Doctoral theses at NTNU, 2025:236

Simon Halvdansson

Extensions of Quantum Harmonic Analysis and Applications to Time-Frequency Analysis

Doctoral thesis

NTNU
Norwegian University of Science and Technology
Thesis for the Degree of Philosophiae Doctor
Faculty of Information Technology and Electrical Engineering
Department of Mathematical Sciences

NTNU
Norwegian University of Science and Technology

Simon Halvdansson

Extensions of Quantum Harmonic Analysis and Applications to Time-Frequency Analysis

Thesis for the Degree of Philosophiae Doctor

Trondheim, June 2025

Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Mathematical Sciences

NTNU

Norwegian University of
Science and Technology

NTNU
Norwegian University of Science and Technology

Thesis for the Degree of Philosophiae Doctor

Faculty of Information Technology and Electrical Engineering
Department of Mathematical Sciences

ISBN 978-82-326-9068-8 (printed ver.)
ISBN 978-82-326-9067-1 (electronic ver.)
ISSN 1503-8181 (printed ver.)
ISSN 2703-8084 (online ver.)

Doctoral theses at NTNU, 2025:236

Printed by NTNU Grafisk senter

Abstract

The focus of this thesis is twofold - we first work on extending the foundations of quantum harmonic analysis and then apply these new tools and others to develop applications within time-frequency analysis. The extensions allow us to apply quantum harmonic analysis to different locally compact groups, quantize functions on some of these groups and convolve measures with operators. Meanwhile, the applications are motivated by numerical realizability and machine learning and are concerned with the properties of time-frequency localization operators and a new family of operators we call time-frequency blurring operators.

Sammendrag

Fokuset i denne avhandlingen er todelt: Først arbeider vi med å utvide grunnlaget for kvanteharmonisk analyse, og deretter anvender vi disse nye verktøyene og andre metoder for å utvikle bruksområder innen tids-frekvensanalyse. Utvidelsene gjør det mulig å bruke kvanteharmonisk analyse på forskjellige lokalt kompakte grupper, kvantisere funksjoner på noen av disse gruppene og konvolvere mål med operatorer. Samtidig er anvendelsene motivert av numerisk realiserbarhet og maskinlæring, og de omhandler egenskapene til tids-frekvens-lokaliseringsoperatorer samt en ny familie operatorer som vi kaller tids-frekvens-uskarphetsoperatorer.

Preface

This thesis is submitted in partial fulfillment of the requirements for the degree of Philosophiae Doctor (PhD) in Mathematical Sciences at the Norwegian University of Science and Technology (NTNU). The research presented here was conducted at the Department of Mathematical Sciences at NTNU. The candidate was supervised by Franz Luef as the main supervisor, and Sigrid Grepstad as the secondary supervisor.

The thesis is comprised of a short introduction and seven research articles, four of which are published with the remaining being preprints. In the introduction we give a brief overview of the background necessary to put the articles in a wider context and also give brief summaries of the main points of each article. The articles are reproduced with minimal typographical changes. References have been consolidated into a single list at the end of the thesis.

Acknowledgments

I have always been told that a PhD is supposed to be an arduous process, including by those significantly smarter than me. While I have not found this to be outright false, it would be an exaggeration to say that I wholeheartedly agree. Reflecting on this, I have arrived at the conclusion that I have had the stars align in a particularly favorable configuration - mostly in the form of the people around me.

From day one, Franz has been nothing but extremely approachable, welcoming and insightful. While always having a good time, Franz has been able to consistently lift my spirits when in doubt about a project, supply good ideas and make connections so far-reaching it makes me wonder if he can stop time and read a book without me noticing whenever I ask him a question. His guidance has been exactly what I have needed.

At the department, I have met so many great people that make it a joy to go to work every day and to see them outside of it. In no too-particular order, I am thankful to Helge, Mathias, Athanasios, Stine, Eirik, Emilie, Erling, Konstantinos, Carlos, Sarah, Markus, Johannes, Anders, Michael, Denis, Manuel and Karl-

Mikael for all the good times and nice talks. I'm sure I'll keep seeing many of you but so far, it's been a pleasure.

There are also those who have shaped my mathematical journey beyond those in Trondheim. Yacin was the first to truly challenge me mathematically and Jan-Fredrik gave me invaluable guidance and many times the hours of advising I was allocated for my master thesis. Without them this thesis might have looked very different or not existed at all. At conferences and during my visit to Vienna I have had the pleasure of meeting many nice people from the time-frequency community. A special thank you to Monika for inviting me to visit and being such a nice host and conversation partner.

Life is about more than work though, and I would be remiss not to thank my friends in Stockholm and beyond who I meet when visiting and when stomping pubs. August, Patrik, Emil, Jakob and Maxi - your friendship is most valued and I always look forward to seeing you. And thanks for playing with me even though I am forever in and of the trenches.

My family's support has been ever-present and most appreciated over the years. Thanks for always encouraging and believing in me.

Looking back at the path that ultimately led me here, I don't think any one person has been as influential as John. For this, his friendship, encouragement, numerous deep discussions and guidance, I am truly grateful.

Finally, Amanda. While I by no means take your love, encouragement and support for granted, I have a hard time imagining life without it. You have genuinely always believed in me and have kept me balanced through the highs and lows. These four years have been infinitely better because of you.

Simon Halvdansson
Trondheim, March 2025

Contents

Part I

Introduction

Chapter 1

Time-Frequency Analysis

Fourier analysis has enabled us to examine the frequencies making up signals since the early 1800s, but when faced with real non-stationary data, computing a Fourier transform is rarely enough to analyze a signal in any depth. It was not until the 1940s that the basic notions of time-varying frequency contents were formalized in Dennis Gabor's *Theory of Communication* [93]. The field of *time-frequency analysis* can be said to be comprised of the tools and methods for carrying out such analysis from the mathematical side. In this chapter, we will go over what the aforementioned basic notions are and outline what modern time-frequency analysis looks like. Throughout, we will aim to keep a relaxed tone and refer the reader to the preliminaries sections of the papers, as well as the standard textbook [107], for proofs.

1.1 Phase space and the short-time Fourier transform

The simplest case of a signal which is amenable to frequency analysis is a real-valued integrable function f on $\mathbb{R}$ which possesses some sort of oscillatory behavior. For such a function we define the *Fourier transform* $\hat{f}$ on $\mathbb{R}$ as the function

$$\hat{f}(\omega) = \mathcal{F}(f)(\omega) = \int_{\mathbb{R}} f(t)e^{-2\pi i\omega t}\,dt.$$

This function is in general complex-valued but its squared modulus $|\hat{f}(\omega)|^2$, often called the *spectrum* of f, is real-valued, non-negative and generally conveys information about which frequencies "make up" the signal f. One way to motivate why the spectrum tells us which frequencies make up f is the inversion formula

$$f(t) = \mathcal{F}^{-1}(\hat{f})(t) = \int_{\mathbb{R}} \hat{f}(\omega)e^{2\pi i\omega t}\,d\omega. \tag{1.1.1}$$

Indeed, if f is a simple trigonometric polynomial such as $f(t) = e^{2\pi i\omega_1 t} + e^{2\pi i\omega_2 t}$ it is clear that the distribution $\hat{f}(\omega) = \delta_{\omega_1}(\omega) + \delta_{\omega_2}(\omega)$, where δ_ω is a Dirac delta, satisfies (1.1.1). This raises the question - what do we lose when we take the square modulus of $\hat{f}(\omega)$? The answer is the *time* of the signal. More precisely, the relation

$$\mathcal{F}(f(\cdot - x))(\omega) = e^{2\pi i x\omega}\hat{f}(\omega)$$

tells us that the spectrum is invariant under translations. In some cases this is fine, but when a signal is used to facilitate communications such as for radio, telephony, telegraphy and television, order is important and we need finer tools for analysis. This was precisely the motivation for Gabor in the 1940s to analyze signals in time and frequency jointly. To do this, we need to split up f in time and compute the Fourier transform for each piece of f. Let f_x denote the function

$$f_x(t) = \chi_{[x-1/2,x+1/2]}(t)f(t),$$

where χ_E is the indicator function of the set E. Then $F(x,\omega) = \hat{f}_x(\omega)$ tells us which frequencies make up f_x, the piece of f which is supported close to x. The choice of multiplying f by $\chi_{[x-1/2,x+1/2]}$ to isolate the part of the signal close to x was clearly arbitrary and we can replace it by $g(\cdot - x)$ where g is a function centered around $t = 0$. For technical reasons which will become clear later, it is more convenient to use $f_x(t) = \overline{g(t-x)}f(t)$ where $\overline{g}$ is the complex conjugate of g. The function F which we have constructed is commonly denoted by $V_g f$ to denote the dependence on g and can be written as

$$V_g f(x,\omega) = \mathcal{F}\big(\overline{g(\cdot - x)}f(\cdot)\big)(\omega) = \int_{\mathbb{R}} f(t)\overline{g(t-x)}e^{-2\pi i\omega t}\,dt. \tag{1.1.2}$$

We say that $V_g f$ is the *short-time Fourier transform* (STFT) of f with respect to g. Note that while the Fourier transform of a function on $\mathbb{R}$ is another function on $\mathbb{R}$, the short-time Fourier transform $V_g f$ is a function on $\mathbb{R}^2$. In this context, we say that $V_g f$ is a function on *phase space* or the *time-frequency space* in the same way that we sometimes say that $\hat{f}$ is a function on frequency space. The STFT is an example of a *time-frequency representation* - we will see more of these later.

We now set out to put the STFT on more solid theoretical footing than the intuition-based motivation above. The translation and modulation operators T_x and M_ω are frequently used in Fourier analysis and are defined as

$$T_x f(t) = f(t-x), \qquad M_\omega f(t) = e^{2\pi i\omega t}f(t).$$

They should be seen as moving a signal in either time or frequency and have the property of commuting up to a unimodular phase factor.

The Fourier transform of an $L^1(\mathbb{R})$ function can be defined using the modulation operator as

$$\hat{f}(\omega) = \langle f, M_\omega 1 \rangle_{L^1, L^\infty}. \tag{1.1.3}$$

Meanwhile the STFT definition (1.1.2) can be written as the $L^2(\mathbb{R})$ inner product

$$V_g f(x, \omega) = \langle f, M_\omega T_x g \rangle$$

for $f, g \in L^2(\mathbb{R})$. Hence the standard Fourier transform can be realized as a special case of the STFT with constant window restricted to the second variable.

For the Fourier transform it is the Plancherel theorem ($\langle f_1, f_2 \rangle = \langle \hat{f}_1, \hat{f}_2 \rangle$) that lends credibility to the spectrum interpretation of $|\hat{f}|^2$ since both $|f|^2$ and $|\hat{f}|^2$ have the same mass. The corresponding theorem for the STFT, *Moyal's identity* states that for $f_1, f_2, g_1, g_2 \in L^2(\mathbb{R})$,

$$\langle V_{g_1} f_1, V_{g_2} f_2 \rangle_{L^2(\mathbb{R}^2)} = \langle f_1, f_2 \rangle \overline{\langle g_1, g_2 \rangle}. \tag{1.1.4}$$

In particular, when g is normalized so that $\|g\|_{L^2} = 1$, $V_g : L^2(\mathbb{R}) \to L^2(\mathbb{R}^2)$ is an isometry. Now just as the spectrum $|\hat{f}|^2$ can be viewed as a probability distribution over $\mathbb{R}$ telling us where the spectrum is localized, the squared modulus $|V_g f|^2$, the *spectrogram*, defines a probability distribution over phase space $\mathbb{R}^2$. Applying a time-frequency shift $M_\omega T_x$ to f translates $V_g f$ but also adds a unimodular phase factor which disappears when looking at the spectrogram so that

$$|V_g(M_\omega T_x f)|^2 = |V_g f(\cdot - (x, \omega))|^2.$$

Time-frequency shifts of this form are frequently written as $\pi(x, \omega) f = M_\omega T_x f$ and are one of the main objects of time-frequency analysis. To highlight that $(x, \omega) \in \mathbb{R}^2$ is just a point in phase space, we often identify $\mathbb{R}^2$ with $\mathbb{C}$ and write z for the point (x, ω).

The smoothness of a function f can be gauged from the decay rate of its Fourier transform. Here the global property of integrability can inform us about the local property of smoothness. In time-frequency analysis, we often use the integrability of the short-time Fourier transform of a function to describe its behavior. We say that f is in the *modulation space* $M^p(\mathbb{R})$ if $V_g f$ is in $L^p(\mathbb{R}^2)$ for g the standard Gaussian.

The last central topic we mention before going into details is the STFT reconstruction formula which follows from Moyal's identity (1.1.4) and states that for g_1, g_2 not orthogonal,

$$f = \frac{1}{\langle g_2, g_1 \rangle} \int_{\mathbb{R}^2} V_{g_1} f(z) \pi(z) g_2 \, dz \tag{1.1.5}$$

weakly.

1.2 Localization operators

From a signal processing perspective, one of the standard things to do with the Fourier transform is to increase/decrease the effect of certain frequencies such as when using an equalizer. This is performed by computing the Fourier transform, multiplying by some function and then taking the inverse Fourier transform, i.e.,

$$f \mapsto \mathcal{F}^{-1}(m \cdot \mathcal{F}(f)). \tag{1.2.1}$$

We call these operators *Fourier multipliers* or, when m is the indicator function of some interval, *band limiting operators*.

This process of multiplying an alternate representation of f with a function before reconstructing it can be applied to the STFT too by introducing an extra function in (1.1.5). We call these operators *localization operators* or *time-frequency multipliers* and they take the form

$$A_m^{g_1,g_2} f = \int_{\mathbb{R}^2} m(z) V_{g_1} f(z) \pi(z) g_2 \, dz$$

where the integral should be interpreted weakly. The function m is called the *mask* or *symbol* of the operator and is often taken to be the indicator function of a set Ω, in which case we write A_Ω^g.

The similarities with (1.2.1) can be realized further by noting that the adjoint of $V_g : L^2(\mathbb{R}) \to L^2(\mathbb{R}^2)$ can be written as

$$V_g^* F = \int_{\mathbb{R}^2} F(z) \pi(z) g \, dz.$$

In view of this, the reconstruction formula (1.1.5) states that

$$\frac{1}{\langle g_1, g_2 \rangle} V_{g_2}^* V_{g_1} f = f$$

so $V_{g_2}^*$ is, up to a constant, the left inverse of V_g and $A_m^{g_1,g_2}$ can be written as

$$A_m^{g_1,g_2} f = V_{g_2}^* (m \cdot V_{g_1} f) \tag{1.2.2}$$

which is similar to (1.2.1).

From now on we will assume that $g_1 = g_2 = g$ and $\|g\|_{L^2} = 1$ which simplifies formulas and guarantees that A_m^g is self-adjoint as long as m is real valued. In this case $V_g^* V_g = I_{L^2(\mathbb{R})}$ is the identity operator on $L^2(\mathbb{R})$. Meanwhile the reverse operation $V_g V_g^* : L^2(\mathbb{R}) \to L^2(\mathbb{R})$ is not as well-behaved because the range of the STFT is not all of $L^2(\mathbb{R}^2)$. This difference to the Fourier transform, which is a bijection on $L^2(\mathbb{R})$, is crucial and is behind much of the depth of time-frequency analysis. We write $V_g(L^2(\mathbb{R})) \subset L^2(\mathbb{R}^2)$ for the range of the STFT and call it the *Gabor space*. The operation $V_g V_g^*$ is then the orthogonal projection onto $V_g(L^2)$ and the localization operator A_m^g is unitarily equivalent to a Toeplitz operator on $V_g(L^2)$, we call these operators *Gabor-Toeplitz operators*.

By the min-max principle, if $(\lambda_k)_k$ and $(h_k)_k$ are the eigenvalues and eigenfunctions of A_m^g, the eigenvalues can be expressed as

$$\lambda_k = \max_{\|f\|_{L^2}=1} \left\{ \int_{\mathbb{R}^2} m(z)|V_g f(z)|^2 \, dz : f \perp h_1, \ldots, h_{k-1} \right\}.$$

Since orthogonal functions have orthogonal short-time Fourier transforms by Moyal's identity (1.1.4), this means that the spectrograms of eigenfunctions try to live near the largest point of m while remaining orthogonal to all previous short-time Fourier transforms. For $m = \chi_\Omega$, we can fit approximately $|\Omega|$ spectrograms inside Ω, each with $\int_{\mathbb{R}^2} |V_g h_k(z)|^2 \, dz = 1$, while later spectrograms are supported primarily outside Ω. Consequently, the first $|\Omega|$ eigenvalues of such operators are close to 1, followed by a *plunge region* after which the remaining eigenvalues are close to 0. Several papers in the thesis investigate this eigenvalue behavior closer.

1.3 Quadratic time-frequency representations

The short-time Fourier transform is linear in its first argument and is therefore said to be a *linear* time-frequency distribution. As we saw earlier, the squared modulus of the STFT, the spectrogram, is useful in signal processing because it can be interpreted as a probability distribution. The spectrogram is just one example of the class of *quadratic* time-frequency distributions which arise from bilinear forms either as

$$Q(f) = B(f, f) \quad \text{or} \quad Q(f) = |B(f, g)|^2$$

for a fixed $g \in L^2(\mathbb{R})$. While quadratic time-frequency distributions have many nice properties, they suffer from so called "ghost terms" because

$$Q(f + g) = Q(f) + 2\,\mathrm{Re}(B(f, g)) + Q(g).$$

A bilinear form which underpins many of the time-frequency distributions we will see later is the *cross-Wigner distribution*, defined as

$$W(f, g)(x, \omega) = \int_{\mathbb{R}} f(x + t/2)\overline{g(x - t/2)}e^{-2\pi i \omega t} \, dt.$$

In the special case where $f = g$ we write $W(f, f) = W(f)$ and drop the word "cross". Wigner distributions generally behave similarly to short-time Fourier transforms due to the relation

$$W(f, g)(x, \omega) = 2e^{4\pi i x\omega} V_{\check{g}} f(2x, 2\omega)$$

where $\check{g}(t) = g(-t)$. Wigner distributions originated in the physics community in the 1930s, long before their connection to time-frequency analysis was uncovered.

There is a large collection of different quadratic time-frequency distributions that are suitable for applications. A simple argument shows that all bilinear maps $Q : L^2(\mathbb{R}) \times L^2(\mathbb{R}) \to C(\mathbb{R}^2)$ which are bounded and respect translations in the sense that

$$Q(\pi(z)f, \pi(z)g)(z') = Q(f, g)(z + z')$$

can be characterized as being of the form

$$Q(f, g)(z) = W(f, g) * \Phi$$

where Φ is some tempered distribution. This class of distributions is called *Cohen's class of quadratic time-frequency distributions* and shows up in several of the papers in this thesis.

1.4 Discrete counterparts

All of the constructions we have discussed so far have been in the continuous setting, however, in applications we need to realize objects discretely. This section is devoted to setting up these discrete constructions and discussing how they relate to their continuous counterparts.

To be precise, there are actually two levels of discrete constructions in time-frequency analysis. In what you might call the hybrid setting, the signal f and window g are still treated as elements of $L^2(\mathbb{R})$ but we sample the STFT on a lattice Λ of the form $A\mathbb{Z}^2$ for some $A \in GL_2(\mathbb{R})$ instead of integrating over $\mathbb{R}^2$. To make sure there is some sort of equivalence between this and the continuous setting, we always assume that the collection $(\pi(\lambda)g)_{\lambda\in\Lambda}$ is a *frame* for $L^2(\mathbb{R})$. This means that there exist constants $A, B > 0$ such that

$$A\|f\|_{L^2}^2 \leq \sum_{\lambda\in\Lambda} |V_g f(\lambda)|^2 \leq B\|f\|_{L^2}^2$$

for all $f \in L^2(\mathbb{R})$. This is an approximate and discrete version of Moyal's identity. Such a frame is called a *Gabor frame* because it is induced by the time-frequency shifts $\pi(\lambda)$. In the special case where $A = B$, we say that the frame is tight which unlocks some additional tools.

Whenever the window g induces a Gabor frame, we know that there exists a *dual window* h such that we have the reconstruction formula

$$f = \sum_{\lambda \in \Lambda} V_g f(\lambda)\pi(\lambda)h. \tag{1.4.1}$$

From this relation we can construct the discrete counterpart to localization operators, *Gabor multipliers*, which are defined as

$$G_m^{g,h} f = \sum_{\lambda \in \Lambda} m(\lambda) V_g f(\lambda)\pi(\lambda)h.$$

When the frame is tight, the dual window is equal to g up to a constant factor and we write $G_m^{g,g} = G_m^g$ for the associated (self-adjoint) operator. Two of the papers in the thesis investigate the relation between A_m^g and G_m^g closer. Specifically we will show how G_m^g converges to A_m^g in a specific operator norm when we increase the density of the lattice and how m can be approximated from the eigenvalues and eigenvectors of G_m^g through a similar procedure as for A_m^g.

In the fully discrete setting, the functions are instead vectors in $\mathbb{R}^N$ which we need to properly normalize to set up integrals and inner products. If the vectors are samples of functions, these quantities will converge to the corresponding continuous integrals and inner products assuming certain regularity conditions on the functions. The fully discrete version of Gabor multipliers are called *frame multipliers* and are used for the numerics in several papers in the thesis.

1.5 Weyl quantization

The phase space of time and frequency which we have been focused on so far is not the only possible notion of phase space. In quantum mechanics, phase space is often that of position and momentum and mathematically they have the same relations as those between time and frequency. In fact, objects such as the Wigner distribution and localization operators were introduced in physics contexts long before their usage in time-frequency analysis [27, 215]. In quantum mechanics, a central theme is that quantities such as position and momentum should not be seen as functions on phase space but rather as operators which act on the wave function of the system.

The transition from classical functions on phase space for position and momentum to the corresponding operators induces a map from functions on $\mathbb{R}^2$ to operators on $L^2(\mathbb{R})$. This map is called *Weyl quantization* and is a bijective isometry to the space of Hilbert-Schmidt operators, $A : L^2(\mathbb{R}^2) \to \mathcal{S}^2(L^2(\mathbb{R}))$. We can define the quantization of $f \in L^2(\mathbb{R}^2)$ as the unique bounded operator A_f which satisfies the weak relation

$$\langle A_f \psi, \phi \rangle_{L^2(\mathbb{R})} = \langle f, W(\phi, \psi) \rangle_{L^2(\mathbb{R}^2)} \tag{1.5.1}$$

for all $\psi, \phi \in L^2(\mathbb{R})$.

1.6 Abstract time-frequency analysis

In abstract harmonic analysis one is in general interested in generalizing Fourier analysis beyond the standard setting of functions on $\mathbb{R}^d$. Time-frequency analysis is influenced by the choice of analyzing functions on $L^2(\mathbb{R})$ but the key object is the time-frequency shifts $\pi : \mathbb{R}^2 \to \mathcal{U}(L^2(\mathbb{R}))$. In this section we will discuss the principal generalization of this as well as a fully general formulation of the ideas of time-frequency analysis.

The theory of wavelets was developed in the 1980s with the idea that instead of taking inner products with time-frequency shifted versions of a window, we should take inner products with *time-scale* shifted windows, called wavelets. These time-scale shifts are of the form

$$U(a,x)f(t) = \frac{1}{\sqrt{a}} f\left(\frac{t-x}{a}\right)$$

for a dilation parameter $a > 0$ and a time parameter $x \in \mathbb{R}$. Consequently the analog of the short-time Fourier transform, the *wavelet transform*, is a function on $\mathbb{R}^+ \times \mathbb{R}$.

The theory still works out so that functions on $L^2(\mathbb{R})$ are analyzed but the windows g need to satisfy the additional *admissibility* condition

$$\int_{\mathbb{R}} \frac{|\hat{g}(\omega)|^2}{|\omega|} \, d\omega < \infty. \tag{1.6.1}$$

In the language of representation theory, the most important property of π and U are that they are unitary square integrable projective irreducible representations of locally compact groups. There is a generalization of Moyal's identity, called the *Duflo-Moore theorem*, which applies in this situation. Let $\sigma : G \to \mathcal{U}(\mathcal{H})$ be a unitary square integrable representation of the group G on the Hilbert space $\mathcal{H}$, the theorem then states that

$$\langle \mathcal{W}_{g_1} f_1, \mathcal{W}_{g_2} f_2 \rangle_{L^2(G)} = \langle f_1, f_2 \rangle_{\mathcal{H}} \overline{\langle \mathcal{D}^{-1} g_1, \mathcal{D}^{-1} g_2 \rangle_{\mathcal{H}}}$$

where $\mathcal{W}_g f(x) = \langle f, \sigma(x)^* g \rangle_{\mathcal{H}}$ is the Wavelet transform, $L^2(G)$ is the space of square integrable functions on G with respect to the right Haar measure and $\mathcal{D}^{-1}$

is the (generally unbounded, densely defined, invertible) *Duflo-Moore* operator, the existence of which is guaranteed by the theorem. The wavelet admissibility condition (1.6.1) can then be stated as the window g being in the domain of the Duflo-Moore operator $\mathcal{D}^{-1}$.

Later work following wavelets has to a large degree been focused on various two-dimensional generalizations of wavelet analysis while on the theoretical side, the topics of representation theory and group theory have become intertwined with that of time-frequency analysis. Much of the motivation also comes from modern physics where representations of these groups play a central role.

Chapter 2

Quantum Harmonic Analysis

Quantum harmonic analysis is a framework which allows us to apply the tools of classical harmonic analysis; translations, convolutions and Fourier transforms, to operators instead of functions. We will introduce the main objects and discuss how they can be induced, present the key results and discuss the connection to time-frequency analysis, without going into proofs or precise statements of results.

2.1 Defining operator convolutions

Suppose we want to define notions of convolutions and Fourier transforms for operators. Obviously we cannot just replace f and g with operators in

$$f * g(x) = \int_{\mathbb{R}} f(y)g(x-y)\,dy, \qquad \hat{f}(\omega) = \int_{\mathbb{R}} f(x)e^{-2\pi i x\omega}\,dx$$

since multiplications, integrals and translations are not operations which are defined for operators. The first step towards this is therefore to come up with analogs of these. The most simple of these is multiplication which we will informally replace by operator composition. For integrals we can take a slightly more algebraic approach. On $L^1(\mathbb{R})$, the integral functional $I : f \mapsto \int_{\mathbb{R}} f(x)\,dx$ is linear, preserves positivity and is faithful in the sense that $I(|f|^2) = 0$ if and only if $f = 0$. Meanwhile for operators, the *trace* operation, defined as

$$\operatorname{tr}(A) = \sum_{n=1}^{\infty} \langle Ae_n, e_n \rangle \tag{2.1.1}$$

for any orthonormal basis $(e_n)_n$ of $L^2(\mathbb{R})$, is uniquely determined by it being linear, preserving positivity, being cyclically invariant and faithful. The cyclic invariance means that $\mathrm{tr}(A_1 \cdots A_n) = \mathrm{tr}(A_{\sigma(1)} \cdots A_{\sigma(n)})$ for any cyclic permutation $\sigma : \{1, \dots, n\} \to \{1, \dots, n\}$.

Just as we can only compute the integral of a function if it is integrable, that is $I(|f|) < \infty$, we can only compute the trace of an operator if it is of *trace-class* meaning that $\mathrm{tr}(|A|) < \infty$. Here $|A| = \sqrt{A^*A}$ is defined through functional calculus and similarly we can define powers of $|A|$ as $|A|^p := (A^*A)^{p/2}$. It is with these notions that the *Schatten p-classes of operators* are defined:

$$\mathcal{S}^p = \left\{A : \mathrm{tr}(|A|^p) < \infty\right\}, \qquad \|A\|_{\mathcal{S}^p} = \mathrm{tr}\left(|A|^p\right)^{1/p}.$$

Equivalently, if we expand A in its singular value decomposition

$$A = \sum_n a_n (f_n^1 \otimes f_n^2)$$

where $(f_n^1 \otimes f_n^2) : g \mapsto \langle g, f_n^2 \rangle f_n^1$ is a rank-one operator and $(a_n)_n$ are the singular values of A, then $\|A\|_{\mathcal{S}^p} = \|(a_n)_n\|_{\ell^p}$. The $p = 1$ and $p = 2$ cases are of special interest and $\mathcal{S}^1$ operators are said to be of *trace-class* while operators in $\mathcal{S}^2$ are called *Hilbert-Schmidt* operators. The interest stems partly from that $\mathcal{S}^2$ is a Hilbert space when equipped with the inner product

$$\langle A, B \rangle_{\mathcal{S}^2} = \mathrm{tr}(AB^*).$$

Just like for the L^p spaces, the dual space of $\mathcal{S}^p$ is $\mathcal{S}^q$ where $\frac{1}{p} + \frac{1}{q} = 1$ for $1 \leq p \leq \infty$. We choose to define the dual pairing $\langle \cdot, \cdot \rangle_{\mathcal{S}^p, \mathcal{S}^q}$ to be sesquilinear, i.e., $\langle A, B \rangle_{\mathcal{S}^p, \mathcal{S}^q} = \mathrm{tr}(AB^*)$ in order to stay consistent with the convention for L^p spaces.

