \varphi(x-z,y)\, u(z+y/2)\, u(z-y/2)^*\, \mathrm{d}z\, \mathrm{d}y, \end{aligned}$$

where $\varphi : \mathbb{R} \times \mathbb{R} \to \mathbb{C}$ is the *time-lag kernel*,

$$(\mathscr{F}^{-1} \otimes I)\phi(x-z,y) = \varphi(x-z,y).$$

Here $z \pm y/2 \in [a,b] \not\ni x$ means $|x-z| > |y|/2$. From this we obtain the condition for the time-locality:

$$\varphi(x-z,y) = 0 \quad \text{whenever} \quad |x-z| > |y|/2.$$

Frequency-locality (dual property to time-locality). Time-frequency distribution $u \mapsto W_\psi[u]$ is *frequency-local* if the supporting intervals are respected as follows:

$$\text{supp}(W_\psi[u]) \subset \mathbb{R} \times [\alpha,\beta] \quad \text{whenever} \quad \text{supp}(\widehat{u}) \subset [\alpha,\beta] \tag{12.12}$$

for all signals u. This means that the doppler-frequency kernel $(I \otimes \mathscr{F})\phi$ satisfies equation $(I \otimes \mathscr{F})\phi(\xi,\eta) = 0$ whenever $|\xi| \leq 2|\eta|$.

Invertibility. Time-frequency distribution $u \mapsto W_\psi[u]$ is *invertible* if $[u]$ (i.e. signal modulo unimodular constants) can be recovered back from $W_\psi[u]$. Notice that $FW_\psi[u] = \phi\,\chi[u]$, where the symplectic Fourier transform F is invertible, and $[u]$ can clearly be recovered from $\chi[u]$. In part of literature, it is erraneously said that the invertibility then means that ϕ does not vanish; but as Cohen in [8] credits Albert Nuttall for showing that $[u]$ can be recovered if ϕ has nowhere dense zero set (i.e. there are no open sets where ϕ would vanish) — then the ratio $(FW_\psi[u](\xi,y))/\phi(\xi,y)$ can be found by taking limits at zeros of ϕ. However, this statement has to be refined a bit: First, each such a zero point has to be of finite order. Second, ϕ should not rapidly decrease at infinity: this actually excludes e.g. the so-called Choi–Williams distributions [6], for which $\phi(\xi,y) = \mathrm{e}^{-(\xi\cdot y)^2/\lambda}$ for constants $\lambda > 0$ — here ϕ never vanishes, but yet the division by ϕ is numerically unstable (just like the inverse heat equation is ill-posed in the sense of Hadamard: there the problem is the frequency-domain division by rapidly decreasing Gaussians).

Unitarity. Time-frequency distribution $u \mapsto C^\phi[u] = W_\psi[u]$ is *unitary* if it satisfies *Moyal's formula*

$$\langle W_\psi(u,v), W_\psi(f,g)\rangle_{L^2(\mathbb{R}\times\widehat{\mathbb{R}})} = \langle u,f\rangle_{L^2(\mathbb{R})}\langle v,g\rangle^*_{L^2(\mathbb{R})}. \tag{12.13}$$

(originally for the Wigner distribution by José Enrique Moyal in [20]). The unitarity is equivalent to that $|\phi(\xi,y)| = 1$ for almost every $(\xi,y) \in \widehat{\mathbb{R}} \times \mathbb{R}$, see e.g. [17] or [14]. Certainly, Moyal's formula is satisfied by the Wigner distribution (corresponding to the Weyl pseudo-differential quantization). Also, Moyal's formula is satisfied by the *Rihaczek transform* [21] (corresponding to the Kohn–Nirenberg pseudo-differential quantization), where $\phi(\xi,y) = \mathrm{e}^{\mathrm{i}\pi\xi\cdot y}$, yielding

$$C^\phi(u,v)(x,\eta) = \mathrm{e}^{-\mathrm{i}2\pi x\cdot\eta}\, u(x)\,\widehat{v}(\eta)^*. \tag{12.14}$$

However, according to Cohen and Janssen [8], Moyal's formula might not be necessary for signal analysis, and it is not really used in quantum mechanics.

Other properties. In the literature, there are many other more or less desirable properties for time-frequency distributions, see e.g. [8] and [9]. Let us still mention *reduced interference property*, that for the scale invariant case $\phi(\xi,y) = f(\xi\cdot y)$ requires that $\lim_{|t|\to\infty} f(t) = 0$. Also, the nature of the zeroes of f plays a role here.

§ 13. Born–Jordan characterization

Which time-frequency transform $Q = \psi * W$ to choose? We shall require:

(1) Q is **scale invariant**.
(2) Q is **time-local**.
(3) Q maps "**comb-to-grid**".

More precisely, these conditions mean:

(1) If $v(x) = \sqrt{\lambda}\, u(\lambda x)$ then $Q[v](x,\eta) = Q[u](\lambda x, \eta/\lambda)$, where $\lambda > 0$.

(2) If $\mathrm{supp}(u) \subset [a,b]$ then $\mathrm{supp}(Q[u]) \subset [a,b] \times \widehat{\mathbb{R}}$.

(3) $Q[\delta_{\mathbb{Z}}] = \delta_{\mathbb{Z}} \otimes \mathbf{1} + \mathbf{1} \otimes \delta_{\mathbb{Z}} - \mathbf{1} \otimes \mathbf{1}$.

Notice that condition (3) here (Dirac delta comb-to-grid) is justified by

$$\begin{aligned} \delta_{\mathbb{Z}}(x) &= \sum_{k \in \mathbb{Z}} \delta_k(x) \\ &= \sum_{k \in \mathbb{Z}} \mathrm{e}^{\mathrm{i}2\pi x \cdot k}. \end{aligned}$$

Spectrograms always violate (1), (2) and (3). We then obtain the following:

Theorem. *Scale invariance, time-locality and "comb-to-grid" property are* **necessary and sufficient** *to characterize the Born–Jordan distribution Q in Cohen's class.*

Moreover, Born–Jordan distribution is $\mathbb{R}$-valued, is stably invertible (mod unimodular constants), is frequency-local, preserves energy, has correct marginals both in time and in frequency, satisfies group delay and instantaneous frequency properties, is easy to discretize, and has same computational complexity as spectrograms do.

However, Born–Jordan transform is not unitary (this is actually really good), and not causal, and the corresponding distributions are not positive.

Born–Jordan characterization, proof idea: Now we have

$$(\mathscr{F} \otimes \mathscr{F}^{-1})(\psi * W[u])(\xi, y) = \phi(\xi, y)\, \chi[u](\xi, y),$$

and we have to show that the ambiguity kernel $\phi = (\mathscr{F} \otimes \mathscr{F}^{-1})\psi$ satisfies $\phi(\xi, y) = \mathrm{sinc}(\xi \cdot y)$. The *ambiguity function* $\chi[u] = (\mathscr{F} \otimes \mathscr{F}^{-1}) W[u]$ is given by

$$\chi[u](\xi, y) = \int_{\mathbb{R}} \mathrm{e}^{-\mathrm{i}2\pi x \cdot \xi}\, u(x + y/2)\, u(x - y/2)^*\, \mathrm{d}x.$$

Gaussian signals have Gaussian ambiguity functions: for example, if $u(x) = \mathrm{e}^{-\pi x^2}$ so that $\widehat{u} = u$, then $\chi[u](\xi, y) = 2^{-1/2} \mathrm{e}^{-\pi(\xi^2 + y^2)/2}$.

First, the dilation-invariance: if $v(x) = \sqrt{\lambda}\, u(\lambda x)$ (where $\lambda > 0$) then $\psi * W[v](x, \eta) = \psi * W[u](\lambda x, \eta/\lambda)$. This means

$$\phi(\xi, y) = \widehat{\varphi}(\xi \cdot y)$$

for some tempered distribution $\varphi \in \mathscr{S}'(\mathbb{R})$.

Second, the time-locality: if $\mathrm{supp}(u) \subset [a,b]$ then $\mathrm{supp}(\psi * W[u]) \subset [a,b] \times \widehat{\mathbb{R}}$. We see that φ must be supported in $[-1/2, 1/2]$. Especially, $\widehat{\varphi}$ extends to an analytic function $\widehat{\varphi} : \mathbb{C} \to \mathbb{C}$, by Schwartz's Paley–Wiener Theorem [25].

Third, the Dirac comb-to-grid property

$$\psi * W[\delta_{\mathbb{Z}}] = \delta_{\mathbb{Z}} \otimes \mathbf{1} + \mathbf{1} \otimes \delta_{\mathbb{Z}} - \mathbf{1} \otimes \mathbf{1}$$

says that

$$\phi\, \chi[\delta_{\mathbb{Z}}] = \delta_{\mathbb{Z}} \otimes \delta_0 + \delta_0 \otimes \delta_{\mathbb{Z}} - \delta_0 \otimes \delta_0,$$

so that especially

$$\widehat{\varphi}(k) = \begin{cases} 1 & \text{if} \quad k = 0, \\ 0 & \text{if} \quad k \in \mathbb{Z} \setminus \{0\}. \end{cases}$$

Notice that sinc-function has these same values on $\mathbb{Z}$, with first order zero at each $k \in \mathbb{Z} \setminus \{0\}$. Hence $\widehat{u} = \widehat{\varphi}/\mathrm{sinc} : \mathbb{C} \to \mathbb{C}$ is an analytic function.

Finally, $\varphi = u * \mathbf{1}_{[-1/2,1/2]}$, yielding

$$\sum_{k\in\mathbb{Z}} \varphi(x+k) = \sum_{k\in\mathbb{Z}} \int_{[-1/2,1/2]} u(x+k-y)\,\mathrm{d}y = \int_{\mathbb{R}} u(x)\,\mathrm{d}x = \widehat{u}(0) = \frac{\widehat{\varphi}(0)}{\mathrm{sinc}(0)} = 1$$

for almost every $x \in \mathbb{R}$. Remembering that φ is supported on the unit interval $[-1/2, 1/2]$, we see that $\varphi(x) = 1$ for almost every $x \in [-1/2, 1/2]$. Thus we obtain

$$\widehat{\varphi}(\xi \cdot y) = \int_{\mathbb{R}} \mathrm{e}^{-\mathrm{i}2\pi x y\cdot\xi}\, \varphi(x)\,\mathrm{d}x = \int_{-1/2}^{1/2} \mathrm{e}^{-\mathrm{i}2\pi x y\cdot\xi}\,\mathrm{d}x = \mathrm{sinc}(\xi \cdot y),$$

and hence $\psi * W[u] = Q[u]$: we end up with the Born–Jordan distribution.

Remark. There are infinitely many Cohen class time-frequency distributions for which exactly two out of conditions (1), (2), (3) would hold.

§ 14. Born–Jordan transform

The Born–Jordan time-frequency transform $Q(u,v) : \mathbb{R} \times \widehat{\mathbb{R}} \to \mathbb{C}$ of Schwartz test functions $u, v \in \mathscr{S}(\mathbb{R})$ is defined by

$$Q(u,v)(x,\eta) := \int_{\mathbb{R}} \mathrm{e}^{-\mathrm{i}2\pi y\cdot\eta} \frac{1}{y} \int_{x-y/2}^{x+y/2} u(t+y/2)\, v(t-y/2)^*\,\mathrm{d}t\,\mathrm{d}y. \tag{14.1}$$

Alternatively, $Q(u,v) = C^{\phi}(u,v) = \psi * W(u,v)$, where $F\psi(\xi,y) = \phi(\xi,y) = \mathrm{sinc}(\xi \cdot y)$. Translating, we may also write the integral transform as

$$Q(u,v)(x,\eta) = \int_{\mathbb{R}} \mathrm{e}^{-\mathrm{i}2\pi y\cdot\eta}\, \frac{1}{y} \int_{x}^{x+y} u(z)\, v(z-y)^*\,\mathrm{d}z\,\mathrm{d}y = \int_0^1 \mathscr{W}_a(u,v)(\eta,x)\,\mathrm{d}a.$$

The *Born–Jordan time-frequency distribution* for signal $u : \mathbb{R} \to \mathbb{C}$ is

$$Q[u] := Q(u,u) : \mathbb{R} \times \widehat{\mathbb{R}} \to \mathbb{R}. \tag{14.2}$$

Interpretation is that $Q[u](x,\eta) \in \mathbb{R}$ is the "energy density" of signal $u : \mathbb{R} \to \mathbb{C}$ at the time-frequency point $(x,\eta) \in \mathbb{R} \times \widehat{\mathbb{R}}$.

Example. The Fourier transform of a function can rarely be found explicitly, and this is the case also for the Born–Jordan transform. Now let us consider the Born–Jordan transform of u at $x = 0$, where $u(t) := \mathrm{e}^{-|t|}$. Then

$$\begin{aligned} Q[u](x,\eta) &= \int_{\mathbb{R}} \mathrm{e}^{-\mathrm{i}2\pi y\cdot\eta} \frac{1}{y} \int_x^{x+y} u(t)\, u(t-y)^* \,\mathrm{d}t\,\mathrm{d}y \\ &\overset{x=0}{=} \int_{\mathbb{R}} \mathrm{e}^{-\mathrm{i}2\pi y\cdot\eta} \frac{1}{y} \int_0^{y} \mathrm{e}^{-|t|}\, \mathrm{e}^{-|t-y|} \,\mathrm{d}t\,\mathrm{d}y \\ &= \int_{\mathbb{R}} \mathrm{e}^{-\mathrm{i}2\pi y\cdot\eta}\, \mathrm{e}^{-|y|} \,\mathrm{d}y \\ &= (1+(2\pi\eta)^2)^{-1}. \end{aligned}$$

Boundedness of density. For all $u, v \in L^2(\mathbb{R})$, the Wigner transform has bound

$$|W(u,v)(x,\eta)| \leq 2\,\|u\|\,\|v\|. \tag{14.3}$$

By the shift-invariance, it is enough to show this boundedness for $(x,\eta) = (0,0)$. Here

$$|W(u,v)(0,0)| = 2\left|\int_{\mathbb{R}} u(z)\, v(-z)^* \,\mathrm{d}z\right| \overset{w(z):=v(-z)}{=} 2\,|\langle u, w\rangle| \leq 2\,\|u\|\,\|v\|.$$

For the Born–Jordan transform $Q(u,v)$,

$$|Q(u,v)(x,\eta)| \leq \pi\,\|u\|\,\|v\| \tag{14.4}$$

for all $(x,\eta) \in \mathbb{R} \times \widehat{\mathbb{R}}$. Especially, the Born–Jordan energy density is bounded by $\|Q[u]\|_{L^\infty} \leq \pi\,\|u\|^2$. By the time-frequency shift-invariance, it is enough to check this for $Q(u,v)(0,0)$. We shall see that $Q(u,v)(0,0) = \langle Au, v\rangle$, where $A : L^2(\mathbb{R}) \to L^2(\mathbb{R})$ is a self-adjoint (symmetric) operator. Let us find this operator:

$$\begin{aligned} Q(u,v)(0,0) &= \int_{\mathbb{R}} \frac{1}{z} \int_{-z/2}^{z/2} u(t+z/2)\, v(t-z/2)^* \,\mathrm{d}t\,\mathrm{d}z \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} K_A(x,y)\, u(y)\, v(x)^* \,\mathrm{d}x\,\mathrm{d}y \quad = \quad \langle Au, v\rangle, \end{aligned}$$

where the Schwartz kernel

$$K_A(x,y) = \begin{cases} |x-y|^{-1} & \text{if } xy < 0, \\ 0 & \text{if } xy \geq 0. \end{cases}$$

Thus

$$\begin{aligned}\langle Au, v\rangle &= \int_{\mathbb{R}}\int_{\mathbb{R}} K_A(x,y)\, u(y)\, v(x)^*\, \mathrm{d}y\, \mathrm{d}x \\ &= \int_{-\infty}^{0}\int_{0}^{\infty} \frac{u(y)\, v(x)^*}{y-x}\, \mathrm{d}y\, \mathrm{d}x + \int_{0}^{\infty}\int_{-\infty}^{0} \frac{u(y)\, v(x)^*}{x-y}\, \mathrm{d}y\, \mathrm{d}x \\ &= \int_{0}^{\infty}\int_{0}^{\infty} \frac{u(y)\, v(-x)^*}{y+x}\, \mathrm{d}y\, \mathrm{d}x + \int_{0}^{\infty}\int_{0}^{\infty} \frac{u(-y)\, v(x)^*}{x+y}\, \mathrm{d}y\, \mathrm{d}x.\end{aligned}$$

From this, applying the Hilbert integral inequality

$$\int_0^\infty\int_0^\infty \frac{|f(x)|\,|g(y)|}{x+y}\, \mathrm{d}x\, \mathrm{d}y \le \pi \left(\int_0^\infty |f(x)|^2\, \mathrm{d}x\right)^{1/2} \left(\int_0^\infty |g(y)|^2\, \mathrm{d}y\right)^{1/2}, \tag{14.5}$$

denoting $u = u_1 + u_2$ and $v = v_1 + v_2$, where

$$u_1(x) = \begin{cases} u(x) & \text{for } x \ge 0, \\ 0 & \text{for } x < 0, \end{cases} \qquad v_1(x) = \begin{cases} v(x) & \text{for } x \ge 0, \\ 0 & \text{for } x < 0, \end{cases}$$

we get

$$|\langle Au, v\rangle| \overset{(14.5)}{\le} \pi\, \|u_2\|\, \|v_1\| + \pi\, \|u_1\|\, \|v_2\| \le \pi\, \|u\|\, \|v\|,$$

where the last inequality followed as $\|u\|^2 = \|u_1\|^2 + \|u_2\|^2$ and $\|v\|^2 = \|v_1\|^2 + \|v_2\|^2$.