There are several viable ways to define translations of operators but for us translations will be based on the time-frequency shifts discussed earlier. Notably these are parametrized by phase space $\mathbb{R}^2$. Specifically, we will define the operator translation α_z as conjugation by time-frequency shifts, that is,

$$\alpha_z(A) = \pi(z) A \pi(z)^*.$$

With multiplication, integrals and translations properly generalized, we move to looking at how convolution can be generalized. Note that we can write the usual convolution operation between $f, g \in L^1(\mathbb{R}^2)$ in two equivalent ways:

$$f * g = \int_{\mathbb{R}^2} f(z) T_z g \, dz, \tag{2.1.2}$$

$$f * g(z) = \int_{\mathbb{R}^2} f(z') T_z \check{g}(z') \, dz'. \tag{2.1.3}$$

Note that (2.1.2) should be interpreted as a Bochner integral taking values in $L^1(\mathbb{R}^2)$ since it is a weighted sum of $T_z g$. If we replace the function g by a trace-class operator S and the translation $T_z g$ by $\alpha_z(S)$, this instead becomes a Bochner integral taking values in $\mathcal{S}^1$, specifically we can set

$$f \star S = \int_{\mathbb{R}^2} f(z)\alpha_z(S)\, dz.$$

This integral is well-defined since $\|\alpha_z(S)\|_{\mathcal{S}^1} = \|S\|_{\mathcal{S}^1}$ and we will call it a *function-operator convolution*. Notably convolving a function with an operator gives a new operator.

Analogously, to generalize (2.1.3), we replace f and g by trace-class operators T and S. Since the integrand now is a trace-class operator, we must replace the Lebesgue integral by a trace. We will use a check $\check{}$ on S to denote conjugation with the parity operator P, defined as $Pf(t) = \check{f}(t) = f(-t)$. This means that we end up with

$$T \star S(z) = \operatorname{tr}(T\alpha_z(\check{S}))$$

as our definition for *operator-operator convolutions*. Here we take special note that convolving two operators gives a function.

Through some standard estimates on Bochner integrals, interpolation and general bounds on traces, we can establish the following two inequalities which mirror the classical Young's convolution inequality for functions

$$\begin{aligned} \|f \star S\|_{\mathcal{S}^r} &\leq \|f\|_{L^p} \|S\|_{\mathcal{S}^q}, \\ \|T \star S\|_{L^r} &\leq \|T\|_{\mathcal{S}^p} \|S\|_{\mathcal{S}^q} \end{aligned} \tag{2.1.4}$$

for $f \in L^p(\mathbb{R}^2)$, $S \in \mathcal{S}^q$, $T \in \mathcal{S}^p$ and $\frac{1}{p} + \frac{1}{q} = 1 + \frac{1}{r}$. The convolutions also share other nice properties with function-function convolutions such as preserving positivity, associativity and preserving integrals in the sense that

$$\operatorname{tr}(f \star S) = \left(\int_{\mathbb{R}^2} f(z)\, dz\right) \operatorname{tr}(S),$$

$$\int_{\mathbb{R}^2} T \star S(z)\, dz = \operatorname{tr}(T)\operatorname{tr}(S).$$

The last property of operator convolutions which we will discuss is that they are adjoints of each other. Specifically, the two mappings

$$\begin{aligned} \mathcal{A}_S &: L^p(\mathbb{R}^2) \to \mathcal{S}^p, \qquad && f \mapsto f \star S, \\ \mathcal{B}_S &: \mathcal{S}^p \to L^p(\mathbb{R}^2), \qquad && T \mapsto T \star S \end{aligned}$$

satisfy $\mathcal{A}_S^* = \mathcal{B}_S$ and $\mathcal{B}_S^* = \mathcal{A}_S$. This means that function-operator convolutions induce operator-operator convolutions and vice-versa. This compatibility should inspire some confidence that these are the "correct" notions of operator convolutions. Later on we will see additional ways to define these convolutions, providing additional evidence that these definitions are canonical.

2.2 Phase space and appropriate Fourier transforms

In the function case, convolutions are closely linked to the Fourier transform. Indeed, via the convolution theorem

$$\mathcal{F}(f * g)(\omega) = \hat{f}(\omega) \cdot \hat{g}(\omega), \tag{2.2.1}$$

convolution is just multiplication in the Fourier domain. We want to find the appropriate Fourier transforms which respect our operator convolutions. The first step in this direction is noting that we really are looking for two Fourier transforms, one acting on functions and another acting on operators. Both of these should map to functions on phase space in order for them to be compatible.

For a Fourier transform for operators, there is strong precedent from quantum mechanics in the form of the *Fourier-Wigner transform*, also known as the Fourier-Weyl transform, Weyl transform and Wigner-Weyl transform. With our time-frequency shifts π we can write it as

$$\mathcal{F}_W(S)(z) = e^{-\pi i x \omega} \operatorname{tr}(\pi(-z)S).$$

A more natural formulation using the *Schrödinger representation* $\rho(z) = e^{-\pi i x \omega}\pi(z)$ is

$$\mathcal{F}_W(S)(z) = \left\langle S, \rho(z) \right\rangle_{\mathcal{S}^1, \mathcal{S}^\infty} = \operatorname{tr}\left(S\rho(z)^*\right)$$

which is similar to the inner product formulation of the Fourier transform (1.1.3). The Fourier-Wigner transform of a trace-class operator S is a function on phase space $\mathbb{R}^2$ and there is a Riemann-Lebesgue lemma; $\mathcal{F}_W(\mathcal{S}^1) \subset C_0(\mathbb{R}^2)$. Moreover, $\mathcal{F}_W$ is unitary and can be extended to a unitary isometry from $\mathcal{S}^2$ to $L^2(\mathbb{R}^2)$.

For a Fourier transform on functions, we could use the standard two-dimensional Fourier transform. However, phase space $\mathbb{R}^2$ should really be seen as a subspace of the Weyl-Heisenberg group which has a symplectic structure. For this reason, we should use the *symplectic Fourier transform*, given by

$$\mathcal{F}_\sigma(f)(z) = \int_{\mathbb{R}^2} f(z')e^{-2\pi i \sigma(z,z')}\, dz' \tag{2.2.2}$$

where $\sigma(z, z') = \omega \cdot x' - \omega' \cdot x$ is the *standard symplectic form* on $\mathbb{R}^2$.

These two Fourier transforms play well with the operator convolutions we have defined in that they satisfy the equalities

$$\mathcal{F}_W(f \star S) = \mathcal{F}_\sigma(f) \cdot \mathcal{F}_W(S),$$
$$\mathcal{F}_\sigma(T \star S) = \mathcal{F}_W(T) \cdot \mathcal{F}_W(S)$$

for $f \in L^1(\mathbb{R}^2)$ and $T, S \in \mathcal{S}^1$, which correspond to (2.2.1).

2.3 Weyl quantization

While we can define Weyl quantization using the weak relation (1.5.1), there is an alternative formulation using the two Fourier transforms we have just introduced which highlights the connection to quantum harmonic analysis. With $A^{-1} = a : S \mapsto a_S$ denoting the dequantization map, we can summarize quantization in a commutative diagram as

$$\begin{tikzcd} \mathcal{S}^2(L^2(\mathbb{R})) \arrow[d, "\mathcal{F}_W"] \arrow[dr, dotted, "a"] & \\ L^2(\mathbb{R}^2) \arrow[r, "\mathcal{F}_\sigma"] & L^2(\mathbb{R}^2). \end{tikzcd}$$

Note that since both $\mathcal{F}_W$ and $\mathcal{F}_\sigma$ are unitary isometries, the full quantization mapping A is also a unitary isometry.

As quantum harmonic analysis investigates the interactions between functions and operators, and Weyl quantization is a map between functions and operators, it is plausible that they should have some nontrivial interactions. The first result in this direction is the fact that operator translation corresponds to translation of the Weyl symbol, i.e.,

$$\alpha_z(S) = A_{T_z a_S}. \tag{2.3.1}$$

Using the linearity of Weyl quantization, it is easy to see that this implies that

$$a_{f \star S} = f * a_S$$

meaning that a function-operator convolution is just a convolution with a Weyl symbol. We can hence write the operator $\mathcal{A}_S$ as $\mathcal{A}_S = AC_{a_S}$ where $A : L^2(\mathbb{R}^2) \to \mathcal{S}^2$ is the quantization map and C_{a_S} is the convolution operator with function a_S. In particular this implies that

$$\mathcal{B}_S = \mathcal{A}_S^* = (AC_{a_S})^* = C_{\check{a}_S} A^{-1} \implies T \star S = a_T * \check{a}_S \tag{2.3.2}$$

since $(C_{a_S})^* = C_{\check{a_S}}$.

Weyl quantization also induces the notion of operator parity $S \mapsto PSP$ from the definition of operator-operator convolutions as $A_{\check{f}} = \check{A_f} = PA_fP$, and turns adjoints into complex conjugation in the sense that $a_{S^*} = \overline{a_S}$.

The weak definition of Weyl quantization (1.5.1) we discussed earlier implies that the quantization of the Wigner distribution $W(f, g)$ is precisely the rank-one operator $(f \otimes g)$.

Paper F deals with generalizing the definitions and properties of Weyl quantization outside the Euclidean setting where we have less structure available to build on.

2.4 Three other ways to define operator convolutions

To further cement that the function-operator and operator-operator convolution definitions we have discussed are the correct ones, we present three additional constructions which all yield the same definitions.

2.4.1 Translations induce convolutions

If we only take our notion of operator translations α_z as given, we can recover the definitions of operator convolutions. The key property of convolutions which we want to generalize is that

$$\delta_z * f = T_z f$$

where δ_z is the Dirac delta measure at z. For function-function convolutions, the representation $\rho : \mathbb{R} \ni x \mapsto T_x \in B(L^1(\mathbb{R}))$ has some nice properties which we can use to extend the action $*_\rho : (x, f) \mapsto T_x f \in L^1(\mathbb{R})$ to pairs of *measures* and functions, $* : M(\mathbb{R}) \times L^1(\mathbb{R}) \to L^1(\mathbb{R})$. This construction yields the standard definition of function-function convolutions.

A similar procedure is detailed in Paper B where we use the representation $\rho : \mathbb{R}^2 \ni z \mapsto \alpha_z$ and the action

$$\star_\rho : \mathbb{R}^2 \times \mathcal{S}^1 \to \mathcal{S}^1, \quad (z, S) \mapsto \alpha_z(S).$$

It is shown that this α has many of the same properties as the ρ we used for function-function convolutions and through some functional analysis we can uniquely extend the action to measures so that

$$\mu \star S = \int_{\mathbb{R}^2} \alpha_z(S)\, d\mu(z).$$

Since $L^1(\mathbb{R}^2)$ can be embedded in $M(\mathbb{R}^2)$, this also provides a definition of function-operator convolutions. Now as we saw earlier, operator-operator convolutions can be defined as the adjoint of the mapping $\mathcal{A}_S : f \mapsto f \star S$ and so we have induced both operator convolutions.

2.4.2 Generalizing the convolution theorem

Instead of taking our definitions of operator convolutions as the starting points, we can start with the Fourier transforms $\mathcal{F}_W, \mathcal{F}_\sigma$ and induce all of quantum harmonic analysis by defining convolution as the map which respects the familiar convolution theorem $\mathcal{F}(f * g)(\omega) = \hat{f}(\omega) \cdot \hat{g}(\omega)$ (2.2.1). To do so, we must assume that function-operator convolutions result in operators, operator-operator convolutions result in functions and that the Fourier-Wigner and symplectic Fourier transforms are the appropriate Fourier transforms for operators and functions respectively. We can then construct convolutions such that the convolution theorem holds meaning that

$$\mathcal{F}_W(f \star S) = \mathcal{F}_\sigma(f) \cdot \mathcal{F}_W(S) \iff f \star S = \mathcal{F}_W^{-1}\big(\mathcal{F}_\sigma(f) \cdot \mathcal{F}_W(S)\big)$$

and

$$\mathcal{F}_\sigma(T \star S) = \mathcal{F}_W(T) \cdot \mathcal{F}_W(S) \iff T \star S = \mathcal{F}_\sigma^{-1}\big(\mathcal{F}_W(T) \cdot \mathcal{F}_W(S)\big).$$

Note that the argument of the inverse Fourier-Wigner transform is integrable when $f \in L^2(\mathbb{R}^2)$ and $S \in \mathcal{S}^2$. Similarly when $T, S \in \mathcal{S}^2$, the argument of the inverse symplectic Fourier transform is also integrable. Extending to other L^p and $\mathcal{S}^p$ spaces can be done through standard methods.

2.4.3 Respecting Weyl quantization

In Section 2.3 we saw how operator convolutions can be computed using Weyl quantization. We can turn this around and instead define

$$T \star S = a_T * a_S, \qquad f \star S = A_{f * a_S}.$$

Weyl quantization as its own topic goes back nearly a hundred years and this is arguably the most "external" way to define operator convolutions as we make no explicit assumptions on what translations or integrals should correspond to for operators.

2.5 The time-frequency connection

So far, the only common object between time-frequency analysis and quantum harmonic analysis has been the time-frequency shifts π. As we saw in the previous section, these are enough to induce operator convolutions. Since these objects also play a key role in time-frequency analysis, we are able to realize some objects from time-frequency analysis as operator convolutions. In particular, we will be able to write localization operators and Cohen's class distributions as function-operator and operator-operator convolutions.

If we write out the weak action of a rank-one function-operator convolution, we see that it is of the form

$$\begin{aligned}\big\langle m \star (g_1 \otimes g_2) f, h \big\rangle &= \int_{\mathbb{R}^2} m(z) \langle \pi(z)(g_1 \otimes g_2)\pi(z)^* f, h \rangle \, dz \\ &= \int_{\mathbb{R}^2} m(z) \langle \pi(z)^* f, g_2 \rangle \langle \pi(z) g_1, h \rangle \, dz \\ &= \int_{\mathbb{R}^2} m(z) V_{g_2} f(z) \overline{V_{g_1} h(z)} \, dz = \big\langle A_m^{g_2, g_1} f, h \big\rangle.\end{aligned}$$

This means that a general function-operator convolution can be written as a sum of localization operators, all with the same symbol. Indeed, if we expand a trace-class operator S in its singular value decomposition, we can see that

$$S = \sum_n s_n (g_n^1 \otimes g_n^2) \implies m \star S = \sum_n s_n A_m^{g_n^2, g_n^1}.$$

These types of operators are sometimes called *mixed-state* localization operators or *multiwindow* localization operators and due to this relation, quantum harmonic analysis allows us to translate results from localization operators to these operators and vice versa.

The simplest operator-operator convolution is between two self-adjoint rank-one operators $(f \otimes f)$ and $(g \otimes g)$. We can compute this as

$$(f \otimes f) \star (g \otimes g)(z) = \operatorname{tr}\big((f \otimes f)\pi(z)(\check{g} \otimes \check{g})\pi(z)^*\big) = |V_{\check{g}} f(z)|^2$$

by expanding the trace. Specifically this implies that $(f \otimes f) \star (\check{g} \otimes \check{g}) = |V_g f|^2$, the spectrogram.

The next step in adding complexity to operator-operator convolutions is considering general convolutions of the form

$$(f \otimes f) \star S = W(f) * \check{a_S}. \tag{2.5.1}$$

Here the equality follows from the rules for Weyl quantizations of convolutions (2.3.2) and the fact that $f \otimes f$ is the quantization of the Wigner distribution $W(f)$.

Notably, this means that all Cohen's class distributions can be realized as operator-operator convolutions. This approach means that we can apply many of the same tools we use for studying the spectrogram to Cohen's class distributions.

Chapter 3

Summary of Papers

3.1 Paper A—Quantum Harmonic Analysis on Locally Compact Groups [115]

By replacing time-frequency shifts π, phase space $\mathbb{R}^2$ and the associated Hilbert space $L^2(\mathbb{R})$ with a unitary representation σ of a locally compact group G on some Hilbert space $\mathcal{H}$, just as we do for the abstract time-frequency analysis discussed in Section 1.6, we can set up a notion of abstract quantum harmonic analysis. This paper takes these objects as its starting point and works out the full theory based on the operator convolutions

$$
\begin{aligned}
f \star_G S &= \int_G f(x)\sigma(x)^* S\sigma(x)\, d\mu_r(x),\\
T \star_G S(x) &= \operatorname{tr}\left(T\sigma(x)^* S\sigma(x)\right).
\end{aligned}
$$

Note that we place the adjoint to the left of the operator when considering the operator translation $\alpha_x(S) = \sigma(x)^* S\sigma(x)$ and that we use the right Haar measure μ_r when integrating over the group G. These conventions help in the locally compact setting by allowing us to avoid the use of the parity operator which is problematic there.

It turns out that most of the basic properties of operator convolutions hold in this setting with a key difference to the time-frequency case being that

$$
\int_G T \star_G S(x)\, d\mu_r(x) = \operatorname{tr}(T) \operatorname{tr}\left(\mathcal{D}^{-1} S \mathcal{D}^{-1}\right).
$$

We say that when the second quantity is finite, or more generally when $\mathcal{D}^{-1} S \mathcal{D}^{-1}$ is trace-class, the operator S is *admissible*.

Having worked out the necessary theory, the paper goes on to generalize deeper parts of quantum harmonic analysis such as Young's inequality (2.1.4), the properties of Cohen's class distributions (2.5.1), the mapping $\mathcal{A}_S : f \mapsto f \star_G S$ as a quantization mapping and the eigenvalues of mixed-state localization operators. As such, it lays the groundwork for future work and is meant to be a standard reference for these results in full generality.

3.2 Paper B—Measure-Operator Convolutions and Applications to Mixed-State Gabor Multipliers [81]

The scheme to define measure-operator convolutions starting from the operator translations discussed in Section 2.4.1 is developed in this paper. After showing that the action $(z, S) \mapsto \alpha_z(S)$ can be uniquely extended to pairs of measures and operators, we show that the resulting map extends function-operator convolutions and that it can be further extended as a map from $\star : M(\mathbb{R}^{2d}) \times \mathcal{S}^p$ to $\mathcal{S}^p$ for $1 \leq p \leq \infty$. In this setting, we also prove that all the standard properties of function-operator convolutions such as the convolution theorem and Young's inequality extend nicely to measure-operator convolutions.

The motivation for defining measure-operator convolutions comes from the lattice setting of Gabor multipliers where discrete sums take the place of continuous integrals. Such operators can be encoded by letting the measure be the sum of Dirac deltas so that for $c \in \ell^1(\Lambda)$,

$$c \star S = \sum_{\lambda \in \Lambda} c(\lambda)\pi(\lambda)S\pi(\lambda)^*.$$

This connection allows us to prove results on Gabor multipliers and the mixed-state generalization, mixed-state Gabor multipliers. The most important set of results we prove are centered around the relation between localization operators and the parametrized discretizations

$$G^g_{m,\alpha,\beta} f = \alpha\beta \sum_{n,k \in \mathbb{Z}} m(n\alpha, k\beta) V_g f(n\alpha, k\beta)\pi(n\alpha, k\beta)g.$$

Specifically, we are able to give conditions on m and g such that

$$\left\| G^g_{m,\alpha,\beta} - A^g_m \right\|_{\mathcal{S}^1} \to 0 \qquad \text{as } \alpha, \beta \to 0, \tag{3.2.1}$$

as well as conditions on the sequences $(m_n)_n, (g_n)_n, (\alpha_n)_n, (\beta_n)_n$ so that

$$\left\| G^{g_n}_{m_n,\alpha_n,\beta_n} - G^g_{m,\alpha,\beta} \right\|_{\mathcal{S}^1} \to 0 \qquad \text{as } n \to \infty.$$

These results are stronger than those that have been available in the literature previously. In particular, (3.2.1) hints that the spectral properties of Gabor multipliers should correspond well to those of localization operators which is important for computational applications.

3.3 Paper C—Weyl Quantization of Exponential Lie Groups for Square Integrable Representations [33]

Over the years, there have been many efforts to develop Weyl quantization, discussed in Section 2.3, beyond the Weyl-Heisenberg group. Instead of looking closely at how specific formulas for the Wigner distribution can be extended, this paper takes the view that it is the symplectic Fourier transform $\mathcal{F}_\sigma$ which should be generalized. This line of work was started in [31] for the affine group and together with one of the authors of that paper, we are able to generalize that approach to a general class of groups.

Specifically, with the Fourier-Wigner transform $\mathcal{F}_W(A)(x) = \mathrm{tr}(A\mathcal{D}\pi(x))$ and the symplectic Fourier transform (2.2.2) replaced by the *Fourier-Kirillov transform* $\mathcal{F}_{\mathrm{KO}}$, we set up the quantization mapping as

$$A_f = \mathcal{F}_W^{-1}(\mathcal{F}_{KO}^{-1}(f)), \qquad a_S = \mathcal{F}_{KO}(\mathcal{F}_W(S)).$$

We show that A and its inverse are unitary isometric bijections before moving to more specific properties of the mapping. The operator translation from quantum harmonic analysis corresponds to translation of the Weyl symbol in the same way as in (2.3.1) and $a_{S^*} = \overline{a_S}$. As a consequence, we have the same form of correspondence between the operator convolutions and Weyl quantization as for the Weyl-Heisenberg group which enables us to use the quantization map for quantum harmonic analysis on more general groups. We also investigate the associated Wigner distributions, the dequantizations of rank-one operators, in detail and find that they shares many properties with the Wigner distributions.

3.4 Paper D—Five Ways to Recover the Symbol of a Non-Binary Localization Operator [118]

The localization operators introduced in Section 1.2 are defined by two objects, the mask m and the window g. This paper aims to answer the question of how we can recover information about m from A_m^g. Previously, Abreu, Gröchenig and Romero [8] had shown that if $m = \chi_\Omega$ and $\sum_{k=1}^\infty \lambda_k^\Omega (h_k^\Omega \otimes h_k^\Omega)$ is the singular value decomposition of A_Ω^g, then χ_Ω can be approximated as

$$\left\| \sum_{k=1}^{\lceil |\Omega| \rceil} |V_g h_k^\Omega|^2 - \chi_\Omega \right\|_{L^1} \leq C_g |\partial \Omega| \tag{3.4.1}$$

for some constant C_g dependent only on g. Romero and Speckbacher [182] later showed that Ω can be estimated by averaging spectrograms of white noise under A_Ω^g using the estimator

$$\frac{1}{K} \sum_{k=1}^{K} |V_g(A_\Omega^g \mathcal{N})(z)|^2 \tag{3.4.2}$$

where K is the number of white noise samples and $\mathcal{N}$ is white noise.

This paper treats the full case when $m \neq \chi_\Omega$ and proposes five new ways to solve this problem, two based on the two methods for indicator functions just mentioned and three completely novel ones. The five estimators are as follows:

1. The *weighted accumulated spectrogram* is similar to the sum in (3.4.1) but with a factor λ_k in each term, i.e.,

   $$\sum_k \lambda_k |V_g h_k^\Omega(z)|^2. \tag{3.4.3}$$

   If the sum is over all k, this quantity is equal to $m * |V_g g|^2$, while for truncated sums we can control the L^1 error.

2. The *weighted accumulated Wigner distribution* is the same thing as (3.4.3) but with the Wigner distribution of the eigenfunction replacing the spectrogram, i.e.,

   $$\sum_k \lambda_k W(h_k)(z).$$

3. Our *white noise estimator* is precisely (3.4.2) and in the paper we show that this estimator works well for general m and determine error bounds.

4. If we sum up the images of each element of an orthonormal basis for $L^2(\mathbb{R})$, we get the *plane tiling estimator*

   $$\sum_n |V_g(A_m^g e_n)(z)|^2$$

   which is a good estimator of m^2 and is equal to the $K \to \infty$ limit of (3.4.2).

5. We can sample $m * |V_g g|^2$ pointwise using the relation

$$V_g(A_m^g(\pi(z)g))(z) = m * |V_g g|^2(z)$$

   which we call the *Gabor projection estimator*. This can be shown to be equivalent to projecting the product of m and $V_g g(\cdot - z)$ onto the Gabor space $V_g(L^2(\mathbb{R}))$ for each z.

We also make MATLAB implementations of all five methods available and evaluate them on a series of examples in the paper.

3.5 Paper E—On Accumulated Spectrograms for Gabor Frames [119]

The accumulated spectrogram result (3.4.1) from [8] was widely influential. Necessarily, any attempt to realize it numerically is based on Gabor multipliers and Gabor frames instead of localization operators. This paper establishes results analogous to those in [8] and the later improvements in [4] for this setting. Specifically, if $(\pi(\lambda)g)_{\lambda\in\Lambda}$ is a Gabor frame with frame bounds A, B and $\sum_{k=1}^{\infty} \lambda_k^{\Omega}(h_k^{\Omega} \otimes h_k^{\Omega})$ is the singular value decomposition of the Gabor multiplier G_{Ω}^g, we use the estimator

$$\rho_\Omega(\lambda) = \frac{1}{\|g\|_{L^2}^2} \sum_{k=1}^{A_\Omega} |V_g h_k^\Omega(\lambda)|^2, \qquad A_\Omega = \left\lceil \frac{\#(\Omega\cap\Lambda)\|g\|_{L^2}^2}{B} \right\rceil.$$

For this estimator, we have the error estimate

$$\|\rho_\Omega - \chi_\Omega\|_{\ell^1(\Lambda)} \leq C_g \#\partial_\Lambda^{r_\Lambda}\Omega + 2\frac{B-A}{B}\#(\Omega\cap\Lambda) + \frac{B}{\|g\|_{L^2}^2}$$

where C_g is a constant depending on g, r_Λ is a constant depending on Λ and $\partial_\Lambda^{r_\Lambda}\Omega$ is a discrete version of the boundary defined as

$$\partial_\Lambda^{r_\Lambda}\Omega = \Lambda \cap (\partial\Omega + B(0, r_\Lambda)).$$

As was done in [4], we also go on to show through an example that this error estimate is sharp in general.

While some of the techniques from [8] and [4] can be reused, large parts of the proof are reworked to account for lattice effects. Together with Paper B, this paper lays some of the groundwork putting Gabor multipliers on the same solid theoretical footing as localization operators.

3.6 Paper F—On a Time-Frequency Blurring Operator with Applications in Data Augmentation [116]

Motivated by the formulation of localization operators as multiplication operators in the Gabor domain (1.2.2), this paper looks at convolution operators in the Gabor domain. Specifically, we consider the *time-frequency blurring operator*

$$B_\mu^g f = V_g^*(\mu * V_g f)$$

where $\mu \in M(\mathbb{R}^2)$ is the *kernel*. One way to think of the action of this operator is as a blurring or spreading of the signal in time-frequency space, especially when μ is something like a Gaussian function.

In the paper we investigate mapping properties between L^p and M^p spaces and show that the operator is non-compact as a bounded operator on $L^2(\mathbb{R})$. However our interest in the operator, apart from its fundamental construction, is motivated by its use as a data augmentation tool for signals. Data augmentation is employed when we are training a machine learning model and are limited by the size of our training data and wish to construct additional examples with the same labels. Standard methods include adding white noise to a signal or applying a localization operator. On a standard speech recognition benchmark, we investigate how the performance of two models improves as we add signals augmented by the time-frequency blurring operator with Gaussian kernel to the training sets. We are able to show statistically significant improvements when using this augmentation method either along or together with other standard methods for both models and various sizes of the training sets.

3.7 Paper G—Empirical Plunge Profiles of Time-Frequency Localization Operators [117]

While the eigenvalues of localization operators are a well-studied topic, specific details of their behavior are sparse. It is only in the case where the symbol is a disk that we have an explicit expression for them. On a very high level, we know that the localization operator A_Ω^g will have approximately $|\Omega|$ eigenvalues close to 1, followed by a short plunge region after which the remaining eigenvalues are close to 0. From the expression for eigenvalues when the symbol is a disk we first conclude that

$$\left|\lambda_k^{B(0,R)} - \frac{1}{2}\operatorname{erfc}\left(\sqrt{2\pi}\frac{k - |B(0,R)|}{|\partial B(0,R)|}\right)\right| = O\left(\frac{1}{R}\right)$$

and then lift this result to show that the same relation holds for any radial symbol Ω with $B(0,R)$ replaced by $R\Omega$.

The core of the paper is the conjecture that this asymptotic holds for *all* symbols Ω, provided the window function g is the standard Gaussian. Beyond the verification of the conjecture in the radial case, the paper investigates the eigenvalues of frame multipliers for a variety of symbols. By computing the area and perimeter of these sets, the discrepancy between erfc and the actual eigenvalues can be computed and it is found that for standard lattice parameters, the L^∞ error is mostly around 1% except in cases where the symbol is exceptionally challenging with a high $\frac{|\partial\Omega|}{|\Omega|}$ ratio.

Part II

Research Papers

Paper A

Quantum Harmonic Analysis on Locally Compact Groups

Simon Halvdansson

Published in
Journal of Functional Analysis,
2023, Volume 285, Issue 8

Paper A

Quantum Harmonic Analysis on Locally Compact Groups

Abstract

On a locally compact group we introduce covariant quantization schemes and analogs of phase space representations as well as mixed-state localization operators. These generalize corresponding notions for the affine group and the Heisenberg group. The approach is based on associating to a square integrable representation of the locally compact group two types of convolutions between integrable functions and trace-class operators. In the case of non-unimodular groups these convolutions only are well-defined for admissible operators, which is an extension of the notion of admissible wavelets as has been pointed out recently in the case of the affine group.

A.1 Introduction

The theory of time-frequency representations, aka phase space representations, and pseudo-differential operators has been extended from the Euclidean framework to various other settings by replacing the Schrödinger representation of the Weyl-Heisenberg group by a unitary representation of a locally compact group, see [143, 160] for recent contributions to this circle of ideas.

A particular class of pseudodifferential operators has been intensively studied in mathematics, quantum mechanics and time-frequency analysis which is known as Toeplitz operators, localization operators and anti-Wick quantization or covariant integral quantization in the respective fields, see [95, 96, 135, 150, 219, 220] for contributions to the non-Euclidean setting.

In [153] the authors have established a link between the theory of localization operators and quantum harmonic analysis on phase space, the latter had been introduced by Werner [211]. In [211] a convolution $f \star S$ between a Lebesgue integrable function f and trace-class operator S, and a convolution, $T \star S$, of two trace-class operators T, S are defined and shown to behave in a manner analogous to the convolution of two functions.

In a series of papers Luef and Skrettingland have demonstrated the merits of viewing localization operators as the convolution of a function and a rank-one operator [154–156]. Furthermore, it has been noticed in [154] that the time-frequency representations associated to the generalization of localization operators, $f \star S$, are Cohen's class distributions defined in terms of S. The formulation of statements in time-frequency analysis in terms of Werner's convolution has turned out to be very fruitful and has been extended to the affine group in [32].

Meanwhile, there has been interest in related problems in the more general setting of square integrable representations of locally compact groups. Examples of this include coorbit spaces [20, 29, 78–80, 181], localization operators [150, 219, 220], covariant integral quantizations [95, 96, 135], reproducing kernel Hilbert spaces [28] and sample reconstruction [88]. Consequently, it is the goal of this paper to set up the theory of quantum harmonic analysis on locally compact groups, inspired by the construction for the affine group in [32] and use it to establish notions and theorems from time-frequency analysis and time-scale analysis in the more general case of locally compact groups .

The main objects in quantum harmonic analysis are the function-operator and operator-operator convolutions, which we define in this paper for a locally compact group G, and a square integrable unitary representation σ on a Hilbert space $\mathcal{H}$ as

$$f \star S = \int_G f(x)\sigma(x)^* S\sigma(x)\, d\mu_r(x), \qquad T \star S(x) = \operatorname{tr}(T\sigma(x)^* S\sigma(x)),$$

where μ_r denotes the right-invariant Haar measure on G.

These definitions are motivated by the idea that the mapping $\alpha_x : S \mapsto \sigma(x)^* S\sigma(x)$ corresponds to a *translation* of an operator and that the trace measures the size of an operator in the same way as the integral measures the size of a function. It turns out that many properties of convolutions such as associativity and a version of Young's inequality hold mutantis mutadis for this type of convolutions, too.

Studying the issue of integrability of operator-operator convolutions leads one to a definition of *admissibility of operators*, generalizing the concept of admissible vectors for the wavelet transform. This criterion turns out to be important to applications and is the main source of discrepancy from the corresponding theory for the Weyl-Heisenberg case where all trace-class operators are admissible.

Most of the novel contributions of this paper are contained in Section A.5 including an uncertainty principle for Cohen's class distributions, results on the distribution of eigenvalues of mixed-state localization operators for the wavelet transform, Berezin-Lieb inequalities for both function-operator and operator-operator convolutions and a version of Wiener's Tauberian theorem giving equivalent conditions for translates of an operator to be dense in the Schatten classes $\mathcal{S}^p$ for $1 \leq p \leq \infty$. Hence, we are able to establish a theory of quantum harmonic analysis on a locally compact group G, which has just one deficiency compared to [211]; the lack of a multiplication theorem for the operator-valued Fourier transform.

Outline

In Section A.2, we go over some preliminaries on operator theory, time-frequency analysis and locally compact groups without discussing quantum harmonic analysis in too much depth. This section can be skipped over if the reader is familiar with works such as [32, 153–155]. Notably, Section A.2.4 lists three examples of square integrable representations which motivate the generalizations in the paper. In Section A.3, we define the three convolutions in quantum harmonic analysis; function-function, function-operator and operator-operator, and establish some elementary properties. Section A.4 is devoted to the construction and properties of admissible operators while Section A.5 goes through applications mainly related to the notion of admissible operators.

Notational conventions

Throughout this article, a general locally compact group will be denoted by G with the zero element denoted by 0_G, general elements denoted by x, y, z and the associated left and right Haar measures written as μ_ℓ and μ_r, respectively. Moreover, $\mathcal{H}$ will denote a Hilbert space, any norm without a subscript will be assumed to be taken in $\mathcal{H}$ and for an operator A, A^* will denote its adjoint. The set $\mathcal{U}(\mathcal{H})$ will denote the set of all unitary operators on $\mathcal{H}$ and the most notable members are unitary square integrable representations σ of G. For $p < \infty$, $\mathcal{S}^p$ will denote the Schatten p-class of operators with singular values in ℓ^p and by $\mathcal{S}^\infty$ we mean $B(\mathcal{H})$, the set of all bounded linear operators on $\mathcal{H}$. We will make extensive use of rank-one operators $\psi \otimes \phi : \xi \mapsto \langle \xi, \phi \rangle$ which we in the bra-ket formalism would write as $\langle \phi | \xi \rangle | \psi \rangle$.

A.2 Preliminaries

In this section we go over some of the preliminaries of quantum harmonic analysis, time-frequency analysis and locally compact groups. The exposition is similar to that in [32, 153, 155] and related works. See also [220] for an introduction more focused on the setting of locally compact groups.

A.2.1 Operator theory

Singular value decomposition

We will frequently make use of the singular value decomposition of a compact operator A which has the form

$$A = \sum_n s_n(A)(\psi_n \otimes \phi_n)$$

where $\{\psi_n\}_n$ and $\{\phi_n\}_n$ are two orthonormal sets in $\mathcal{H}$, $(s_n(A))_n$ is a sequence converging to zero and $\psi \otimes \phi$ is the *rank-one operator*, defined by $(\psi \otimes \phi)(\xi) = \langle \xi, \phi \rangle \psi$. The sum converges in the strong topology of $B(\mathcal{H})$ and the numbers $s_n(A)$ are the eigenvalues of $\sqrt{A^*A}$ and are called the singular values of A.

In case A is a positive, compact operator, then we can take $\phi_n = \psi_n$ and the singular values $s_n(A)$ agree with the eigenvalues $\lambda_n(A)$ of A. Hence, we have the following spectral decomposition of A:

$$A = \sum_n \lambda_n(A)(\psi_n \otimes \psi_n).$$

Schatten classes of operators

The Schatten class $\mathcal{S}^p$ is the space of compact operators with singular values in ℓ^p for $p \in [1, \infty]$. In particular, the space $\mathcal{S}^1$ is referred to as the space of *trace-class* operators and $\mathcal{S}^2$ as the space of *Hilbert-Schmidt* operators. It is a non-trivial fact that $\mathcal{S}^p$ is a Banach space for any $1 \leq p \leq \infty$ and that $\mathcal{S}^2$ is a Hilbert space with the inner product $\langle S, T \rangle_{\mathcal{S}^2} = \mathrm{tr}(ST^*)$. For a trace-class operator $S \in \mathcal{S}^1$, we define the *trace* of S as

$$\mathrm{tr}(S) = \sum_n \langle Se_n, e_n \rangle$$

where $\{e_n\}_n$ is an orthonormal basis. This quantity is finite and independent of the chosen orthonormal basis. Similarly, it turns out that the Schatten p-norm of $S \in \mathcal{S}^p$ can be shown to be equal to $\|S\|_{\mathcal{S}^p}^p = \mathrm{tr}(|S|^p)$ for $1 \leq p < \infty$ where $|S|$ is the absolute value of S.

In the same way as for L^p-spaces of functions, we define duality brackets for conjugate $\mathcal{S}^p$ spaces as

$$\langle A, B \rangle_{\mathcal{S}^p, \mathcal{S}^q} = \mathrm{tr}(AB)$$

where $A \in \mathcal{S}^p, B \in \mathcal{S}^q$ and $\frac{1}{p} + \frac{1}{q} = 1$.

Vector valued integration

In defining convolutions between functions and operators, we will need to integrate operator valued functions $H : G \to B(\mathcal{H})$ which are of the form $H(x) = f(x)F(x)$ where $f \in L^1_r(G)$ and $F : G \to B(\mathcal{H})$ are measurable, bounded and strongly continuous. The operator-valued integral of H is then defined weakly as

$$\left\langle \left(\int_G f(x)F(x)\, d\mu \right) \psi, \phi \right\rangle = \int_G f(x) \langle F(x)\psi, \phi \rangle \, d\mu$$

for $\psi, \phi \in \mathcal{H}$ where μ is a measure on G. For more on this matter of defining operators via integrals, known as Bochner integration, see the discussion in [153, Sec. 2.3].

A.2.2 Time-frequency analysis

We briefly introduce some of the main objects of time-frequency analysis. For a more thorough introduction, see e.g. [53, 107].

Short-time Fourier transform

Perhaps the most classical tool of time-frequency analysis is the following *time-frequency representation*: Given $\psi, \varphi \in L^2(\mathbb{R}^d)$, the *short-time Fourier transform* (STFT) of ψ with respect to the *window* φ is the function

$$V_\varphi \psi(x, \omega) = \int_{\mathbb{R}^d} \psi(t) \overline{\varphi(t - x)} e^{-2\pi i \omega \cdot t} \, dt$$

on $\mathbb{R}^{2d}$. It has shown to be useful to consider the STFT to be induced by the *(projective) representation* $\pi(x, \omega) = M_\omega T_x$ of the *Weyl-Heisenberg group* where

$$T_x f(t) = f(t - x), \qquad M_\omega f(t) = e^{2\pi i \omega t} f(t).$$

In this notation we can write the STFT as $V_\varphi \psi(x, \omega) = \langle \psi, \pi(x, \omega)\varphi \rangle$. The STFT is a member of $L^2(\mathbb{R}^{2d})$ which can be seen by an application of *Moyal's identity*

$$\left\langle V_{\varphi_1}\psi_1, V_{\varphi_2}\psi_2 \right\rangle_{L^2(\mathbb{R}^{2d})} = \langle \psi_1, \psi_2 \rangle \overline{\langle \varphi_1, \varphi_2 \rangle}.$$

Often in applications the square of the modulus of the STFT, the *spectrogram*, is used because it possesses many nice properties such as non-negativity. It is a *quadratic* time-frequency representation.

Wigner distribution

Another quadratic time-frequency distribution is the *Wigner distribution*, introduced by Wigner [215] in the 1930's. Given two functions $\psi, \phi \in L^2(\mathbb{R}^d)$, the cross-Wigner distribution of ψ and ϕ is given by

$$W(\psi, \phi)(x, \omega) = \int_{\mathbb{R}^d} \psi(t + x/2)\overline{\phi(t - x/2)}e^{-2\pi i \omega \cdot t}\, dt.$$

When $\psi = \phi$, we simply write $W(\psi, \psi) = W(\psi)$ and call it the Wigner distribution.

The Wigner distribution has numerous applications in engineering, mathematics and physics as a time-frequency representation. In addition, it is also the "dual" object to Weyl quantization, a well-known *quantization scheme* that associates operators to functions, via

$$\langle L_f \psi, \phi \rangle = \langle f, W(\phi, \psi) \rangle$$

and the map $f \mapsto L_f$ is called the *Weyl transform*, aka Weyl quantization.

Cohen's class of quadratic time-frequency distributions

Cohen's class provides a nice class of covariant quadratic time-frequency distributions, including the spectrogram, scalogram and Wigner distribution. It consists of all functions of the form

$$Q_\Phi(\psi, \phi) = W(\psi, \phi) * \Phi$$

where Φ is a function or tempered distribution. It turns out that many of the properties of Cohen's class distributions are determined by the Weyl transform of Φ and later in the paper, we will define Cohen's class of not just time-frequency distributions but distributions with different underlying groups using an operator as a replacement for the Weyl transform of Φ. For more on Cohen's class of time-frequency distributions, see [39, 155].

Localization operators

Localization operators are classically defined with respect to the short-time Fourier transform as the operator valued integral

$$A_f^{\varphi_1, \varphi_2}(g) = \int_{\mathbb{R}^{2d}} f(x, \omega) V_{\varphi_1} g(x, \omega) \pi(x, \omega) \varphi_2 \, dx\, d\omega.$$

The function f is referred to as the *mask*, *multiplier* or *filter* and is often taken to be the indicator function of some compact set. In that case, the localization operator is written as $A_{\chi_\Omega}^{\varphi_1, \varphi_2} = A_\Omega^{\varphi_1, \varphi_2}$. Localization operators of the above form were originally introduced by I. Daubechies in [52].

A.2.3 Locally compact groups

Much of abstract harmonic analysis is carried out on locally compact groups because they possess many of the properties we need to define convolutions and other objects. For more on the general theory of harmonic analysis on locally compact groups, see [85, Chap. 2] and [87].

Left and right Haar measures

Given a locally compact group G, there always exist two Radon measures, the left Haar measure μ_ℓ and the right Haar measure μ_r. The left (right) Haar measure is left (right) invariant, meaning that $\mu_\ell(xE) = \mu_\ell(E)$ ($\mu_r(Ex) = \mu_r(E)$) for $x \in G$ and $E \subset G$. Both measures are equivalent in the sense that they are related via

$$\mu_r(E) = \mu_\ell(E^{-1}), \qquad d\mu_r(x) = \Delta_G(x^{-1})\, d\mu_\ell(x), \qquad d\mu_\ell(xy) = \Delta_G(y)\, d\mu_\ell(x)$$

where the function $\Delta_G : G \to (0, \infty)$ is called the *modular function*. If $\Delta_G \equiv 1$, the group is said to be *unimodular*. When discussing L^p-integrable functions $f : G \to \mathbb{C}$ with respect to the left and right Haar measures, we write

$$L^p_\ell(G) = L^p(G, d\mu_\ell), \qquad L^p_r(G) = L^p(G, d\mu_r)$$

while for $p = \infty$, we simply write $L^\infty(G)$.

Weyl-Heisenberg group

The Weyl-Heisenberg group $\mathbb{H}^n = (\mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}, \cdot_{\mathbb{H}^n})$ is equipped with the group operation

$$(x, \omega, t) \cdot_{\mathbb{H}^n} (x', \omega', t') = \left(x + x', \omega + \omega', t + t' + \frac{1}{2}(x'\omega - x\omega')\right),$$

which should be compared with the composition rule for time-frequency shifts

$$(T_x M_\omega)(T_{x'} M_{\omega'}) = e^{2\pi i x' \cdot \omega} T_{x+x'} M_{\omega+\omega'}. \tag{A.2.1}$$

We are interested in the *projective* representation

$$\pi : \mathbb{R}^{2d} \to \mathcal{U}(L^2(\mathbb{R}^d)), \qquad \pi(x, \omega) = T_x M_\omega$$

of the Weyl-Heisenberg group for which (A.2.1) is the Mackey induced representation.

Wavelet transform

The wavelet transform is a *time-scale representation* based on taking the inner product of a signal and translations and dilations of some window function. The dilation operator D_a is defined for positive a as

$$D_a f(y) = \frac{1}{\sqrt{a}} f\left(\frac{y}{a}\right)$$

and hence the wavelet transform has the form

$$W_\phi \psi(x,a) = \left\langle \psi, T_x D_a \phi \right\rangle = \frac{1}{\sqrt{a}} \int_{\mathbb{R}} \psi(t) \overline{\phi\left(\frac{t-x}{a}\right)}\, dt.$$

There exists a version of Moyal's identity for the wavelet transform, often referred to simply as the *orthogonality relation*

$$\langle V_{\phi_1}\psi_1, V_{\phi_2}\psi_2 \rangle = \langle \psi_1, \psi_2 \rangle \langle \mathcal{D}^{-1}\phi_1, \mathcal{D}^{-1}\phi_2 \rangle$$

where $\mathcal{D}$ denotes the Duflo-Moore operator of the affine group and is defined by

$$\widehat{\mathcal{D}^{-1}\phi}(\omega) = \frac{\hat{\phi}(\omega)}{\sqrt{|\omega|}};$$

and ϕ_1, ϕ_2 are two admissible wavelets. The square of the wavelet transform is referred to as the *scalogram* in analogy to the spectrogram and used similarly.

Affine group

The affine group Aff $= (\mathbb{R} \times \mathbb{R}^+, \cdot_{\mathrm{Aff}})$ has the group operation

$$(x,a) \cdot_{\mathrm{Aff}} (y,b) := (x + ay, ab)$$

which coincides with the relation

$$(T_x D_a)(T_y D_b) = T_x T_{ay} D_a D_b = T_{x+ay} D_{ab}$$

between translation and dilation operators. Thus the representation $\pi(x,a) = T_x D_a$ of the affine group induces the wavelet transform discussed above. Moreover, it is easy to see that $\pi(x,a) = T_x D_a$ is a unitary representation of Aff on $L^2(\mathbb{R})$. Often, when dealing with wavelet analysis we are only interested in *analytic signals* which are L^2-functions for which $\hat{f}(\omega) = 0$ for $\omega < 0$ because the above representation is irreducible on the Hardy space of the real line. Using the Plancherel theorem, we can do everything on the Fourier side where our underlying Hilbert space becomes

$\mathcal{H} = L^2(\mathbb{R}^+)$ which will be the case for the remainder of this paper when discussing quantum harmonic analysis on the affine group.

A quick calculation shows that the inverse $(x,a)^{-1}$ of an element $(x,a) \in \mathrm{Aff}$ is given by $\left(\frac{-x}{a}, \frac{1}{a}\right)$. The affine group is an example of a non-abelian and non-unimodular group since the left and right Haar measures are given by

$$d\mu_\ell(x,a) = \frac{dx\,da}{a^2}, \qquad d\mu_r(x,a) = \frac{dx\,da}{a},$$

respectively and so in particular, $L^p_r(\mathrm{Aff}) = L^p\left(\mathbb{R} \times \mathbb{R}^+, \frac{dx\,da}{a}\right)$.

We will have use for dilates of sets in the affine group and so motivated by the group operation specified above, we define the scaling function $\Gamma_R : \mathrm{Aff} \to \mathrm{Aff}$ with parameter $R > 0$ as

$$\Gamma_R(x,a) = (Rx, a^R), \quad \Gamma_R^{-1}(x,a) = \left(\frac{x}{R}, a^{1/R}\right), \quad R\Omega = \left\{\Gamma_R(x,a), (x,a) \in \Omega\right\}.$$

This scaling relation is natural for the right Haar measure in the sense that it is the only one which is the identity for $R = 1$ and for which $d\mu_r(\Gamma_R(x,a)) = C \cdot d\mu_r(x,a)$ for some non-zero constant C. In particular, the constant C has the value R^2 which can be verified directly and consequently,

$$d\mu_r(\Gamma_R(x,a)) = R^2\, d\mu_r(x,a), \qquad \mu_r(R\Omega) = R^2 \mu_r(R\Omega).$$

For technical reasons, we will need to discuss convergence of sequences in Aff as well as neighborhoods. To that end, we define the following distance function

$$d_r^{\mathrm{Aff}}\big((x,a),(y,b)\big) = |x-y| + \left|\ln \frac{a}{b}\right|.$$

It is chosen mostly for convenience but has the nice property that the distance from (x,a) and $(0,1) = 0_{\mathrm{Aff}}$ is given by $|x| + |\ln(a)|$ which is the sum of the horizontal and vertical distance when integrating with respect to the right Haar measure $\frac{dx\,da}{a}$.

Using this distance function, we define the following type of balls in Aff:

$$B_r^{\mathrm{Aff}}\big((x,a),\delta\big) = \left\{(y,b) \in \mathrm{Aff} : d_r^{\mathrm{Aff}}\big((x,a),(y,b)\big) < \delta\right\}.$$

Additional properties of the affine group relevant to quantum harmonic analysis are discussed in [32].

Abstract harmonic analysis

The approach for the Weyl-Heisenberg and affine groups described above can be generalized to the locally compact setting. Here we let σ denote a square integrable unitary representation of a locally compact group G and write $\mathcal{H}$ for the underlying Hilbert space so that $\sigma : G \to \mathcal{U}(\mathcal{H})$. This view is more closely connected to quantum mechanics and representation theory than time-frequency and time-scale analysis but many of the arguments work in the same way. Partly due to this connection to physics, G is referred to as *phase space*.

A.2.4 Motivating examples of square integrable representations

Since the main contribution of this paper is setting up quantum harmonic analysis on general locally compact groups, we present two motivating examples on which the results apply and hint at their generalizations to locally compact groups.

Shearlet group

The shearlet group represents an attempt to extend the wavelet transform to two-dimensional inputs. The dilations and one-dimensional translations of the wavelet transform are here replaced by asymmetric dilations, shears and two-dimensional translations using the *parabolic scaling matrix* A_a and the *shear matrix* S_s, given by

$$A_a = \begin{pmatrix} a & 0 \\ 0 & \sqrt{a} \end{pmatrix}, \qquad S_s = \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix}.$$

The associated square integrable unitary representation can be written as

$$\pi(a,s,x)\psi(t) = T_x D_{S_s A_a}\psi(t) = a^{-3/4}\psi\big(A_a^{-1}S_s^{-1}(t-x)\big)$$

and it induces a group operation on the *shearlet group* $\mathbb{S} = (\mathbb{R}^+ \times \mathbb{R} \times \mathbb{R}^2, \cdot_{\mathbb{S}})$,

$$(a,s,x) \cdot_{\mathbb{S}} (a',s',x') = (aa', s + s'\sqrt{a}, x + S_s A_a x').$$

The left and right Haar measures associated to the shearlet group can be computed to be

$$d\mu_\ell(a,s,x) = \frac{da\,ds\,dx}{a^3}, \qquad d\mu_r(a,s,x) = \frac{da\,ds\,dx}{a}.$$

For more on the shearlet group as well references for the statements above, see [46, 47, 111, 144]. The shearlet group and associated transform has been generalized to higher dimensions which is also based on a square integrable representation, see [48].

Similitude group

The perhaps most straight-forward generalization of the wavelet transform to two-dimensional signals comes in the form of what is sometimes referred to as the *two-dimensional wavelet transform* which is induced by the unitary representation

$$\pi(a, x, \theta)\psi(t) = a^{-1}\psi\Big(\tau_{-\theta}\Big(\frac{t-x}{a}\Big)\Big)$$

of the *similitude group* $\mathrm{SIM}(2) = (\mathbb{R}^+ \times \mathbb{R}^2 \rtimes \mathrm{SO}(2), \cdot_{\mathrm{SIM}(2)})$ where $\tau_\theta \in \mathrm{SO}(2)$ is a rotation which acts as

$$\tau_\theta(x, y) = (x\cos\theta - y\sin\theta, x\sin\theta + y\cos\theta).$$

The similitude group SIM(2) is equipped with the group operation

$$(a, x, \tau_\theta) \cdot_{\mathrm{SIM}(2)} (a', x', \tau_{\theta'}) = (aa', b + a\tau_\theta x', \tau_{\theta+\theta'}).$$

Just as for shearlet group, the left and right Haar measures of SIM(2) depend on the dilation parameter and are given by

$$d\mu_\ell(a, x, \theta) = \frac{da\, dx\, d\theta}{a^3}, \qquad d\mu_r(a, x, \theta) = \frac{da\, dx\, d\theta}{a}.$$

More on role of the similitude group in the two-dimensional wavelet transform can be found in [14, 16, 49, 186]. Similitude groups can be seen as a specific case of the affine group on $\mathbb{R}^d$ which generalizes the affine group by replacing dilations with multiplications by matrices in $\mathrm{GL}(\mathbb{R}^d)$ and normalized by the determinant. This also corresponds to a square integrable representation [144].

Affine Poincaré group

Another approach to two-dimensional wavelet transforms comes in the form of the affine Poincaré group $\mathcal{P}_{\mathrm{aff}}$ which consists of translations, zooming and hyperbolic rotations. More specifically, the group law takes the form

$$(b, a, \vartheta) \cdot_{\mathcal{P}_{\mathrm{Aff}}} (b', a', \vartheta') = (b + a\Lambda_\vartheta b', aa', \vartheta + \vartheta'), \qquad \Lambda_\vartheta = \begin{pmatrix} \cosh\vartheta & \sinh\vartheta \\ \sinh\vartheta & \cosh\vartheta \end{pmatrix}$$

Since we have the same sort of zooming and rotational system as in the shearlet and similitude groups, the affine Poincaré group has the same left and right Haar measures given by

$$d\mu_\ell(b, a, \vartheta) = \frac{db\, da\, d\vartheta}{a^3}, \qquad d\mu_r(b, a, \vartheta) = \frac{db\, da\, d\vartheta}{a}.$$

The natural square integrable representation of $\mathcal{P}_{\mathrm{Aff}}$ is given by

$$\pi(b, a, \vartheta)\psi(t) = \frac{1}{a}\psi\left(\frac{1}{a}\Lambda_\vartheta(t - b)\right)$$

and it can be decomposed as the direct sum of four irreducible representations on orthogonal subsets of $L^2(\mathbb{R}^2)$.

An introduction to the affine Poincaré group can be found in [17, Sec. 7.4] and some recent work extending concepts from time-frequency analysis to $\mathcal{P}_{\mathrm{Aff}}$ is available in [50]. The group also has applications in physics in the context of Minkowski spacetime [15, Sec. 16.2.4].

A.3 Operator convolutions

In this section we introduce the three types of convolutions we deal with: function-function, function-operator and operator-operator and prove some elementary properties and bounds. All of the definitions generalize those in [32, 153, 211] and the proofs are similar with the exception of Proposition A.3.5. The function-function convolutions are standard but we write them down to be clear about the right Haar measure convention.

Definition A.3.1. For $f, g \in L^1_r(G)$, the *convolution* $f *_G g$ is defined as

$$f *_G g(x) = \int_G f(y)g(xy^{-1})\, d\mu_r(y).$$

The following standard estimate follows by Young's inequality.

Proposition A.3.2. Let $f \in L^1_r(G)$ and $g \in L^p_r(G)$ for $1 \leq p \leq \infty$. Then

$$\|f *_G g\|_{L^p_r(G)} \leq \|f\|_{L^1_r(G)} \|g\|_{L^p_r(G)}.$$

A.3.1 Function-operator convolutions

Inspired by the notion of a shift for operators of the form

$$\alpha_x(S) = \sigma(x)^* S \sigma(x)$$

which moves a function in phase space by x, applies S and then moves it back by x^{-1}, we have the following definition for function-operator convolutions.

Definition A.3.3. Let $f \in L^1_r(G)$ and $S \in \mathcal{S}^1$, then the *convolution* $f \star_G S$ is defined as the operator on $\mathcal{H}$ given by

$$f \star_G S = \int_G f(x)\alpha_x(S)\, d\mu_r(x).$$

This operator acts weakly in the way described in Section A.2.1, i.e. as

$$\left\langle f \star_G S\psi, \phi \right\rangle = \int_G f(x)\langle \alpha_x(S)\psi, \phi\rangle\, d\mu_r(x)$$

and we define $S \star_G f = f \star_G S$.

Remark A.3.4. In an upcoming paper by the author in collaboration with Feichtinger and Luef, the function-operator convolution defined above is realized as a special case of measure-operator convolutions and then the description of the weak action becomes a theorem, not a definition.

The following boundedness property of function-operator convolutions is an important result which will be used extensively. It is somewhat analogous to the $p = 1$ case of Proposition A.3.2 and the corresponding statement for the $p > 1$ range is proved in Section A.4.3.

Proposition A.3.5. Let $f \in L^1_r(G)$ and $S \in \mathcal{S}^1$. Then we have

$$\|f \star_G S\|_{\mathcal{S}^1} \leq \|f\|_{L^1_r(G)} \|S\|_{\mathcal{S}^1}.$$

Proof. We control the trace-class norm of $f \star_G S$ by bounding $|\langle f \star_G S, T\rangle|$ for $T \in B(\mathcal{H})$ with $\|T\|_{B(\mathcal{H})} = 1$ as

$$\begin{aligned}
|\langle f \star_G S, T\rangle| &= |\operatorname{tr}\left((f \star_G S)T^*\right)| \\
&= \left|\sum_n \left\langle (f \star_G S)e_n, Te_n \right\rangle\right| \\
&\leq \sum_n \int_G |f(x)||\langle \alpha_x(S)e_n, Te_n\rangle| \, d\mu_r(x) \\
&= \int_G |f(x)| \sum_n |\langle T^* \alpha_x(S)e_n, e_n\rangle| \, d\mu_r(x)
\end{aligned}$$

where we used in the last step Tonelli's Theorem. For each x, the above sum can be bounded by $\|S\|_{\mathcal{S}^1}$ using [40, Thm. 18.11]. Hence

$$|\langle f \star_G S, T\rangle| \leq \int_G |f(x)| \|S\|_{\mathcal{S}^1} \, d\mu_r(x) = \|f\|_{L^1_r(G)} \|S\|_{\mathcal{S}^1}$$

as desired. □

In the same way that the integral over a function-function convolution can be decoupled, the trace of a function-operator convolution may be written as a product in the following way.

Proposition A.3.6. Let $f \in L^1_r(G)$ and $S \in \mathcal{S}^1$. Then

$$\operatorname{tr}(f \star_G S) = \operatorname{tr}(S) \int_G f(x) \, d\mu_r(x).$$

Proof. We compute

$$\operatorname{tr}(f \star_G S) = \sum_n \langle (f \star S)e_n, e_n\rangle = \sum_n \int_G f(x)\langle \alpha_x(S)e_n, e_n\rangle \, d\mu_r(x).$$

By Tonelli's Theorem we have that

$$\begin{aligned}\sum_n \int_G |f(x)\langle \sigma(x)^* S\sigma(x)e_n, e_n\rangle| \, d\mu_r(x) &= \int_G \sum_n |f(x)\langle S\sigma(x)e_n, \sigma(x)e_n\rangle| \, d\mu_r(x) \\ &\leq \int_G |f(x)| \, d\mu_r(x) \sum_n |\langle S\sigma(x)e_n, \sigma(x)e_n\rangle|\end{aligned}$$

where the first factor is finite by the integrability of f. The finiteness of the second factor follows by [40, Prop. 18.9]. We can now finish the computation as

$$\begin{aligned}\operatorname{tr}(f \star_G S) &= \int_G f(x) \sum_n \langle S\sigma(x)e_n, \sigma(x)e_n\rangle \, d\mu_r(x) \\ &= \operatorname{tr}(S) \int_G f(x) \, d\mu_r(x).\end{aligned}$$

□

Lastly we show that function-operator convolutions preserve positivity.