Born–Jordan and Hilbert transforms. Let Hu be the Hilbert transform of u, i.e. the principal value integral

$$Hu(x) = \text{p.v.}\ \frac{1}{\pi}\int_{\mathbb{R}} \frac{u(y)}{y-x}\, \mathrm{d}y = \frac{1}{\pi} \lim_{0<\varepsilon\to 0} \int_{|y|>\varepsilon} \frac{u(y)}{y-x}\, \mathrm{d}y.$$

For $u \in L^1(\mathbb{R})$, define $S[u] \in L^1(\mathbb{R})$ by

$$S[u](y) := \int_0^y u(t)\, u(t-y)^*\, \mathrm{d}t.$$

Here $\|S[u]\|_{L^1} \le \|u\|_{L^1}^2$, and we have

$$Q[u](x,\eta) = \pi\, H(S[M(\eta)u])(x).$$

Inversion formula. It is obvious that the Cohen class time-frequency distributions lose a bit information about signals. Namely, already the Wigner distribution is insensitive to the global phase of the signal: if $\lambda \in \mathbb{C}$ then $W[\lambda u] = |\lambda|^2\, W[u]$. This is not an essential issue, however. In quantum mechanics, wave functions u are identified modulo unimodular constants λ. For the real-valued signals (such as in acoustics), the only unimodular possibilities are $\lambda = +1$ and $\lambda = -1$, and here $+u$ and $-u$ are indistinguishable by ear: it does not matter to turn the sound wave upside down. Born–Jordan picture $Q[u]$ gives u modulo unimodular constants:

$$Q[u](x,\eta) = (I \otimes \mathscr{F}) Ru(x,\eta),$$

where Born–Jordan autocorrelation $R[u]$ is given by

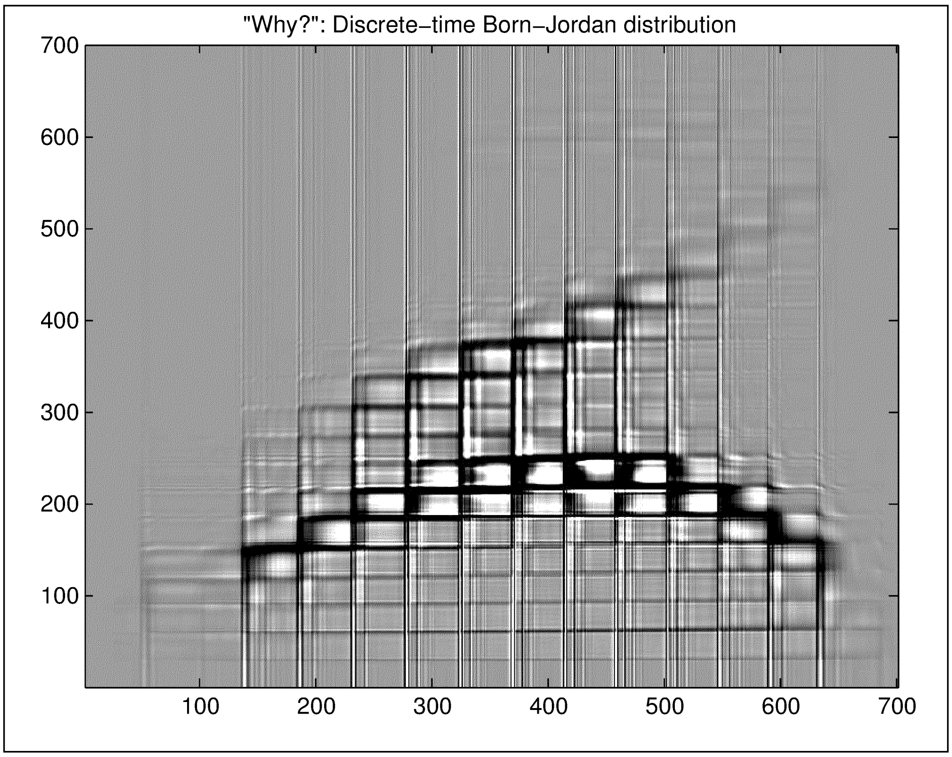
$$\begin{aligned} R[u](x,y) &= (I \otimes \mathscr{F}^{-1}) Q[u](x,y), \\ R[u](x,y) &\overset{y\to 0^+}{\longrightarrow} |u(x)|^2. \end{aligned}$$

For smooth real u, $u(x) \neq 0$ (one fixed x),

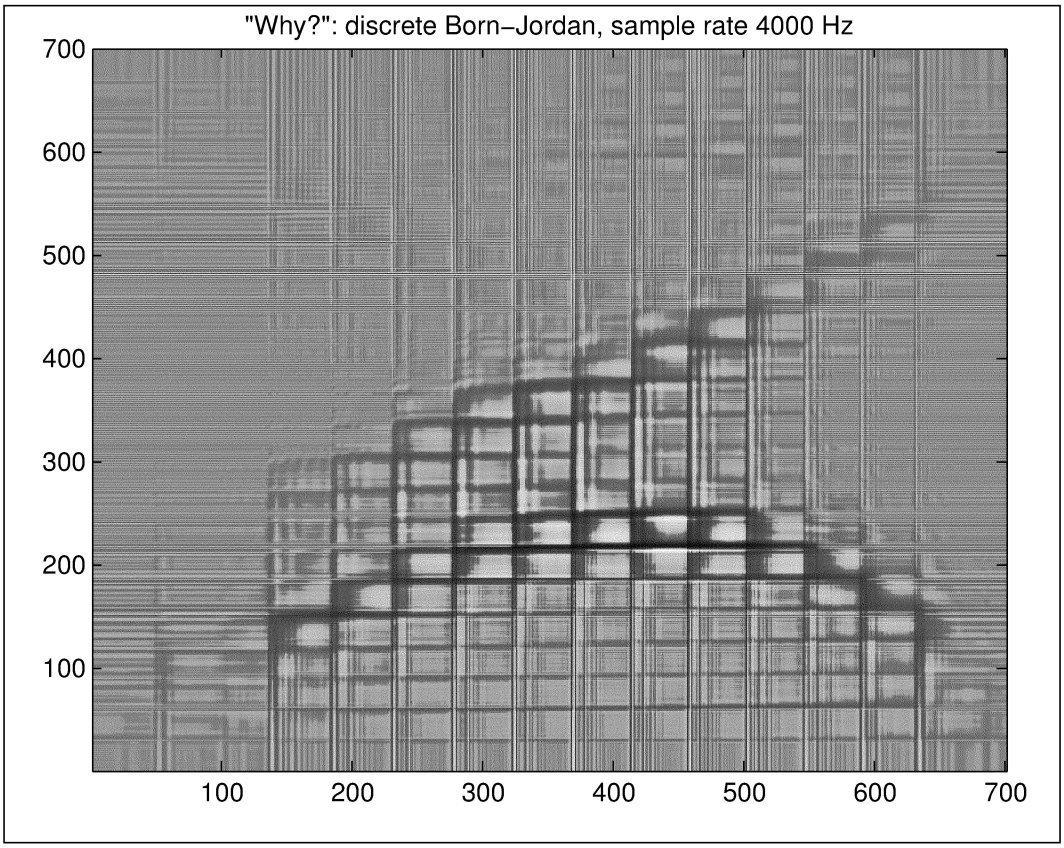
$$u(x+h) = \frac{h}{u(x)} \sum_{k=0}^{\infty} \partial_1 R[u](x-kh,|h|).$$

The reader may write out the inversion formula in the complex-valued case.

Example 14.1. Spectrograms become fuzzy (losing information) thanks to doubly suffering from the Heisenberg uncertainty, as spectrograms are convolution of two Wigner distributions. Also, clarity of spectrograms is inferior partly because of choosing a time-analysis window, but again this is not a problem for Born–Jordan distribution: there is no window to choose. In the following Born–Jordan pictures, with the familiar signal of a man speaking the word "Why?" (excerpt from [15]), the horizontal lines represent "whistling sounds", whereas the vertical lines are "snapping sounds". Again, the gray-scale colors correspond to the energy density: darker gray means higher positive density, and lighter gray to *negative* densities! The total energy is naturally always non-negative.



Just above, in the first of these pictures there was the discrete-time Born–Jordan distribution, and in the second picture we presented the fully periodized discrete Born–Jordan distribution:



From a numerical point of view, it requires as much effort to compute any spectrogram or Born–Jordan distribution (or basically other Cohen class time-frequency distributions, should one want to do so). In the sequel, the fine resolution of Born–Jordan distribution enables designing sharp time-frequency operations on signals. Notice also that signal u can be stably recovered (modulo unimodular constant) from its Born–Jordan distribution $Q[u]$: for any constant $\lambda \in \mathbb{C}$ it holds that $Q[\lambda u] = |\lambda|^2 Q[u]$, but from $Q[u]$ we can recover λu for some constant $\lambda \in \mathbb{C}$ for which $|\lambda| = 1$ (that is, in the case of a real signal u, we shall recover either $+u$ or $-u$) — this loss of information suffices in practise.

Recalling the well-known properties of Born–Jordan. So, Born–Jordan transform is a symmetric time-frequency transform: it is easy to see that $Q[u]$ is real-valued even for complex valued u. Also the time and frequency shifts work as they should: if

$$v(x) := u(x-y) \quad \text{and} \quad w(x) := \mathrm{e}^{\mathrm{i}2\pi x\cdot\xi}\, u(x),$$

then

$$\begin{aligned} Q[v](x,\eta) &= Q[u](x-y,\eta), \\ Q[w](x,\eta) &= Q[u](x,\eta-\xi). \end{aligned}$$

Notice that the Dirac delta (respectively complex exponential) distributions have vertical (respectively horizontal) Dirac delta lines as their Born–Jordan distributions:

$$Q[\delta_{x_0}](x,\eta) = \delta_{x_0}(x),$$
$$Q[e_{\eta_0}](x,\eta) = \delta_{\eta_0}(\eta),$$

where $e_{\eta_0}(x) := \mathrm{e}^{\mathrm{i}2\pi x\cdot\eta_0}$. Furthermore, the simple interference behaviour is examplified by the following: if $\alpha < \beta$, then

$$Q[\lambda e_\alpha + \mu e_\beta](x,\eta) = |\lambda|^2\delta_\alpha(\eta) + |\mu|^2\delta_\beta(\eta) + 2\,\mathrm{Re}\,(\lambda\,\mu^*\,e_{\alpha-\beta}(x))\,\frac{\mathbf{1}_{[\alpha,\beta]}(\eta)}{\beta-\alpha},$$

where $\lambda,\mu \in \mathbb{C}$ are constants — namely, here the "interference"

$$2\,\mathrm{Re}\,(\lambda\,\mu^*\,e_{\alpha-\beta}(x))\,\frac{\mathbf{1}_{[\alpha,\beta]}(\eta)}{\beta-\alpha}$$

is uniformly spread in the wide frequency strip $[\alpha,\beta]$, but nowhere else. Let us put strong emphasis on these two facts: first, thanks to this confinement to the $[\alpha,\beta]$ strip, this time-frequency transform is frequency-local (Fourier-dual to time-local) — second, the uniform spread to a wide strip guarantees good cancellation of oscillations in interferences of complicated signals. For the Born–Jordan transform, the marginal distributions are natural:

$$\int_{\mathbb{R}} Q[u](x,\eta)\,\mathrm{d}x = |\widehat{u}(\eta)|^2,$$
$$\int_{\widehat{\mathbb{R}}} Q[u](x,\eta)\,\mathrm{d}\eta = |u(x)|^2.$$

Hence, the energy is obtained from

$$\int_{\widehat{\mathbb{R}}}\int_{\mathbb{R}} Q[u](x,\eta)\,\mathrm{d}x\,\mathrm{d}\eta \;=\; \|u\|^2.$$

Property

$$Q[\widehat{u}](\eta,x) = Q[u](-x,\eta)$$

suggests that the Fourier transform turns the time-frequency plane by 90 degrees (that is, $\pi/2$ radians). Even though this follows easily from $\phi(\xi,y) = \mathrm{sinc}(\xi\cdot y)$, let us verify this by a direct calculation:

Proposition. *For Born–Jordan transform, Fourier transform turns the time-frequency plane by the right angle: more precisely,*

$$Q(\widehat{u},\widehat{v})(\eta,x) = Q(u,v)(-x,\eta). \tag{14.6}$$

Especially, $Q[\widehat{u}](\eta,x) = Q[u](-x,\eta)$*.*

Proof. Here,

$$\begin{aligned}
& Q(\widehat{u},\widehat{v})(\eta,x) \\
&= \int_{\widehat{\mathbb{R}}} \mathrm{e}^{-\mathrm{i}2\pi\xi\cdot x}\,\frac{1}{\xi}\int_{\eta-\xi/2}^{\eta+\xi/2} \widehat{u}(\tau+\xi/2)\,\widehat{v}(\tau-\xi/2)^*\,\mathrm{d}\tau\,\mathrm{d}\xi \\
&= \int_{\widehat{\mathbb{R}}} \mathrm{e}^{-\mathrm{i}2\pi x\cdot\xi}\,\frac{1}{\xi}\int_{\eta-\xi/2}^{\eta+\xi/2} \int_{\mathbb{R}}\int_{\mathbb{R}} u(a)\,v(b)^*\,\mathrm{e}^{-\mathrm{i}2\pi a\cdot(\tau+\xi/2)}\,\mathrm{e}^{+\mathrm{i}2\pi b\cdot(\tau-\xi/2)}\,\mathrm{d}a\,\mathrm{d}b\,\mathrm{d}\tau\,\mathrm{d}\xi.
\end{aligned}$$

By change $(a,b)\mapsto(t+y/2,t-y/2)$ of variables, we get

$$\begin{aligned}
& Q(\widehat{u},\widehat{v})(\eta,x) \\
&= \int_{\widehat{\mathbb{R}}} \mathrm{e}^{-\mathrm{i}2\pi x\cdot\xi}\,\frac{1}{\xi}\int_{\eta-\xi/2}^{\eta+\xi/2} \int_{\mathbb{R}}\int_{\mathbb{R}} u(t+y/2)\,v(t-y/2)^*\,\mathrm{e}^{-\mathrm{i}2\pi(t\cdot\xi+\tau\cdot y)}\,\mathrm{d}t\,\mathrm{d}y\,\mathrm{d}\tau\,\mathrm{d}\xi.
\end{aligned}$$

Here

$$\begin{aligned}
& \int_{\widehat{\mathbb{R}}} \mathrm{e}^{-\mathrm{i}2\pi x\cdot\xi}\,\frac{1}{\xi}\int_{\eta-\xi/2}^{\eta+\xi/2} \mathrm{e}^{-\mathrm{i}2\pi(t\cdot\xi+\tau\cdot y)}\,\mathrm{d}\tau\,\mathrm{d}\xi \\
&= \int_{\widehat{\mathbb{R}}} \mathrm{e}^{-\mathrm{i}2\pi(x+t)\cdot\xi}\,\frac{1}{\xi}\int_{\eta-\xi/2}^{\eta+\xi/2} \mathrm{e}^{-\mathrm{i}2\pi\tau\cdot y}\,\mathrm{d}\tau\,\mathrm{d}\xi \\
&= \mathrm{e}^{-\mathrm{i}2\pi y\cdot\eta}\int_{\widehat{\mathbb{R}}} \mathrm{e}^{-\mathrm{i}2\pi(x+t)\cdot\xi}\,\frac{\mathrm{e}^{\mathrm{i}\pi y\cdot\xi}-\mathrm{e}^{-\mathrm{i}\pi y\cdot\xi}}{\mathrm{i}2\pi y\cdot\xi}\,\mathrm{d}\xi \\
&= \mathrm{e}^{-\mathrm{i}2\pi y\cdot\eta}\,\frac{1}{|y|}\,\mathbf{1}_{[-x-|y|/2,-x+|y|/2]}(t),
\end{aligned}$$

where $\mathbf{1}_{[a,b]}$ is the characteristic function of interval $[a,b]$ (for the present discussion, it does not matter what happens in sets of measure zero, e.g. whether we consider open or closed intervals). This amounts to

$$\begin{aligned}
Q(\widehat{u},\widehat{v})(\eta,x) &= \int_{\mathbb{R}} \mathrm{e}^{-\mathrm{i}2\pi y\cdot\eta}\,\frac{1}{y}\int_{-x-y/2}^{-x+y/2} u(t+y/2)\,v(t-y/2)^*\,\mathrm{d}t\,\mathrm{d}y \\
&= Q(u,v)(-x,\eta)
\end{aligned}$$

proving the result. □

A conjugate bilinear time-frequency transformation. For $u,v\in\mathscr{S}(\mathbb{R})$, let us define $B(u,v):\mathbb{R}\times\widehat{\mathbb{R}}\to\mathbb{C}$ by

$$B(u,v)(x,\eta) := \int_{-\infty}^{x}\int_{-\infty}^{\eta} \widehat{u}(\omega)\,\mathrm{e}^{\mathrm{i}2\pi t\cdot\omega}\,v(t)^*\,\mathrm{d}\omega\,\mathrm{d}t.$$

Then it is easy to show that

$$B:\mathscr{S}(\mathbb{R})\times\mathscr{S}(\mathbb{R})\to L^\infty\cap C^\infty(\mathbb{R}\times\widehat{\mathbb{R}}), \tag{14.7}$$

and that this also extends to a conjugate bilinear mapping

$$B : L^2(\mathbb{R}) \times L^2(\mathbb{R}) \to L^\infty \cap C(\mathbb{R} \times \widehat{\mathbb{R}}). \tag{14.8}$$

Let us also denote $B[u] := B(u, u)$.

Proposition. *Born–Jordan distribution $Q[u] = Q(u, u)$ is related to $B[u]$ by*

$$Q[u] = 4\pi \operatorname{Im}(B[u]). \tag{14.9}$$

Proof. The mixed partial derivatives of the Born–Jordan distribution $Q[u]$ give

$$\begin{aligned}
\partial_x \partial_\eta Q[u](x,\eta) &= \int_{\mathbb{R}} \mathrm{e}^{-\mathrm{i}2\pi y\cdot\eta}\,(-\mathrm{i}2\pi y)\frac{1}{y}\int_x^{x+y} u(t)\,u(t-y)^*\,\mathrm{d}t\,\mathrm{d}y \\
&= -\mathrm{i}2\pi \int_{\mathbb{R}} \mathrm{e}^{-\mathrm{i}2\pi y\cdot\eta}\,[u(x+y)\,u(x)^* - u(x)\,u(x-y)^*]\,\mathrm{d}y \\
&= 4\pi \operatorname{Im}\left(\widehat{u}(\eta)\,\mathrm{e}^{\mathrm{i}2\pi x\cdot\eta}\,u(x)^*\right) \\
&= 4\pi \operatorname{Im}\left(\partial_x\partial_\eta B[u](x,\eta)\right).
\end{aligned}$$

From this, $Q[u] = 4\pi \operatorname{Im}(B[u])$ for all $u \in \mathscr{S}(\mathbb{R})$ follows, as both $Q[u]$ and $B[u]$ vanish when x and η tend to $-\infty$. □

Many old known results like continuity $Q : L^2(\mathbb{R}) \to L^2 \cap C(\mathbb{R}\times\widehat{\mathbb{R}})$ are now obvious by the properties of $u \mapsto B[u]$. One of the benefits of $u \mapsto B[u]$ is that it gives a better understanding of how Born–Jordan distribution behaves with respect to differentiation:

$$B[u'](x,\eta) = \left[(2\pi\eta)^2 + \mathrm{i}2\pi\eta\,\frac{\partial}{\partial x}\right] B[u](x,\eta). \tag{14.10}$$

Proposition. *Let $u \in L^2(\mathbb{R})$ have support in $[a,b]$. Then*

$$(x,\eta) \mapsto Q[u](x,\eta)$$

is supported in $[a,b]\times\mathbb{R}$, is C^{k+1} at x if $u \in C^k$ at x, is C^∞ in η.