Lemma A.3.7. *If $f \in L^1_r(G)$ is non-negative and $S \in \mathcal{S}^1$ is positive, then so is $f \star_G S$.*

Proof. We verify this directly as

$$\begin{aligned}\langle (f \star_G S)\phi, \phi\rangle &= \int_G f(x)\langle \alpha_x(S)\phi, \phi\rangle \, d\mu_r(x) \\ &= \int_G f(x)\langle S\sigma(x)\phi, \sigma(x)\phi\rangle \, d\mu_r(x) \geq 0.\end{aligned}$$

□

A.3.2 Operator-operator convolutions

A central theme in quantum harmonic analysis is that when replacing functions by operators, integrals should be replaced by traces. This motivates the following definition of operator-operator convolutions.

Definition A.3.8. Let $T \in \mathcal{S}^1$ and $S \in B(\mathcal{H})$, then the *convolution* $T \star_G S$ is the function on G given by

$$T \star_G S(x) = \operatorname{tr}(T\alpha_x(S)).$$

The following lemma is an example of an operator-operator convolution.

Lemma A.3.9. *For $\psi, \phi \in \mathcal{H}$ and $S \in B(\mathcal{H})$ we have*

$$(\psi \otimes \phi) \star_G S(x) = \left\langle S\sigma(x)\psi, \sigma(x)\phi \right\rangle.$$

Proof. We compute

$$\begin{aligned}(\psi \otimes \phi) \star_G S(x) &= \operatorname{tr}\left((\psi \otimes \phi)\alpha_x(S)\right) \\ &= \sum_n \left\langle \langle \alpha_x(S)e_n, \phi\rangle \psi, e_n \right\rangle \\ &= \langle \psi, \sigma(x)^* S^* \sigma(x)\phi\rangle \\ &= \langle S\sigma(x)\psi, \sigma(x)\phi\rangle.\end{aligned}$$

□

Lemma A.3.10. *Let $S \in \mathcal{S}^1$ and $\{\xi_n\}_n$ be an orthonormal basis of $\mathcal{H}$, then*

$$\sum_n (\xi_n \otimes \xi_n) \star_G S(x) = \operatorname{tr}(S).$$

Proof. This follows directly from Lemma A.3.9 and the fact that $\{\sigma(x)\xi_n\}_n$ is an orthonormal basis. □

We also have the following estimate which follows by standard properties of trace norms, see [40, Thm. 18.11 (g)] for a proof.

Lemma A.3.11. *Let $T \in \mathcal{S}^1$ and $S \in B(\mathcal{H})$. Then*

$$\|T \star_G S\|_{L^\infty(G)} \leq \|T\|_{\mathcal{S}^1} \|S\|_{B(\mathcal{H})}.$$

As to be expected, operator-operator convolutions preserve positivity in the same way as usual function-function convolutions and function-operator convolutions in Lemma A.3.7.

Lemma A.3.12. *Let $T \in \mathcal{S}^1$ and $S \in B(\mathcal{H})$ both be positive. Then*

$$T \star_G S(x) \geq 0 \qquad \text{for all } x \in G.$$

Proof. We expand T in its singular value decomposition and compute the trace with the same basis to find

$$\begin{aligned}T \star_G S(x) = \operatorname{tr}(T\alpha_x(S)) &= \sum_n \left\langle \sum_m \lambda_m (e_m \otimes e_m)(\alpha_x(S)e_n), e_n \right\rangle \\ &= \sum_{n,m} \lambda_m \langle (\sigma(x)^* S \sigma(x))e_n, e_m\rangle \langle e_m, e_n\rangle \\ &= \sum_n \lambda_n \langle S\sigma(x)e_n, \sigma(x)e_n\rangle \geq 0,\end{aligned}$$

where we used that the eigenvalues of a positive operator are non-negative. □

Lastly, we show that all the convolutions introduced in this section are associative in an appropriate manner.

Proposition A.3.13. Let $f, g \in L^1_r(G)$, $T \in \mathcal{S}^1$, and let S be a bounded operator on $\mathcal{H}$. Then the following compatibility relations hold

$$
\begin{aligned}
(f \star_G T) \star_G S &= f *_G (T \star_G S),\\
f \star_G (g \star_G T) &= (f *_G g) \star_G T.
\end{aligned}
$$

Proof. We proceed by direct computation

$$
\begin{aligned}
f *_G (T \star_G S) &= \int_G f(y) \operatorname{tr}\left(T\sigma(xy^{-1})^* S\sigma(xy^{-1})\right) d\mu_r(y)\\
&= \int_G f(y) \operatorname{tr}\left(\sigma(y)^* T\sigma(y)\sigma(x)^* S\sigma(x)\right) d\mu_r(y)\\
&= \operatorname{tr}\left(\left(\int_G f(y)\alpha_y(T)\, d\mu_r(y)\right)\alpha_x(S)\right)\\
&= \operatorname{tr}\left((f \star_G T)\alpha_x(S)\right)\\
&= (f \star_G T) \star_G S.
\end{aligned}
$$

For the other equality, we have

$$
\begin{aligned}
(f *_G g) \star_G T &= \int_G \left(\int_G f(x)g(zx^{-1})\, d\mu_r(x)\right)\alpha_z(T)\, d\mu_r(z)\\
&= \int_G \int_G f(x)g(zx^{-1})\alpha_z(T)\, d\mu_r(z)\, d\mu_r(x)
\end{aligned}
$$

and so applying the change of variables $y = zx^{-1}$, we find

$$
\begin{aligned}
(f *_G g) \star_G T &= \int_G \int_G f(x)g(y)\alpha_{yx}(T)\, d\mu_r(y)\, d\mu_r(x)\\
&= \int_G f(x)\alpha_x \left(\int_G g(y)\alpha_y(T)\, d\mu_r(y)\right) d\mu_r(x)\\
&= \int_G f(x)\alpha_x \left(g \star_G T\right)\, d\mu_r(x)\\
&= f \star_G (g \star_G T).
\end{aligned}
$$

□

We have not established all the mapping properties of function-operator and operator-operator convolutions between $L^p_r(G)$ and $\mathcal{S}^p$ spaces, since this requires some more preparation which is contained in the next section, in particular A.4.3.

A.4 Admissibility of operators

A.4.1 Integrability of operator-operator convolutions

As seen in [32], to generalize some results on operator convolutions to non-unimodular groups, we need to introduce a notion of *admissibility of operators*. The definition is motivated by the desire for $T \star_G S$ to be integrable. Before stating such a result, we recall the following classical theorem from [64] which we write out using the right Haar measure convention.

Theorem A.4.1 (Duflo-Moore)**.** *Let $(\sigma, \mathcal{H})$ be a square integrable, irreducible, unitary representation of a locally compact group G. Then there exists a unique, possibly unbounded, densely defined, positive, closed, self-adjoint operator $\mathcal{D}^{-1}$: $\mathcal{H} \to \mathcal{H}$ with densely defined inverse $\mathcal{D}$ such that:*

(i) The admissible vectors $\psi \in \mathcal{H}$ are exactly those vectors in the domain of $\mathcal{D}^{-1}$.

(ii) For $\phi_1, \phi_2 \in \mathcal{H}$ and $\psi_1, \psi_2 \in \mathrm{Dom}(\mathcal{D}^{-1})$, the following orthogonality relation holds

$$\int_G \langle \phi_1, \sigma(x)^* \psi_1 \rangle \overline{\langle \phi_2, \sigma(x)^* \psi_2 \rangle} \, d\mu_r(x) = \langle \phi_1, \phi_2 \rangle \overline{\langle \mathcal{D}^{-1} \psi_1, \mathcal{D}^{-1} \psi_2 \rangle}. \tag{A.4.1}$$

(iii) The following covariance relation holds

$$\sigma(x) \mathcal{D} \sigma(x)^* = \sqrt{\Delta_G(x)} \mathcal{D}. \tag{A.4.2}$$

We can lift this result to the operator setting by taking the singular value decompositions of two operators and apply the above result to the rank-one situations.

Theorem A.4.2. *Let $S \in \mathcal{S}^1$ satisfy $\mathcal{D} S \mathcal{D} \in B(\mathcal{H})$. For any $T \in \mathcal{S}^1$ we have that $T \star_G \mathcal{D} S \mathcal{D} \in L^1_r(G)$ with*

$$\|T \star_G \mathcal{D} S \mathcal{D}\|_{L^1_r(G)} \leq \|T\|_{\mathcal{S}^1} \|S\|_{\mathcal{S}^1}$$

and

$$\int_G T \star_G \mathcal{D} S \mathcal{D}(x) \, d\mu_r(x) = \mathrm{tr}(T) \, \mathrm{tr}(S). \tag{A.4.3}$$

Proof. The proof consists of two steps.
Step 1: We consider first the case where $S = \psi \otimes \phi$ for $\psi, \phi \in \mathcal{H}$. Then

$\mathcal{D}S\mathcal{D} = \mathcal{D}\psi \otimes \mathcal{D}\phi$ and since $\mathcal{D}S\mathcal{D}$ is bounded by assumption, it in particular holds that $\psi, \phi \in \mathrm{Dom}(\mathcal{D})$. To see this, note that

$$\mathcal{D}S\mathcal{D}(f) = \langle f, \mathcal{D}\phi\rangle \mathcal{D}\psi$$

and so since this is bounded for all f, $\mathcal{D}\psi$ and $\mathcal{D}\phi$ are also bounded.

We now compute

$$\begin{aligned} T \star_G \mathcal{D}S\mathcal{D}(x) &= \mathrm{tr}\left(T\sigma(x)^*(\mathcal{D}\psi \otimes \mathcal{D}\phi)\sigma(x)\right) \\ &= \sum_n \left\langle T\sigma(x)^*(\mathcal{D}\psi \otimes \mathcal{D}\phi)\sigma(x)e_n, e_n\right\rangle \\ &= \sum_n \langle \sigma(x)e_n, \mathcal{D}\phi\rangle\langle T\sigma(x)^* D\psi, e_n\rangle \\ &= \langle T\sigma(x)^*\mathcal{D}\psi, \sigma(x)^*\mathcal{D}\phi\rangle. \end{aligned}$$

Since $T \in \mathcal{S}^1$, we can expand it using its singular value decomposition $T = \sum_n t_n \xi_n \otimes \eta_n$ which allows us to write

$$\begin{aligned} T \star_G \mathcal{D}S\mathcal{D}(x) &= \sum_n t_n \langle (\xi_n \otimes \eta_n)(\sigma(x)^*\mathcal{D}\psi), \sigma(x)^*\mathcal{D}\phi\rangle \\ &= \sum_n t_n \langle \sigma(x)^*\mathcal{D}\psi, \eta_n\rangle\langle \xi_n, \sigma(x)^*\mathcal{D}\phi\rangle. \end{aligned}$$

Now each term is of the form in the Duflo-Moore orthogonality relation (A.4.1). We can therefore proceed by integrating each term after bounding the result as

$$\begin{aligned} &\int_G \left|\langle \sigma(x)^*\mathcal{D}\psi, \eta_n\rangle\langle \xi_n, \sigma(x)^*\mathcal{D}\phi\rangle\right| d\mu_r(x) \\ &\quad \leq \left(\int_G |\langle \sigma(x)^*\mathcal{D}\psi, \eta_n\rangle|^2 \, d\mu_r(x)\right)^{1/2} \left(\int_G |\langle \xi_n, \sigma(x)^*\mathcal{D}\phi\rangle|^2 \, d\mu_r(x)\right)^{1/2} \\ &\quad = \|\eta_n\| \|\mathcal{D}^{-1}\mathcal{D}\psi\| \|\xi_n\| \|\mathcal{D}^{-1}\mathcal{D}\phi\| = \|\psi\|\|\phi\|. \end{aligned}$$

Since $T \in \mathcal{S}^1$, $(t_n)_n$ is summable and we can move the integral inside to deduce that

$$\|T \star_G \mathcal{D}S\mathcal{D}\|_{L^1_r(G)} \leq \|T\|_{\mathcal{S}^1} \|\psi\|\|\phi\|.$$

We can now establish (A.4.3) by moving the integral inside the sum and using the Duflo-Moore orthogonality relation (A.4.1) which yields

$$\int_G T \star_G \mathcal{D}S\mathcal{D}(x) \, d\mu_r(x) = \sum_n t_n \langle \xi_n, \eta_n\rangle\langle \phi, \psi\rangle = \mathrm{tr}(T)\langle \phi, \psi\rangle.$$

Step 2: We now move to considering S as in the theorem. By compactness, we can consider its singular value decomposition which is of the form

$$S = \sum_n s_n(\psi_n \otimes \phi_n).$$

Hence we have

$$\begin{aligned} T \star_G \mathcal{D}S\mathcal{D} &= T \star_G \mathcal{D}\left(\sum_n s_n\psi_n \otimes \phi_n\right)\mathcal{D} \\ &= \sum_n s_n T \star_G \mathcal{D}(\psi_n \otimes \phi_n)\mathcal{D} \\ &= \sum_n s_n T \star_G (\mathcal{D}\psi_n \otimes \mathcal{D}\phi_n) \end{aligned}$$

where we are allowed to move the sum outside since the outer sum is uniformly convergent by Lemma A.3.11. We can also estimate the norm as

$$\begin{aligned} \|T \star_G \mathcal{D}S\mathcal{D}\|_{L^1_r(G)} &\leq \sum_n |s_n| \|T \star_G (\mathcal{D}\psi_n \otimes \mathcal{D}\phi_n)\|_{L^1_r(G)} \\ &= \|T\|_{\mathcal{S}^1} \|S\|_{\mathcal{S}^1}. \end{aligned}$$

For (A.4.3), the same method used in the first step yields the desired conclusion. □

Because we typically integrate $T \star_G S$ in applications, this theorem is more useful when considering the operator $\mathcal{D}^{-1}S\mathcal{D}^{-1}$ which leads us to make the following definition.

Definition A.4.3. Let $S \neq 0$ be a bounded operator on $\mathcal{H}$ that maps $\mathrm{Dom}(\mathcal{D})$ into $\mathrm{Dom}(\mathcal{D}^{-1})$. We say that S is *admissible* if the composition $\mathcal{D}^{-1}S\mathcal{D}^{-1}$ is bounded on $\mathrm{Dom}(\mathcal{D}^{-1})$ and extends to a trace-class operator $\mathcal{D}^{-1}S\mathcal{D}^{-1} \in \mathcal{S}^1$.

We can now restate Theorem A.4.2 using the above definition.

Corollary A.4.4. Let S be an admissible operator and $T \in \mathcal{S}^1$. Then $T \star_G S \in L^1_r(G)$ with

$$\|T \star_G S\|_{L^1_r(G)} \leq \|T\|_{\mathcal{S}^1} \|\mathcal{D}^{-1}S\mathcal{D}^{-1}\|_{\mathcal{S}^1}$$

and

$$\int_G T \star_G S(x)\, d\mu_r(x) = \mathrm{tr}(T)\,\mathrm{tr}(\mathcal{D}^{-1}S\mathcal{D}^{-1}).$$

Example A.4.5. As hinted at by the proof of Theorem A.4.2, we can recover the orthogonality relation (A.4.1) by choosing S and T as rank-one operators. To make this explicit, choose

$$T = \phi_1 \otimes \phi_2, \qquad S = \psi_2 \otimes \psi_1$$

where ψ_1, ψ_2 are admissible vectors in the sense of Theorem A.4.1, i.e. $\psi_1, \psi_2 \in \mathrm{Dom}(\mathcal{D}^{-1})$. Then S is admissible since

$$\|\mathcal{D}^{-1} S \mathcal{D}^{-1}\|_{\mathcal{S}^1} = \|\mathcal{D}^{-1}\psi_2 \otimes \mathcal{D}^{-1}\psi_1\|_{\mathcal{S}^1} \leq \|\mathcal{D}^{-1}\psi_2\|\|\mathcal{D}^{-1}\psi_1\| < \infty.$$

Corollary A.4.4 combined with Lemma A.3.9 now allows us to compute $T \star_G S(x)$ as

$$\begin{aligned}\int_G T \star_G S(x)\, d\mu_r(x) &= \int_G \langle \phi_1, \sigma(x)^* \psi_1 \rangle \overline{\langle \phi_2, \sigma(x)^* \psi_2 \rangle}\, d\mu_r(x) \\ &= \mathrm{tr}(\phi_1 \otimes \phi_2)\, \mathrm{tr}(\mathcal{D}^{-1}\psi_2 \otimes \mathcal{D}^{-1}\psi_1) \\ &= \langle \phi_1, \phi_2 \rangle \overline{\langle \mathcal{D}^{-1}\psi_1, \mathcal{D}^{-1}\psi_2 \rangle}\end{aligned}$$

which is exactly (A.4.1).

By demanding stronger conditions on both S and T we can deduce the following corollary.

Corollary A.4.6. Let S and T be admissible trace-class operators on $\mathcal{H}$, then the convolution $T \star_G S$ satisfies $T \star_G S \in L^1_r(G) \cap L^1_\ell(G)$ and

$$\int_G T \star_G S(x)\, d\mu_r(x) = \mathrm{tr}(T)\, \mathrm{tr}(\mathcal{D}^{-1} S \mathcal{D}^{-1}),$$
$$\int_G T \star_G S(x)\, d\mu_\ell(x) = \mathrm{tr}(S)\, \mathrm{tr}(\mathcal{D}^{-1} T \mathcal{D}^{-1}).$$

Proof. The first equality and $T \star_G S \in L^1_r(G)$ is the statement of Corollary A.4.4. The two corresponding statements for the left Haar measure follow by making the change of variables $x \mapsto x^{-1}$ and using that $\sigma(x^{-1})^* = \sigma(x)$, $d\mu_r(x^{-1}) = d\mu_\ell(x)$ and $T \star_G S(x^{-1}) = S \star_G T(x)$. □

A.4.2 Conditions for admissibility

In this section we go over some useful conditions for admissibility, all of which are generalizations of results in [32]. The first result shows how admissible functions are related to admissible operators.

Proposition A.4.7. A rank-one operator $S = \psi \otimes \phi$ for non-zero $\psi, \phi \in \mathcal{H}$ is an admissible operator if and only if ψ and ϕ are admissible functions.

Proof. If $S = \psi \otimes \phi$ is admissible, then $\mathcal{D}^{-1} S \mathcal{D}^{-1}$ is trace-class and in particular bounded. Hence for $\xi \in \mathrm{Dom}(\mathcal{D}^{-1})$

$$\|\mathcal{D}^{-1} S \mathcal{D}^{-1} \xi\| = |\langle \mathcal{D}^{-1}\xi, \phi \rangle| \|\mathcal{D}^{-1}\psi\| = |\langle \xi, \mathcal{D}^{-1}\phi \rangle| \|\mathcal{D}^{-1}\psi\| < C\|\xi\|$$

which implies that both ψ and ϕ are in $\mathrm{Dom}(\mathcal{D}^{-1})$. The converse was shown in Example A.4.5. □

The next proposition characterizes positive admissible operators.

Proposition A.4.8. Let S be a non-zero positive compact operator with spectral decomposition

$$S = \sum_n s_n(\xi_n \otimes \xi_n).$$

Then S is admissible if and only if each ξ_n is admissible and

$$\sum_n s_n \|\mathcal{D}^{-1}\xi_n\|^2 < \infty.$$

Proof. We first treat the case where S is admissible. Let $\xi \in \mathcal{H}$ with $\|\xi\| = 1$. Then by Lemma A.3.9 and linearity,

$$\begin{aligned}
(\xi \otimes \xi) \star_G S(x) &= \sum_n s_n \big\langle (\xi_n \otimes \xi_n)\sigma(x)\xi, \sigma(x)\xi \big\rangle \\
&= \sum_n s_n |\langle \sigma(x)\xi, \xi_n \rangle|^2.
\end{aligned}$$

By integrating the above and using the monotone convergence theorem we find

$$\begin{aligned}
\int_G \xi \otimes \xi \star_G S(x)\, d\mu_r(x) &= \sum_n s_n \int_G |\langle \sigma(x)\xi, \xi_n \rangle|^2\, d\mu_r(x) \\
&= \sum_n s_n \|\xi\|^2 \|\mathcal{D}^{-1}\xi_n\|^2
\end{aligned}$$

where we used the Duflo-Moore relation (A.4.1) in the last step. This sum is finite since the integral can be bounded using Corollary A.4.4, the fact that S is admissible and $\|\xi\| = 1$.

For the other direction, assume that each ξ_n is admissible and $\sum_n s_n \|\mathcal{D}^{-1}\xi_n\|^2 < \infty$. It is then clear that

$$\sum_n s_n (\mathcal{D}^{-1}\xi_n) \otimes (\mathcal{D}^{-1}\xi_n) \tag{A.4.4}$$

is trace-class by an application of the triangle inequality. It is however not clear that the above is equal to $\mathcal{D}^{-1}S\mathcal{D}^{-1}$ and that S maps $\mathrm{Dom}(\mathcal{D})$ into $\mathrm{Dom}(\mathcal{D}^{-1})$ when the sum defining S is infinite. For a fixed $\psi \in \mathcal{H}$, the partial sums

$$(S\psi)_M = \sum_{n=1}^{M} s_n \langle \psi, \xi_n \rangle \xi_n$$

converge to $S\psi$ as $M \to \infty$ by definition and moreover, $(S\psi)_M$ is admissible for each M.

The corresponding sequence of partial sums $\mathcal{D}^{-1}(S\psi)_M$ also converges in $\mathcal{H}$ since

$$\begin{aligned}\sum_n s_n|\langle\psi,\xi_n\rangle|\|\mathcal{D}^{-1}\xi_n\| &\leq \left(\sum_n |\langle\psi,\xi_n\rangle|^2\right)^{1/2}\left(\sum_n s_n^2\|\mathcal{D}^{-1}\xi_n\|^2\right)^{1/2}\\ &\leq C(S)\|\psi\|\left(\sum_n s_n\|\mathcal{D}^{-1}\xi_n\|^2\right)^{1/2}\end{aligned}$$

where $C(S)$ is some constant depending only on S. By Theorem A.4.1, the Duflo-Moore operator $\mathcal{D}^{-1}$ is closed and hence $S\psi$ is admissible with

$$\mathcal{D}^{-1}S\psi = \sum_n s_n\langle\psi,\xi_n\rangle\mathcal{D}^{-1}\xi_n. \tag{A.4.5}$$

Now to show that $\mathcal{D}^{-1}S\mathcal{D}^{-1}$ is bounded on $\mathrm{Dom}(\mathcal{D}^{-1})$, let $\phi\in\mathrm{Dom}(\mathcal{D}^{-1})$ and note that by (A.4.5),

$$\begin{aligned}\mathcal{D}^{-1}S\mathcal{D}^{-1}\phi &= \sum_n s_n\langle\mathcal{D}^{-1}\phi,\xi_n\rangle\mathcal{D}^{-1}\xi_n\\ &= \sum_n s_n\langle\phi,\mathcal{D}^{-1}\xi_n\rangle\mathcal{D}^{-1}\xi_n\\ &= \left(\sum_n s_n(\mathcal{D}^{-1}\xi_n)\otimes(\mathcal{D}^{-1}\xi_n)\right)\phi\end{aligned}$$

which we recognize as (A.4.4). This equality can be extended from the dense subspace $\mathrm{Dom}(\mathcal{D}^{-1})$ to all of $\mathcal{H}$ by density since

$$\|\mathcal{D}^{-1}S\mathcal{D}^{-1}\phi\| \leq \sum_n s_n\|\phi\|\|\mathcal{D}^{-1}\xi_n\|^2 \leq \|\phi\|\sum_n s_n\|\mathcal{D}^{-1}\xi_n\|^2 < \infty$$

by Cauchy-Schwarz and assumption. □

As a corollary, we have the following converse of Corollary A.4.4.

Corollary A.4.9. Let T be a non-zero positive trace-class operator and let S be a non-zero positive compact operator. If

$$\int_G T\star_G S(x)\,d\mu_r(x) < \infty,$$

then S is admissible with

$$\mathrm{tr}(\mathcal{D}^{-1}S\mathcal{D}^{-1}) = \frac{1}{\mathrm{tr}(T)}\int_G T\star_G S(x)\,d\mu_r(x). \tag{A.4.6}$$

In particular, if S is a non-zero, positive trace-class operator, then S is admissible if and only if $S\star_G S\in L^1_r(G)$.

Proof. Since T and S both are positive trace-class operators, we can expand them in their singular value decompositions with $S = \sum_n s_n \xi_n \otimes \xi_n$ and move the sum outside the integral using Tonelli's theorem to obtain

$$\int_G T \star_G S(x)\, d\mu_r(x) = \operatorname{tr}(T) \sum_n s_n \|\mathcal{D}^{-1}\xi_n\|^2 < \infty$$

where we used the Duflo-Moore orthogonality relation (A.4.1) to compute the resulting inner integral. By Proposition A.4.8, we conclude that S is admissible. Lastly Corollary A.4.4 yields the equality (A.4.6). $\square$

Admissibility of S does not automatically imply admissibility of $f \star_G S$ for $f \in L^p_r(G)$ as illustrated by the following proposition.

Proposition A.4.10. Suppose $f \in L^1_\ell(G) \cap L^1_r(G)$ be a non-zero and non-negative function. If S is a positive, admissible trace-class operator on $\mathcal{H}$, then so is $f \star_G S$ with

$$\operatorname{tr}(\mathcal{D}^{-1}(f \star_G S)\mathcal{D}^{-1}) = \int_G f(x)\, d\mu_\ell(x) \operatorname{tr}(\mathcal{D}^{-1} S \mathcal{D}^{-1}). \tag{A.4.7}$$

Proof. That $f \star_G S$ is trace-class follows from Proposition A.3.5 while positivity follows from Lemma A.3.7. Equation (A.4.7) follows if we can show that for T non-zero, positive and trace-class,

$$\int_G T \star_G (f \star_G S)(y)\, d\mu_r(y) = \operatorname{tr}(T) \int_G f(x)\, d\mu_r(x) \operatorname{tr}(\mathcal{D}^{-1} S \mathcal{D}^{-1}) \tag{A.4.8}$$

by applying Corollary A.4.9 to $f \star_G S$ which also yields admissibility of $f \star_G S$. To show (A.4.8), we note that

$$\begin{aligned} T \star_G (f \star_G S)(y) &= \operatorname{tr}\left(T\alpha_y\left(\int_G f(x)\alpha_x(S)\, d\mu_r(x)\right)\right) \\ &= \int_G f(x) \operatorname{tr}\left(T\alpha_{xy}(S)\right) d\mu_r(x) \\ &= \int_G f(x) T \star_G S(xy)\, d\mu_r(x). \end{aligned}$$

From here we can integrate $T \star_G (f \star_G S)$ to find

$$\begin{aligned} \int_G T \star_G (f \star_G S)(y)\, d\mu_r(y) &= \int_G \int_G f(x) T \star_G S(x)\, d\mu_r(xy)\, d\mu_r(y) \\ &= \int_G f(x) \int_G T \star_G S(xy)\, d\mu_r(y)\, d\mu_r(x) \\ &= \int_G f(x)\, d\mu_\ell(x) \operatorname{tr}(T) \operatorname{tr}(\mathcal{D}^{-1} S \mathcal{D}^{-1}) \end{aligned}$$

where we used the change of variables $z = xy$ and the relations between the left and right Haar measure from Section A.2.3. □

A.4.3 Interpolated convolution mapping properties

With the machinery of admissible operators in place, we can establish the remaining promised mapping properties from Section A.3 and generalize the inequalities

$$\|f \star_G S\|_{\mathcal{S}^1} \leq \|f\|_{L^1_r(G)} \|S\|_{\mathcal{S}^1},$$
$$\|T \star_G S\|_{L^\infty(G)} \leq \|T\|_{B(\mathcal{H})} \|S\|_{\mathcal{S}^1}$$

from Proposition A.3.5 and Lemma A.3.11 to all $\mathcal{S}^p$ and L^p-spaces. These results are generalizations of [32, Prop. 4.16, Lem. 4.17, Prop. 4.18] which treat the affine case and [153, Prop. 4.2] in the Weyl-Heisenberg case.

Proposition A.4.11. Let $1 \leq p \leq \infty$ and let q be its conjugate exponent given by $\frac{1}{p} + \frac{1}{q} = 1$. If $S \in \mathcal{S}^p, T \in \mathcal{S}^q$ and $f \in L^1_r(G)$, then the following holds:

(i) $f \star_G S \in \mathcal{S}^p$ with $\|f \star_G S\|_{\mathcal{S}^p} \leq \|f\|_{L^1_r(G)} \|S\|_{\mathcal{S}^p}$.

(ii) $T \star_G S \in L^\infty(G)$ with $\|T \star_G S\|_{L^\infty(G)} \leq \|S\|_{\mathcal{S}^p} \|T\|_{\mathcal{S}^q}$.

Proof. The first inequality follows from [129, Prop. 1.2.2] for $p < \infty$ while the $p = \infty$ case can be deduced by an elementary estimate on the weak action $\langle f \star_G S\psi, \phi\rangle$. Meanwhile the second inequality follows from [190, Thm. 2.8]. □

Item (i) above should be seen as a version of Young's inequality for function-operator convolutions and shows that $\mathcal{S}^p$ is a Banach module over $L^1_r(G)$ via the mapping $(f, S) \mapsto f \star_G S$. We now turn our attention to the case where f and operator-operator convolutions are in $L^p_r(G)$ for $p \neq 1$. For interpolation purposes, we will first need the following lemma.

Lemma A.4.12. *Let $S \in \mathcal{S}^1$ and $f \in L^\infty(G)$. Define the operator $f \star_G \mathcal{D}S\mathcal{D}$ weakly for $\psi, \phi \in \mathrm{Dom}(\mathcal{D})$ by*

$$\langle f \star_G \mathcal{D}S\mathcal{D}\psi, \phi\rangle = \int_G f(x)\langle S\mathcal{D}\sigma(x)\psi, \mathcal{D}\sigma(x)\phi\rangle \, d\mu_r(x).$$

Then $f \star_G \mathcal{D}S\mathcal{D}$ uniquely extends to a bounded linear operator on $\mathcal{H}$ satisfying

$$\|f \star_G \mathcal{D}S\mathcal{D}\|_{B(\mathcal{H})} \leq \|f\|_{L^\infty(G)} \|S\|_{\mathcal{S}^1}.$$

In particular, if R is an admissible operator, then $f \star_G R \in B(\mathcal{H})$ with

$$\|f \star_G R\|_{B(\mathcal{H})} \leq \|f\|_{L^\infty(G)} \|\mathcal{D}^{-1}R\mathcal{D}^{-1}\|_{\mathcal{S}^1}.$$

Proof. Using equation (A.4.2), we can rewrite $\langle f \star_G \mathcal{D}S\mathcal{D}\psi, \phi\rangle$ as

$$\begin{aligned}\langle f \star_G \mathcal{D}S\mathcal{D}\psi, \phi\rangle &= \int_G f(x)\langle S\sigma(x)\mathcal{D}\psi, \sigma(x)\mathcal{D}\phi\rangle \, d\mu_\ell(x) \\ &= \int_G f(x^{-1})\langle S\sigma(x)^*\mathcal{D}\psi, \sigma(x)^*\mathcal{D}\phi\rangle \, d\mu_r(x) \\ &= \int_G f(x^{-1})(S \star_G (\mathcal{D}\psi \otimes \mathcal{D}\phi))(x) \, d\mu_r(x).\end{aligned}$$

To bound this, we note that $\mathcal{D}\psi \otimes \mathcal{D}\phi$ is an admissible operator and so by an elementary estimate of the integral and the use of Corollary A.4.4, we deduce that

$$\begin{aligned}|\langle f \star_G \mathcal{D}S\mathcal{D}\psi, \phi\rangle| &\leq \|f\|_{L^\infty(G)}\|S\|_{\mathcal{S}^1}\|\mathcal{D}^{-1}(\mathcal{D}\psi \otimes \mathcal{D}\phi)\mathcal{D}^{-1}\|_{\mathcal{S}^1} \\ &= \|f\|_{L^\infty(G)}\|S\|_{\mathcal{S}^1}\|\psi\|\|\phi\|.\end{aligned}$$

That $f \star_G \mathcal{D}S\mathcal{D}$ extends uniquely follows from the denseness of $\mathrm{Dom}(\mathcal{D})$ in $\mathcal{H}$. □

Proposition A.4.13. Let $1 \leq p \leq \infty$ and let q be its conjugate exponent given by $\frac{1}{p} + \frac{1}{q} = 1$. Suppose $S \in \mathcal{S}^1$ is admissible, $T \in \mathcal{S}^p$ and $f \in L^p_r(G)$.

(i) $f \star_G S \in \mathcal{S}^p$ with $\|f \star_G S\|_{\mathcal{S}^p} \leq \|f\|_{L^p_r(G)}\|S\|^{1/p}_{\mathcal{S}^1}\|\mathcal{D}^{-1}S\mathcal{D}^{-1}\|^{1/q}_{\mathcal{S}^1}$.

(ii) $T \star_G S \in L^p_r(G)$ with $\|T \star_G S\|_{L^p_r(G)} \leq \|T\|_{\mathcal{S}^p}\|S\|^{1/q}_{\mathcal{S}^1}\|\mathcal{D}^{-1}S\mathcal{D}^{-1}\|^{1/p}_{\mathcal{S}^1}$.

Proof. Item (i) follows by complex interpolation between the $p = 1$ case from Proposition A.3.5 and the $p = \infty$ case which follows from Lemma A.4.12.

Similarly, item (ii) follows by interpolating between the $p = 1$ case from Corollary A.4.4 and the $p = \infty$ case of Lemma A.3.11. □

With the interpolation results established, we can prove a generalization of [153, Thm. 4.7], previously also noted in [211] and [24] in the Weyl-Heisenberg case.

Proposition A.4.14. Let $S \in \mathcal{S}^1$ be admissible and define the maps

$$\begin{aligned}\mathcal{A}_S &: L^p_r(G) \to \mathcal{S}^p, \quad f \mapsto f \star_G S, \\ \mathcal{B}_S &: \mathcal{S}^p \to L^p_r(G), \quad T \mapsto T \star_G S\end{aligned}$$

for $1 \leq p \leq \infty$. Then both maps are bounded and the adjoint of $\mathcal{A}_S : L^p_r(G) \to \mathcal{S}^p$ is given by $(\mathcal{A}_S)^* = \mathcal{B}_S : \mathcal{S}^q \to L^q_r(G)$ where $\frac{1}{p} + \frac{1}{q} = 1$ while the adjoint of $\mathcal{B}_S : \mathcal{S}^p \to L^p_r(G)$ is given by $(\mathcal{B}_S)^* = \mathcal{A}_S : L^q_r(G) \to \mathcal{S}^q$.

Proof. Boundedness of the mappings follows from Proposition A.4.13. We compute the adjoint of $\mathcal{A}_S$ by considering the duality brackets

$$\left\langle (\mathcal{A}_S)^* T, f \right\rangle_{L^p_r(G), L^q_r(G)} = \left\langle T, \mathcal{A}_S f \right\rangle_{\mathcal{S}^p, \mathcal{S}^q}$$

for $T \in \mathcal{S}^p$ and $f \in L^q_r(G)$. The statement in the other direction for $(\mathcal{B}_S)^*$ will then follow by the same argument.

First, assume that $T \in \mathcal{S}^1$ and $f \in L^1_r(G)$. It then holds that

$$\begin{aligned}
\left\langle T, \mathcal{A}_S f \right\rangle &= \operatorname{tr}\left(T \mathcal{A}_S f\right) \\
&= \sum_n \left\langle T(f \star_G S) e_n, e_n \right\rangle \\
&= \sum_n \int_G f(x) \left\langle T \alpha_x(S) e_n, e_n \right\rangle d\mu_r(x).
\end{aligned}$$

To justify the use of Fubini's Theorem on the above, we use Tonelli's Theorem to note that

$$\sum_n \int_G \left| f(x) \left\langle T \alpha_x(S) e_n, e_n \right\rangle \right| d\mu_r(x) = \int_G |f(x)| \sum_n \left| \left\langle T \alpha_x(S) e_n, e_n \right\rangle \right| d\mu_r(x).$$

The operator $T\sigma(x)^* S \sigma(x)$ is trace-class and the Hilbert-Schmidt norm of $\alpha_x(S)$ is independent of x and so the proof of [40, Prop. 18.9] yields that the sum is uniformly bounded. By the integrability of f, we deduce that the entire quantity is finite. We can hence apply Fubini's Theorem to obtain

$$\begin{aligned}
\left\langle T, \mathcal{A}_S f \right\rangle_{\mathcal{S}^p, \mathcal{S}^q} &= \int_G f(x) \sum_n \left\langle T \alpha_x(S) e_n, e_n \right\rangle d\mu_r(x) \\
&= \int_G f(x) T \star S(x) \, d\mu_r(x) = \left\langle \mathcal{B}_S T, f \right\rangle_{L^p_r(G), L^q_r(G)}.
\end{aligned}$$

By the density of $\mathcal{S}^1$ and $L^1_r(G)$, this result can be extended to all of $\mathcal{S}^p$ and $L^q_r(G)$ while for $p = \infty$ the result holds by duality. □

Lastly we show how we can weaken the conditions on Proposition A.3.5 but still get that the function-operator convolution is a compact operator which generalizes part of [156, Lem. 2.3]. We remind the reader that $L^0(G)$ is the set of all $L^\infty(G)$ functions which vanish at infinity while $\mathcal{K}$ is the set of compact operators.

Corollary A.4.15. Let $f : G \to \mathbb{C}$ and $S \in B(\mathcal{H})$ satisfy one of the following

(i) $f \in L^0(G)$ and $S \in \mathcal{S}^1$ is admissible,

(ii) $f \in L^1_r(G)$ and $S \in \mathcal{K}$.

Then $f \star_G S \in \mathcal{K}$.

Proof. Both alternatives follow by approximating the less well behaved object and showing that the approximations are compact and converge to $f \star_G S$ in the operator norm. For (i), let $f_n = f\chi_{B(0,n)}$ so that f_n has compact support and $\|f_n - f\|_{L^\infty(G)} \to 0$ as $n \to \infty$. Then $f_n \in L^1_r(G)$ which implies that $f_n \star_G S \in \mathcal{S}^1$ by Proposition A.3.5 and so $f_n \star_G S$ is compact. Lemma A.4.12 then yields that

$$\|f \star_G S - f_n \star_G S\|_{B(\mathcal{H})} \leq \|f - f_n\|_{L^\infty(G)} \|\mathcal{D}^{-1} S \mathcal{D}^{-1}\|_{\mathcal{S}^1} \to 0$$

as $n \to \infty$, which yields the desired conclusion.

For (ii), since S is compact, it is the limit in operator norm of a sequence of finite rank operators $(S_n)_n \subset \mathcal{S}^1$. By Proposition A.3.5, $f \star_G S_n$ is compact. Thus by Proposition A.4.11 with $p = \infty$,

$$\|f \star_G S - f \star_G S_n\|_{B(\mathcal{H})} \leq \|f\|_{L^1_r(G)} \|S - S_n\|_{B(\mathcal{H})} \to 0$$

as $n \to \infty$ and so $f \star S$ is also the limit in operator norm of a sequence of compact operators, hence it is compact. ◻

A.5 Applications

A.5.1 Cohen's class distributions

Cohen's class of time-frequency distributions have a clear generalization to locally compact groups via quantum harmonic analysis and their integrability will in this section be shown to be connected to admissibility of an associated operator. In [154] it was shown that, with the Weyl-Heisenberg group as the underlying group, any Cohen's class distribution can be written as in the following definition using the Weyl transform from Section A.2.2.

Definition A.5.1. A bilinear map $Q : \mathcal{H} \times \mathcal{H} \to L^\infty(G)$ is said to belong to the *Cohen's class* if $Q = Q_S$ for some $S \in B(\mathcal{H})$ where

$$Q_S(\psi, \phi)(x) = (\psi \otimes \phi) \star_G S(x) = \langle S\sigma(x)\psi, \sigma(x)\phi \rangle. \tag{A.5.1}$$

We write $Q_S(\psi, \psi) = Q_S(\psi)$.

Cohen's class distributions should be thought of as generalizations of the spectrogram and scalogram as illustrated by the following example.

Example A.5.2. When S is the rank-one operator $S = \xi \otimes \eta$,

$$Q_S(\psi, \phi)(x) = \langle \xi, \sigma(x)\phi \rangle \overline{\langle \eta, \sigma(x)\psi \rangle}.$$

In particular, when $S = \xi \otimes \xi$,

$$Q_S(\psi)(x) = |\langle \xi, \sigma(x)\psi|^2 = |\langle \psi, \sigma(x)^* \xi \rangle|^2$$

which reduces down to the spectrogram in the Weyl-Heisenberg case and the scalogram in the affine case after replacing x with x^{-1}. Note also that by the linearity of the mapping $S \mapsto Q_S$, if S had been a finite rank operator, Q_S would have been a sum of functions of the above form.

Example A.5.3. Because of the formalism we have set up, we can easily compute the Cohen's class distribution corresponding to the operator $f \star_G S$ using Proposition A.3.13 as

$$Q_{f \star_G S}(\psi, \phi) = (\psi \otimes \phi) \star_G (f \star_G S) = f *_G ((\psi \otimes \phi) \star_G S) = f *_G Q_S(\psi, \phi).$$

From here we can deduce some elementary properties of Q_S, generalizing the results in [32, Prop. 6.9] and [154, Prop. 7.2, 7.3, 7.5].

Proposition A.5.4. Let $S \in B(\mathcal{H})$. Then for $\psi, \phi \in \mathcal{H}$ the following properties hold:

(i) The function $Q_S(\psi, \phi)$ satisfies

$$\|Q_S(\psi, \phi)\|_{L^\infty(G)} \leq \|S\|_{B(\mathcal{H})} \|\psi\| \|\phi\|.$$

(ii) If S is admissible, then $Q_S(\psi, \phi) \in L^1_r(G)$ and

$$\int_G Q_S(\psi, \phi)(x)\, d\mu_r(x) = \langle \psi, \phi \rangle \operatorname{tr}(\mathcal{D}^{-1} \mathcal{S} \mathcal{D}^{-1}).$$

(iii) If S is trace-class and ψ, ϕ are admissible, then $Q_S(\psi, \phi) \in L^1_\ell(G)$ and

$$\int_G Q_S(\psi, \phi)(x)\, d\mu_\ell(x) = \langle \mathcal{D}^{-1}\psi, \mathcal{D}^{-1}\phi \rangle \operatorname{tr}(\mathcal{S}).$$

(iv) We have the covariance property

$$Q_S(\sigma(x)\psi, \sigma(x)\phi)(y) = Q_S(\psi, \phi)(yx) \tag{A.5.2}$$

for all $x, y \in G$.

(v) The function $Q_S(\psi)$ is (real-valued) positive for all $\psi \in \mathcal{H}$ if and only if S is (self-adjoint) positive.

Proof. Item (i) is a direct consequence of Lemma A.3.11 or alternatively the Cauchy-Schwarz inequality while item (ii) follows from Corollary A.4.4 and item (iii) follows by an argument similar to that in the proof of Corollary A.4.6. Item (iv) follows by a straight-forward calculation. Item (v) is clear from the definition (A.5.1). □

Remark A.5.5. In [154], Cohen's class distributions Q_S on the Weyl-Heisenberg group are said to have the *correct total energy property* if they satisfy

$$\int_{\mathbb{R}^{2d}} Q_S(\psi, \phi)(x, \omega)\, dx\, d\omega = \langle \psi, \phi \rangle.$$

This corresponds to S being admissible with admissibility constant $\operatorname{tr}(\mathcal{D}^{-1} S \mathcal{D}^{-1}) = 1$ by item (ii) above.

It also turns out that under rather loose conditions, any bilinear map on $\mathcal{H} \times \mathcal{H}$ is a Cohen's class distribution as shown in the following proposition generalizing [107, Thm. 4.5.1] in the Weyl-Heisenberg case and [32, Prop. 6.11] in the affine case.

Proposition A.5.6. Let $Q : \mathcal{H} \times \mathcal{H} \to L^\infty(G)$ be a bilinear map. Assume that for all $\psi, \phi \in \mathcal{H}$ we know that $Q(\psi, \phi)$ is a continuous function on G that satisfies the covariance property (A.5.2) and the estimate

$$|Q(\psi, \phi)(0_G)| \leq C \|\psi\| \|\phi\|$$

for some constant C. Then there exists a unique bounded operator $S \in B(\mathcal{H})$ such that $Q = Q_S$.

Proof. By the Riesz representation theorem, there exists a bounded operator S such that

$$\langle S\psi, \phi \rangle = Q(\psi, \phi)(0_G)$$

and that $Q = Q_S$ now follows from the covariance relation (A.5.2). □

In [154, Thm. 7.6], positive Cohen's class distributions with the correct total energy property were characterized as (possibly infinite) convex combinations of spectrograms in the Weyl-Heisenberg case. An analogous result holds in the locally compact setting.

Theorem A.5.7. *If $S \in \mathcal{S}^1$ is positive, then there exists an orthonormal basis $\{\varphi_n\}_n$ and a sequence $\{\lambda_n\}_n$ of non-negative numbers with $\sum_n \lambda_n = \mathrm{tr}(S)$ such that*

$$Q_S(\psi)(x) = \sum_n \lambda_n |\langle \psi, \sigma(x)^* \varphi_n \rangle|^2$$

where the convergence in the sum is uniform for any $\psi \in \mathcal{H}$.

Proof. Since S is trace-class and positive, it can be expanded in its singular value decomposition

$$S = \sum_n \lambda_n (\varphi_n \otimes \varphi_n)$$

where $\sum_n \lambda_n = \mathrm{tr}(S)$ by a theorem due to Lidskii [190]. This allows us to write

$$\begin{aligned} Q_S(\psi)(x) &= (\psi \otimes \psi) \star_G \sum_n \lambda_n \varphi_n \otimes \varphi_n \\ &= \sum_n \lambda_n (\psi \otimes \psi) \star_G (\varphi_n \otimes \varphi_n) \\ &= \sum_n \lambda_n |\langle \psi, \sigma(x)^* \varphi_n \rangle|^2. \end{aligned}$$

That the convergence is uniform follows by Lemma A.3.11 applied to $(\psi \otimes \psi) \star_G (\varphi_n \otimes \varphi_n)$. □

The following proposition highlights a connection between the operator $f \star_G S$ and the Cohen's class distribution Q_S. It is a straight-forward generalization of the Weyl-Heisenberg situation described in [154, Prop. 8.2] and the affine case considered in [32, Prop. 6.12].

Proposition A.5.8. Let S be a positive, compact operator on $\mathcal{H}$ and let $f \in L^1_r(G)$ be a non-negative function. Then $f \star_G S$ is a positive, compact operator. Denote by $(\lambda_n)_n$ its eigenvalues in non-decreasing order with associated orthogonal eigenvectors $(\phi_n)_n$. Then

$$\lambda_n = \max_{\|\psi\|=1} \left\{ \int_G f(x) Q_S(\psi)(x)\, d\mu_r(x) : \psi \perp \phi_k \text{ for } k = 1, \ldots, n-1 \right\}.$$

Proof. Positivity of $f \star_G S$ follows from Lemma A.3.7 while compactness follows from Corollary A.4.15 (ii). The eigenvalue equality can be seen as a consequence of Courant's minimax theorem [148, Thm. 28.4] upon noting that

$$\langle f \star_G S\psi, \psi \rangle = \int_G f(x) Q_S(\psi)(x)\, d\mu_r(x)$$

which can be seen as a consequence of (A.5.1). □

Since we have an L^∞ bound on $Q_S(\psi)$ from Proposition A.5.4, we can formulate an uncertainty principle in the same way as [154, Cor. 7.7] and [107, Prop. 3.3.1] does for the Weyl-Heisenberg situation which says that if much of the mass of Q_S is concentrated in a subset $\Omega \subset G$, Ω cannot be too small.

Corollary A.5.9. Let $S \in B(\mathcal{H})$ and Ω be a compact subset of G such that

$$\int_\Omega |Q_S(\psi)(x)|\, d\mu_r(x) \geq (1-\varepsilon)\|S\|_{B(\mathcal{H})} \tag{A.5.3}$$

for some $\psi \in \mathcal{H}$ with $\|\psi\| = 1$. Then,

$$\mu_r(\Omega) \geq 1-\varepsilon.$$

Proof. Using (A.5.3), an elementary integral estimate and property (i) of Proposition A.5.4, we obtain

$$\begin{aligned}(1-\varepsilon)\|S\|_{B(\mathcal{H})} &\leq \int_\Omega |Q_S(\psi)(x)|\, d\mu_r(x)\\ &\leq \|Q_S(\psi)\|_{L^\infty(G)}\mu_r(\Omega)\\ &\leq \|S\|_{B(\mathcal{H})}\mu_r(\Omega)\end{aligned}$$

which yields the desired inequality upon dividing out $\|S\|_{B(\mathcal{H})}$. □

Example A.5.10. Consider the case where $\mathcal{H} = L^2(\mathbb{R}^2)$ and $G = \mathbb{S}$, the shearlet group. If we let $\mathcal{SH}_\varphi\psi$ denote the shearlet transform of $\psi \in L^2(\mathbb{R}^2)$ with respect to the normalized window φ, the above corollary applied to a subset $\Omega \subset \mathbb{S}$ reads as

$$\int_\Omega |\mathcal{SH}_\varphi\psi(a,s,x)|^2 \frac{da\, ds\, dx}{a^3} \geq 1-\varepsilon \implies \int_\Omega \frac{da\, ds\, dx}{a} \geq 1-\varepsilon$$

where we used the result of Example A.5.2 and the left and right Haar measure relations from Section A.2.3.

A.5.2 Mixed-state localization operators

The localization operators introduced in Section A.2.2 have an equivalent formulation using quantum harmonic analysis for which the generalization to locally compact groups is clear.

Definition A.5.11. Let $f \in L^1_r(G)$ and $\varphi_1, \varphi_2 \in \mathcal{H}$. We then define the *localization operator* $A_f^{\varphi_1,\varphi_2}$ on $\mathcal{H}$ as the operator

$$A_f^{\varphi_1,\varphi_2} = f \star_G (\varphi_1 \otimes \varphi_2).$$

By Proposition A.3.5, all localization operators are trace-class operators.

Remark A.5.12. The rank-one case of the interpolation results in Section A.4.3 can be found stated for localization operators in [220, Chap. 13] which notably investigates localization operators on locally compact groups.

Borrowing some terminology from quantum mechanics, the operator $\varphi_1 \otimes \varphi_2 = \varphi \otimes \varphi$ describes a *pure state* of a system while a positive trace-class operator which does not have rank one describes a *mixed state* since it is the limit of a linear combination of pure states. This leads us to the following two definitions, discussed in more depth in [154, 155] for the Weyl-Heisenberg case.

Definition A.5.13. A positive trace-class operator that is admissible with

$$\operatorname{tr}(\mathcal{D}^{-1} S \mathcal{D}^{-1}) = 1$$

is said to be a *density operator*.

Definition A.5.14. Let S be a density operator and $f \in L^1_r(G)$. Then the *mixed-state localization operator* corresponding to S and f is defined as $f \star_G S$.

Remark A.5.15. It is possible to view localization operators as induced by a broader class of functions and operators and investigate the properties of those localization operators as in [42]. For the purposes of this paper we restrict our attention to localization operators as specified by the definitions above.

Remark A.5.16. Mixed-state localization operators show up as the natural analogue of localization operators for the operator wavelet transform defined in Section A.5.4 below where the density operator condition is equivalent to the operator wavelet transform being an isometry.

We are especially interested in the case where $f = \chi_\Omega$ where Ω is a compact subset of G and S is a density operator in which case we refer to $\chi_\Omega \star_G S$ as the mixed-state localization operator corresponding to S and Ω. By the positivity and compactness of $\chi_\Omega \star_G S$, we can write the singular value decomposition as

$$\chi_\Omega \star_G S = \sum_k \lambda_k^\Omega (h_k^\Omega \otimes h_k^\Omega)$$

where the eigenvalues $(\lambda_k^\Omega)_k$ are ordered decreasingly.

In [155], we have the following result on the distributions of the eigenvalues of mixed-state localization operators in the Weyl-Heisenberg case, essentially saying that the number of eigenvalues of $\chi_\Omega \star_G S$ close to 1 scales as the size of the compact subset Ω provided S is a density operator. This was originally proved in the rank-one case in [74].

Theorem A.5.17 ([155, Thm. 4.4])**.** *Let S be a density operator on $L^2(\mathbb{R}^d)$, let $\Omega \subset \mathbb{R}^{2d}$ be a compact domain and fix $\delta \in (0,1)$. If $\{\lambda_k^{R\Omega}\}_k$ are the non-zero eigenvalues of $\chi_{R\Omega} \star_{\mathbb{H}^n} S$, then*

$$\frac{\#\{k : \lambda_k^{R\Omega} > 1-\delta\}}{|R\Omega|} \to 1 \quad \text{as } R \to \infty.$$

The proof of this theorem uses of approximate identities which show up as a consequence of the scaling $R\Omega$. In the locally compact setting we cannot consider dilations and therefore settle for proving the corresponding statement for the affine group.

Theorem A.5.18. *Let S be a density operator on $L^2(\mathbb{R}^+)$, let $\Omega \subset \mathit{Aff}$ be a compact domain and fix $\delta \in (0,1)$. If $\{\lambda_k^{R\Omega}\}_k$ are the non-zero eigenvalues of $\chi_{R\Omega} \star_{\mathit{Aff}} S$, then*

$$\frac{\#\{k : \lambda_k^{R\Omega} > 1-\delta\}}{\operatorname{tr}(S)\mu_r(R\Omega)} \to 1 \quad \text{as } R \to \infty. \tag{A.5.4}$$

Note that the scaled set $R\Omega$ is defined as in Section A.2.3 using the scaling function

$$\Gamma_R : \mathrm{Aff} \to \mathrm{Aff}, \qquad \Gamma_R(x,a) = (Rx, a^R), \qquad \Gamma_R^{-1}(x,a) = \left(\frac{x}{R}, a^{1/R}\right)$$

where $R\Omega = \{\Gamma_R(x,a), (x,a) \in \Omega\}$.

Before starting the proof, we will need to establish some auxiliary lemmas, some of which we state in the locally compact setting for the sake of generality.

Lemma A.5.19. *Let S be a density operator and $\Omega \subset G$ a compact domain. Then the eigenvalues $\{\lambda_k^{\Omega}\}_k$ of $\chi_\Omega \star_G S$ satisfy $0 \le \lambda_k^\Omega \le 1$.*

Proof. All of the eigenvalues are non-negative and real-valued since $\chi_\Omega \star_G S$ is a positive operator by Lemma A.3.7. For the upper limit, we have

$$\begin{aligned}\lambda_k^\Omega &= \left\langle (\chi_\Omega \star_G S) h_k^\Omega, h_k^\Omega \right\rangle = \int_G \chi_\Omega(x) \left\langle \sigma(x)^* S \sigma(x) h_k^\Omega, h_k^\Omega \right\rangle d\mu_r(x) \\ &\le \int_G Q_S(h_k^\Omega)(x)\, d\mu_r(x) = \left\langle h_k^\Omega, h_k^\Omega \right\rangle = 1\end{aligned}$$

where we used Proposition A.5.4 (ii). □

Lemma A.5.20. *Let S be a density operator, then the function $\tilde{S} = S \star_G S$ is non-negative, nonzero at 0_G and has total integral 1 with respect to both the left and right Haar measure.*

Proof. Non-negativity of $\tilde{S}$ follows from Lemma A.3.12 and plugging in $T = S$ and $x = 0_G$ in its proof yields

$$\tilde{S}(0_G) = \sum_n \lambda_n^2 > 0$$

where $(\lambda_n)_n$ are the eigenvalues of S. Meanwhile the last statement follows from Corollary A.4.6 and S being a density operator. □

The following technical lemma on approximations of the identity is standard in the Weyl-Heisenberg case but requires some work in the affine case. Note that the interior in the formulation below refers to the ball defined in Section A.2.3 as

$$B_r^{\mathrm{Aff}}\big((x,a),\delta\big) = \big\{(y,b) \in \mathrm{Aff} : d_r^{\mathrm{Aff}}\big((x,a),(y,b)\big) < \delta\big\}$$

where

$$d_r^{\mathrm{Aff}}\big((x,a),(y,b)\big) = |x-y| + \left|\ln\frac{a}{b}\right|.$$

Lemma A.5.21. *Let $\phi \in L_r^1(\mathrm{Aff})$ be non-negative, have total integral 1 with respect to both the left and right Haar measure and be nonzero at $(0,1)$. Moreover, let $\Omega \subset \mathrm{Aff}$ be compact and (y,b) be a point in the interior of Ω, then*

$$\lim_{R\to\infty} R^2 \int_\Omega \phi\big((\Gamma_R(x,a))(\Gamma_R(y,b))^{-1}\big)\, d\mu_r(x,a) = 1.$$

Proof. Using the change of variables $(z,c) = \Gamma_R(x,a)$ and $(w,u) = \Gamma_R(y,b)$ and the elementary observation $\Omega \subset \mathrm{Aff}$, we see that

$$\begin{aligned}\lim_{R\to\infty} R^2 \int_\Omega &\phi\big((\Gamma_R(x,a))(\Gamma_R(y,b))^{-1}\big)\, d\mu_r(x,a)\\ &\leq \int_{\mathrm{Aff}} \phi\big((z,c)(w,u)^{-1}\big)\, d\mu_r(z,c) = 1.\end{aligned}$$

We devote the remainder of the proof to showing that the limit is greater than or equal to 1 also. Choose δ so small that $B_r^{\mathrm{Aff}}\big((y,b),\delta\big) \subset \Omega$, then

$$\begin{aligned}\lim_{R\to\infty} R^2 \int_\Omega &\phi\big((\Gamma_R(x,a))(\Gamma_R(y,b))^{-1}\big)\, d\mu_r(x,a)\\ &\geq \lim_{R\to\infty} R^2 \int_{B_r^{\mathrm{Aff}}((y,b),\delta)} \phi\big((\Gamma_R(x,a))(\Gamma_R(y,b))^{-1}\big)\, d\mu_r(x,a)\\ &= \lim_{R\to\infty} \int_{\{(z,c):\Gamma_R^{-1}(z,c)\in B_r^{\mathrm{Aff}}((y,b),\delta)\}} \phi\big((z,c)(w,u)^{-1}\big)\, d\mu_r(z,c),\end{aligned}$$

where we once again used the change of variables $(z,c) = \Gamma_R(x,a)$ and $(w,u) = \Gamma_R(y,b)$. We claim that for each point (z,c), there exists an R so large that $\Gamma_R^{-1}((z,c)) \in B_r^{\mathrm{Aff}}((y,b),\delta)$ meaning that the region we are integrating over grows to all of Aff as $R \to \infty$. The desired conclusion will then follow by the monotone convergence theorem. Indeed, a quick computation using that $\Gamma_R^{-1}(x,a) = \left(\frac{x}{R}, a^{1/R}\right)$ yields

$$\begin{aligned}\Gamma_R^{-1}(z,c) \in B_r^{\mathrm{Aff}}\big(\Gamma_R^{-1}(w,u),\delta\big) &\iff d_r^{\mathrm{Aff}}\big(\Gamma_R^{-1}(z,c),\Gamma_R^{-1}(w,u)\big) < \delta \\ &\iff \left|\frac{z-w}{R}\right| + \left|\ln\frac{c^{1/R}}{u^{1/R}}\right| = \frac{1}{R}\left(|z-w| + \left|\ln\frac{c}{u}\right|\right) < \delta\end{aligned}$$

which clearly is true for large enough R. Hence

$$\begin{aligned}\lim_{R\to\infty}\int_{\{(z,c):\Gamma_R^{-1}(z,c)\in B_r^{\mathrm{Aff}}((y,b),\delta)\}} &\phi((z,c)(w,u)^{-1})\, d\mu_r(z,c) \\ &= \int_{\mathrm{Aff}} \phi(zw^{-1})\, d\mu_r(z,c) = 1\end{aligned}$$

which yields the desired conclusion. □

The following lemma generalizes [154, Prop. 6.1] and [155, Lem. 4.2].