Proof. The support property is just the time-locality, which could also be verified by a direct calculation. The infinite smoothness in the frequency variable follows, since by the Paley–Wiener Theorem, the Fourier transform of a compactly supported square-integrable function is analytic. Finally, from

$$\partial_\eta \partial_x^{k+1} Q[u](x,\eta) = -\mathrm{i}2\pi \int_{\mathbb{R}} \mathrm{e}^{-\mathrm{i}2\pi y\cdot\eta}\,\partial_x^k\,[u(x+y)\,u(x)^* - u(x)\,u(x-y)^*]\,\mathrm{d}y$$

we see that there is a gain of one extra degree of smoothness in time. □

§ 15. Born–Jordan quantization

In the Born–Jordan quantization $\sigma \mapsto A_\sigma$, the Born–Jordan pseudo-differential operator $A = A_\sigma : \mathscr{S}(\mathbb{R}) \to \mathscr{S}(\mathbb{R})$ with symbol $\sigma : \mathbb{R} \times \widehat{\mathbb{R}} \to \mathbb{C}$ is defined by the L^2-duality

$$\langle u, Av \rangle_{L^2(\mathbb{R})} := \langle Q(u,v), \sigma \rangle_{L^2(\mathbb{R}\times\widehat{\mathbb{R}})}. \tag{15.1}$$

Here formally

$$Au(x) = \int_{\mathbb{R}} K_A(x,y)\, u(y)\, \mathrm{d}y, \tag{15.2}$$

where the Schwartz kernel K_A formally satisfies

$$K_A(x,y) = \frac{1}{y-x} \int_x^y \int_{\widehat{\mathbb{R}}} \mathrm{e}^{\mathrm{i}2\pi(x-y)\cdot\eta} \sigma(t,\eta)\, \mathrm{d}\eta\, \mathrm{d}t. \tag{15.3}$$

A_σ is a *Born–Jordan localization operator* if the symbol σ is the characteristic function $\mathbf{1}_E : \mathbb{R} \times \widehat{\mathbb{R}} \to \mathbb{R}$ of a nice enough set $E \subset \mathbb{R} \times \widehat{\mathbb{R}}$. Then the *spreading representation* is

$$\mu : \widehat{\mathbb{R}} \times \mathbb{R} \to \mathbb{C}$$

where

$$\mu(\xi,y) = \operatorname{sinc}(\xi \cdot y)\, F\mathbf{1}_E(\xi,y) = \operatorname{sinc}(\xi\cdot y)\, (\mathscr{F} \otimes \mathscr{F}^{-1})\mathbf{1}_E(\xi,y),$$

and its symplectic inverse Fourier transform $\lambda := F^{-1}\mu = (\mathscr{F}^{-1} \otimes \mathscr{F})\mu$.

A *time-frequency symbol* is a nice enough function $\sigma : \mathbb{R} \times \widehat{\mathbb{R}} \to \mathbb{C}$. Our next task is to design integral operator A_σ such that we obtain “best possible Born–Jordan approximation”

$$Q[A_\sigma v](x,\eta) \approx \sigma(x,\eta)\, Q[v](x,\eta)$$

for “all” signals v at all time-frequencies $(x,\eta) \in \mathbb{R} \times \widehat{\mathbb{R}}$. Operator A_σ is defined by formula

$$\langle u, A_\sigma v \rangle := \langle Q(u,v), \sigma \rangle \tag{15.4}$$

for signals $u, v \in L^2(\mathbb{R})$ (the L^2-inner product on the left for signals in time, the L^2-inner product on the right for functions in time-frequency). Here

$$\begin{aligned}
\langle u, A_\sigma v \rangle &= \langle Q(u,v), \sigma \rangle \\
&= \int_{\widehat{\mathbb{R}}} \int_{\mathbb{R}} Q(u,v)(z,\eta)\, \sigma(z,\eta)^*\, \mathrm{d}z\, \mathrm{d}\eta \\
&= \int_{\widehat{\mathbb{R}}} \int_{\mathbb{R}} \int_{\mathbb{R}} \mathrm{e}^{-\mathrm{i}2\pi w\cdot\eta}\, \frac{1}{w} \int_{z-\frac{w}{2}}^{z+\frac{w}{2}} u(\tilde{x} + \frac{w}{2})\, v(\tilde{x} - \frac{w}{2})^*\, \mathrm{d}\tilde{x}\, \mathrm{d}w\, \sigma(z,\eta)^*\, \mathrm{d}z\, \mathrm{d}\eta \\
&= \int_{\mathbb{R}} u(x) \left[\int_{\widehat{\mathbb{R}}} \int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}2\pi(x-y)\cdot\eta}\, v(y)\, \frac{1}{y-x} \int_x^y \sigma(z,\eta)\, \mathrm{d}z\, \mathrm{d}y\, \mathrm{d}\eta \right]^* \mathrm{d}x.
\end{aligned}$$

Hence

$$A_\sigma v(x) = \int_{\widehat{\mathbb{R}}} \int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}2\pi(x-y)\cdot\eta} u(y)\, a(x,y,\eta)\, \mathrm{d}y\, \mathrm{d}\eta, \tag{15.5}$$

where *amplitude* $a : \mathbb{R} \times \widehat{\mathbb{R}} \to \mathbb{C}$ is defined by $a(x,x,\eta) := \sigma(x,\eta)$ and for case $x \neq y$ as follows:

$$a(x,y,\eta) = \frac{1}{y-x} \int_x^y \sigma(z,\eta)\, \mathrm{d}z. \tag{15.6}$$

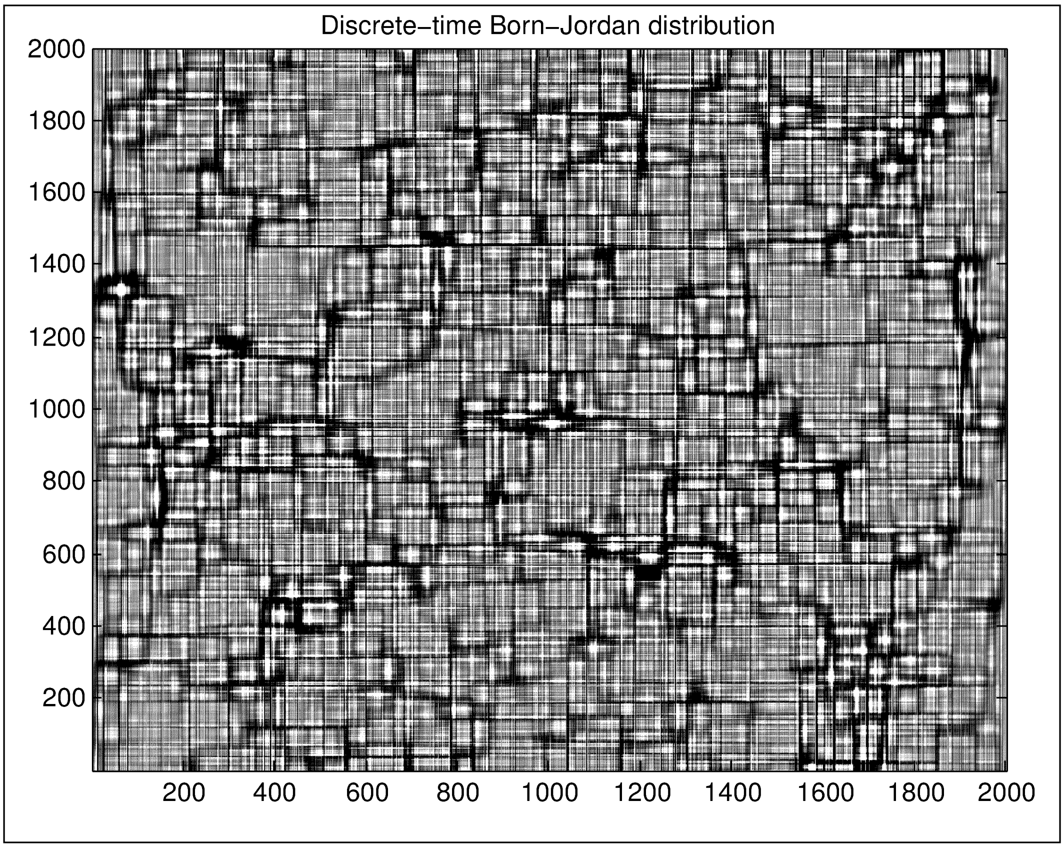
We obtain

$$A_\sigma u(x) = \int_{\mathbb{R}} K_{A_\sigma}(x,y)\, u(y)\, \mathrm{d}y,$$

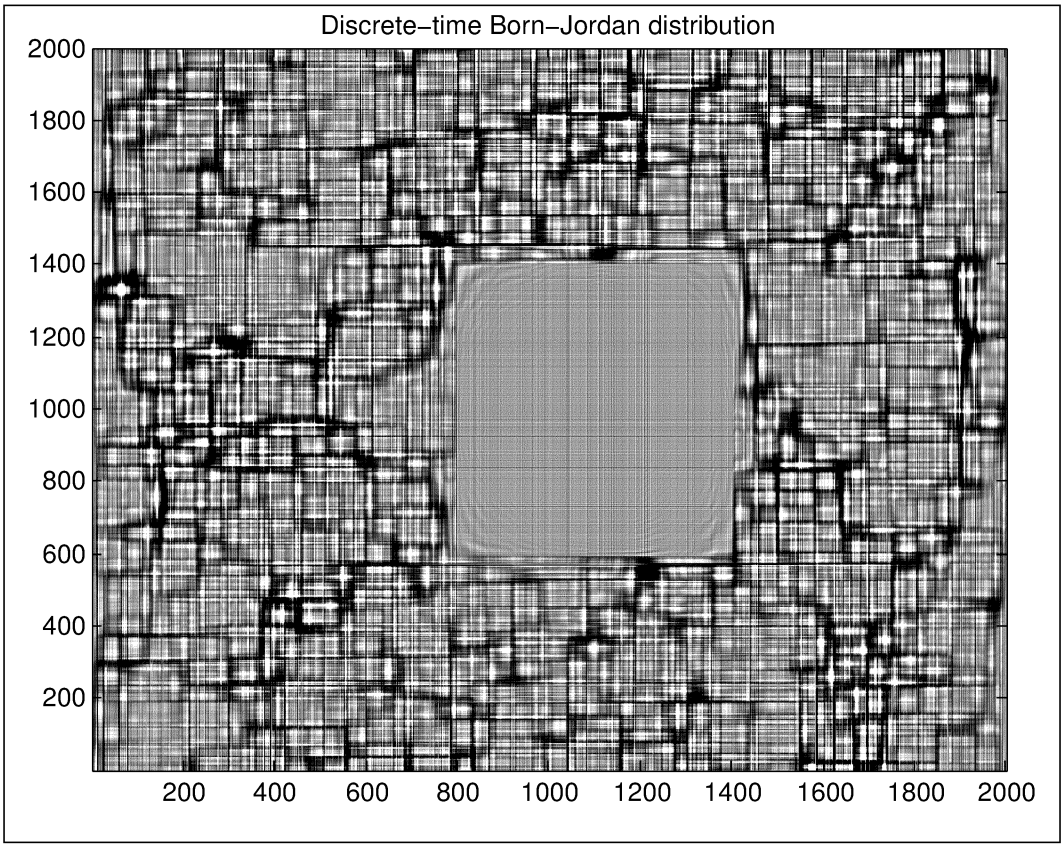
where nice enough integral operator A_σ has kernel function $K_{A\sigma} : \mathbb{R} \times \widehat{\mathbb{R}} \to \mathbb{C}$ satisfying $K_{A_\sigma}(x,x) = \int_{\widehat{\mathbb{R}}} \sigma(x,\eta)\, \mathrm{d}\eta$, and in case of $x \neq y$:

$$K_{A_\sigma}(x,y) = \frac{1}{y-x} \int_x^y \int_{\widehat{\mathbb{R}}} \mathrm{e}^{\mathrm{i}2\pi(x-y)\cdot\eta}\, \sigma(z,\eta)\, \mathrm{d}\eta\, \mathrm{d}z. \tag{15.7}$$

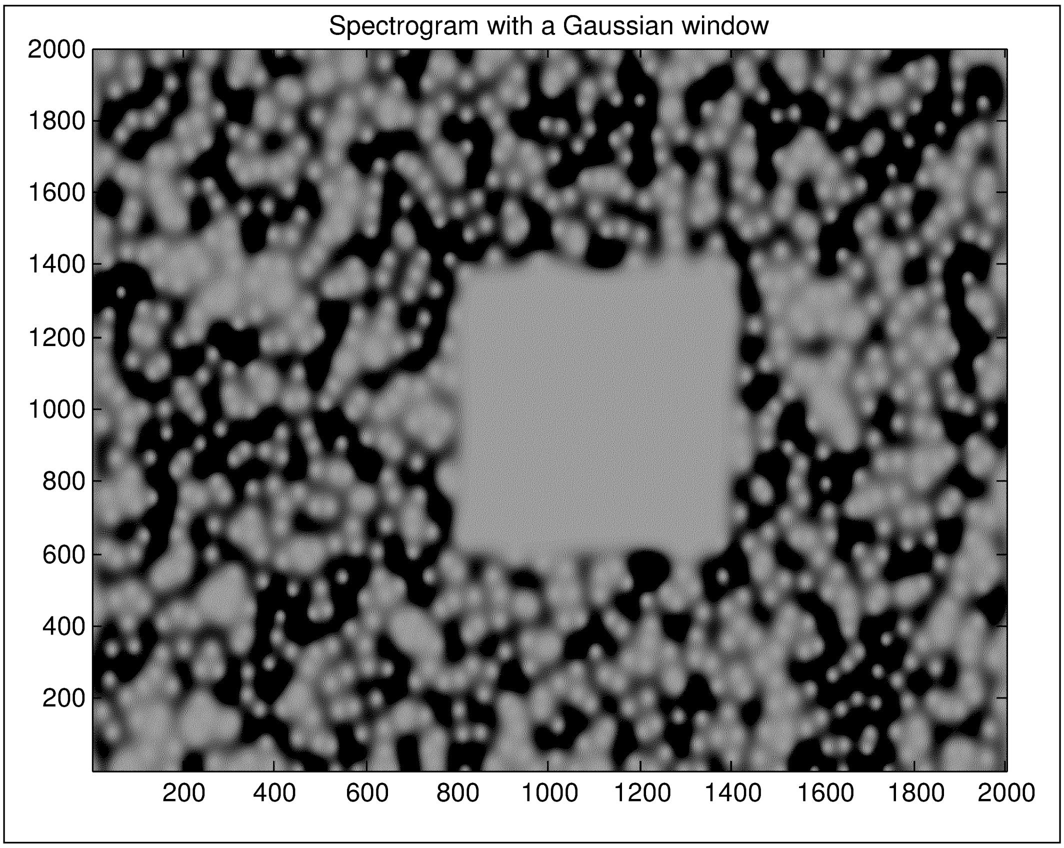
Example 15.1. The following picture shows the discrete-time Born–Jordan distribution $Q[u]$ for the noise u generated by the `rand` command in `Matlab`:



Now, let us choose the time-frequency symbol $\sigma : \mathbb{R} \times \widehat{\mathbb{R}} \to \mathbb{C}$ so that $\sigma(x, \eta) = 0$ inside a certain time-frequency rectangle, in whose complement $\sigma(x, \eta) = 1$. Then the filtered signal $A_\sigma u$ has the following Born–Jordan distribution $Q[A_\sigma u]$:



The chosen original rectangle is clearly visible in the picture above. For comparison, here is a Gaussian spectrogram corresponding to the previous Born–Jordan distribution:



Example 15.2. Let us assume the time-invariance $\sigma(x,\eta) = \widehat{g}(\eta)$ for all $(x,\eta) \in \mathbb{R} \times \widehat{\mathbb{R}}$. Of course, then $A_\sigma u = g * u$, because

$$\begin{aligned} A_\sigma u(x) &= \int_{\widehat{\mathbb{R}}} \int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}2\pi(x-y)\cdot\eta}\, u(x)\, \frac{1}{y-x} \int_x^y \widehat{g}(\eta)\, \mathrm{d}z\, \mathrm{d}y\, \mathrm{d}\eta \\ &= \int_{\widehat{\mathbb{R}}} \int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}2\pi(x-y)\cdot\eta}\, u(y)\, \widehat{g}(\eta)\, \mathrm{d}y\, \mathrm{d}\eta \\ &= \int_{\widehat{\mathbb{R}}} \mathrm{e}^{\mathrm{i}2\pi x\cdot\eta}\, \widehat{u}(\eta)\, \widehat{g}(\eta)\, \mathrm{d}\eta \ = \ g * u(x). \end{aligned}$$

Example 15.3. Let us assume the frequency-invariance $\sigma(x,\eta) = f(x)$ for all $(x,\eta) \in \mathbb{R} \times \widehat{\mathbb{R}}$. Then the amplitude $a(x,y,\eta) = b(x,y)$, so that $A_\sigma u = f\,u$, because

$$\begin{aligned} A_\sigma u(x) &= \int_{\widehat{\mathbb{R}}} \int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}2\pi(x-y)\cdot\eta}\, u(y)\, b(x,y)\, \mathrm{d}y\, \mathrm{d}\eta \\ &= u(x)\, b(x,x) \ = \ f(x)\, u(x). \end{aligned}$$

Example 15.4. It is possible to compute fast in the special case $\sigma = f \otimes \widehat{g}$: Suppose $\sigma(t,\eta) = f(t)\, \widehat{g}(\eta)$ is “nice enough”. Let

$$\Phi(x) := \int_p^x f(t)\, \mathrm{d}t$$

for some $p \in \mathbb{R}$, and $v(z) := -g(z)/z$ (for $z \neq 0$). Then

$$\begin{aligned} A_\sigma u(x) &= \int_{\widehat{\mathbb{R}}} \int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}2\pi(x-y)\cdot\eta}\, u(y)\, \frac{1}{y-x} \int_x^y \sigma(t,\eta)\,\mathrm{d}t\,\mathrm{d}y\,\mathrm{d}\eta \\ &= \int_{\mathbb{R}} \left[\frac{g(x-y)}{y-x} \int_x^y f(t)\,\mathrm{d}t \right] u(y)\,\mathrm{d}y \\ &= v * (\Phi u)(x) - \Phi(x)(v * u)(x) \\ &= [C_v, M_\Phi] u(x), \end{aligned}$$

where

$$C_v u = v * u, \quad M_\Phi u(x) = \Phi(x) u(x).$$

In discretized computations, multiplications are naturally fast, but so are the convolutions thanks to the fast Fourier transform. Thereby computing $A_\sigma u$ is fast in this case.