Lemma A.5.22. *Let Ω be a compact subset of G and let $S \in \mathcal{S}^1$ be a positive operator. Then*

(i) $\mathrm{tr}(\chi_\Omega \star_G S) = \mathrm{tr}(S)\mu_r(\Omega)$.

(ii) $\mathrm{tr}((\chi_\Omega \star_G S)^2) = \int_\Omega\int_\Omega \tilde{S}(xy^{-1})\, d\mu_r(x)\, d\mu_r(y)$.

(iii) If $\{\lambda_k^\Omega\}_k$ are the eigenvalues of $\chi_\Omega \star_G S$ counted with algebraic multiplicity, then

$$\sum_k \lambda_k^\Omega = \mathrm{tr}(S)\mu_r(\Omega).$$

Proof. (i) This follows from Proposition A.3.6.

(ii) By an argument analogous to that in the proof of Proposition A.3.13 and properties of the Haar measure,

$$S \star_G (\chi_\Omega \star_G S)(x) = \int_\Omega \tilde{S}(yx)\, d\mu_r(y).$$

Hence,

$$\begin{aligned}\operatorname{tr}((\chi_\Omega \star_G S)^2) &= (\chi_\Omega \star_G S) \star_G (\chi_\Omega \star_G S)(0_G) \\ &= \chi_\Omega * \left(\int_\Omega \tilde{S}(y\,\cdot)\,d\mu_r(y)\right)(0_G) \\ &= \int_\Omega \int_\Omega \tilde{S}(xy^{-1})\,d\mu_r(x)\,d\mu_r(y).\end{aligned}$$

(iii) This follows from the fact that the sum of the eigenvalues counted with algebraic multiplicity is equal to the trace [190] and Proposition A.3.6.

□

From here we can follow the proof in [155], which in turn follows the proof in [8], by first generalizing [155, Lem. 4.3] to the locally compact setting.

Lemma A.5.23. *Let S be a density operator, $\Omega \subset G$ be compact and fix $\delta \in (0,1)$. Then*

$$\begin{aligned}&\left|\#\left\{k : \lambda_k^\Omega > 1-\delta\right\} - \operatorname{tr}(S)\mu_r(\Omega)\right| \\ &\qquad \le \max\left\{\frac{1}{\delta}, \frac{1}{1-\delta}\right\}\left|\int_\Omega \int_\Omega \tilde{S}(xy^{-1})\,d\mu_r(x)\,d\mu_r(y) - \operatorname{tr}(S)\mu_r(\Omega)\right|.\end{aligned}$$

Proof. Define the function

$$G(t) = \begin{cases} -t & \text{if } 0 \le t \le 1-\delta, \\ 1-t & \text{if } 1-\delta < t \le 1 \end{cases}$$

and consider the operator

$$G(\chi_\Omega \star_G S) = \sum_k G(\lambda_k^\Omega) h_k^\Omega \otimes h_k^\Omega$$

which is well defined since $0 \le \lambda_k^\Omega \le 1$ by Lemma A.5.19. The sequence $\left\{G(\lambda_k^\Omega)\right\}_k$ is absolutely summable since $\left\{\lambda_k^\Omega\right\}_k$ is summable and $|G(\lambda_k^\Omega)| = \lambda_k^\Omega$ for large enough k. It therefore follows that

$$\operatorname{tr}(G(\chi_\Omega \star_G S)) = \sum_k G(\lambda_k^\Omega) = \#\left\{k : \lambda_k^\Omega > 1-\delta\right\} - \operatorname{tr}(S)\mu_r(\Omega)$$

using Lemma A.5.22 (iii) and the above. Hence

$$\left|\#\left\{k : \lambda_k^\Omega > 1-\delta\right\} - \operatorname{tr}(S)\mu_r(\Omega)\right| = \left|\operatorname{tr}\left(G(\chi_\Omega \star_G S)\right)\right| \le \operatorname{tr}\left(|G|(\chi_\Omega \star_G S)\right).$$

To bound this, we need an estimate on $|G(t)|$. We wish to decide the constant A such that the polynomial $At(1-t)$ is always greater than $|G(t)|$. Note that it suffices to make sure that this is the case around $t = 1-\delta$ by the concavity of $At(1-t)$ and so we can set

$$A = \sup_{t\in[0,1]} \frac{|G(t)|}{t(1-t)} = \max\left\{\frac{1}{\delta}, \frac{1}{1-\delta}\right\}$$

yielding

$$|G(t)| \leq \max\left\{\frac{1}{\delta}, \frac{1}{1-\delta}\right\}(t-t^2).$$

With the bounds on $|G(t)|$ established, it follows that

$$\left|\#\left\{k : \lambda_k^\Omega > 1-\delta\right\} - \mathrm{tr}(S)\mu_r(\Omega)\right| \leq \max\left\{\frac{1}{\delta}, \frac{1}{1-\delta}\right\} \mathrm{tr}\left(\chi_\Omega \star_G S - (\chi_\Omega \star_G S)^2\right)$$

and we obtain the desired conclusion by applying Lemma A.5.22 to the right hand side. □

Proof of Theorem A.5.18. By Lemma A.5.23, we have the bound

$$\begin{aligned}
&\left|\#\left\{k : \lambda_k^{R\Omega} > 1-\delta\right\} - \mathrm{tr}(S)\mu_r(R\Omega)\right| \\
&\leq \max\left\{\frac{1}{\delta}, \frac{1}{1-\delta}\right\}\left|\int_{R\Omega}\int_{R\Omega} \tilde{S}\big((x,a)(y,b)^{-1}\big)\, d\mu_r(x,a)\, d\mu_r(y,b) - \mathrm{tr}(S)\mu_r(R\Omega)\right|.
\end{aligned}$$

Hence by dividing by $\mathrm{tr}(S)\mu_r(R\Omega)$ and setting $\phi = \tilde{S}/\mathrm{tr}(S)$ which has total integral 1 by Corollary A.4.4 and the fact that S is a density operator, we get

$$\begin{aligned}
&\left|\frac{\#\left\{k : \lambda_k^{R\Omega} > 1-\delta\right\}}{\mathrm{tr}(S)\mu_r(R\Omega)} - 1\right| \\
&\quad\leq \max\left\{\frac{1}{\delta}, \frac{1}{1-\delta}\right\}\left|\frac{1}{\mu_r(R\Omega)}\int_{R\Omega}\int_{R\Omega} \phi\big((x,a)(y,b)^{-1}\big)\, d\mu_r(x,a)\, d\mu_r(y,b) - 1\right| \\
&\quad= \max\left\{\frac{1}{\delta}, \frac{1}{1-\delta}\right\}\left|\int_{R\Omega}\frac{1}{\mu_r(R\Omega)}\left(\int_{R\Omega} \phi\big((x,a)(y,b)^{-1}\big)\, d\mu_r(x,a) - 1\right) d\mu_r(y,b)\right| \\
&\quad\leq \max\left\{\frac{1}{\delta}, \frac{1}{1-\delta}\right\}\int_{R\Omega}\frac{1}{\mu_r(R\Omega)}\left|\int_{R\Omega} \phi\big((x,a)(y,b)^{-1}\big)\, d\mu_r(x,a) - 1\right| d\mu_r(y,b).
\end{aligned}$$

We now focus on showing that this approaches zero as $R \to \infty$. The change of variables $(x,a) = \Gamma_R(z,c)$ and $(y,b) = \Gamma_R(w,u)$ yields

$$\begin{aligned}
&\int_{R\Omega}\frac{1}{\mu_r(R\Omega)}\left|\int_{R\Omega} \phi((x,a)(y,b)^{-1})\, d\mu_r(x,a) - 1\right| d\mu_r(y,b) \\
&\quad= R^2\int_\Omega \frac{1}{\mu_r(R\Omega)}\left|R^2\int_\Omega \phi\big((\Gamma_R(z,c))(\Gamma_R(w,u))^{-1}\big)\, d\mu_r(z,c) - 1\right| d\mu_r(w,u).
\end{aligned}$$

From here we can move the $R \to \infty$ limit inside the integral by the compactness of Ω and the bound $\left|R^2 \int_\Omega \phi((\Gamma_R(z,c))(\Gamma_R(w,u))^{-1})\, d\mu_r(z,c) - 1\right| \leq 2$. To conclude that the outer integral is zero, we need to make the integrand small even when making the change of variables back to $(x,a),(y,b)$ which requires that $\left|R^2 \int_\Omega \phi((\Gamma_R(z,c))(\Gamma_R(w,u))^{-1})\, d\mu_r(z,c) - 1\right|$ vanishes as $R \to \infty$. This is the contents of Lemma A.5.21 and so we are done. □

Remark A.5.24. The formulation of Theorem A.5.18 can be generalized to the locally compact setting as

$$\frac{\#\left\{k : \lambda_k^{\Omega_n} > 1 - \delta\right\}}{\mathrm{tr}(S)\mu_r(\Omega_n)} \to 1 \quad \text{as } n \to \infty$$

where $(\Omega_n)_n$ is an exhausting sequence of G. Unfortunately, the proof does not fully carry over to this formulation but as long as an approximate identity lemma analogous to Lemma A.5.21 can be proved the full statement should follow as well. We therefore expect generalizations of Theorem A.5.18 to homogeneous groups or stratified Lie groups to hold.

For an example of how density operators can be constructed, see the following proposition.

Proposition A.5.25. Let ψ be an admissible vector such that $\|\mathcal{D}\psi\| = 1$. Then $S = \psi \otimes \psi$ is a density operator. Moreover, if $f \in L^1_\ell(G)$ is non-negative with $\|f\|_{L^1_\ell(G)} = 1$ and S is a density operator, then $f \star_G S$ is also a density operator.

Proof. The first part is clear from the definition of density operators and the second part follows by Lemma A.3.7 and Proposition A.4.10. □

Density operators can also be constructed from linear combinations of admissible functions as in Proposition A.4.8.

Remark A.5.26. The construction carried out for the affine group in this section can just as easily be applied to the shearlet group by defining the dilation function

$$\Gamma_R : \mathbb{S} \to \mathbb{S}, \qquad \Gamma_R(a,s,x) = \left(a^R, \frac{s}{R}, \frac{x}{R}\right),$$

which has the property $\mu_r(R\Omega) = R^4\mu_r(\Omega)$, and the distance function

$$d_r^{\mathbb{S}}((a,s,x),(b,t,y)) = \left|\ln \frac{a}{b}\right| + |s-t| + |x-y|.$$

The proof in Lemma A.5.21 then works in the same way with obvious modifications.

A.5.3 Covariant integral quantizations

Covariant integral quantizations on G are maps Γ_S given by

$$\Gamma_S(f) = f \star_G S$$

which include mixed-state localization operators as a special case. They have been studied by Gazeau and collaborators, motivated by applications in physics [95–97].

The following result generalizes [211, Prop. 3.2 (3)] and [32, Prop. 6.4].

Proposition A.5.27. Let T be a trace-class operator on $\mathcal{H}$. Then

$$\Gamma_{\mathcal{D}T\mathcal{D}}(1) = 1 \star_G \mathcal{D}T\mathcal{D} = \operatorname{tr}(T) I_{\mathcal{H}}.$$

Proof. Let $\psi, \phi \in \operatorname{Dom}(D)$, then using Lemma A.4.12,

$$\begin{aligned}
\langle 1 \star_G \mathcal{D}T\mathcal{D}\psi, \phi\rangle &= \int_G \langle T\mathcal{D}\sigma(x)\psi, \mathcal{D}\sigma(x)\phi\rangle \, d\mu_r(x) \\
&= \int_G (\psi \otimes \phi) \star_G (\mathcal{D}T\mathcal{D})(x) \, d\mu_r(x) \\
&= \operatorname{tr}(\psi \otimes \phi) \operatorname{tr}(T) \\
&= \langle \psi, \phi\rangle \operatorname{tr}(T)
\end{aligned}$$

by Lemma A.3.9 and Theorem A.4.2. $\square$

From the above proposition we see that when $\operatorname{tr}(T) = 1$, $\Gamma_{\mathcal{D}T\mathcal{D}}(1)$ is the identity operator which has the following resolution of identity as a consequence

$$I_{\mathcal{H}} = \int_G \alpha_x\big(\mathcal{D}T\mathcal{D}\big) \, d\mu_r(x) \implies \langle \psi, \phi\rangle = \int_G \big\langle \mathcal{D}T\mathcal{D}\sigma(x)\psi, \sigma(x)\phi\big\rangle \, d\mu_r(x)$$

It turns out that this property together with a few more uniquely determines linear maps from $L^\infty(G)$ to $B(\mathcal{H})$ which is the contents of the following theorem from [137] which generalizes the formulations in [154, Thm. 6.2] and [32, Thm. 6.5].

Theorem A.5.28. *Let $\Gamma : L^\infty(G) \to B(\mathcal{H})$ be a linear map satisfying*

1. *Γ sends positive functions to positive operators,*
2. *$\Gamma(1) = I_{\mathcal{H}}$,*
3. *Γ is continuous from the weak* topology on $L^\infty(G)$ (as the dual space of $L^1_r(G)$) to the weak* topology on $B(\mathcal{H})$,*
4. *$\sigma(x)^*\Gamma(f)\sigma(x) = \Gamma(R_x^{-1} f)$,*

where $R_x f(y) = f(yx)$ is the right-translation map. Then there exists a unique positive trace-class operator T with $\mathrm{tr}(T) = 1$ *such that*

$$\Gamma(f) = f \star_G \mathcal{D}T\mathcal{D}.$$

Remark A.5.29. That $\Gamma(f) = f \star_G \mathcal{D}T\mathcal{D}$ satisfies all of these properties can be verified directly.

Proof. The proof is a matter of translating our situation to one described in a remark in [135] which references [137] for the full proof. This follows by first considering the bijection $\Gamma \mapsto \Gamma_l$ where $\Gamma_l(f) = \Gamma(\check{f})$ and $\check{f}(x) = f(x^{-1})$. To translate the result back into the form $f \star_G \mathcal{D}T\mathcal{D}$ we can use equation (A.4.2) from Theorem A.4.1. □

By Proposition A.4.14, the adjoint of $\Gamma_S = \mathcal{A}_S$ is given by the map $T \mapsto T \star_G S$ and so covariant integral quantizations can be seen as inducing operator-operator convolutions as well.

Remark A.5.30. Werner refers to mappings of the kind discussed above as *positive correspondence rules* in [211].

A.5.4 Operator wavelet transforms

In [194], the STFT of an element of $L^2(\mathbb{R}^d)$ with respect to a Hilbert-Schmidt operator is defined and a generalization of Moyal's identity is proved which mirrors the Duflo-Moore theorem. Later in [59], this approach was generalized by letting the STFT also act on operators in what they call the *operator STFT*. In this section we generalize both of these constructions to the locally compact setting.

Wavelet transform with operator window

Definition A.5.31. Let S be a bounded operator on $\mathcal{H}$ such that S^*S is admissible and $\psi \in \mathcal{H}$. Then the *wavelet transform with operator window* $\mathfrak{W}_S(\psi)$ of ψ with window to S is defined as the function

$$\mathfrak{W}_S(\psi)(x) = S\sigma(x)\psi \in \mathcal{H}$$

for $x \in G$.

Remark A.5.32. Our definition differs from that in [59, 194] in that we don't take the adjoint of σ to stay in line with the right Haar measure convention of this paper. Note also that the condition that S^*S is admissible corresponds to S being a Hilbert-Schmidt operator in the unimodular case.

The wavelet transform with operator window should be considered as an element of the Hilbert space $L^2_r(G, \mathcal{H})$ of equivalence classes of elements $\Psi : G \mapsto \mathcal{H}$ such that

$$\|\Psi\|_{L^2_r(G,\mathcal{H})} = \left(\int_G \|\Psi(x)\|^2_{\mathcal{H}} \, d\mu_r(x) \right)^{1/2} < \infty$$

where $\Psi \sim \Phi$ if $\Psi(x) = \Phi(x)$ in $\mathcal{H}$ for μ_r-a.e. $x \in G$. The inner product in this space is given by

$$\langle \Psi, \Phi \rangle_{L^2_r(G,\mathcal{H})} = \int_G \big\langle \Psi(x), \Phi(x) \big\rangle_{\mathcal{H}} \, d\mu_r(x).$$

We are now ready to prove the following orthogonality relation, which in particular shows that the transform indeed is an element of $L^2_r(G, \mathcal{H})$.

Proposition A.5.33. Let $S_1, S_2 \in B(\mathcal{H})$ be such that $S_2^* S_1$ is admissible and $\psi_1, \psi_2 \in \mathcal{H}$. Then

$$\big\langle \mathfrak{W}_{S_1}\psi_1, \mathfrak{W}_{S_2}\psi_2 \big\rangle_{L^2_r(G,\mathcal{H})} = \langle \psi_1, \psi_2 \rangle \langle S_1 \mathcal{D}^{-1}, S_2 \mathcal{D}^{-1} \rangle_{\mathcal{S}^2}.$$

In particular, $\mathfrak{W}_S \psi \in L^2_r(G, \mathcal{H})$ for $S \in B(\mathcal{H})$ such that $S^* S$ is admissible and $\psi \in \mathcal{H}$.

Proof. By writing

$$\begin{aligned} \big\langle \mathfrak{W}_{S_1}\psi_1(x), \mathfrak{W}_{S_2}\psi_2(x) \big\rangle &= \big\langle S_1 \sigma(x) \psi_1, S_2 \sigma(x) \psi_2 \big\rangle \\ &= \big\langle \alpha_x(S_2^* S_1) \psi_1, \psi_2 \big\rangle \\ &= ((\psi_1 \otimes \psi_2) \star_G S_2^* S_1)(x) \end{aligned}$$

using Lemma A.3.9, we see that this can be integrated using Corollary A.4.4 to yield

$$\begin{aligned} \big\langle \mathfrak{W}_{S_1}\psi_1, \mathfrak{W}_{S_2}\psi_2 \big\rangle_{L^2_r(G,\mathcal{H})} &= \operatorname{tr}(\psi_1 \otimes \psi_2) \operatorname{tr}(\mathcal{D}^{-1} S_2^* S_1 \mathcal{D}^{-1}) \\ &= \langle \psi_1, \psi_2 \rangle \operatorname{tr}(S_1 \mathcal{D}^{-1} (S_2 \mathcal{D}^{-1})^*) \\ &= \langle \psi_1, \psi_2 \rangle \langle S_1 \mathcal{D}^{-1}, S_2 \mathcal{D}^{-1} \rangle_{\mathcal{S}^2} \end{aligned}$$

as desired. □

Remark A.5.34. Note that when $S_1 = \xi \otimes \phi_1, S_2 = \xi \otimes \phi_2$ and ξ is normalized, we recover the familiar Duflo-Moore relation

$$\big\langle \mathfrak{W}_{S_1}\psi_1, \mathfrak{W}_{S_2}\psi_2 \big\rangle = \langle \psi_1, \psi_2 \rangle \overline{\langle \mathcal{D}^{-1}\phi_1, \mathcal{D}^{-1}\phi_2 \rangle}.$$

Moreover, when

$$S_1 = \sum_{n=1}^{N} \xi \otimes \phi_n, \qquad S_2 = \sum_{m=1}^{M} \xi \otimes \eta_m$$

for some normalized ξ, $S_2^* S_1$ is admissible precisely when ϕ_n and η_m are admissible for each n, m since

$$S_2^* S_1 = \sum_{n=1}^{N} \sum_{m=1}^{M} \eta_m \otimes \phi_n$$

by Proposition A.4.7 characterizing admissible rank-one operators.

Operator wavelet transform

Definition A.5.35. Let S be a bounded operator on $\mathcal{H}$ such that S^*S is admissible and $T \in \mathcal{HS}$, then the *operator wavelet transform* of T with respect to S is defined as

$$\mathfrak{W}_S T(x) = S\sigma(x)T.$$

Note that $\mathfrak{W}_S T$ maps elements of G to operators on $\mathcal{H}$. In fact, we will show that $\mathfrak{W}_S T$ is an element of the Hilbert space $L^2_r(G, \mathcal{HS})$ provided S satisfies the same admissibility criterion as for the wavelet transform with operator window. The inner product in this space is given by

$$\langle A, B \rangle_{L^2_r(G,\mathcal{HS})} = \int_G \langle A(x), B(x) \rangle_{\mathcal{HS}} \, d\mu_r(x).$$

As in the wavelet transform with operator window case, we still have a version of Moyal's identity which generalizes [59, Prop. 3.4].

Proposition A.5.36. Let S_1, S_2 be such that $S_2^* S_1$ is admissible and $T, R \in \mathcal{HS}$. Then

$$\left\langle \mathfrak{W}_{S_1} T, \mathfrak{W}_{S_2} R \right\rangle_{L^2_r(G,\mathcal{HS})} = \langle T, R \rangle_{\mathcal{HS}} \langle S_1 \mathcal{D}^{-1}, S_2 \mathcal{D}^{-1} \rangle_{\mathcal{HS}}.$$

In particular, $\mathfrak{W}_S T \in L^2_r(G, \mathcal{HS})$ for $S \in B(\mathcal{H})$ such that S^*S is admissible and $T \in \mathcal{HS}$.

Proof. We compute

$$\begin{aligned}
\left\langle \mathfrak{W}_{S_1} T, \mathfrak{W}_{S_2} R \right\rangle_{L^2_r(G,\mathcal{HS})} &= \int_G \left\langle \mathfrak{W}_{S_1} T(x), \mathfrak{W}_{S_2} R(x) \right\rangle_{\mathcal{HS}} d\mu_r(x) \\
&= \int_G \operatorname{tr}\left(S_1 \sigma(x) T R^* \sigma(x)^* S_2^* \right) d\mu_r(x) \\
&= \int_G ((TR^*) \star_G (S_2^* S_1))(x) \, d\mu_r(x) \\
&= \operatorname{tr}(TR^*) \operatorname{tr}(\mathcal{D}^{-1} S_2^* S_1 \mathcal{D}^{-1}) \\
&= \langle T, R \rangle_{\mathcal{HS}} \langle S_1 \mathcal{D}^{-1}, S_2 \mathcal{D}^{-1} \rangle_{\mathcal{HS}}
\end{aligned}$$

where we in the second to last step used Corollary A.4.4. □

Remark A.5.37. More structure such as the Toeplitz operators discussed in [59] carry over to the locally compact setting with similar modifications as in the above but in the interest of brevity we leave this be.

A.5.5 A Berezin-Lieb inequality

The Berezin-Lieb inequality as investigated in [138, 152, 211] can be seen as a generalization of Corollary A.4.4. We present here a generalization to locally compact groups with the proof partially based on a proof for the Weyl-Heisenberg case in [60].

Theorem A.5.38. *Fix a positive $T \in \mathcal{S}^1$ and let $S \in \mathcal{S}^1$ be admissible. If Φ is a non-negative, convex and continuous function on a domain containing the spectrum of $\mathrm{tr}(S)T$ and the range of $T \star_G S$, then*

$$\int_G \Phi \circ (T \star_G S)(x)\, d\mu_r(x) \leq \mathrm{tr}\left(\Phi(\mathrm{tr}(S)T)\right)\frac{\mathrm{tr}(\mathcal{D}^{-1}S\mathcal{D}^{-1})}{\mathrm{tr}(S)}$$

where $\Phi(S)$ is defined by the functional calculus. Similarly, if $S \in \mathcal{S}^1$ is positive and admissible, $f \in L^\infty(G)$ is non-negative and Φ is a non-negative, convex and continuous function on a domain containing the spectrum of $f \star_G S$ and the range of $\mathrm{tr}(\mathcal{D}^{-1}S\mathcal{D}^{-1})f$, then

$$\mathrm{tr}(\Phi(f \star_G S)) \leq \frac{\mathrm{tr}(S)}{\mathrm{tr}(\mathcal{D}^{-1}S\mathcal{D}^{-1})} \int_G \Phi\big(\mathrm{tr}(\mathcal{D}^{-1}S\mathcal{D}^{-1})f(x)\big)\, d\mu_r(x).$$

Proof. Expanding T in its singular value decomposition and using Lemma A.3.9, we obtain

$$\begin{aligned}
(T \star_G S)(x) &= \sum_n \lambda_n((\xi_n \otimes \xi_n) \star_G S)(x) \\
&= \sum_n \lambda_n \langle S\sigma(x)\xi_n, \sigma(x)\xi_n\rangle \\
&= \sum_n \mathrm{tr}(S)\lambda_n \frac{\langle S\sigma(x)\xi_n, \sigma(x)\xi_n\rangle}{\mathrm{tr}(S)}.
\end{aligned}$$

Since $\{\xi_n\}_n$ is an orthonormal basis, so is $\{\sigma(x)\xi_n\}_n$ and so for each x we can view the above as the integral over $(\mathrm{tr}(S)\lambda_n)_n$ with respect to a discrete probability measure. Applying Jensen's inequality, we find that

$$\begin{aligned}\Phi \circ (T \star_G S)(x) &= \Phi\left(\sum_n \operatorname{tr}(S)\lambda_n \frac{\langle S\sigma(x)\xi_n, \sigma(x)\xi_n\rangle}{\operatorname{tr}(S)}\right)\\ &\leq \sum_n \Phi(\operatorname{tr}(S)\lambda_n)\frac{1}{\operatorname{tr}(S)}\langle S\sigma(x)\xi_n, \sigma(x)\xi_n\rangle\end{aligned}$$

for each x. Integrating both sides yields

$$\begin{aligned}\int_G \Phi \circ (T \star_G S)(x)\, d\mu_r(x) &\leq \int_G \sum_n \Phi(\operatorname{tr}(S)\lambda_n)\frac{1}{\operatorname{tr}(S)}\langle S\sigma(x)\xi_n, \sigma(x)\xi_n\rangle\\ &= \sum_n \Phi(\operatorname{tr}(S)\lambda_n)\frac{1}{\operatorname{tr}(S)}\int_G (\xi_n \otimes \xi_n) \star_G S(x)\, d\mu_r(x)\\ &= \sum_n \Phi(\operatorname{tr}(S)\lambda_n)\frac{\operatorname{tr}(\mathcal{D}^{-1}S\mathcal{D}^{-1})}{\operatorname{tr}(S)}\end{aligned}$$

where we used Tonelli, Lemma A.3.9 and Corollary A.4.4.

For the function statement, we use that $f \star_G S$ is a positive operator by Lemma A.3.7 to expand $f \star_G S$ as $f \star_G S = \sum_n \lambda_n(\xi_n \otimes \xi_n)$ which yields

$$\Phi(f \star_G S) = \sum_n \Phi(\lambda_n)\xi_n \otimes \xi_n.$$

Now before taking the trace of this, note that by the above,

$$\big\langle \Phi(f \star_G S)\xi_n, \xi_n\big\rangle = \Phi(\lambda_n) = \Phi\big(\langle (f \star_G S)\xi_n, \xi_n\rangle\big)$$

and so

$$\begin{aligned}\operatorname{tr}(\Phi(f \star_G S)) &= \sum_n \big\langle \Phi(f \star_G S)\xi_n, \xi_n\big\rangle\\ &= \sum_n \Phi\big(\langle (f \star_G S)\xi_n, \xi_n\rangle\big)\\ &= \sum_n \Phi\left(\int_G f(x)\langle S\sigma(x)\xi_n, \sigma(x)\xi_n\rangle\, d\mu_r(x)\right)\\ &= \sum_n \Phi\left(\int_G \operatorname{tr}(\mathcal{D}^{-1}S\mathcal{D}^{-1}) f(x)\frac{\langle S\sigma(x)\xi_n, \sigma(x)\xi_n\rangle}{\operatorname{tr}(\mathcal{D}^{-1}S\mathcal{D}^{-1})}\, d\mu_r(x)\right).\end{aligned}$$

Now for each n, $\frac{\langle S\sigma(x)\xi_n, \sigma(x)\xi_n\rangle}{\operatorname{tr}(\mathcal{D}^{-1}S\mathcal{D}^{-1})}\, d\mu_r(x)$ can be viewed as a probability measure since the inner product $\langle S\sigma(x)\xi_n, \sigma(x)\xi_n\rangle$ integrates to $\operatorname{tr}(\mathcal{D}^{-1}S\mathcal{D}^{-1})$ by Lemma

A.3.9 and Corollary A.4.4. We can therefore apply Jensen's inequality to get

$$\begin{aligned}
\operatorname{tr}(\Phi(f \star_G S)) &\leq \sum_n \int_G \Phi\big(\operatorname{tr}(\mathcal{D}^{-1}S\mathcal{D}^{-1})f(x)\big) \frac{\langle S\sigma(x)\xi_n, \sigma(x)\xi_n\rangle}{\operatorname{tr}(\mathcal{D}^{-1}S\mathcal{D}^{-1})}\, d\mu_r(x) \\
&= \int_G \Phi\big(\operatorname{tr}(\mathcal{D}^{-1}S\mathcal{D}^{-1})f(x)\big) \frac{\sum_n \langle S\sigma(x)\xi_n, \sigma(x)\xi_n\rangle}{\operatorname{tr}(\mathcal{D}^{-1}S\mathcal{D}^{-1})}\, d\mu_r(x) \\
&= \frac{\operatorname{tr}(S)}{\operatorname{tr}(\mathcal{D}^{-1}S\mathcal{D}^{-1})} \int_G \Phi\big(\operatorname{tr}(\mathcal{D}^{-1}S\mathcal{D}^{-1})f(x)\big)\, d\mu_r(x)
\end{aligned}$$

where we used Tonelli to change the order of integration and summation. □

Remark A.5.39. A result similar to the rank-one case of the second inequality in the above theorem can be found in [220, Thm. 14.11].