Non-injectivity. Let $0 \neq g \in \mathscr{S}(\mathbb{R})$ vanish in a neighborhood of the origin: $g(z) = 0$ when $z \approx 0$. Define the non-zero symbol $\sigma : \mathbb{R} \times \widehat{\mathbb{R}} \to \mathbb{C}$ by

$$\sigma(t,\eta) := \int_{\mathbb{R}} \mathrm{e}^{-\mathrm{i}2\pi z\cdot\eta}\, \mathrm{e}^{\mathrm{i}2\pi t/z}\, g(z)\,\mathrm{d}z.$$

Then the Schwartz kernel K_A of the Born–Jordan operator $A = A_\sigma$ vanishes nearby the (x,y)-diagonal, because

$$\begin{aligned} K_A(x,y) &= \frac{1}{y-x} \int_x^y \int_{\widehat{\mathbb{R}}} \mathrm{e}^{\mathrm{i}2\pi(x-y)\cdot\eta}\, \sigma(t,\eta)\,\mathrm{d}\eta\,\mathrm{d}t \\ &= \frac{1}{y-x} \int_x^y \int_{\widehat{\mathbb{R}}} \int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}2\pi(x-y)\cdot\eta}\, \mathrm{e}^{-\mathrm{i}2\pi z\cdot\eta}\, \mathrm{e}^{\mathrm{i}2\pi t/z}\, g(z)\,\mathrm{d}z\,\mathrm{d}\eta\,\mathrm{d}t \\ &= \frac{1}{y-x} \int_x^y \mathrm{e}^{\mathrm{i}2\pi t/(x-y)}\,\mathrm{d}t\, g(x-y), \end{aligned}$$

where $g(x-y) = 0$ when $x \approx y$. If $x \neq y$, we have $K_A(x,y) = 0$, because

$$\frac{1}{y-x} \int_x^y \mathrm{e}^{\mathrm{i}2\pi t/(x-y)}\,\mathrm{d}t = \frac{\mathrm{e}^{\mathrm{i}2\pi x/(x-y)} - \mathrm{e}^{\mathrm{i}2\pi y/(x-y)}}{\mathrm{i}2\pi} = \mathrm{e}^{\mathrm{i}2\pi x/(x-y)}\, \frac{1-\mathrm{e}^{-\mathrm{i}2\pi}}{\mathrm{i}2\pi} = 0.$$

So here $A_\sigma = 0$: thereby the Born–Jordan quantization is not injective.

Born–Jordan singular integral operators. For Born–Jordan operator A_σ, the Schwartz distribution kernel $K = K_{A_\sigma}$ is given by

$$K(x,y) = \frac{1}{y-x} \int_x^y \int_{\widehat{\mathbb{R}}} \mathrm{e}^{\mathrm{i}2\pi(x-y)\cdot\eta}\, \sigma(t,\eta)\,\mathrm{d}\eta\,\mathrm{d}t.$$

Here operator A_σ is symmetric if σ real-valued. Operator A_σ preserves real-valued signals if

$$\forall t, \eta : \ \sigma(t,\eta)^* = \sigma(t,-\eta).$$

If σ is smooth and compactly supported then K is smooth and rapidly decaying away from the $x = y$ diagonal:

$$(y-x)^\alpha \partial_x^\beta \partial_y^\gamma K(x,y) = \partial_x^\beta \partial_y^\gamma \int_0^1 \int_{\widehat{\mathbb{R}}} \mathrm{e}^{\mathrm{i}2\pi(x-y)\cdot\eta} \, (\mathrm{i}2\pi\partial_\eta)^\alpha \sigma(\tau x + (1-\tau)y, \eta) \, \mathrm{d}\eta \, \mathrm{d}\tau.$$

Actually K is smooth and rapidly decaying away from the $x = y$ diagonal if $\sigma \in S^m_{\rho,\delta}(\mathbb{R} \times \widehat{\mathbb{R}})$, where $\rho > 0$.

Boundedness properties. Let us consider L^2-boundedness issues for the Born–Jordan quantization $\sigma \mapsto A_\sigma$. Let $\|A_\sigma\|$ denote the L^2-operator norm

$$\|A_\sigma\|_{L^2 \to L^2} = \sup \left\{ \|A_\sigma u\| : \ u \in L^2(\mathbb{R}), \ \|u\| \le 1 \right\}.$$

Recall that in the Born–Jordan quantization, we have

$$A_\sigma u(x) = \int_{\widehat{\mathbb{R}}} \int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}2\pi(x-y)\cdot\eta} \, u(y) \, \frac{1}{y-x} \int_x^y \sigma(t,\eta) \, \mathrm{d}t \, \mathrm{d}y \, \mathrm{d}\eta.$$

Smoothness of σ not necessary for boundedness $A_\sigma : L^2(\mathbb{R}) \to L^2(\mathbb{R})$: For instance, if $\sigma(x,\eta) = f(x)\,\widehat{g}(\eta)$ where f is continuous bounded and $g \ge 0$ integrable then

$$\|A_\sigma\| \le \|\sigma\|_{L^\infty} :$$

here

$$\begin{aligned}
\|A_\sigma u\|^2 &= \int \left| \int g(x-y) \, \frac{1}{y-x} \int_x^y f(t) \, \mathrm{d}t \, u(y) \, \mathrm{d}y \right|^2 \mathrm{d}x \\
&\le \|f\|_{L^\infty}^2 \int \left[\int g(x-y) \, |u(y)| \, \mathrm{d}y \right]^2 \mathrm{d}x \\
&\le \|f\|_{L^\infty}^2 \, \|\widehat{g}\|_{L^\infty}^2 \, \|u\|_{L^2}^2 \\
&= \|\sigma\|_{L^\infty}^2 \, \|u\|^2.
\end{aligned}$$

In general, as $A_\sigma = A_{\mathrm{Re}(\sigma)} + \mathrm{i}A_{\mathrm{Im}(\sigma)}$ and $(A_\sigma)^* = A_{\sigma^*}$, it suffices to study boundedness properties in the case of symmetric operators A_σ (i.e. real-valued symbols σ). For symmetric A_σ,

$$\begin{aligned}
\|A_\sigma\| &= \sup_{u \in L^2 : \ \|u\| \le 1} |\langle u, A_\sigma u \rangle| \\
&= \sup_{u \in L^2 : \ \|u\| \le 1} |\langle Q[u], \sigma \rangle| \\
&\le \sup_{u \in L^2 : \ \|u\| \le 1} \|Q[u]\|_X \, \|\sigma\|_{X'},
\end{aligned}$$

where X' is the dual of a Banach space X. For instance, $X = L^2(\mathbb{R})$ gives boundedness, if $\sigma \in L^2(\mathbb{R} \times \widehat{\mathbb{R}})$; in this case, actually A_σ is even a Hilbert–Schmidt operator with $\|A_\sigma\|_{HS} \leq \|\sigma\|_{L^2}$: especially, compactly supported essentially bounded symbols σ give rise to Hilbert–Schmidt operators A_σ on $L^2(\mathbb{R})$. An open problem is whether $X = L^1(\mathbb{R})$ would do: does $\|Q[u]\|_{L^1} \leq C\,\|u\|^2$? Then namely $\|A_\sigma\| \leq C\,\|\sigma\|_{L^\infty}$ would hold.

§ 16. Comparing Born–Jordan to Wigner

How do Born–Jordan and Wigner time-frequency transforms differ? Remember the characterization of the Born–Jordan energy density $u \mapsto Q[u]$:

(1) Q is **scale invariant**.
(2) Q is **time-local**.
(3) Q maps "**comb-to-grid**":

$$Q[\delta_{\mathbb{Z}}] = \delta_{\mathbb{Z}} \otimes \mathbf{1} + \mathbf{1} \otimes \delta_{\mathbb{Z}} - \mathbf{1} \otimes \mathbf{1}.$$

What are the corresponding properties for the Wigner energy density $u \mapsto W[u]$?

(1') W is **scale invariant**.
(2') W is **time-local**.
(3') $W[\delta_{\mathbb{Z}}] = \delta_{\mathbb{Z}} \otimes \delta_{\mathbb{Z}} + \delta_{\mathbb{Z}+1/2} \otimes \delta_{\mathbb{Z}} + \delta_{\mathbb{Z}} \otimes \delta_{\mathbb{Z}+1/2} - \delta_{\mathbb{Z}+1/2} \otimes \delta_{\mathbb{Z}+1/2}$

$$= \delta_{\mathbb{Z}/2} \otimes \delta_{\mathbb{Z}/2} - \delta_{\mathbb{Z}+1/2} \otimes \delta_{\mathbb{Z}+1/2}$$

Apparently, conditions (3) and (3′) are quite different. It is important to notice that spectrograms have none of these properties (no scale invariance, no time-locality, no comb-to-grid, ...).

Let $e_\omega(x) = \mathrm{e}^{\mathrm{i}2\pi x\cdot\omega}$. For $\alpha < \beta$,

$$\begin{aligned} W[\lambda e_\alpha + \mu e_\beta](x,\eta) &= |\lambda|^2 \delta_\alpha(\eta) + |\mu|^2 \delta_\beta(\eta) + 2\,\mathrm{Re}\,(\lambda\,\mu^*\, e_{\alpha-\beta}(x))\ \delta_0(\eta - \frac{\alpha+\beta}{2}), \\ Q[\lambda e_\alpha + \mu e_\beta](x,\eta) &= |\lambda|^2 \delta_\alpha(\eta) + |\mu|^2 \delta_\beta(\eta) + 2\,\mathrm{Re}\,(\lambda\,\mu^*\, e_{\alpha-\beta}(x))\ \mathbf{1}_{[\alpha,\beta]}(\eta)/(\beta-\alpha). \end{aligned}$$

The interference terms here hint that there is basically no hope for the Wigner distribution to cancel interferences, whereas the Born–Jordan distribution uniformly smooths out the interference to a low-amplitude oscillation in a wide time-frequency strip.

Let $u(x) = \mathrm{e}^{\mathrm{i}2\pi\omega x^2/2}$. That is, $u : \mathbb{R} \to \mathbb{C}$ is a linear chirp signal having instantaneous frequency ωx at time $x \in \mathbb{R}$. Then

$$\begin{aligned} W[u](x,\eta) &= \delta_0(\eta - \omega x), \\ Q[u](x,\eta) &= \widehat{v}(\eta - \omega x), \end{aligned}$$

where $v(y) = \operatorname{sinc}(\omega y^2)$, i.e. for $y \neq 0$

$$v(y) = \frac{\sin(\pi\omega y^2)}{\pi\omega y^2}.$$

While $Q[u]$ is not a sharp line here, it is still quite well concentrated.

Amplitude $a : \mathbb{R} \times \mathbb{R} \times \widehat{\mathbb{R}} \to \mathbb{C}$ defines operator $\mathrm{Op}(a)$ by

$$\mathrm{Op}(a)u(x) := \int_{\widehat{\mathbb{R}}} \int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}2\pi(x-y)\cdot\eta}\, u(y)\, a(x,y,\eta)\, \mathrm{d}y\, \mathrm{d}\eta.$$

For instance,

$$X^{n-\ell} D^m X^\ell u(x) = \int_{\widehat{\mathbb{R}}} \int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}2\pi(x-y)\cdot\eta}\, u(y) \left(x^{n-\ell}\, \eta^m\, y^\ell\right) \mathrm{d}y\, \mathrm{d}\eta,$$

where “Heisenberg commutator” $[D, X] = (\mathrm{i}2\pi)^{-1}$,

$$Xu(x) := x\, u(x),\ Du(x) := (\mathrm{i}2\pi)^{-1}\, u'(x).$$

From symbol $\sigma : \mathbb{R} \times \widehat{\mathbb{R}} \to \mathbb{C}$, we get (Weyl–)Wigner amplitude

$$a = a_{WW}(x,y,\eta) = \sigma(\frac{x+y}{2}, \eta)$$

and Born–Jordan amplitude

$$a = a_{BJ}(x,y,\eta) = \frac{1}{y-x} \int_x^y \sigma(t,\eta)\, \mathrm{d}t.$$

In reasonable quantizations, polynomial symbols should correspond to differential operators; especially, symbol $(t,\eta) \mapsto t^n$ should correspond to the multiplication operator $u \mapsto X^n u$ for $X^n u(x) := x^n\, u(x)$, and symbol $(t,\eta) \mapsto \eta^m$ should correspond to the differential operator $u \mapsto D^m u$ where $D = (\mathrm{i}2\pi)^{-1}\partial$ for derivative $\partial u = u'$.

Let $\sigma(t,\eta) = t^n \eta^m$ with $m, n \in \{0, 1, 2, 3, \cdots\}$. Then we have amplitudes

$$a_{WW}(x,y,\eta) = \left(\frac{x+y}{2}\right)^n \eta^m = \eta^m \frac{1}{2^n} \sum_{\ell=0}^{n} \binom{n}{\ell} x^{n-\ell}\, y^\ell,$$

$$a_{BJ}(x,y,\eta) = \left(\frac{1}{y-x} \int_x^y t^n\, \mathrm{d}t\right) \eta^m = \eta^m \frac{1}{n+1} \sum_{\ell=0}^{n} x^{n-\ell}\, y^\ell$$

— roughly speaking, this states that

$$xy \not\sim yx \text{ for Weyl} - \text{Wigner},$$

$$xy \sim yx \text{ for Born} - \text{Jordan}.$$

Let us show that

$$\mathrm{Op}(a_{WW}) \neq \mathrm{Op}(a_{BJ}) \Longleftrightarrow m, n \geq 2.$$

Write $\partial := \partial/\partial x$. Let $m, n \in \mathbb{N}_0$. Then to the symbol $(t,\eta) \mapsto t^n \eta^m$ corresponds the operator

$$\begin{aligned}\sum_{\ell=0}^{n} c_{nm\ell} X^{n-\ell} \partial^m X^\ell &= \sum_{\ell=0}^{n} \sum_{k=0}^{m} c_{nm\ell} \binom{m}{k} \ell^{(k)} X^{n-k} \partial^{m-k} \\ &= \sum_{k=0}^{m} \left(c_{nm\ell}\, \ell^{(k)} \right) \binom{m}{k} X^{n-k} \partial^{m-k}.\end{aligned}$$

The data

$$\left\{ \sum_{\ell=0}^{n} c_{nm\ell}\, \ell^{(k)} :\ 0 \leq k \leq n \right\}$$

uniquely determines coefficients $c_{nm\ell} \in \mathbb{C}$.

Here, $c_{nm\ell} = 1/(n+1)$ for the Born–Jordan quantization, and $c_{nm\ell} = 2^{-n}\binom{n}{\ell}$ for the Weyl quantization. Let us check when the Born–Jordan and Weyl quantizations give different operators for the same polynomial symbol $(t,\eta) \mapsto t^n \eta^m$. Notice that for Born–Jordan

$$\frac{1}{n+1} \sum_{\ell=0}^{n} \ell^{(k)} = \frac{1}{n+1} \sum_{\ell=0}^{n} \frac{1}{k+1} \triangle_\ell \ell^{(k+1)} = \frac{1}{n+1} \frac{1}{k+1} (n+1)^{(k+1)} = \frac{1}{k+1} n^{(k)},$$

and for Weyl

$$\begin{aligned}\frac{1}{2^n} \sum_{\ell=0}^{n} \binom{n}{\ell} \ell^{(k)} &= \frac{1}{2^n} \sum_{\ell=0}^{n} \binom{n}{\ell} \left(\frac{\mathrm{d}}{\mathrm{d}x} \right)^k x^\ell \Bigg|_{x=1} \\ &= \frac{1}{2^n} \left(\frac{\mathrm{d}}{\mathrm{d}x} \right)^k (x+1)^n \Bigg|_{x=1} = \frac{1}{2^n} n^{(k)} (x+1)^{n-k} \Bigg|_{x=1} \\ &= 2^{-k} n^{(k)}.\end{aligned}$$

Here, $k+1 = 2^k$ if and only if $k \in \{0,1\}$ (clearly, if $k \geq 2$ then $2^k = \sum_{j=0}^{k-1} 2^j + 1 > k+1$).

§ 17. Discretizing Born–Jordan transform

From the continuous time $x \in \mathbb{R}$, we can move to numerical computation by sampling and periodizing (in whichever order):

$$\begin{array}{ccc} \mathbb{R} & \xrightarrow{periodize} & \mathbb{R}/\mathbb{Z} \\ sample \downarrow & & \downarrow sample \\ \mathbb{Z} & \xrightarrow{periodize} & \mathbb{Z}/N\mathbb{Z}. \end{array}$$

At each intermediate level, we have naturally related familiar Fourier transforms: Fourier integral on $\mathbb{R}$, Fourier series on $\mathbb{R}/\mathbb{Z}$, Fourier coefficients on $\mathbb{Z}$, DFT (FFT) on $\mathbb{Z}/N\mathbb{Z}$.

Notation for discrete-time Fourier analysis. Let us denote d-dimensional torus by $\mathbb{T}^d := \mathbb{R}^d/\mathbb{Z}^d$. Finitely supported functions $u : \mathbb{Z}^d \to \mathbb{C}$ have Fourier transform $\mathcal{F}u = \widehat{u} : \mathbb{T}^d \to \mathbb{C}$ given by

$$\widehat{u}(\eta) := \sum_{y\in\mathbb{Z}^d} \mathrm{e}^{-\mathrm{i}2\pi y\cdot\eta}\, u(y). \tag{17.1}$$

Here, $\widehat{u}$ is a trigonometric polynomial. The corresponding inverse Fourier transform $\mathcal{F}^{-1}$ is given by

$$u(y) = \int_{\mathbb{T}^d} \mathrm{e}^{+\mathrm{i}2\pi y\cdot\eta}\, \widehat{u}(\eta)\,\mathrm{d}\eta. \tag{17.2}$$

As it is well-known, Fourier transform extends here to an isomorphism

$$\mathcal{F} : \mathscr{S}(\mathbb{Z}^d) \to C^\infty(\mathbb{T}^d),$$

where $\mathscr{S}(\mathbb{Z}^d)$ denotes the space of those functions $u : \mathbb{Z}^d \to \mathbb{C}$ for which $|u(y)| \to 0$ rapidly when $|y| \to \infty$, and $C^\infty(\mathbb{T}^d)$ denotes the space of infinitely smooth complex-valued functions on the torus. Also, Fourier transform extends to an isomorphism

$$\mathcal{F} : \ell^2(\mathbb{Z}^d) \to L^2(\mathbb{T}^d),$$

where Hilbert space $\ell^2(\mathbb{Z}^d)$ has inner product given by

$$\langle u, v\rangle_{\ell^2(\mathbb{Z}^d)} = \sum_{y\in\mathbb{Z}^d} u(y)\, v(y)^*, \tag{17.3}$$

and Hilbert space $L^2(\mathbb{T}^d)$ has inner product given by

$$\langle U, V\rangle_{L^2(\mathbb{T}^d)} = \int_{\mathbb{T}^d} U(\eta)\, V(\eta)^*\,\mathrm{d}\eta \tag{17.4}$$

In the sequel, to keep notation simple, we shall concentrate on the one-dimensional case $d = 1$.

Discrete-time Born–Jordan transform. For functions $u, v : \mathbb{Z} \to \mathbb{C}$, define *discrete-time Born–Jordan cross-correlation* $R(u, v) : \mathbb{Z}\times\mathbb{Z} \to \mathbb{C}$ by

$$R(u, v)(x, 0) := u(x)\, v(x)^*, \tag{17.5}$$

$$R(u, v)(x, -y) := R(v, u)(x, +y)^*, \tag{17.6}$$

where for $y > 0$ we set

$$R(u,v)(x,y) := \frac{1}{y}\sum_{t=0}^{y-1} u(x+t)\, v(x+t-y)^*. \tag{17.7}$$

Here, to achieve more symmetric expression, we could have chosen

$$\frac{u(x+y)v(x)^* + u(x)v(x-y)^*}{2y} + \frac{1}{y}\sum_{t=1}^{y-1} u(x+t)\, v(x+t-y)^*;$$

however, there is no practical reason to use such complicated expression. Let us call $R[u] := R(u,u)$ the *discrete-time Born–Jordan autocorrelation* of u. For $u, v \in \mathscr{S}(\mathbb{Z})$, let $Q(u,v) : \mathbb{Z} \times \mathbb{T} \to \mathbb{C}$ be defined by

$$Q(u,v) := (I \otimes \mathcal{F}) R(u,v),$$

where I is the identity operator $u \mapsto u$. In other words,

$$Q(u,v)(x,\eta) = \sum_{y\in\mathbb{Z}} \mathrm{e}^{-\mathrm{i}2\pi y\cdot\eta}\, R(u,v)(x,y). \tag{17.8}$$

We call the conjugate bilinear mapping

$$(u,v) \mapsto Q(u,v)$$

the *discrete-time Born–Jordan transform*, and $Q[u] := Q(u,u)$ is called the *discrete-time Born–Jordan distribution* of $u \in \mathscr{S}(\mathbb{Z})$. Notice that $Q(v,u) = Q(u,v)^*$, so that $Q[u]$ is automatically real-valued:

$$Q[u] : \mathbb{Z} \times \mathbb{T} \to \mathbb{R}.$$

For $Q[u](x,\eta)$, variable $x \in \mathbb{Z}$ will be called *time* and variable $\eta \in \mathbb{T}$ will be called *frequency*. Notice that if $u(x) = 0$ for $x < a$ and $x > b$ then $Q[u](x,\eta) = 0$ for $x < a$ and $x > b$.

Theorem. *If $u, v \in \mathscr{S}(\mathbb{Z})$ then $R(u,v) \in \mathscr{S}(\mathbb{Z}\times\mathbb{Z})$ and $Q(u,v) \in C^\infty(\mathbb{Z}\times\mathbb{T})$. If $u, v \in \ell^2(\mathbb{Z})$ then $R(u,v) \in \ell^2(\mathbb{Z}\times\mathbb{Z})$ and $Q(u,v) \in L^2(\mathbb{Z}\times\mathbb{T})$, with*

$$\|R(u,v)\|_{\ell^2(\mathbb{Z}\times\mathbb{Z})} \le \|u\|_{\ell^2(\mathbb{Z})}\, \|v\|_{\ell^2(\mathbb{Z})}, \tag{17.9}$$

$$\|Q(u,v)\|_{L^2(\mathbb{Z}\times\mathbb{T})} \le \|u\|_{\ell^2(\mathbb{Z})}\, \|v\|_{\ell^2(\mathbb{Z})}. \tag{17.10}$$

Proof. Let $u, v \in \mathscr{S}(\mathbb{Z})$. Then for each $M > 0$ there exist constants $c_M < \infty$ such that $|u(x)|, |v(x)| \leq c_M \langle x \rangle^{-M}$ for all $x \in \mathbb{Z}$, where $\langle 0 \rangle := 1$ and $\langle x \rangle = |x|$ for $x \neq 0$. Therefore

$$\begin{aligned} |R(u,v)(x,0)| = |u(x)|\,|v(x)| &\leq c_M^2 \langle x \rangle^{-2M}, \\ |R(u,v)(x,-y)| = |R(v,u)(x,y)|, \end{aligned}$$

and if $y > 0$ then

$$|R(u,v)(x,y)| \leq \frac{1}{y} \sum_{t=0}^{y-1} |u(x+t)|\,|v(x+t-y)| \leq \frac{1}{y} \sum_{t=0}^{y-1} c_M^2 \langle x+t \rangle^{-M} \langle x+t-y \rangle^{-M}.$$

From this we obtain

$$|R(u,v)(x,y)| \leq c_M^2 \langle x \rangle^{-M/2} \langle y \rangle^{-M/2},$$

showing $R(u,v) \in \mathscr{S}(\mathbb{Z} \times \mathbb{Z})$. Next,

$$\begin{aligned} \|R(u,v)\|^2_{\ell^2(\mathbb{Z}\times\mathbb{Z})} &= \sum_{x\in\mathbb{Z}} \sum_{y\in\mathbb{Z}} |R(u,v)(x,y)|^2 \\ &= \sum_{x\in\mathbb{Z}} |u(x)|^2\,|v(x)|^2 + \sum_{x\in\mathbb{Z}} \sum_{y=1}^{\infty} \frac{1}{y^2} \left| \sum_{t=0}^{y-1} u(x+t)\,v(x+t-y)^* \right|^2 + \\ &\quad + \sum_{x\in\mathbb{Z}} \sum_{y=1}^{\infty} \frac{1}{y^2} \left| \sum_{t=0}^{y-1} v(x+t)\,u(x+t-y)^* \right|^2 \\ &\leq \sum_{x\in\mathbb{Z}} \sum_{y\in\mathbb{Z}} |u(x)|^2\,|v(y)|^2 \\ &= \|u\|^2_{\ell^2(\mathbb{Z})}\,\|v\|^2_{\ell^2(\mathbb{Z})}. \end{aligned}$$

Fourier transform is an isometric isomorphism $\ell^2(\mathbb{Z}) \to L^2(\mathbb{T})$, yielding

$$\|Q(u,v)\|_{L^2(\mathbb{Z}\times\mathbb{T})} = \|R(u,v)\|_{\ell^2(\mathbb{Z}\times\mathbb{Z})},$$

proving the last formula in theorem. □

Examples. Let $\delta_p : \mathbb{Z} \to \mathbb{C}$ be the Kronecker delta at $p \in \mathbb{Z}$, i.e. $\delta_p(p) = 1$ and $\delta_p(x) = 0$ for $x \neq p$. Then $R(\delta_p, \delta_p)(x,y) = \delta_p(x)\delta_0(y)$, leading to

$$Q[\delta_p](x,\eta) = \delta_p(x).$$

Let $e_\alpha : \mathbb{Z} \to \mathbb{C}$ be defined by $e_\alpha(x) := \mathrm{e}^{\mathrm{i}2\pi x\cdot\alpha}$, where $\alpha \in \mathbb{T}$. Then $R(e_\alpha, e_\alpha)(x,y) = \mathrm{e}^{\mathrm{i}2\pi y\cdot\alpha}$, leading to

$$Q[e_\alpha](x,\eta) = \delta_\alpha(\eta),$$

where $\delta_\alpha \in \mathcal{D}(\mathbb{T})$ is the Dirac delta distribution at $\alpha \in \mathbb{T}$.

Theorem. *Born–Jordan has the marginal properties*

$$\int_{\mathbb{T}} Q[u](x,\eta)\,\mathrm{d}\eta = |u(x)|^2, \tag{17.11}$$

$$\sum_{x\in\mathbb{Z}} Q[u](x,\eta) = |\widehat{u}(\eta)|^2. \tag{17.12}$$

Proof. First,

$$\int_{\mathbb{T}} Q[u](x,\eta)\,\mathrm{d}\eta = \int_{\mathbb{T}} \sum_{y\in\mathbb{Z}} \mathrm{e}^{-\mathrm{i}2\pi y\cdot\eta}\, R[u](x,y)\,\mathrm{d}\eta = R[u](x,0) = |u(x)|^2,$$

and second,

$$\begin{aligned}
\sum_{x\in\mathbb{Z}} Q[u](x,\eta) &= \sum_{x\in\mathbb{Z}} \sum_{y\in\mathbb{Z}} \mathrm{e}^{-\mathrm{i}2\pi y\cdot\eta}\, R[u](x,y) \\
&= \sum_{y\in\mathbb{Z}} \mathrm{e}^{-\mathrm{i}2\pi y\cdot\eta} \sum_{x\in\mathbb{Z}} R[u](x,y) \\
&= \sum_{y\in\mathbb{Z}} \mathrm{e}^{-\mathrm{i}2\pi y\cdot\eta} \sum_{z\in\mathbb{Z}} u(z)\, u(z-y)^* \\
&= \sum_{z\in\mathbb{Z}} \mathrm{e}^{-\mathrm{i}2\pi z\cdot\eta}\, u(z) \sum_{y\in\mathbb{Z}} \mathrm{e}^{\mathrm{i}2\pi (z-y)\cdot\eta}\, u(z-y)^* \\
&= |\widehat{u}(\eta)|^2.
\end{aligned}$$

□

Interference terms. For $\lambda,\mu\in\mathbb{C}$ and $u,v\in\mathscr{S}(\mathbb{Z})$, notice the quadratic behavior

$$\begin{aligned}
Q[\lambda u+\mu v] &= Q(\lambda u+\mu v, \lambda u+\mu v) \\
&= Q(\lambda u,\lambda u) + Q(\mu v,\mu v) + Q(\lambda u,\mu v) + Q(\mu v,\lambda u) \\
&= |\lambda|^2 Q[u] + |\mu|^2 Q[v] + 2\,\mathrm{Re}\left(\lambda\,\mu^*\, Q(u,v)\right).
\end{aligned}$$

In signal processing, terms of type $Q[u], Q[v]$ are sometimes called auto-terms, and terms of type $Q(u,v), Q(v,u)$ are interference terms (or ghost terms); both of these types are important in understanding time-frequency properties of signals. For instance, if $p<q$ then

$$R(\delta_p,\delta_q)(x,y) = \begin{cases} (q-p)^{-1} & \text{when } p<x\leq q \text{ and } y=p-q, \\ 0 & \text{otherwise,} \end{cases}$$

leading to

$$Q(\delta_p,\delta_q)(x,\eta) = \begin{cases} (q-p)^{-1}\,\mathrm{e}^{\mathrm{i}2\pi(q-p)\cdot\eta} & \text{when } p<x\leq q, \\ 0 & \text{otherwise,} \end{cases}$$

Hence for $p < q$, we have

$$Q[\lambda\delta_p + \mu\delta_q](x,\eta) = |\lambda|^2\delta_p(x) + |\mu|^2\delta_q(x) + ghost_1(x,\eta), \tag{17.13}$$

where $ghost_1(x,\eta) \neq 0$ only if $p < x \leq q$, and when $\lambda, \mu \in \mathbb{R}$ we have

$$ghost_1(x,\eta) = 2\lambda\mu\frac{\cos(2\pi(q-p)\cdot\eta)}{q-p}. \tag{17.14}$$

Discrete-time Born–Jordan quantization. Next we introduce *discrete-time Born–Jordan quantization* $\sigma \mapsto A_\sigma$, where *discrete-time symbol Born–Jordan symbol* σ is function $\sigma : \mathbb{Z} \times \mathbb{T} \to \mathbb{C}$ satisfying some conditions (depending on application), and linear operator A_σ maps functions $g : \mathbb{Z} \to \mathbb{C}$ to functions of similar type. This *discrete-time Born–Jordan pseudo-differential operator* A_σ is defined by the Hilbert space duality

$$\langle u, A_\sigma v\rangle_{\ell^2(\mathbb{Z})} := \langle Q(u,v), \sigma\rangle_{L^2(\mathbb{Z}\times\mathbb{T})}, \tag{17.15}$$

where

$$\langle Q(u,v), \sigma\rangle_{L^2(\mathbb{Z}\times\mathbb{T})} = \sum_{x\in\mathbb{Z}} \int_{\mathbb{T}} Q(u,v)(x,\eta)\,\sigma(x,\eta)^*\,\mathrm{d}\eta.$$

Theorem. *Discrete-time Born–Jordan pseudo-differential operator A_σ satisfies*

$$A_\sigma v(x) = \sum_{y\in\mathbb{Z}} v(y) \int_{\mathbb{T}} \mathrm{e}^{\mathrm{i}2\pi(x-y)\cdot\eta}\, a(x,y,\eta)\,\mathrm{d}\eta,$$

where amplitude $a : \mathbb{Z}\times\mathbb{Z}\times\mathbb{T} \to \mathbb{C}$ is given by

$$\begin{aligned} a(x,x,\eta) &= \sigma(x,\eta), \\ a(y,x,\eta) &= a(x,y,\eta), \end{aligned}$$

and for $x < y$

$$a(x,y,\eta) = \frac{1}{y-x}\sum_{t=x}^{y-1}\sigma(t,\eta).$$

Proof. Straight-forward calculation yields

$$\begin{aligned}
&\langle u, A_\sigma v\rangle_{\ell^2(\mathbb{Z})} \\
&= \langle Q(u,v), \sigma\rangle_{L^2(\mathbb{Z}\times\mathbb{T})} \\
&= \sum_{z\in\mathbb{Z}} \int_{\mathbb{T}} Q(u,v)(z,\eta)\,\sigma(z,\eta)^*\,\mathrm{d}z\,\mathrm{d}\eta \\
&= \sum_{z\in\mathbb{Z}} \int_{\mathbb{T}} \sum_{w\in\mathbb{Z}} \mathrm{e}^{-\mathrm{i}2\pi w\cdot\eta}\, R(u,v)(z,w)\,\sigma(z,\eta)^*\,\mathrm{d}\eta \\
&= \sum_{z\in\mathbb{Z}} \int_{\mathbb{T}} \Bigg(u(z)\,v(z)^* + \sum_{w=1}^{\infty} \Bigg[\mathrm{e}^{-\mathrm{i}2\pi w\cdot\eta}\,\frac{1}{w}\sum_{t=0}^{w-1} u(z+t)\,v(z+t-w)^* + \\
&\qquad + \mathrm{e}^{+\mathrm{i}2\pi w\cdot\eta}\,\frac{1}{w}\sum_{t=0}^{w-1} v(z+t)\,u(z+t-w)^* \Bigg]\Bigg)\,\sigma(z,\eta)^*\,\mathrm{d}\eta \\
&= \sum_{x\in\mathbb{Z}} u(x) \left[\sum_{y\in\mathbb{Z}} \int_{\mathbb{T}} \mathrm{e}^{\mathrm{i}2\pi(x-y)\cdot\eta}\, v(y)\, a(x,y,\eta)\,\mathrm{d}\eta \right]^* .
\end{aligned}$$

Hence A_σ satisfies

$$A_\sigma v(x) = \sum_{y\in\mathbb{Z}} v(y) \int_{\mathbb{T}} \mathrm{e}^{\mathrm{i}2\pi(x-y)\cdot\eta}\, a(x,y,\eta)\,\mathrm{d}\eta,$$

where amplitude $a : \mathbb{Z}\times\mathbb{Z}\times\mathbb{T}\to\mathbb{C}$ is of the claimed form. □

Examples of symbols and operators. Born–Jordan quantization behaves nicely for multiplication and convolution operators: Namely, if $\sigma(t,\eta) = \varphi(t)$ then

$$A_\sigma u(x) = \varphi(x)\,u(x),$$

and if $\sigma(t,\eta) = \widehat{\psi}(\eta)$ then

$$A_\sigma u(x) = \sum_{y\in\mathbb{Z}} \psi(x-y)\,u(y) = \psi * u(x).$$

While symbol-to-operator quantization $\sigma\mapsto A_\sigma$ is clearly linear, it is not injective in the Born–Jordan case. For instance $A_0 = 0$, but it is easy to check that $A_\sigma = 0$ if

$$\sigma(t,\eta) = (-1)^t\,\mathrm{e}^{\mathrm{i}2\pi 2\cdot\eta}.$$

Kronecker delta functions $\delta_p : \mathbb{Z}\to\mathbb{C}$ provide an orthonormal basis for $\ell^2(\mathbb{Z})$; if linear operator $A : \ell^2(\mathbb{Z})\to\ell^2(\mathbb{Z})$ maps δ_{p_0} to δ_{q_0}, and $A(\delta_p) = 0$ when $p\neq p_0$, then $A = A_\sigma$ for

$$\sigma(t,\eta) = \varphi(t)\,\mathrm{e}^{\mathrm{i}2\pi(p_0-q_0)\cdot\eta},$$

for any $\varphi : \mathbb{Z} \to \mathbb{C}$ satisfying

$$\frac{1}{q_0 - p_0} \sum_{t=p}^{q_0-1} \varphi(t) = \begin{cases} 1, & \text{if } p = p_0, \\ 0, & \text{if } p \neq p_0. \end{cases}$$

Born–Jordan symbols $\sigma : \mathbb{Z} \times \mathbb{T} \to \mathbb{C}$ of linear operators $A : \mathscr{S}(\mathbb{Z}) \to \mathscr{S}(\mathbb{Z})$ can be built out of this example.

Inversion. Noticing that $Q[\lambda u] = |\lambda|^2 Q[u]$, it is obvious that $u \mapsto Q[u]$ cannot be invertible. However, from $Q[u]$ we can find λu where $|\lambda| = 1$: First, here $R[u] = (I \otimes \mathcal{F}^{-1}) Q[u]$ and $R[u](x,0) = |u(x)|^2$. Now suppose $u(x) \to 0$ as $|x| \to \infty$ (for instance, in applications u might belong to $\ell^2(\mathbb{Z})$). If $u(x_0) \neq 0$ then for all $y \neq 0$ we have

$$u(x_0 + y) = \frac{y}{u(x_0)^*} \sum_{k=0}^{\infty} \triangle_1 R[u](x_0 - ky, y), \tag{17.16}$$

where the partial difference operator $\triangle_1$ is defined by

$$\triangle_1 R[u](x,y) := R[u](x+1,y) - R[u](x,y). \tag{17.17}$$

In other words, (17.16) returns u provided that we know a point value $u(x_0) \neq 0$.