A.5.6 Wiener's Tauberian theorem

In this section we extend a theorem originally proved in [136] which can also be found in [153]. It can be viewed as a descendant to Wiener's classical Tauberian theorem [214] and is similar in spirit to the ideal formulation in [149]. To state it, we first need to introduce the concept of regular functions and operators.

Definition A.5.40. A function $g \in L^p_r(G)$ is said to be p-regular if

$$\overline{\operatorname{span}\left\{g(\cdot\, x^{-1})\right\}_{x\in G}} = L^p_r(G).$$

Similarly, an operator $S \in \mathcal{S}^p$ is said to be p-regular if

$$\overline{\operatorname{span}\left\{\sigma(x)^* S \sigma(x)\right\}_{x\in G}} = \mathcal{S}^p.$$

We now state the theorem before a short discussion.

Theorem A.5.41. *Assume that there exists an admissible operator $R \in \mathcal{S}^1$ such that $R \star_G R$ is regular, let $S \in \mathcal{S}^p$ be admissible, $1 \leq p \leq \infty$ and let q be the conjugate exponent of p. Then the following are equivalent:*

1. *S is p-regular,*
2. *If $f \in L^q_r(G)$ and $f \star_G S = 0$, then $f = 0$,*
3. *$\mathcal{S}^p \star_G S$ is dense in $L^p_r(G)$,*
4. *If $T \in \mathcal{S}^q$ and $T \star_G S = 0$, then $T = 0$,*
5. *$L^p_r(G) \star_G S$ is dense in $\mathcal{S}^p$,*

6. *$S \star_G S$ is p-regular,*

7. *For any regular $T \in \mathcal{S}^1, T \star_G S$ is p-regular.*

Since for $1 \leq p \leq p' \leq \infty$, we have the inclusion $\mathcal{S}^p \subset \mathcal{S}^{p'}$, it suffices to find an admissible 1-regular operator to establish existence of p-regular operators for any $p \geq 1$. In the Weyl-Heisenberg case, the operator $S = \varphi_0 \otimes \varphi_0$ where φ_0 is the standard Gaussian is a regular operator as is verified directly in [153]. It is not as easy to find such an operator in the locally compact case as the closest we have is the indicator function on a compact neighborhood of the origin for which the proof method does not translate.

Apart from the existence of a regular operator, the proof in the Weyl-Heisenberg case from [153] carries over with minimal modifications to account for the use of the right Haar measure as opposed to the left Haar measure, the change of the underlying group and the requirement of admissibility. We show three implications in detail, the first of which is notable because it requires Proposition A.4.14 and motivates the requirement for S to be admissible, and leave the remaining required modifications from the proof in [153] to the reader.

Proof. (2) $\iff$ (3): By Proposition A.4.14, the mapping $\mathcal{A}_S : f \mapsto f \star S$ is adjoint to $\mathcal{B}_S : T \mapsto T \star_G S$ and by [185, Thm. 4.12], denseness of the image of a mapping is equivalent to injectiveness of its adjoint.

(4) $\iff$ (5): This follows by the same argument as the above but with the roles of $\mathcal{A}_S$ and $\mathcal{B}_S$ reversed.

(2) $\implies$ (4): Assume that $T \star_G S = 0$ for some $T \in \mathcal{S}^q$, then $A \star_G T \star_G S = 0$ for any $A \in \mathcal{S}^1$ which implies that $A \star T = 0$ for all $A \in \mathcal{S}^1$ by (2). In particular,

$$A \star_G T(0) = \operatorname{tr}(AT) = \langle A, T^* \rangle = 0 \quad \text{for all } A \in \mathcal{S}^1$$

and hence $T = 0$ as desired.

Again, the remainder of the implications follows with similar small modifications from [153] which we leave to the interested reader. □

Acknowledgements

The author wishes to extend their gratitude to Eirik Berge and Stine Marie Berge for helpful discussions and to Franz Luef for reading drafts of the paper as well as guidance.

Paper B

Measure-Operator Convolutions and Applications to Mixed-State Gabor Multipliers

Hans G. Feichtinger, Simon Halvdansson and Franz Luef

Published in
Sampling Theory, Signal Processing, and Data Analysis,
2024, Volume 22

Paper B

Measure-Operator Convolutions and Applications to Mixed-State Gabor Multipliers

Abstract

For the Weyl-Heisenberg group, convolutions between functions and operators were defined by Werner as a part of a framework called quantum harmonic analysis. We show how recent results by Feichtinger can be used to extend this definition to include convolutions between measures and operators. Many properties of function-operator convolutions carry over to this setting and allow us to prove novel results on the distribution of eigenvalues of mixed-state Gabor multipliers and derive a version of the Berezin-Lieb inequality for lattices. New results on the continuity of Gabor multipliers with respect to lattice parameters, masks and windows as well as their ability to approximate localization operators are also derived using this framework.

B.1 Introduction

In recent years the framework of quantum harmonic analysis, which was introduced by R. Werner in [211], has been successfully applied to many problems in time-frequency analysis and operator theory [90, 153, 155, 156]. The most central operations of the framework are the *function-operator* and *operator-operator* convolutions which generalize classical convolutions and basic objects in time-frequency analysis such as localization operators and Cohen's class distributions. In this paper, we will be focusing on the extension of the function-operator convolution, defined for $f \in L^1(\mathbb{R}^{2d})$ and a trace-class operator S via the Bochner integral

$$f \star S = \int_{\mathbb{R}^{2d}} f(z)\pi(z)S\pi(z)^* \, dz, \tag{B.1.1}$$

where $\pi(z)\psi(t) = M_\omega T_x\psi(t) = e^{2\pi i\omega\cdot t}\psi(t-x)$ is a time-frequency shift by $z = (x,\omega) \in \mathbb{R}^{2d}$.

For certain applications in time-frequency analysis, one would like to generalize this definition to the case of measures. We will show that it suffices to define this convolution $\star$ for point-measures δ_z: Namely, $\delta_z \star S = \pi(z)S\pi(z)^*$, in order to extend it to the general case, which is given weakly by

$$\mu \star S = \int_{\mathbb{R}^{2d}} \pi(z)S\pi(z)^* \, d\mu(z)$$

for S in the Schatten p-class of operators for $1 \leq p \leq \infty$. Notably, this implies (B.1.1) when μ is absolutely continuous with respect to Lebesgue measure.

Our definition of measure-operator convolution is based on the framework developed by Feichtinger in [72, 73], which allows us to avoid the use of Bochner integrals mentioned in [30] and thus several theorems from quantum harmonic analysis follow immediately from the construction, notably a version of Young's theorem and the Fourier multiplication theorem.

With the relevant definitions and basic properties in place, we explore applications of measure-operator convolutions to the setting of lattices in Section B.4 where we are able to recover classical results as well as new results on the eigenvalues of mixed-state Gabor multipliers and a version of the Berezin-Lieb inequality. Here we find that the correct setting for doing quantum harmonic analysis is requiring that the window operator S is compatible with the lattice Λ in the sense that

$$A\|\psi\|^2 \leq \sum_{\lambda\in\Lambda} Q_S(\psi)(\lambda) \leq B\|\psi\|^2 \qquad \text{for all } \psi \in L^2(\mathbb{R}^d),$$

where Q_S is a Cohen's class distribution. This setting was originally explored by Skrettingland in [193] but the above-mentioned results are novel as is the approach to interactions with measure-operator convolutions.

Finally, we treat the case of varying the lattice Λ in Section B.4.5. In this setting, we are able to establish two notable results. First, we show that for a lattice $\alpha\mathbb{Z}^d \times \beta\mathbb{Z}^d$, the associated (suitably normalized) Gabor multipliers approximate localization operators in the trace-class norm under some conditions on the mask. Moreover, we are also able to establish a result on how Gabor multipliers continuously depend on the lattice parameters, the mask and the windows provided the mask convergence is in the Wiener amalgam space $W(C_0,\ell^1)(\mathbb{R}^{2d})$.

Notation:

For $1 \leq p \leq \infty$, $\mathcal{S}^p$ denotes the Schatten p-class of operators on $L^2(\mathbb{R}^d)$ with singular values in ℓ^p where we use the convention that $\mathcal{S}^\infty = \mathcal{L}(L^2)$, the space of bounded linear operators on $L^2(\mathbb{R}^d)$. A check on an operator will denote conjugation by the parity operator $P(\psi)(t) = \psi(-t)$, i.e., $\check{T} = PTP$, while for measures, $\check{\mu}(E) = \mu(-E)$ where $-E = \{-x : x \in E\}$. For a discrete set E, we will write $|E|$ for its cardinality and for a non-discrete set we will use the same notation $|E|$ for the Lebesgue measure. Given a set $\Omega \subset \mathbb{R}^{2d}$ and $R \in \mathbb{R}$, we will write $R\Omega$ for the dilated set $\{R\omega : \omega \in \Omega\}$, Ω^c for the complement $\mathbb{R}^{2d} \setminus \Omega$ and χ_Ω for the indicator function of the set Ω. Norms $\|\cdot\|$ and inner products $\langle \cdot, \cdot \rangle$ without subscripts will always be understood to be the ones for $L^2(\mathbb{R}^d)$. The rank-one operator $\psi \otimes \phi$ is defined as $(\psi \otimes \phi)(f) = \langle f, \phi \rangle \psi$, for $\phi, \psi, f \in L^2(\mathbb{R}^d)$.

B.2 Preliminaries

In this section we go over the relevant preliminaries on time-frequency analysis and quantum harmonic analysis which should be well-known to readers familiar with other works on quantum harmonic analysis such as [153, 155, 156, 192]. The last subsection, Section B.2.4 is solely based on the two papers [72, 73] by Feichtinger and the results are later used extensively in defining measure-operator convolutions in Section B.3.

B.2.1 Time-frequency analysis

We briefly introduce some of the main concepts of time-frequency analysis which will be used in Section B.4. For a more thorough introduction, the reader is referred to [107].

The short-time Fourier transform

Given a *signal* $\psi \in L^2(\mathbb{R}^d)$ and a *window* $\varphi \in L^2(\mathbb{R}^d)$, the *short-time Fourier transform* (STFT) of ψ with respect to φ is defined on $\mathbb{R}^{2d}$

$$V_\varphi \psi(z) = \langle \psi, \pi(z)\varphi \rangle = \int_{\mathbb{R}^d} \psi(t)\overline{\varphi(t-x)} e^{-2\pi i \omega \cdot t}\, dt, \quad z = (x, \omega).$$

The interpretation is that $V_\varphi \psi(z)$ measures the time-frequency content of ψ at $z = (x, \omega)$ where x is the time and ω is the frequency.

The classical *Moyal's identity* shows how the STFT respects inner products and the L^2 energy of its constituents as

$$\left\langle V_{\varphi_1}\psi_1, V_{\varphi_2}\psi_2 \right\rangle_{L^2(\mathbb{R}^{2d})} = \langle \psi_1, \psi_2 \rangle \overline{\langle \varphi_1, \varphi_2 \rangle}. \tag{B.2.1}$$

Lastly we mention that the squared modulus of the STFT, $|V_\varphi\psi|^2$, referred to as the *spectrogram*, is a real-valued byproduct of the STFT which is often used in applications and is an example of a *quadratic* time-frequency distribution.

Cohen's class of quadratic time-frequency distributions

There is a large collection of different quadratic time-frequency distributions and those which satisfy some basic desirable properties can be characterized as being of the form

$$Q_\Phi(\psi,\phi) = W(\psi,\phi) * \Phi, \quad W(\psi,\phi)(z) = \int_{\mathbb{R}^d} \psi(t+x/2)\overline{\phi(t-x/2)}e^{-2\pi i\omega\cdot t}\,dt,$$

where W is the (cross) Wigner distribution and Φ is a tempered distribution. Such bilinear distributions are said to belong to *Cohen's class of quadratic time-frequency distributions*.

Localization operators

A non-trivial consequence of Moyal's identity (B.2.1) is that any function $\psi \in L^2(\mathbb{R}^d)$ can be reconstructed from its STFT $V_\varphi\psi$ as a weak integral:

$$\psi = \int_{\mathbb{R}^{2d}} V_\varphi\psi(z)\pi(z)\varphi\,dz. \tag{B.2.2}$$

Localization operators weight this reconstruction by some function $m : \mathbb{R}^{2d} \to \mathbb{C}$, sometimes taken to be an indicator function, as

$$A_m^\varphi\psi = \int_{\mathbb{R}^{2d}} m(z)V_\varphi\psi(z)\pi(z)\varphi\,dz \tag{B.2.3}$$

and we write $A_{\chi_\Omega}^\varphi = A_\Omega^\varphi$ when $m = \chi_\Omega$. The interpretation is that $A_m^\varphi\psi$ should be approximately concentrated in $\mathrm{supp}(m)$ in phase space. These operators have been extensively studied due to applications in physics, signal processing and the theory of pseudodifferential operators [104, 109, 199] and were first introduced in this form by I. Daubechies in [51].

Gabor frames

In applications, instead of considering the continuous function $V_\varphi\psi$ on $\mathbb{R}^{2d}$ we often use the discrete analogue $(V_\varphi\psi(\lambda))_{\lambda\in\Lambda}$ where Λ is a lattice on $\mathbb{R}^{2d}$, i.e., $\Lambda = A\mathbb{Z}^{2d}, A \in GL(2d,\mathbb{R})$. This makes computations feasible but we lose a few of the nice results we have established in the continuous setting. Of course, we want as many properties as possible to carry over from the continuous case to this setting. It turns out that the correct requirement for this is the *frame condition*:

$$A\|\psi\|^2 \leq \sum_{\lambda\in\Lambda} |\langle\psi, \pi(\lambda)\varphi\rangle|^2 \leq B\|\psi\|^2 \qquad \text{for all } \psi \in L^2(\mathbb{R}^d), \tag{B.2.4}$$

where A and B are finite real constants. If (B.2.4) holds, the collection $\{\pi(\lambda)\varphi\}_{\lambda\in\Lambda}$ is said to be a *Gabor frame with frame bounds* A, B. Important special cases are tight frames, meaning that $A = B$, or equivalently that the frame operator, i.e. the mapping

$$\psi \mapsto \sum_{\lambda\in\Lambda} V_\varphi\psi(\lambda)\pi(\lambda)\varphi \tag{B.2.5}$$

is a scalar multiple of the identity, analogous to the reconstruction formula (B.2.2) in the continuous case.

Gabor multipliers

As for localization operators, (B.2.5) suggests inserting weights before reconstruction in the tight case. More generally, if the frame condition (B.2.4) is satisfied, there exists a *dual window* γ such that we have a reconstruction of ψ using an unconditionally convergent series expansion in $L^2(\mathbb{R}^d)$:

$$\psi = \sum_{\lambda\in\Lambda} V_\varphi\psi(\lambda)\pi(\lambda)\gamma \qquad \text{for all } \psi \in L^2(\mathbb{R}^d).$$

Gabor multipliers can then be defined with the help of a *mask* or *(upper) symbol*, i.e., a weight function $m : \Lambda \to \mathbb{C}$, as follows

$$G^{\varphi,\gamma}_{m,\Lambda}\psi = \sum_{\lambda\in\Lambda} m(\lambda)V_\varphi\psi(\lambda)\pi(\lambda)\gamma.$$

Wiener amalgam spaces and bounded uniform partitions of unity

In order to describe the global behavior of a local property of a function *Wiener amalgam spaces* [67, 69, 123] are a natural choice. For $1 \leq p, q < \infty$ the original construction on $\mathbb{R}$ selects measurable functions such that the following norm is finite:

$$\|F\|_{W(L^p,\ell^q)} = \left(\sum_{n\in\mathbb{Z}} \left(\int_n^{n+1} |F(t)|^p \, dt\right)^{q/p}\right)^{1/q}. \tag{B.2.6}$$

Here p controls the local behavior of the function while q controls global summability. There are many generalizations of Wiener amalgam spaces. We will make use of the characterization of these space (on $\mathbb{R}^{2d}$) using *bounded uniform partitions of unity* (BUPU's). Given an index set I, a BUPU is a collection of (continuous) functions $\Psi = \{\psi_i\}_{i\in I}$ such that

(i) $\sum_{i\in I} \psi_i(x) = 1, \quad x \in \mathbb{R}^{2d}$,

(ii) $\sup_{i\in I} \|\psi_i\|_{L^\infty} < \infty$,

(iii) There exists a compact set $U \subset \mathbb{R}^{2d}$ with nonempty interior and points z_i such that $\mathrm{supp}(\psi_i) \subset U + z_i$ for all $i \in I$,

(iv) For each compact $K \subset \mathbb{R}^{2d}$,

$$\sup_{i\in I} \left|\left\{ j \in I : K + z_i \cap K + z_j \neq \emptyset \right\}\right| < \infty,$$

this is sometimes called the *bounded overlap property*.

A trivial construction of a BUPU is the collection of indicator functions of translated fundamental domains of a lattice $\Lambda = \alpha\mathbb{Z}^d \times \beta\mathbb{Z}^d$, $\psi_{n,k} = \chi_{[\alpha n,\alpha(n+1))\times[\beta k,\beta(k+1))}$ or better more intricate constructions with continuous or smooth functions, e.g. B-splines.

For our generalization of the Wiener amalgam spaces considered above, we will consider functions on $\mathbb{R}^{2d}$ which are locally in some Banach space $(B, \|.\|_B)$. The norm can then be written as

$$\|F\|_{W(B,\ell^q)} = \left(\sum_{i\in I} \|F\psi_i\|_B^q \right)^{1/q}.$$

A central property of these norms to be used later is the fact that different BUPU's Ψ give equivalent norms, [123]. For $p = \infty$ obvious modifications to (B.2.6) apply.

B.2.2 Operator theory

The main tool from operator theory will be the *singular value decomposition* of a compact operator A, which can be written as

$$A = \sum_n s_n(A)(\psi_n \otimes \phi_n),$$

where $\{\psi_n\}_n$ and $\{\phi_n\}_n$ are orthonormal sequences in $L^2(\mathbb{R}^d)$ and $(s_n(A))_n$ is a sequence of non-negative, decreasing numbers, the *singular values*, converging to zero. When the operator A is self-adjoint, we can take $\phi_n = \psi_n$ and the singular values are the eigenvalues of A (up to a sign). Otherwise they are just the eigenvalues of the positive operator $|A| = \sqrt{A^*A}$.

The Schatten p-class of operators for $p < \infty$ is the Banach space of compact operators with singular values in ℓ^p. In many ways these spaces behave as the L^p spaces with the *trace*, defined as $\mathrm{tr}(A) = \sum_n \langle Ae_n, e_n\rangle$ for some orthonormal basis $\{e_n\}_n$ of $L^2(\mathbb{R}^d)$, replacing the integral so that $\|A\|_{\mathcal{S}^p} = \mathrm{tr}(|A|^p)^{1/p}$. These norms are monotonic in p in the sense that if $1 \leq p \leq p' \leq \infty$,

$$\|S\|_{\mathcal{S}^\infty} \leq \|S\|_{\mathcal{S}^{p'}} \leq \|S\|_{\mathcal{S}^p} \leq \|S\|_{\mathcal{S}^1}.$$

The cases $p = 1, 2$ are of special relevance with the $\mathcal{S}^1$ operators being referred to as *trace-class* and $\mathcal{S}^2$ being the *Hilbert-Schmidt* operators. The $\mathcal{S}^p$ - $\mathcal{S}^q$ (sesquilinear) duality brackets for the Schatten classes may be given by

$$\langle A, B\rangle_{\mathcal{S}^p,\mathcal{S}^q} = \mathrm{tr}(AB^*). \tag{B.2.7}$$

For more on this class of operators, see [190]. By this duality the bounded linear operators form the dual space to $\mathcal{S}^1$.

B.2.3 Quantum harmonic analysis

In this section we recall some classical properties of quantum harmonic analysis without proofs so that they can be compared to the corresponding results for measure-operator convolutions. The reader familiar with quantum harmonic analysis can safely skip over this section. For proofs, see [153, 211].

Operator convolutions

A central idea in quantum harmonic analysis is that just as functions can be moved around using translations T_x, so can operators in phase space using a representation π, via the operator translation:

$$\alpha_z(S) = \pi(z)S\pi(z)^*. \tag{B.2.8}$$

Together with the idea that traces are the proper substitute for integrals in $\mathcal{S}^p$, this leads to the following definition of function-operator and operator-operator convolutions which form the key ingredients of QHA as introduced by R. Werner:

$$f \star S = \int_{\mathbb{R}^{2d}} f(z)\alpha_z(S)\,dz, \qquad T \star S(z) = \mathrm{tr}\left(T\alpha_z(\check{S})\right). \tag{B.2.9}$$

Note in particular that $f \star S$ is an operator on $L^2(\mathbb{R}^d)$ while $T \star S$ is a function on $\mathbb{R}^{2d}$. We also define convolutions between an operator and a function as $S \star f := f \star S$. Below we collect some properties of these convolutions and later show how they relate to time-frequency analysis in Sections B.2.3 and B.2.3.

Proposition B.2.1. Let $f, g \in L^1(\mathbb{R}^{2d})$, $S \in \mathcal{S}^p$, $T \in \mathcal{S}^q$ for $1 \leq p, q, r \leq \infty$ with $\frac{1}{p} + \frac{1}{q} = 1 + \frac{1}{r}$ and R be a compact operator, then

(i) $f \star S$ is positive if f is non-negative and S is positive,

(ii) $T \star S$ is non-negative if T and S are positive,

(iii) $\operatorname{tr}(f \star S) = \operatorname{tr}(S) \int_{\mathbb{R}^{2d}} f(z)\, dz$ for $p = 1$,

(iv) $(f \star S)^* = \bar{f} \star S^*$,

(v) $(f \star S) \star T = f * (S \star T)$,

(vi) $(f * g) \star S = f \star (g \star S)$,

(vii) $f \star R$ is compact,

(viii) $\|f \star S\|_{\mathcal{S}^p} \leq \|f\|_{L^1} \|S\|_{\mathcal{S}^p}$,

(ix) $\|T \star S\|_{L^r} \leq \|S\|_{\mathcal{S}^p} \|T\|_{\mathcal{S}^q}$,

(x) $\int_{\mathbb{R}^{2d}} T \star S(z)\, dz = \operatorname{tr}(T) \operatorname{tr}(S)$ if $p = q = 1$.

We also mention that both convolutions are commutative, function-operator convolutions by definition while for operator-operator convolutions this can be manually verified.

Fourier transforms

There is also an associated Fourier transform for operators, the *Fourier-Wigner transform*, which generalizes the classical convolutions theorem in two different ways. It is defined for $S \in \mathcal{S}^1$ as a function on $\mathbb{R}^{2d}$ given by

$$\mathcal{F}_W(S)(z) = e^{-\pi i x \omega} \operatorname{tr}(\pi(-z)S), \tag{B.2.10}$$

where $z = (x, \omega)$. The terminology comes from Folland [84] but was first used for operators by Werner in [211] and has since been used extensively in quantum harmonic analysis [153, 156].

Before stating the convolution theorems, we also need to introduce the symplectic Fourier transform which is essentially a rotated Fourier transform on $\mathbb{R}^{2d}$, defined as

$$\mathcal{F}_\sigma(f)(z) = \int_{\mathbb{R}^{2d}} f(z') e^{-2\pi i \sigma(z,z')}\, dz',$$

where $z = (x, \omega)$, $z' = (x', \omega')$ and $\sigma(z, z') = \omega x' - \omega' x$ is the standard symplectic form. With this, we have all the notation we need for the convolution theorem in place.

Proposition B.2.2. Let $f \in L^1(\mathbb{R}^{2d})$ and $S, T \in \mathcal{S}^1$, then

$$\mathcal{F}_W(f \star S) = \mathcal{F}_\sigma(f)\mathcal{F}_W(S),$$
$$\mathcal{F}_\sigma(T \star S) = \mathcal{F}_W(T)\mathcal{F}_W(S).$$

Cohen's class distributions

In [154], the Cohen's class distributions discussed in Section B.2.1 above were characterized as the bilinear maps $Q : L^2(\mathbb{R}^d) \times L^2(\mathbb{R}^d) \to L^\infty(\mathbb{R}^{2d})$ such that $Q = Q_S$ for some $S \in \mathcal{L}(L^2(\mathbb{R}^d))$ where

$$Q_S(\psi, \phi)(z) = (\psi \otimes \phi) \star \check{S}(z) = \left\langle \check{S}\pi(z)^*\psi, \pi(z)^*\phi \right\rangle.$$

Moreover, when S is positive and trace-class, we can expand S in its singular value decomposition as $S = \sum_n \lambda_n(\varphi_n \otimes \varphi_n)$ and write

$$Q_S(\psi)(z) = \sum_n \lambda_n |\langle \psi, \pi(z)\varphi_n \rangle|^2.$$

by [154, Theorem 7.6].

We also mention that there is a reconstruction of identity related to Cohen's class distributions which generalizes (B.2.2), first shown in [211, Proposition 3.2 (3)]. It states that if S is trace-class with $\operatorname{tr}(S) = 1$,

$$\psi = \int_{\mathbb{R}^{2d}} \pi(z)S\pi(z)^*\psi \, dz, \tag{B.2.11}$$

which in particular implies that $1 \star S = \operatorname{tr}(S)I_{L^2}$.

Mixed-state localization operators

Inspired by the identity (B.2.11) and the fact that the case where S is a rank-one operator precisely corresponds to classical localization operators $A_m^\varphi = m \star (\varphi \otimes \varphi)$, *mixed-state* localization operators were introduced in [154]. These correspond to letting S be an arbitrary trace-class operator, i.e., an operator of the form $f \star S$. The terminology is borrowed from physics where each rank-one term in the singular value decomposition of a trace-class operators can be seen as a *pure state* while a general trace-class operator corresponds to a *mixing* of these states.

B.2.4 A functional-analytic approach to convolutions

In the interest of completeness, we recall the key definitions and theorems from [72, 73] which we later use to set up measure-operator convolutions and deduce their elementary properties. For the necessary background on group representations, we refer to the book by Folland [85].

We first go through the results in [73], starting with some relevant definitions.

Definition B.2.3. A mapping $\rho : G \to \mathcal{L}(B)$ is called an *isometric representation* of a group G on a Banach space B if the mapping ρ is linear, preserves identities, is a group homomorphism, i.e., satisfies

$$\rho(xy) = \rho(x) \circ \rho(y) \qquad \text{for all } x, y \in G,$$

and if each of the operators are isometric on B meaning that

$$\|\rho(x)S\|_B = \|S\|_B \qquad \text{for all } x \in G \text{ and } S \in B.$$

Moreover, if the mapping $x \mapsto \rho(x)S$ is continuous from G to B in the sense that

$$\lim_{x \to 0} \|\rho(x)S - S\|_B = 0,$$

we say that the representation ρ is *strongly continuous*.

In [188], pairings (B, ρ) are called *abstract homogeneous Banach spaces*. The usual *homogeneous Banach spaces* are Banach spaces of locally integrable functions endowed with the regular representation (by translations). For us, B will be the space of trace-class operatos $\mathcal{S}^1$ and $\rho(z) = \alpha_z$ will be the operator translations (B.2.8).

We will use the notation $\bullet_\rho : G \times B \to B$ for the action of the representation and also define the representation for point measures as

$$\delta_x \bullet_\rho S := x \bullet_\rho S = \rho(x)S. \tag{B.2.12}$$

Recall that a *Banach algebra* is a Banach space $(A, \|\cdot\|_A)$ together with a continuous operation $* : A \times A \to A$. We can endow a Banach algebra with the following additional structure.

Definition B.2.4. A Banach space $(B, \|\cdot\|_B)$ is a *Banach module* over a Banach algebra $(A, *, \|\cdot\|_A)$ if one has a mapping $(a, b) \mapsto a \bullet b$ from $A \times B$ to B that is bilinear and associative, i.e.,

(i) $\|a \bullet b\|_B \leq \|a\|_A \|b\|_B$,

(ii) $a_1 \bullet (a_2 \bullet b) = (a_1 * a_2) \bullet b$.

If moreover

$$\overline{\operatorname{span}(A \bullet B)} = B,$$

the Banach module is said to be *essential*.

For us, the Banach algebra A will be the convolutions algebra of bounded measures later on.

We are now ready to state the main theorem we will use which is a reformulation of [73, Theorem 2]. In the sequel, $M(G)$ denotes the space of bounded complex signed measures on the locally compact abelian group G with $\|\mu\|_M = |\mu|(G)$, viewed as a Banach algebra with respect to convolution, see [72].