Computational issues. Above, we were considering discrete-time signals. For $u \in \mathscr{S}(\mathbb{R})$ and $h > 0$, define $u_h \in \mathscr{S}(\mathbb{Z})$ by $u_h(j) := u(jh)$. Then we may approximate the Born–Jordan transform of $u, v \in \mathscr{S}(\mathbb{R})$ by

$$\begin{aligned} Q(u,v)(jh,\eta) &= \int_{\mathbb{R}} \mathrm{e}^{-\mathrm{i}2\pi y\cdot\eta}\, R(u,v)(jh,y)\,\mathrm{d}y \\ &\approx h \sum_{k\in\mathbb{Z}} \mathrm{e}^{-\mathrm{i}2\pi (kh)\cdot\eta}\, R(u_h,v_h)(j,k) \\ &= h\, Q(u_h,v_h)(j,[\eta h]), \end{aligned}$$

where $[\eta h] = \eta h + \mathbb{Z} \in \mathbb{T}$. Naturally, in practise we handle only signals with finite support, computing discrete Fourier transforms by FFT (Fast Fourier transform). It should be noted that then computing discrete Born–Jordan transform has lower complexity than computing spectrograms related to short-time Fourier transforms. However, choosing window for short-time Fourier transform is somewhat arbitrary and heavily influences the corresponding spectrogram; there is no window to choose in Born–Jordan case.

On convergence of discretization: From

$$R[u](0,y) = \frac{1}{y}\int_0^y u(t)\,u(t-y)^* = \int_0^1 u(\tau y)\,u((\tau-1)y)^*\,\mathrm{d}\tau$$

we see that

$$\left|\frac{\mathrm{d}}{\mathrm{d}y}R(0,y)\right| \le \|u'\|_{L^\infty}\,\|u\|_{L^\infty}. \tag{17.18}$$

Now suppose $u \in C^2(\mathbb{R})$ such that $u(x) = 0$ for $|x| > L$. Let $Q_h[u]$ denote the trapezoidal rule approximation to $Q[u]$, where the step-size $h = L/N$ with $N \in \mathbb{Z}^+$. Then

$$\begin{aligned}
& |Q[u](0,0) - Q_h[u](0,0)| \\
= & \left|\int_{-L}^{L} R[u](0,y) - h\sum_{\ell=-N}^{N-1} R_h[u](0,\ell)\right| \\
\le & \sum_{\ell=-N}^{N-1}\left[\int_{\ell h}^{(\ell+1)h} |R[u](0,y) - R[u](0,\ell h)|\,\mathrm{d}y + \int_{\ell h}^{(\ell+1)h}\left|R[u](0,\ell h) - \sum_{\cdots}\cdots\right|\,\mathrm{d}y\right] \\
\overset{(17.18)}{\le} & \sum_{\ell=-N}^{N-1}\left[\frac{h}{2}\,\|u'\|_{L^\infty}\,\|u\|_{L^\infty} + \int_{\ell h}^{(\ell+1)h}\frac{(\ell h)^2}{12\ell^2}2\left(\|u\|_{L^\infty}\,\|u''\|_{L^\infty} + \|u'\|^2_{L^\infty}\right)\,\mathrm{d}y\right] \\
= & \; h\,L\left(\|u'\|_{L^\infty}\,\|u\|_{L^\infty} + \frac{1}{3}\,\|u\|_{L^\infty}\,\|u''\|_{L^\infty} + \frac{1}{3}\,\|u'\|^2_{L^\infty}\right).
\end{aligned}$$

In real-life situations, u rarely vanishes at a point, so relative phases of values can be effectively found from

$$R[u](x,1) = u(x)\,u(x-1)^*.$$

Discretizing frequency is pretty straightforward: For instance,

$$\begin{aligned}
A_\sigma v(x) &= \sum_{y\in\mathbb{Z}}\int_{\mathbb{T}} \mathrm{e}^{\mathrm{i}2\pi(x-y)\cdot\eta}\,v(y)\,a(x,y,\eta)\,\mathrm{d}\eta \\
&\approx \sum_{y\in\mathbb{Z}}\frac{1}{N}\sum_{k=1}^{N}\mathrm{e}^{\mathrm{i}2\pi(x-y)\cdot k/N}\,v(y)\,a(x,y,k/N) \\
&= \sum_{y\in\mathbb{Z}}\left[\frac{1}{N}\sum_{k=1}^{N}\mathrm{e}^{\mathrm{i}2\pi(x-y)\cdot k/N}\,a(x,y,k/N)\right]v(y).
\end{aligned}$$

On discrete ghost terms: Now, if $\alpha,\beta \in \mathbb{T}$ and $\alpha \ne \beta$ then

$$Q[\lambda e_\alpha + \mu e_\beta](x,\eta) = |\lambda|^2\delta_\alpha(\eta) + |\mu|^2\delta_\beta(\eta) + ghost_2(x,\eta) \tag{17.19}$$

where

$$R(e_\alpha, e_\beta)(x,0) = \mathrm{e}^{\mathrm{i}2\pi x\cdot(\alpha-\beta)}$$

and for $y \neq 0$

$$R(e_\alpha, e_\beta)(x,y) = \frac{\mathrm{e}^{\mathrm{i}2\pi y\cdot\beta} - \mathrm{e}^{\mathrm{i}2\pi y\cdot\alpha}}{\mathrm{i}2\pi y}\,\mathrm{i}2\pi \frac{\mathrm{e}^{\mathrm{i}2\pi x\cdot(\alpha-\beta)}}{1-\mathrm{e}^{\mathrm{i}2\pi(\alpha-\beta)}}.$$

Why did we write $R(e_\alpha, e_\beta)$ like this? Suppose $-1/2 < \alpha_0 < \beta_0 < 1/2$, where $\alpha = \alpha_0 + \mathbb{Z}$ and $\beta = \beta_0 + \mathbb{Z}$. Let $\mathbf{1}_{[\alpha_0,\beta_0]} : \mathbb{R} \to \mathbb{C}$ be the characteristic function of interval $[\alpha_0, \beta_0]$, and define $\chi : \mathbb{T} \to \mathbb{C}$ its periodization given by

$$\chi(\eta) = \sum_{k\in\mathbb{Z}} \mathbf{1}_{[\alpha_0,\beta_0]}(\eta - k).$$

Now

$$Q(e_\alpha, e_\beta)(x,\eta) = \left[\chi(\eta) - (\beta_0-\alpha_0) + \frac{1-\mathrm{e}^{\mathrm{i}2\pi(\alpha-\beta)}}{\mathrm{i}2\pi}\right]\mathrm{i}2\pi\frac{\mathrm{e}^{\mathrm{i}2\pi x\cdot(\alpha-\beta)}}{1-\mathrm{e}^{\mathrm{i}2\pi(\alpha-\beta)}}.$$

When $\lambda, \mu \in \mathbb{R}$, this leads to
(17.20)

$$ghost_2(x,\eta) = 2\lambda\mu\left[(\chi(\eta) - (\beta_0-\alpha_0))\,(-2\pi)\,\mathrm{Im}\left(\frac{\mathrm{e}^{\mathrm{i}2\pi x\cdot(\alpha-\beta)}}{1-\mathrm{e}^{\mathrm{i}2\pi(\alpha-\beta)}}\right) + \cos(2\pi x\cdot(\alpha-\beta))\right].$$

Fully discrete time-frequency analysis. Let $N \in \mathbb{Z}^+$. For periodic discrete signals $u, v : \mathbb{Z}/N\mathbb{Z} \to \mathbb{C}$, it is possible to define discrete analogues of the Cohen class time-frequency transforms in a straight-forward way, when N is odd. Then for instance the requirement for the correct time and frequency marginals mimics the continuous time case. We shall investigate these properties in future works. Again, FFT is the key ingredient in the computations.

§ 18. Periodic Born–Jordan transform

Let us now consider periodic signals $\widehat{u} \in C^\infty(\mathbb{T})$, which are Fourier dual to discrete-time signals $u \in \mathscr{S}(\mathbb{Z})$:

$$\widehat{u}(\eta) = \sum_{x\in\mathbb{Z}} \mathrm{e}^{-\mathrm{i}2\pi x\cdot\eta}\,u(x),$$
$$u(x) = \int_{\mathbb{T}} \mathrm{e}^{+\mathrm{i}2\pi x\cdot\eta}\,\widehat{u}(\eta)\,\mathrm{d}\eta.$$

Let us define the *periodic Born–Jordan transform* $Q(\widehat{u}, \widehat{v}) : \mathbb{T}\times\mathbb{Z} \to \mathbb{C}$ of periodic signals $\widehat{u}, \widehat{v} \in C^\infty(\mathbb{T})$ via discrete-time Born–Jordan transformation such that

$$Q(\widehat{u}, \widehat{v})(\eta, x) := Q(u,v)(-x,\eta). \tag{18.1}$$

Denote also $Q[\widehat{u}] := Q(\widehat{u}, \widehat{u})$. Loosely speaking, (18.1) can be interpreted as turning the "time-frequency plane" $\mathbb{T} \times \mathbb{Z}$ by 90 degrees to $\mathbb{Z} \times \mathbb{T}$. Test function space $C^\infty(\mathbb{T})$ is dense in $L^2(\mathbb{T})$, which has orthonormal basis $\{e_p : \ p \in \mathbb{Z}\}$, where

$$e_p(\eta) := \mathrm{e}^{\mathrm{i}2\pi\eta\cdot p}. \tag{18.2}$$

Notice that $e_p = \widehat{\delta_{-p}}$, where $\delta_{-p} : \mathbb{Z} \to \mathbb{C}$ is the Kronecker delta at $-p \in \mathbb{Z}$. Hence by calculations in the previous section, we obtain $Q[e_p](\eta, x) = \delta_p(\eta)$, and for $p > q$ that

$$Q(e_p, e_q)(\eta, x) = e_{p-q}(\eta)\, \frac{\mathbf{1}_{[q,p)}(x)}{p-q} = \begin{cases} (p-q)^{-1}\, e_{p-q}(\eta) & \text{when } q \le x < p, \\ 0 & \text{otherwise.} \end{cases} \tag{18.3}$$

Hence for $p > q$, we have

$$Q[\lambda e_p + \mu e_q](\eta, x) = |\lambda|^2 \delta_p(x) + |\mu|^2 \delta_q(x) + ghost_2(\eta, x), \tag{18.4}$$

where $ghost_2(\eta, x) \neq 0$ only if $q \le x < p$, and when $\lambda, \mu \in \mathbb{R}$ we have

$$ghost_2(\eta, x) = 2\lambda\mu \frac{\cos(2\pi\eta \cdot (p-q))}{p-q}. \tag{18.5}$$

Let $H^s(\mathbb{T})$ be the Sobolev space of order $s \in \mathbb{R}$: this Hilbert space is the completion of $C^\infty(\mathbb{T})$ with respect to the norm

$$\|\widehat{u}\|_{H^s(\mathbb{T})} = \left[\sum_{y \in \mathbb{Z}} \langle y \rangle^{2s} |u(y)|^2 \right]^{1/2},$$

where $\langle 0 \rangle = 1$ and $\langle y \rangle = |y|$ for $y \neq 0$.

Theorem. *Let $\widehat{u} \in H^s(\mathbb{T})$, where $s > 1/2$. Then*

$$(\eta \mapsto Q[\widehat{u}](\eta, x)) \in H^{s-1/2}(\mathbb{T}), \tag{18.6}$$

$$\left| Q[\widehat{u}](\eta, x) - |u(x)|^2 \right| \le c_s \|\widehat{u}\|_{H^s(\mathbb{T})}, \tag{18.7}$$

where constant $c_s < \infty$ depends only on s.

Proof. Because $s > 1/2$,

$$\eta \mapsto \widehat{u}(\eta) = \sum_{p \in \mathbb{Z}} u(p)\, e_{-p}(\eta)$$

is continuous. Notice that

$$\begin{aligned} Q[\widehat{u}](\eta, x) &= \sum_{p \in \mathbb{Z}} |u(p)|^2\, Q(e_p, e_p)(\eta, x) + \sum_{p,q:\ p \neq q} u(p)\, u(q)^*\, Q(e_p, e_q)(\eta, x) \\ &= |u(x)|^2 + ghost(\eta, x), \end{aligned}$$

where

$$ghost(\eta, x) = \sum_{p,q:\ p\neq q} u(p)\, u(q)^*\, e_{p-q}(\eta)\, \frac{\mathbf{1}_{[\min\{p,q\},\max\{p,q\})}(x)}{|p-q|}.$$

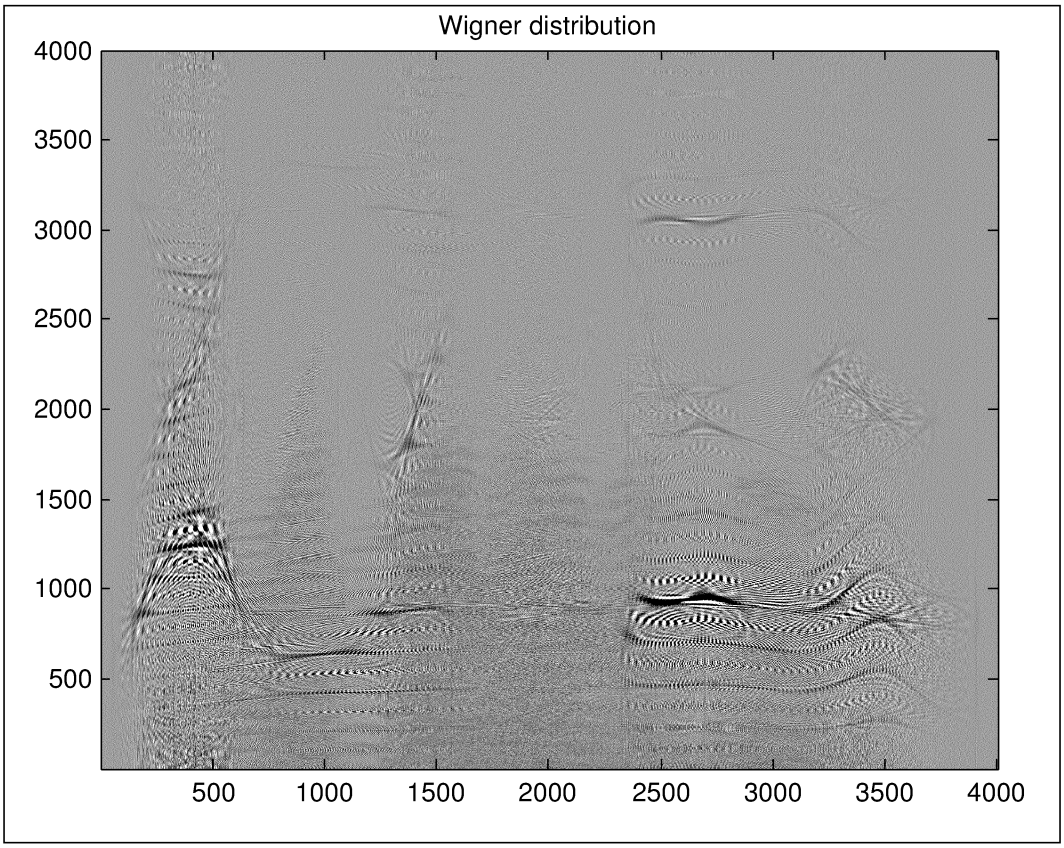
For $0 < t < s$, $u \in H^s(\mathbb{T})$ and $\varphi_x(\eta) := Q[\widehat{u}](\eta, x)$, this means

$$\begin{aligned}
\|\varphi_x\|^2_{H^r(\mathbb{T})} &= |u(x)|^2 + \sum_{k\in\mathbb{Z}:k\neq 0} |k|^{2r} \left| \sum_{p,q:p-q=k} u(p)\, u(q)^*\, \frac{\mathbf{1}_{[\min\{p,q\},\max\{p,q\})}(x)}{|k|} \right|^2 \\
&\leq |u(x)|^2 + c_{u,s} \sum_{k\in\mathbb{Z}} \langle k\rangle^{2r-2} \left(\sum_{p\in\mathbb{Z}} \langle p\rangle^{-s} \langle p-k\rangle^{-t} \right)^2 \\
&\leq |u(x)|^2 + c_{u,s,t} \sum_{k\in\mathbb{Z}} \langle k\rangle^{2(r-1-t)} \left(\sum_{p\in\mathbb{Z}} \langle p\rangle^{t-s} \right)^2 .
\end{aligned}$$

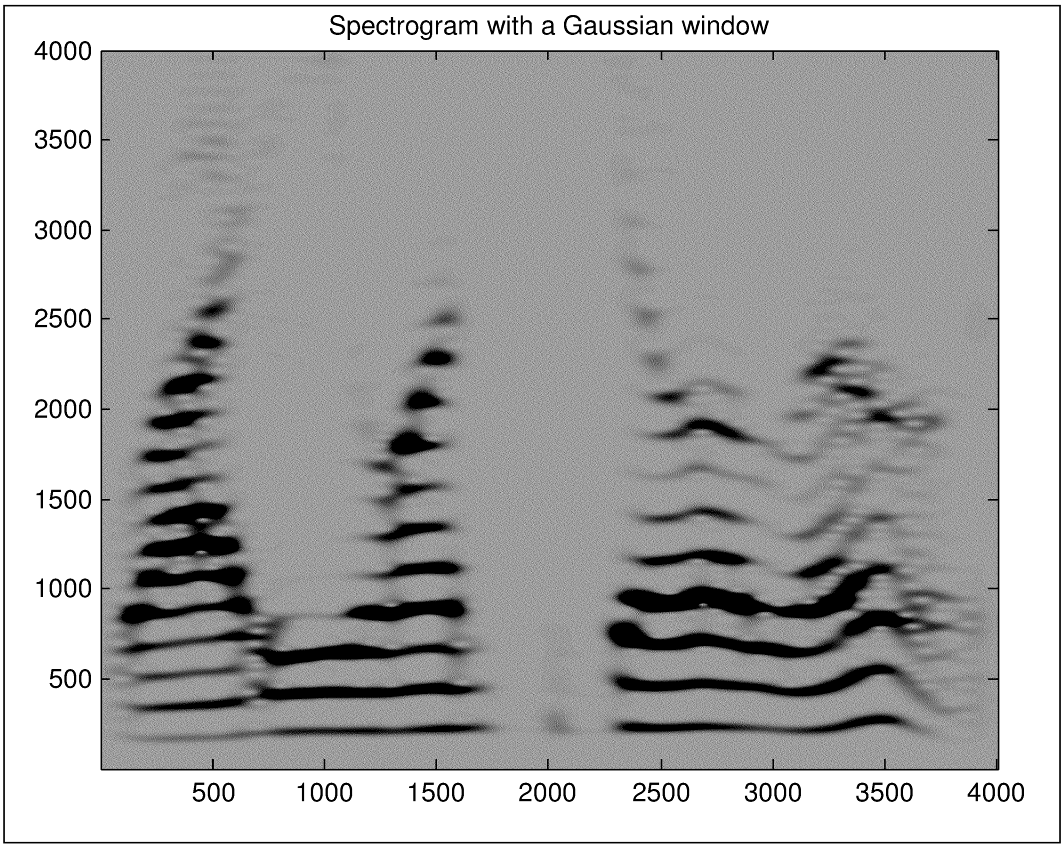
Thus $\varphi_x \in H^r(\mathbb{T})$ if $0 < t$, $2(r-1-t) < -1$ and $t - s < -1$. Hence $r < s - 1/2$. □

§ 19. Examples of discrete-time time-frequency distributions

In this Section, we shall see various discrete-time time-frequency distributions $P[u] = P(u, u)$ for same human speech signal u (a man speaking "Why do you want to go alone?", extracted from the signal in [15]). First, here is the Wigner distribution:

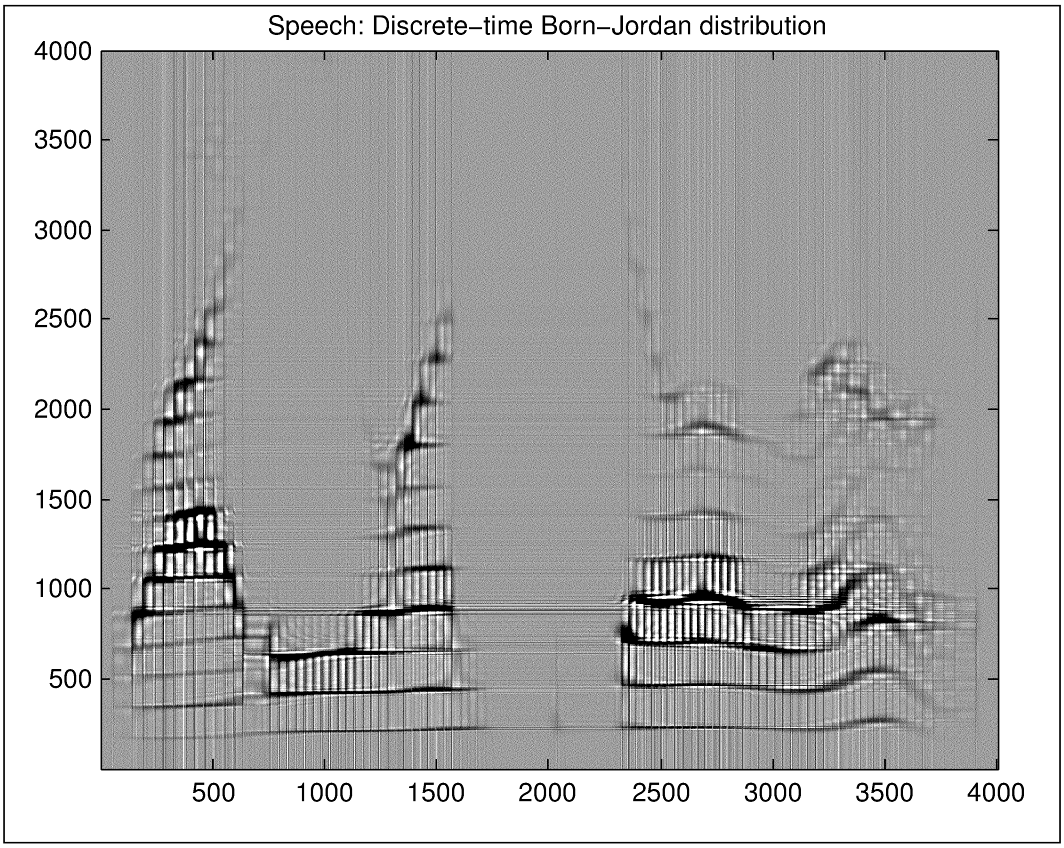


Even though the Wigner distribution has mathematically nice properties (e.g. there is no loss of information, as $[u]$ can be recovered from $W[u]$), it is very sensitive to noise (it has strong interference terms), as can be seen here: therefore the Wigner distribution is often of no practical use. On the other hand, spectrograms are not sensitive to noise, but they lose the information by smoothing too much: the following spectrogram with a Gaussian window can be thought as a melted-down version of the Wigner distribution (this claim can be made precise by studying a suitably normalized heat equation in the plane):

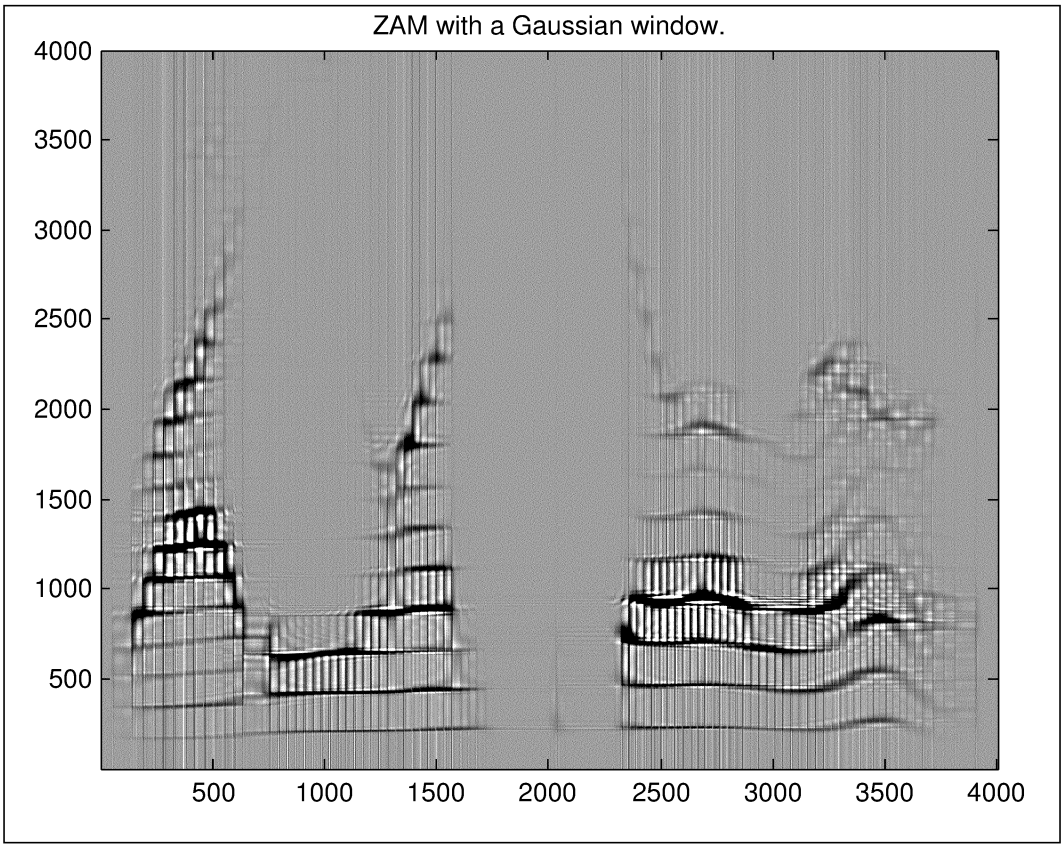


Moreover, we must remember that the choice of (a more-or-less arbitrary) time-analysis window in the Short-Time Fourier Transform heavily affects the shape of the spectrogram.

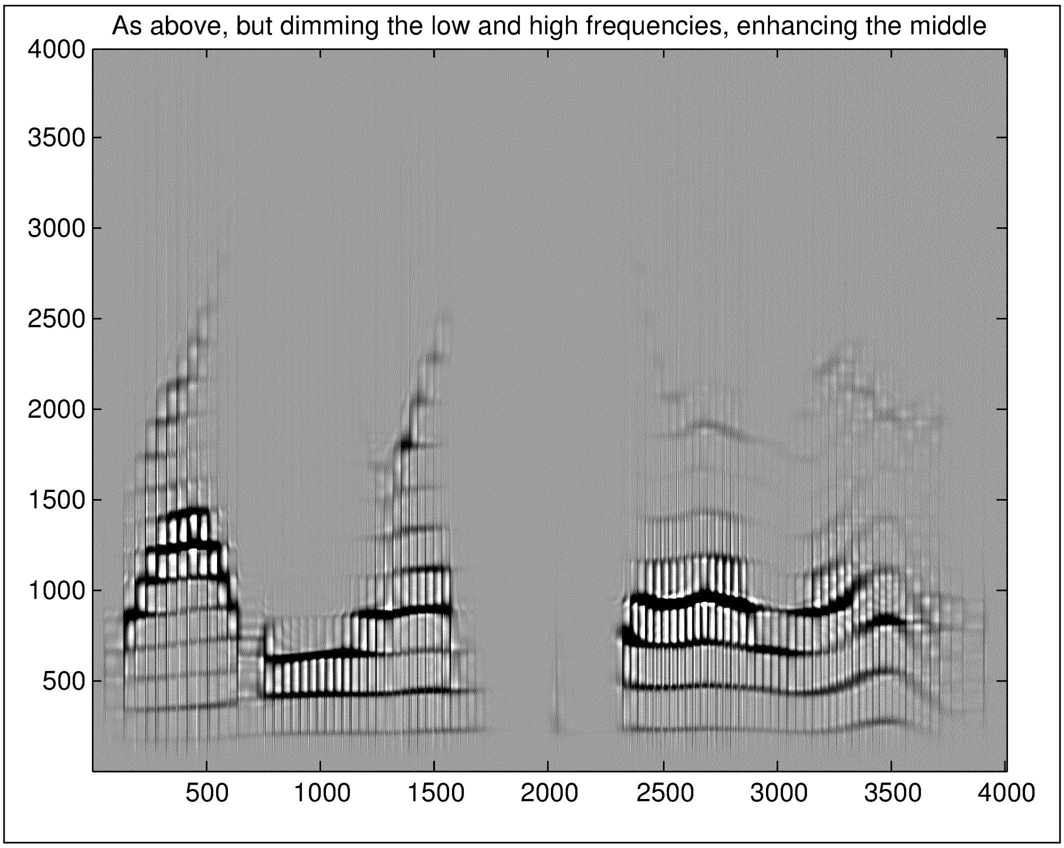
However, we do not have to lose information while reducing noise-sensitivity (reducing interferences). The Born–Jordan distribution exemplifies this. Recall that the Born–Jordan transform was characterized by the three natural properties (scale invariance, time-locality, comb-to-grid property). Here we see the outcome, the Born–Jordan distribution $Q[u]$:



Occasionally it is claimed that there are still interferences in the Born–Jordan distribution. It is perhaps not even meaningful to exactly define what intereferences mean. Here in this picture above, "interferences" could be the geometrically sharp horizontal and vertical lines. Especially in the case of the sharp horizontal lines here, there are rapid oscillations between positive and negative values, effectively in average almost zero: by a little bit of smoothing in the next picture, we display a Gaussian lag-weighted Born–Jordan distribution, also called a time-frequency distribution in the Zhao–Atlas–Marks family:

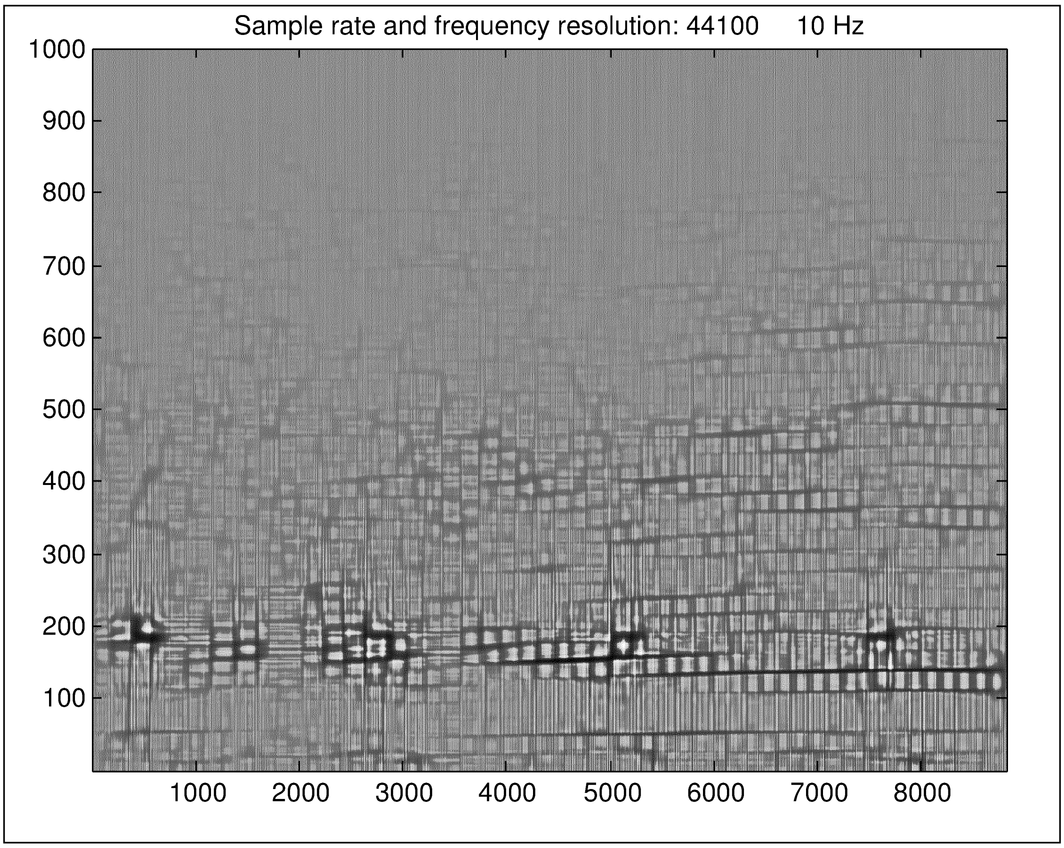


This previous time-frequency distribution, however, already loses a little of the information about the signal. Nevertheless, so does the human hearing, as the low and high frequencies are badly perceived. In the next picture, there is a qualitative attempt to mimic the loss of accuracy in the low and high frequencies:



§ 20. Discrete-time Born–Jordan examples

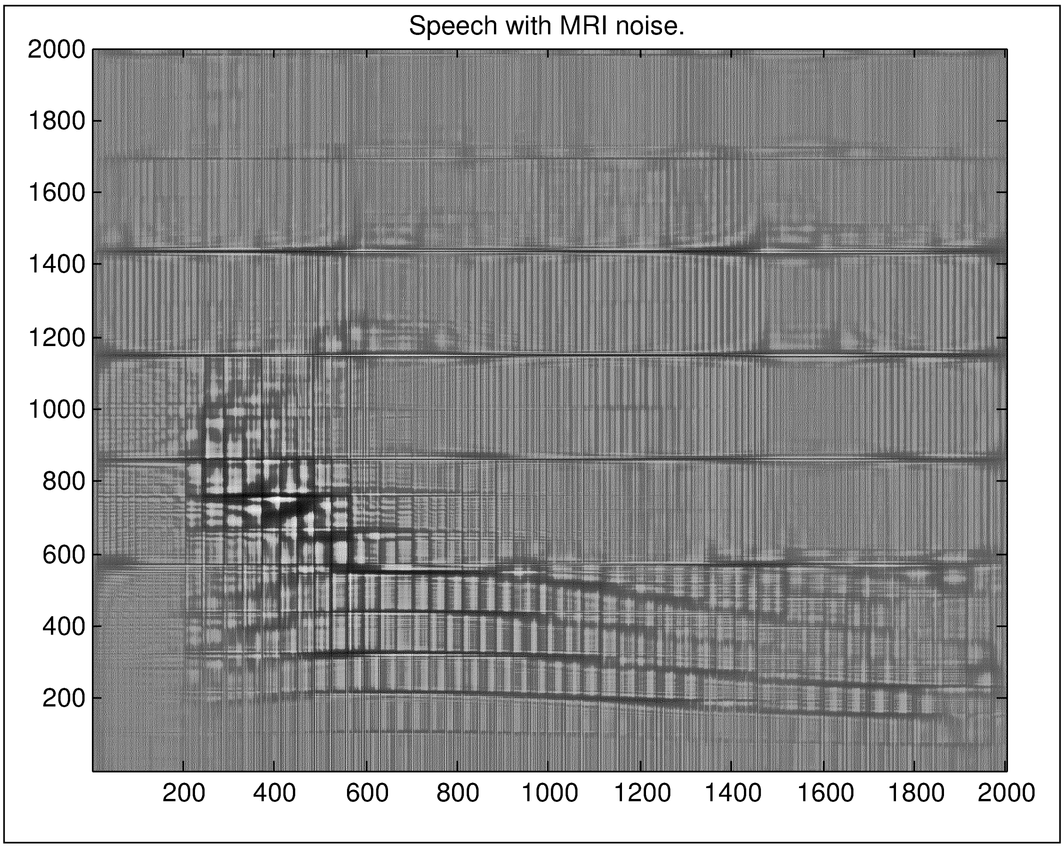
Example: Whale sounds. In the following picture, there is the Born–Jordan distribution for a Beluga whale sound from [19]:



In such signals, there are simultaneously fast and slow spectral developments, which are rather troubling for spectrograms: Vertical features (snapping sounds, quick transients) would require very short time-analysis windows wiping out the horizontal features (whistling sounds), and vice versa.

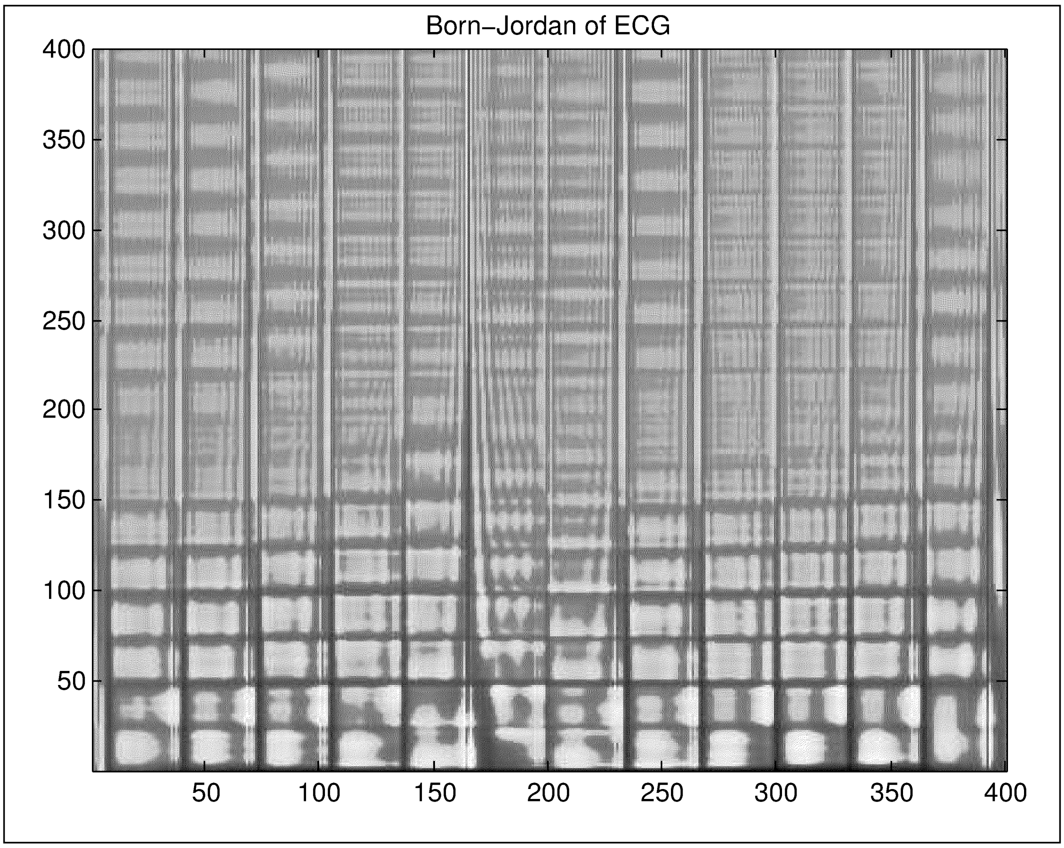
Of course, instead of spectrograms of time-frequency analysis (given by the Short-Time Fourier Transform), we could try to use scaleograms of time-scale analysis (given by the wavelet transform). However, there still would be analogous problems with the Heisenberg uncertainty, and the more-or-less arbitrary choice of the mother wavelet would affect drastically the shape of the scaleograms. Moreover, it is good to remember that the Born–Jordan distribution is automatically also scale invariant!

Example: Speech and MRI. Let us consider human speech with heavy noise coming from the MRI (Magnetic Resonance Imaging) as recorded by the research group "Speech & Math" lead by Dr. Jarmo Malinen [1]. In the complete signal, the doctoral student Mr. Juha Kuortti speaks the Finnish sentence "*Ruusu varoo laavaa*". In the following Born–Jordan picture with low sampling rate of 5512 Hz, we depict the first syllable "Ruu":

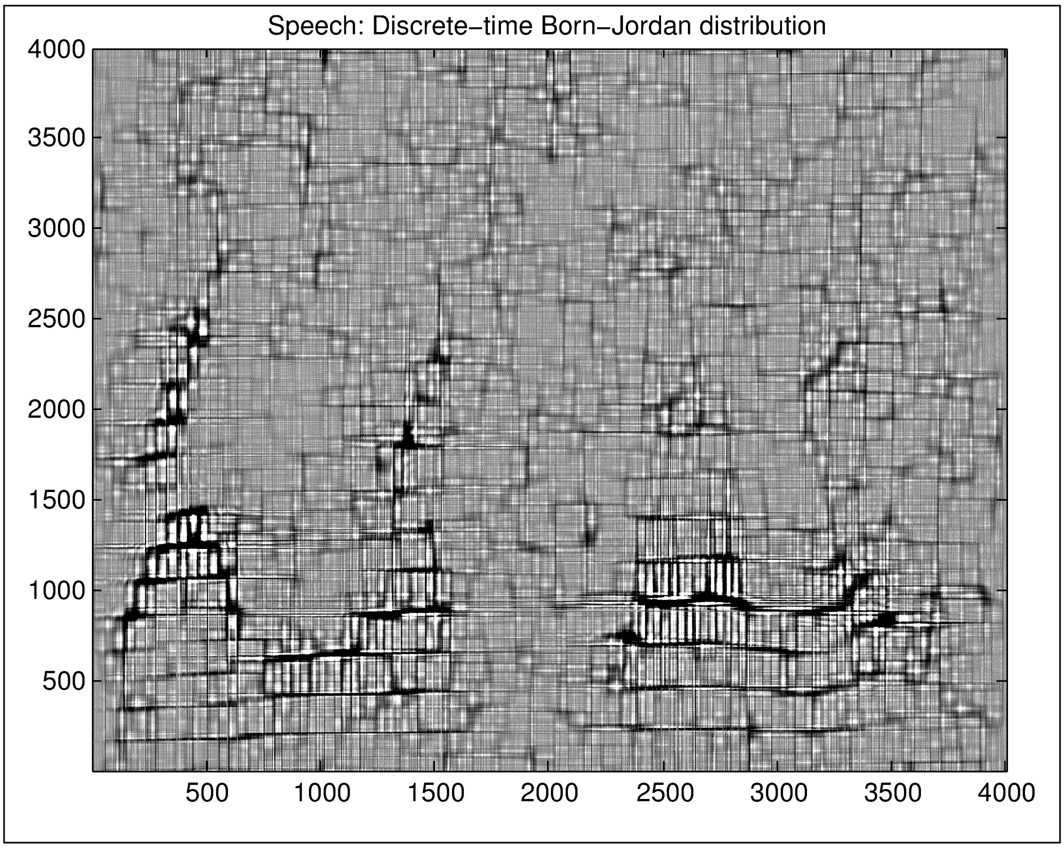


In the picture above, there are essentially two different grid-like patterns of rectangles of unit time-frequency area: the narrow and tall rectangles coming from the MRI, and the other rectangles coming from the speech. With this knowledge of time-frequency localization, it is possible to separate the speech from the MRI noise in a sharp fashion.

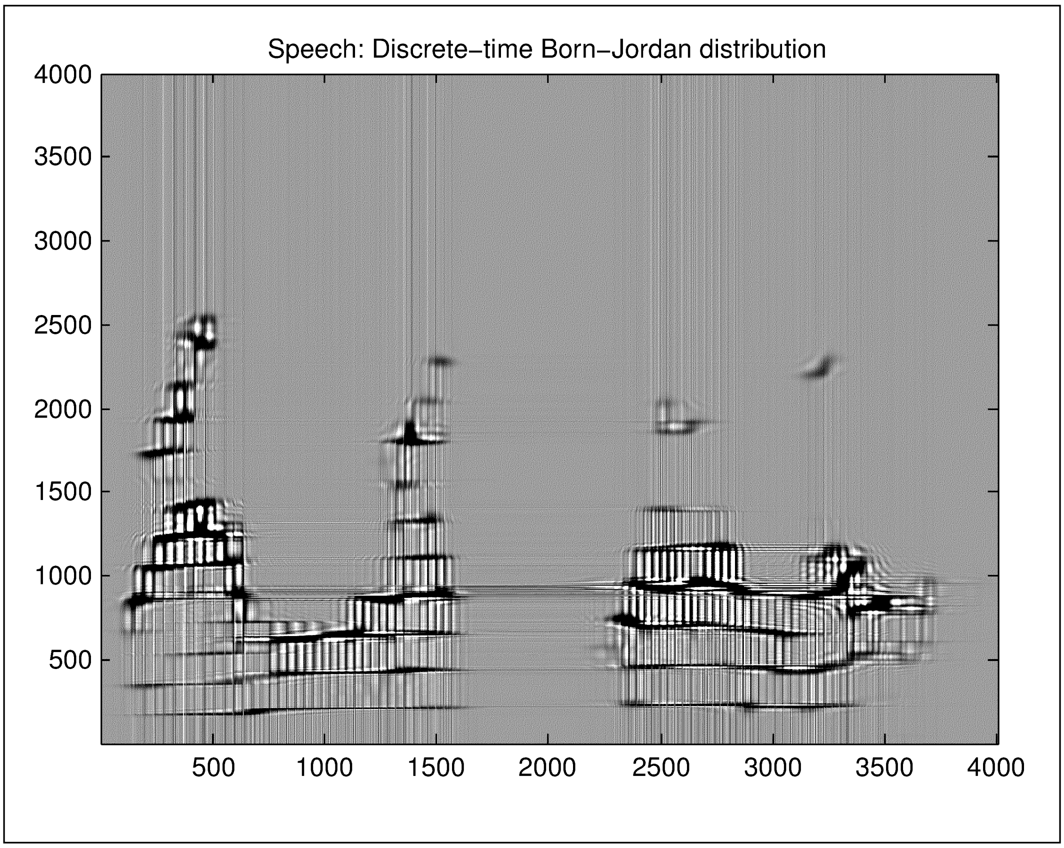
Example: ECG diagnostics. As another medical application, let us consider electrocardiogram data (ECG). Simplified a bit, a healthy heart should produce a strong regular grid-like Born-Jordan distribution. We chose an excerpt from the public MIT-BIH Arrhythmia Database ([18], [13]), deliberately wiping the signal to zero both in the past and in the future (thereby making the left and right ends of the picture unreliable: however, there is basically very little distortion there). Here, there is unusual activity around the time index 170:



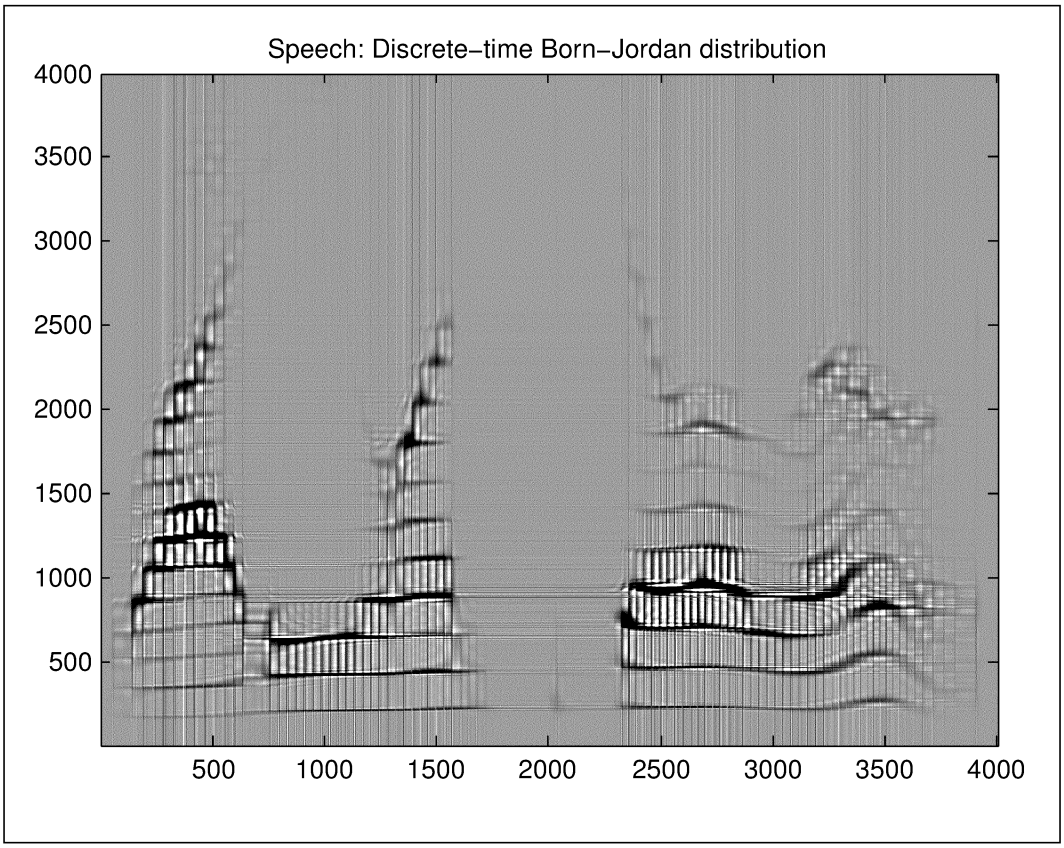
Linear phase-preserving denoising. Now let us consider sampled speech, a male voice asking "**Why do you want to go alone?**" (from [15]). The original sampling was at 8000 Hz, but we sample at only 4000 Hz, taking 4200 samples (i.e. 1050 milliseconds). Moreover, we add heavy random noise (`rand-.5` in `Matlab`, with energy equal to the original speech). The following three pictures show the Born–Jordan energy densities for the following sounds. First, the noisy original:



Then we apply two Born–Jordan localizations. In each of these localizations, the symbol is a characteristic function of a planar set, which is computed from simple natural conditions. These two simple conditions basically search those time-frequency regions where the energy density is "large-enough". We obtain the enhanced filtered signal with the following Born–Jordan energy density:



Of course, here the quality of the signal has suffered due to the heavy noise. For comparison, here is the Born–Jordan distribution of the original clean signal, without adding artificial noise:



We would have obtained much better signal reconstruction, had we exploited the information that there is a speech signal in the background. Such applications will be considered in future articles.

§ 21. Closing remarks

Born–Jordan distribution is a "well-known but poorly understood" member of the Cohen class time-frequency distributions, introduced by Leon Cohen in 1966, building on the 1925 quantum matrix mechanics of Heisenberg, Born and Jordan: we should remember that the Born–Jordan quantization is the only correct quantization for Heisenberg's matrix mechanics.

So, why Born–Jordan is not used that much yet? Superficially it just looks like "one approach out of infinitely many". Spectrograms are likely the most used Cohen class time-frequency distributions, and their positivity may partly explain their popularity, even though they destroy information; looking at most acoustic spectrograms, it seems that researchers favor longish time analysis windows, mostly missing "the vertical lines" of the time-frequency behavior. Of course, the power of tradition is also strong, if people have grown to use spectrograms. In the literature, there are also occasional mistakes about the Cohen class properties, e.g. misunderstanding the computational complexity and some analytic properties.

Above, we saw that the Born–Jordan distribution provides a reasonable alternative to spectrograms. The characterization of the Born–Jordan distribution among the Cohen class distributions shows its Fourier analytic naturality. The Born–Jordan transform offers noise-robust yet information-preserving pictures of good clarity, with no arbitrary window to choose for analysis. The computational complexity for the spectrograms is the same as for the Born–Jordan distribution, and can be implemented by `Matlab` with the usual routines.

Yet there is much to investigate in the Fourier analysis of the Born–Jordan transform, both in the continuous and in the discrete time cases. We shall continue working in these directions in the future papers.

References

[1] D. Aalto, O. Aaltonen, R.-P. Happonen, P. Jääsaari, A. Kivelä, J. Kuortti, J.-M. Luukinen, J. Malinen, T. Murtola, R. Parkkola, J. Saunavaara, T. Soukka, and M. Vainio: Large scale data acquisition of simultaneous MRI and speech. *Applied Acoustics* **83** (2014), 64–75.

[2] P. Boggiatto, G. De Donno and A. Oliaro, Time-frequency representations of Wigner type and pseudo-differential operators, *Trans. Amer. Math. Soc.*, **362** (2010), 4955–4981.

[3] P. Boggiatto, E. Carypis and A. Oliaro, Windowed-Wigner representations in the Cohen class and uncertainty principles, *Geom. Anal.*, **4** (2011), 1753–1779.

[4] M. Born and P. Jordan, Zur Quantenmechanik, *Zeitschrift für Physik*, **34** (1925), 858–888.

[5] M. Born, W. Heisenberg and P. Jordan, Zur Quantenmechanik II, *Zeitschrift für Physik*, **35** (1926), 557–615.

[6] H. I. Choi and W. J. Williams, Improved time-frequency representation of multicomponent signals using exponential kernels, *IEEE Trans. Acoust., Speech, Signal Processing*, **37** (1989), 862–871.

[7] L. Cohen, Generalized phase-space distribution functions, *J. Mathematical Phys.*, **7** (1966), 781–786.

[8] L. Cohen, Time-frequency distributions — A review, *Proc. IEEE*, **77** (1989), 941–981.

[9] L. Cohen, *Time-Frequency Analysis*, Prentice Hall, 1995.

[10] W. A. Fedak and J. J. Prentis, The 1925 Born and Jordan paper "On quantum mechanics", *Am. J. Phys.*, **77** (2009), 128–139.

[11] G. B. Folland, *Harmonic Analysis in Phase Space*, Princeton Univ. Press, 1989.

[12] D. Gabor, Theory of communication, *J. IEE (London)*, **93** (III) (1946), 429–457.

[13] A. L. Goldberger AL, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. Ch. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. *Circulation* **101**, 23 (June 13, 2000), e215-e220. `http://circ.ahajournals.org/cgi/content/full/101/23/e215`

[14] K. Gröchenig, *Foundations of Time-Frequency Analysis*, Birkhäuser, 2001.

[15] P. C. Hansen, speech signal `alone.wav` at the web page `http://www.imm.dtu.dk/~pcha/Projekter/tsvd.html`

[16] W. Heisenberg, Über quantentheoretische Umdeutung kinematischer und mechanischer Beziehungen, *Zeitschrift für Physik*, **33** (1925), 879–893.

[17] A. J. E. M. Janssen, On the locus and spread of pseudo-density functions in the time-frequency plane, *Philips J. Res.*, **37** (1982), 79–110.

[18] G. B. Moody and R. G. Mark. The impact of the MIT-BIH Arrhythmia Database. *IEEE Eng in Med and Biol* **20**, 3 (May–June 2001), 45–50. Database at `https://www.physionet.org/physiobank/database/mitdb/`

[19] R. Montgomery, signal *Beluga Sound 1* at the web page `http://www.soundboard.com/sb/drmontgomery`

[20] J. E. Moyal, Quantum mechanics as a statistical theory, *Proc. Cambridge Phil. Soc.*, **45** (1949), 99–124.

[21] W. Rihaczek, Signal energy distribution in time and frequency, *IEEE Trans. Informat. Theory*, **14** (1968), 369–374.

[22] W. Rudin, *Functional Analysis*, McGraw-Hill, 1973.

[23] W. Rudin, *Real and Complex Analysis*, McGraw-Hill, 1987.

[24] M. Ruzhansky and V. Turunen, *Pseudo-Differential Operators and Symmetries*, Birkhäuser, 2010.

[25] L. Schwartz, Transformation de Laplace des distributions, *Comm. Sém. Math. Univ. Lund [Medd. Lunds Univ. Mat. Sem.],* (1952), 196–206.

[26] M. Taketani and M. Nagasaki, *The Formation and Logic of Quantum Mechanics. Vol. III: The Establishment and Logic of Quantum Mechanics*, World Scientific, 2000.

[27] J. Ville, Théorie et applications de la notion de signal analytique, *Câbles et Transmission*, **2** (1948), 61–74.

[28] H. Weyl, Quantenmechanik und Gruppentheorie, *Zeitschrift für Physik*, **46** (1927), 1–46.

[29] H. Weyl, *Gruppentheorie und Quantenmechanik*, S. Hirzel, 1928.

[30] E. P. Wigner, On the quantum correction for thermo-dynamic equilibrium, *Phys. Rev. Letters,* **40** (1932), 749–759.

[31] P. M. Woodward, *Probability and Information Theory with Applications to Radar*, Pergamon Press, 1953.