Theorem B.2.5. *Let $(B, \|\cdot\|_B)$ be a Banach space and ρ a strongly continuous and isometric representation of a locally compact abelian group G. Then B is a Banach module over the Banach algebra $(M(G), \|\cdot\|_M)$ with respect to a mapping $\bullet_\rho : (\mu, S) \mapsto \mu \bullet_\rho S$ that extends the action of the discrete measure $\delta_x \bullet_\rho S = \rho(x)S$ and satisfies the estimate*

$$\|\mu \bullet_\rho S\|_B \le \|\mu\|_{M(G)} \|S\|_B, \qquad \mu \in M(G),\ S \in B.$$

We interpret such a mapping $\bullet_\rho$ as a (generalized) convolution between the measure μ and the operator S.

Remark B.2.6. By identifying $L^1(G)$ with a closed subspace (ideal) of $M(G)$, this also defines a convolution between functions and elements of B.

We moreover have a uniqueness theorem for this extension which is [73, Theorem 6].

Theorem B.2.7. *Let $(B, \|\cdot\|_B)$ be a Banach space and ρ a strongly continuous and isometric representation of a locally compact abelian group G. Then there is a unique w^*-continuous (for bounded and tight families) extension of $\bullet_\rho$ to $(M(G), \|\cdot\|_M)$, i.e., a bounded, bilinear mapping*

$$(\mu, S) \mapsto \mu \bullet_\rho S, \qquad \mu \in M(G),\ S \in B,$$

which turns $(B, \|\cdot\|_B)$ into an essential $L^1(G)$-module.

The module action of $L^1(G)$ is obtained by restriction of the action of $M(G)$. Since $L^1(G)$ does not contain a unit element in the non-discrete case it is interesting to observe that the module B is essential by the following corollary.

Corollary B.2.8 ([73, Corollary 1])**.** Given the situation in Theorem B.2.5 and a bounded approximate unit $(g_\alpha)_{\alpha \in I}$ in $(L^1(G), \|\cdot\|_{L^1})$, we have that

$$\lim_{\alpha \to \infty} \|g_\alpha \bullet_\rho S - S\|_B = 0 \qquad \text{for all } S \in B.$$

For applications, we often want an explicit form of this convolution, which is going to be given in a weak formulation. In the original proof of Theorem B.2.5, the convolution is defined as the limit

$$\mu \star S = \lim_{|\Psi| \to 0} D_\Psi \mu \bullet_\rho S, \tag{B.2.13}$$

where Ψ is a BUPU of G, discussed in Section B.2.1, and $D_\Psi : \mu \mapsto \sum_{i\in I} \mu(\psi_i)\delta_{z_i}$ is a *discretization operator*. To compute it, we can use the following lemma which is specific to the Euclidean case.

Lemma B.2.9 ([72, Lemma 7]). *Let $f \in C_0(\mathbb{R}^{2d})$, then*

$$\lim_{|\Psi|\to 0} D_\Psi \mu(f) = \mu(f) = \int_{\mathbb{R}^{2d}} f(z)\, d\mu(z).$$

Lastly we cover a theorem which we will use in Section B.4.5, concerning the continuity of the action $\bullet_\rho$. Its formulation is based on tight and bounded nets of measures $(\mu_\alpha)_{\alpha\in I} \subset M(G)$ indexed by some set I. We remind the reader that nets are a generalization of sequences and all relevant facts can be found in [73, Section 7]. When we say that a net is bounded we mean that it is bounded uniformly in the measure norm and tightness means that for any $\varepsilon > 0$, we can find a function $k \in C_c(G)$ such that $\|\mu_\alpha - k \cdot \mu_\alpha\|_M < \varepsilon$ for all $\alpha \in I$.

Theorem B.2.10 ([73, Theorem 5]). *Let ρ be a strongly continuous, isometric representation of the locally compact group G on the Banach space $(B, \|\cdot\|_B)$ and $(\mu_\alpha)_{\alpha\in I}$ a bounded and tight net in $(M(G), \|\cdot\|_M)$ with $\mu_0 = w^* - \lim_{\alpha\to\infty} \mu_\alpha$. Then one has*

$$\lim_{\alpha\to\infty} \|\mu_\alpha \bullet_\rho S - \mu_0 \bullet_\rho S\|_B = 0 \qquad \text{for all } S \in B.$$

B.3 Definition and properties of measure-operator convolutions

In this section we first show that the results recapitulated in Section B.2.4 apply to measures on $\mathbb{R}^{2d}$ and the Schatten p-class of operators (for $p < \infty$) on $L^2(\mathbb{R}^d)$ and then deduce elementary properties of the resulting operators, mostly using Theorem B.3.3 below giving the weak action of measure-operator convolutions.

B.3.1 Defining measure-operator convolutions for trace-class operators

Our main task in this section is to show that the assumptions of Theorem B.2.5 are satisfied for the representation

$$z \mapsto \rho(z), \qquad \rho(z)S = \pi(z)S\pi(z)^* \tag{B.3.1}$$

of the time-frequency plane $\mathbb{R}^{2d}$ on the space of trace-class operators $\mathcal{S}^1$. In fact, we have

Lemma B.3.1. *The mapping ρ defined by* (B.3.1) *is a strongly continuous and isometric representation of* $\mathbb{R}^{2d}$.

Proof. We first verify linearity, identity preservation and associativity of ρ as

$$\begin{aligned}\rho(z)(\alpha S_1 + \beta S_2) &= \pi(z)(\alpha S_1 + \beta S_2)\pi(z)^* = \alpha\rho(z)S_1 + \beta\rho(z)S_2,\\ \rho(0)S &= \pi(0)S\pi(0)^* = S,\\ \rho(z_1)\big(\rho(z_2)S\big) &= \pi(z_1)\big(\pi(z_2)S\pi(z_2)^*\big)\pi(z_1)^* = \pi(z_1z_2)S\pi(z_1z_2)^* = \rho(z_1z_2)S.\end{aligned}$$

The remaining continuity property follows from [25, Lemma 2.1]. □

We now obtain the following corollary by applying Theorem B.2.5 and Theorem B.2.7.

Corollary B.3.2. If we identify the map $\bullet_\rho : \mathbb{R}^{2d}\times\mathcal{S}^1 \to \mathcal{S}^1,\ (z, S) \mapsto \pi(z)S\pi(z)^*$ with $(\delta_z, S) \mapsto \pi(z)S\pi(z)^*$, there exists an extension to $M(\mathbb{R}^{2d})\times\mathcal{S}^1$ which satisfies

$$\|\mu \bullet_\rho S\|_{\mathcal{S}^1} \le \|\mu\|_M\|S\|_{\mathcal{S}^1} \qquad \text{for all } \mu \in M(\mathbb{R}^{2d}) \text{ and } S \in \mathcal{S}^1,$$

and makes $\mathcal{S}^1$ a Banach module over $M(\mathbb{R}^{2d})$. Moreover, this extension is unique among extensions which turn $\mathcal{S}^1$ into an essential $L^1(\mathbb{R}^{2d})$-module upon embedding $L^1(\mathbb{R}^{2d})$ in $M(\mathbb{R}^{2d})$.

From here on we write $\star$ for $\bullet_\rho$ and call it the *measure-operator convolution* or *function-operator convolution* if the measure is absolutely continuous with respect to Lebesgue measure. We also make the definition $S \star \mu := \mu \star S$ so that we can convolve with measures from both sides.

Next we show how we can get an explicit formula for the weak action of a measure-operator convolution. This is useful for computations and shows that our approach appropriately extends the function-operator convolutions of quantum harmonic analysis defined with Bochner integrals which can be found in [30, 153, 211].

Theorem B.3.3. *Let $S \in \mathcal{S}^1$ and $\mu \in M(\mathbb{R}^{2d})$. Then*

$$\big\langle(\mu \star S)\psi, \phi\big\rangle = \int_{\mathbb{R}^{2d}} \big\langle\pi(z)S\pi(z)^*\psi, \phi\big\rangle\, d\mu(z) \qquad \textit{for all } \psi, \phi \in L^2(\mathbb{R}^d).$$

Proof. Firstly, $\mu \star S$ is defined as the limit of a sequence of $\mathcal{S}^1$ operators which converge in the $\mathcal{S}^1$ norm by Theorem B.2.5. In particular, they converge in the operator norm. Hence, using (B.2.13), we have

$$\begin{aligned}
\big\langle (\mu \star S)\psi, \phi \big\rangle &= \left\langle \left(\lim_{|\Psi|\to 0} D_\Psi \mu \star S \right) \psi, \phi \right\rangle \\
&= \lim_{|\Psi|\to 0} \big\langle (D_\Psi \mu \star S)\psi, \phi \big\rangle \\
&= \lim_{|\Psi|\to 0} \left\langle \left(\sum_{i\in I} \mu(\psi_i)\delta_{z_i} \star S \right) \psi, \phi \right\rangle \\
&= \lim_{|\Psi|\to 0} \sum_{i\in I} \mu(\psi_i) \big\langle \pi(z_i) S \pi(z_i)^* \psi, \phi \big\rangle \\
&= \lim_{|\Psi|\to 0} D_\Psi \mu \left(\big\langle \pi(\cdot) S \pi(\cdot)^* \psi, \phi \big\rangle \right) \\
&= \int_{\mathbb{R}^{2d}} \big\langle \pi(z) S \pi(z)^* \psi, \phi \big\rangle \, d\mu(z),
\end{aligned}$$

where we used Lemma B.2.9 in the last step. To justify this we still need to show that $\langle \pi(\cdot) S \pi(\cdot)^* \psi, \phi \rangle \in C_0(\mathbb{R}^{2d})$. The continuity follows from π being a representation and S being a bounded operator. For vanishing at infinity, by an application of Cauchy-Schwarz we deduce that it suffices to show that

$$\|\pi(z) S \pi(z)^* \psi\| = \|S \pi(z)^* \psi\| \to 0 \quad \text{as } z \to \infty.$$

Since S is a compact operator, this follows if we can show that $\pi(z)^*\psi$ goes to zero weakly as $z \to \infty$. This is equivalent to the STFT vanishing at infinity by

$$\langle \pi(z)^* \psi, \phi \rangle = \langle \psi, \pi(z)\phi \rangle = V_\phi \psi(z),$$

which is a well-known fact [107]. □

B.3.2 Extending to the Schatten classes

Having defined measure-operator convolutions for trace-class operators, we wish to extend the definition to $\mathcal{S}^p$ for $1 < p \leq \infty$. The first step in this direction is defining $\mu \star T$ for $T \in \mathcal{S}^\infty$ using Schatten duality in a way that generalizes the function-operator case (see [211] and [153, Proposition 4.2]) as

$$\big\langle \mu \star T, S \big\rangle_{\mathcal{S}^\infty, \mathcal{S}^1} = \big\langle \mu, S \star \check{T}^* \big\rangle_{M, C_b}.$$

Here S is taken to be in $\mathcal{S}^1$ and $C_b(\mathbb{R}^{2d})$ is the space of continuous bounded functions. The fact that the quantity in the right hand side is finite follows from that operator-operator convolutions map into $C_b(\mathbb{R}^{2d})$ [211]. To show that $\mu \star T \in \mathcal{S}^\infty$, we estimate

$$\begin{aligned}\|\mu \star T\|_{\mathcal{S}^\infty} &= \sup_{\|S\|_{\mathcal{S}^1}=1} \left|\left\langle \mu \star T, S\right\rangle_{\mathcal{S}^\infty, \mathcal{S}^1}\right| \\ &= \sup_{\|S\|_{\mathcal{S}^1}=1} \left|\left\langle \mu, S \star \check{T}^*\right\rangle_{M, C_b}\right| \\ &\leq \sup_{\|S\|_{\mathcal{S}^1}=1} \|\mu\|_M \|S \star \check{T}^*\|_{L^\infty} \leq \|\mu\|_M \|T\|_{\mathcal{S}^\infty},\end{aligned}$$

where we used Proposition B.2.1 (ix) for this last step. By a version of Riesz-Thorin interpolation for Schatten classes [190, Theorem 2.10], we immediately get the following result on the Schatten boundedness.

Proposition B.3.4. Let $\mu \in M(\mathbb{R}^{2d})$ and $S \in \mathcal{S}^p$ for $1 \leq p \leq \infty$, then

$$\|\mu \star S\|_{\mathcal{S}^p} \leq \|\mu\|_M \|S\|_{\mathcal{S}^p}.$$

We can also go further than this and extend the weak formulation from Theorem B.3.3 to $\mathcal{S}^p$ for $1 \leq p < \infty$. The proof is a standard approximation argument but we write it out for completeness.

Corollary B.3.5. Let $S \in \mathcal{S}^p$ for $1 \leq p < \infty$ and $\mu \in M(\mathbb{R}^{2d})$, then

$$\left\langle (\mu \star S)\psi, \phi \right\rangle = \int_{\mathbb{R}^{2d}} \left\langle \pi(z) S \pi(z)^* \psi, \phi \right\rangle d\mu(z) \qquad \text{for all } \psi, \phi \in L^2(\mathbb{R}^d).$$

Proof. We approximate $S \in \mathcal{S}^p$ by a sequence of finite rank operators S_N consisting of the first N terms of the spectral decomposition of S so that $S_N \in \mathcal{S}^1$ and $S_N \to S \in \mathcal{S}^p$ as $N \to \infty$ since $p < \infty$. In particular, $S_N \to S$ in $\mathcal{L}(L^2)$ also in this case.

By Cauchy-Schwarz, $\langle (\mu \star S)\psi, \phi \rangle$ exists and our goal is to show that it is equal to the integral in the formulation of the corollary. The first step in this direction is noting that $\langle (\mu \star S_N)\psi, \phi \rangle \to \langle (\mu \star S)\psi, \phi \rangle$ as $N \to \infty$ by Cauchy-Schwarz and the fact that $S_N \to S$ in $\mathcal{L}(L^2)$. Next we use this to write

$$\begin{aligned}&\left| \left\langle (\mu \star S)\psi, \phi \right\rangle - \int_{\mathbb{R}^{2d}} \left\langle \pi(z) S \pi(z)^* \psi, \phi \right\rangle d\mu(z) \right| \\ &\quad = \left| \lim_{N\to\infty} \left\langle (\mu \star S_N)\psi, \phi \right\rangle - \int_{\mathbb{R}^{2d}} \left\langle \pi(z) S \pi(z)^* \psi, \phi \right\rangle d\mu(z) \right| \\ &\quad \leq \lim_{N\to\infty} \int_{\mathbb{R}^{2d}} \left| \left\langle [\pi(z) S_N \pi(z)^* - \pi(z) S \pi(z)^*]\psi, \phi \right\rangle \right| d|\mu|(z) \\ &\quad = \int_{\mathbb{R}^{2d}} \lim_{N\to\infty} \left| \left\langle [\pi(z) S_N \pi(z)^* - \pi(z) S \pi(z)^*]\psi, \phi \right\rangle \right| d|\mu|(z),\end{aligned}$$

where used Theorem B.3.3 for the inequality and moved the limit inside the integral using the dominated convergence theorem with dominating function $2\|S\|_{\mathcal{L}(L^2)}\|\psi\|_{L^2}\|\phi\|_{L^2}$ which is integrable by the boundedness of μ. From here our desired conclusion follows as the integrand can be seen to be zero upon employing Cauchy-Schwarz and using elementary properties of π as well as the fact that $S_N \to S$ in $\mathcal{L}(L^2)$ as $N \to \infty$. □

Lastly for $p = \infty$ we *define* the weak action $\langle(\mu \star S)\psi, \phi\rangle$ to agree with the integral formulations in Theorem B.3.3 and Corollary B.3.5 since we cannot make a denseness argument. This is precisely what is done for function-operator convolutions in [153].

B.3.3 Properties of measure-operator convolutions

From the weak formulation of measure-operator convolutions in Corollary B.3.5 and the extension to $p = \infty$, we can deduce generalizations of many of the properties from Section B.2. Specifically we are able to generalize relevant parts of Proposition B.2.1.

Proposition B.3.6. Let $\mu \in M(\mathbb{R}^{2d})$, $S \in \mathcal{S}^p$, $T \in \mathcal{S}^q$ for $1 \leq p \leq \infty$ where $\frac{1}{p} + \frac{1}{q} = 1$ and R be a compact operator. Then

(i) $\mu \star S$ is positive if μ and S are positive,

(ii) $(\mu \star S)^* = \bar{\mu} \star S^*$,

(iii) $(\mu \star S)\check{} = \check{\mu} \star \check{S}$,

(iv) $\operatorname{tr}(\mu \star S) = \mu(\mathbb{R}^{2d}) \operatorname{tr}(S)$ if $p = 1$,

(v) $\mu \star R$ is compact,

(vi) $(\mu \star S) \star T = \mu * (S \star T)$,

(vii) $(\mu * \nu) \star S = \mu \star (\nu \star S)$.

Proof. Item (i): Using the weak formulation we verify

$$\big\langle(\mu \star S)\psi, \psi\big\rangle = \int_{\mathbb{R}^{2d}} \big\langle S\pi(z)^*\psi, \pi(z)^*\psi\big\rangle \, d\mu(z) \geq 0,$$

since the integral of a non-negative function with respect to a positive measure is non-negative.

Item (ii): We verify the relation weakly as

$$\begin{aligned}
\left\langle (\mu \star S)^* \psi, \phi \right\rangle &= \left\langle \psi, (\mu \star S)\phi \right\rangle = \overline{\left\langle (\mu \star S)\phi, \psi \right\rangle} \\
&= \overline{\int_{\mathbb{R}^{2d}} \langle \pi(z) S \pi(z)^* \phi, \psi \rangle \, d\mu(z)} \\
&= \int_{\mathbb{R}^{2d}} \overline{\langle \pi(z) S \pi(z)^* \phi, \psi \rangle} \, d\bar{\mu}(z) \\
&= \int_{\mathbb{R}^{2d}} \langle \pi(z) S^* \pi(z)^* \psi, \phi \rangle \, d\bar{\mu}(z) \\
&= \left\langle (\bar{\mu} \star S^*)\psi, \phi \right\rangle.
\end{aligned}$$

Item (iii): Using the same method we find

$$\begin{aligned}
\langle P(\mu \star S) P \psi, \phi \rangle &= \langle (\mu \star S) P\psi, P\phi \rangle \\
&= \int_{\mathbb{R}^{2d}} \langle \pi(z) S \pi(z)^* P\psi, P\phi \rangle \, d\mu(z) \\
&= \int_{\mathbb{R}^{2d}} \langle \pi(-z) PSP \pi(-z)^* \psi, \phi \rangle \, d\mu(z) \\
&= \left\langle (\check{\mu} \star \check{S})\psi, \phi \right\rangle .
\end{aligned}$$

The calculation in the middle step is detailed in [191, Lemma 3.2 (5)].
Item (iv): We compute

$$\begin{aligned}
\operatorname{tr}(\mu \star S) &= \sum_n \int_{\mathbb{R}^{2d}} \langle S \pi(z)^* e_n, \pi(z)^* e_n \rangle \, d\mu(z) \\
&= \int_{\mathbb{R}^{2d}} \operatorname{tr}(S) \, d\mu(z) = \mu(\mathbb{R}^{2d}) \operatorname{tr}(S),
\end{aligned}$$

where the use of Fubini is justified by $\mu \in M(\mathbb{R}^{2d})$ and [40, Proposition 18.9].
Item (v): Since R is compact, it is the limit in operator norm of a sequence of finite rank operators $(R_n)_n \subset \mathcal{S}^1$. For each n, $\mu \star R_n$ is compact by Corollary B.3.2, hence

$$\|\mu \star R - \mu \star R_n\|_{\mathcal{L}(L^2)} \leq \|\mu\|_M \|R - R_n\|_{\mathcal{L}(L^2)},$$

where we used Proposition B.3.4 with $p = \infty$. We conclude that $\mu \star R$ is the limit in operator norm of a sequence of compact operators and hence it too is compact.

Item (vi): We verify

$$\begin{aligned}((\mu \star S) \star T)(z) &= \operatorname{tr}\left((\mu \star S)\pi(z)\check{T}\pi(z)^*\right) \\ &= \sum_n \langle(\mu \star S)\pi(z)\check{T}\pi(z)^* e_n, e_n\rangle \\ &= \sum_n \int_{\mathbb{R}^{2d}} \left\langle \pi(y)S\pi(y)^*\pi(z)\check{T}\pi(z)^* e_n, e_n \right\rangle d\mu(y) \\ &= \int_{\mathbb{R}^{2d}} (S \star T)(z-y)\, d\mu(y) \\ &= (\mu * (S \star T))(z),\end{aligned}$$

where the use of Fubini is justified by Hölder's inequality for the Schatten p-classes of operators.
Item (vii): For $p = 1$, this follows from the module structure of $\mathcal{S}^1$ over $M(\mathbb{R}^{2d})$ established in Theorem B.2.5 while for $1 < p < \infty$ it follows by approximating S by finite rank operators as in the proof of Corollary B.3.5. However for $p = \infty$ we need to do the full calculation of looking at the weak action and expanding both sides. A quick calculation shows that both of these quantities are equal to

$$\int_{\mathbb{R}^{2d}} \int_{\mathbb{R}^{2d}} \langle \pi(z)\pi(y)S\pi(y)^*\pi(z)^*\psi, \phi\rangle\, d\mu(z)\, d\nu(y),$$

which yields the desired conclusion. □

Next we show how the convolution property of the Fourier-Wigner transform from (B.2.10) carries over to the measure-operator convolution setting.

Proposition B.3.7. Let $\mu \in M(\mathbb{R}^{2d})$ and $S \in \mathcal{S}^1$, then

$$\mathcal{F}_W(\mu \star S) = \mathcal{F}_\sigma(\mu)\mathcal{F}_W(S).$$

Proof. We follow the same arguments as Werner's original proof in [211], assuming that $\mathcal{F}_W(S)$ is integrable, by first computing

$$\begin{aligned}\mathcal{F}_W(\mu \star S)(z) &= e^{-\pi i x\cdot\omega} \operatorname{tr}(\pi(-z)(\mu \star S)) \\ &= e^{-\pi i x\cdot\omega} \sum_n \langle \pi(-z)(\mu \star S)(e_n), e_n\rangle \\ &= e^{-\pi i x\cdot\omega} \sum_n \int_{\mathbb{R}^{2d}} \langle \pi(-z)\pi(z')S\pi(z')^* e_n, e_n\rangle\, d\mu(z') \\ &= e^{-\pi i x\cdot\omega} \int_{\mathbb{R}^{2d}} \operatorname{tr}\left(\pi(-z)\pi(z')S\pi(z')^*\right) d\mu(z'),\end{aligned}$$

where we used Theorem B.3.3 to move to an integral formulation and justified the change of order of summation and integration by Fubini using the integrability of $\mathcal{F}_W(S)$. From here we use the relation $\operatorname{tr}\left(\pi(-z)\pi(z')S\pi(z')^*\right) = e^{-2\pi i\sigma(z,z')}\operatorname{tr}(\pi(-z)S)$ which can be verified directly as is done in [153] to deduce

$$\begin{aligned}\mathcal{F}_W(\mu \star S)(z) &= e^{-\pi i x\cdot\omega}\int_{\mathbb{R}^{2d}} e^{-2\pi i\sigma(z,z')}\operatorname{tr}\left(\pi(-z)S\right)\,d\mu(z')\\ &= \mathcal{F}_\sigma(\mu)(z)\mathcal{F}_W(S)(z).\end{aligned}$$

The case where $\mathcal{F}_W(S)$ is not integrable can now be deduced by density. □

By specializing Theorem B.2.10 to measure-operator convolutions, we obtain the following corollary.

Corollary B.3.8. Let $(\mu_\alpha)_{\alpha\in I}$ be a bounded and tight net in $M(\mathbb{R}^{2d})$ with $\mu_0 = w^* - \lim_{\alpha\to\infty}\mu_\alpha$, then

$$\lim_{\alpha\to\infty}\|\mu_\alpha\star S - \mu_0\star S\|_{\mathcal{S}^1} = 0 \qquad \text{for all } S\in\mathcal{S}^1.$$

This is a non-trivial extension to the $M(\mathbb{R}^{2d})$ continuity of measure-operator convolutions of Corollary B.3.2 which we will need in Section B.4.5.

B.4 Applications to lattice convolutions

Much of applied harmonic analysis and time-frequency analysis is carried out on lattices. There is therefore inherent value in translating results to this setting. This was the aim of the recent paper [192] by Skrettingland, which introduced quantum harmonic analysis on lattices. Using our results on measure-operator convolutions, we can deduce several results from [192] immediately from Corollary B.3.2.

Before diving into the full technical aspects of measure-operator convolutions, we clarify the links to [192] and [60] which explicitly and implicitly use lattice convolutions. Later we discuss a natural generalization of Gabor frames which sets the stage for a mixed-state analogue of Gabor multipliers that mirror the relation between localization operators and mixed-state localization operators. From this generalized notion of Gabor frames we are also able to deduce a version of the Berezin-Lieb inequality on lattices. Lastly we recall a result from [73] which will allow us to prove the promised result on approximating localization operators by Gabor multipliers and the continuity of Gabor multipliers in Section B.4.5.

B.4.1 Generalities

In [192], *lattice convolutions* are defined between sequences $c \in \ell^1(\Lambda)$, where Λ is a lattice, and operators $S \in \mathcal{S}^p$ as

$$c \star_\Lambda S = \sum_{\lambda \in \Lambda} c(\lambda)\pi(\lambda)S\pi(\lambda)^* = \underbrace{\left(\sum_{\lambda \in \Lambda} c(\lambda)\delta_\lambda\right)}_{=:\mu_c} \star S. \tag{B.4.1}$$

Following standard terminology for the case of ordinary convolution, one could also call this a *semi-discrete twisted convolution*, or interpret it as the action of the discrete (here bounded) measure on the operators S via the (extended) version of the convolution. As sequences in $\ell^1(\Lambda)$ can be isometrically embedded in $M(\mathbb{R}^{2d})$ as discrete measures, we immediately get the bound

$$\|c \star_\Lambda S\|_{\mathcal{S}^p} \leq \|c\|_{\ell^1(\Lambda)} \|S\|_{\mathcal{S}^p} \tag{B.4.2}$$

from Proposition B.3.4. Other properties, such as the commutativity of measure-operator convolutions, allow us to recover results from [192] without much work.

Gabor multipliers and more generally *mixed-state* Gabor multipliers (following the terminology of [154]) are defined in [192] as $\mu_c \star S$ and they appropriately generalize the classical notion of Gabor multipliers on lattices discussed in Section B.2.1 in the case where S is rank-one. See also the early work [70] on this subject in the context of LCA groups.

In [60], quantum harmonic analysis is used to develop tools for investigating systems of functions. A central tool is the *data operator* $S_\mathcal{D}$ associated to the (indexed) *data set* $\mathcal{D} = \{f_n\}_n$ in $L^2(\mathbb{R}^d)$ given by

$$S_\mathcal{D} = \sum_n f_n \otimes f_n.$$

Motivated by the desire to increase the size of the data set for machine learning purposes, the Ω-*augmentation* of $\mathcal{D}$ with respect to a compact set $\Omega \subset \mathbb{R}^d$ is defined as

$$\mathcal{D}_\Omega = \left\{ \frac{\pi(z)}{|\Omega|^{1/2}} f_n \,:\, f_n \in \mathcal{D},\, z \in \Omega \right\}, \tag{B.4.3}$$

which leads to the augmented data operator

$$S_{\mathcal{D}_\Omega} = \frac{1}{|\Omega|} \int_\Omega \sum_n \pi(z) f_n \otimes \pi(z) f_n \, dz = \frac{1}{|\Omega|} \chi_\Omega \star S_\mathcal{D}. \tag{B.4.4}$$

Real-world data augmentation in machine learning is of course based on discrete time-frequency shifts in (B.4.3) meaning that the set Ω should be discrete. However, the formulation (B.4.4) is inherently continuous since we are integrating over Ω and can only be realized as a classical function-operator convolution. To solve this problem, we can use measure-operator convolutions to define a version of (B.4.4) which works for discrete Ω as

$$S_{\mathcal{D}_\Omega} = \left(\frac{1}{|\Omega|} \sum_{\omega \in \Omega} \delta_\omega \right) \star S_{\mathcal{D}}.$$

Not all results from [60] carry over directly to this discrete setting but we will treat one result on the eigenvalues of mixed-state Gabor multipliers in particular that [60] relies on, originally proved in [75] for Gabor multipliers. To do so, we merge the setup in [60] with the one in [192] and consider the set Ω restricted to a lattice Λ.

B.4.2 Mixed-state Gabor frames

In the same way that the frame condition (B.2.4) is what is needed to discretize the reconstruction formula (B.2.2), we will see that the reconstruction formula (B.2.11) only holds in the lattice setting when (S, Λ) satisfies a version of the following condition, originally discussed in [193].

Definition B.4.1. A pair (S, Λ) where $S \in \mathcal{S}^1$ is positive and Λ is a lattice in $\mathbb{R}^{2d}$ is said to be a *mixed-state Gabor frame with frame constants A, B* if

$$A\|\psi\|^2 \leq \sum_{\lambda \in \Lambda} Q_S(\psi)(\lambda) \leq B\|\psi\|^2 \qquad \text{for all } \psi \in L^2(\mathbb{R}^d) \tag{B.4.5}$$

for some $A, B > 0$. Moreover, if $A = B$, the frame is said to be *tight* and the common quantity $A = B$ is referred to as the *frame constant*.

Technically, if $S = \sum_n s_n(\varphi_n \otimes \varphi_n)$, the above definition is equivalent to the collection $\left\{\sqrt{s_n}\pi(\lambda)\varphi_n\right\}_{\lambda,n}$ being an (infinite) *multi-window* Gabor frame, introduced in [223]. This connection is made clear in the following example which shows how tight mixed-state Gabor frames can be constructed as linear combinations of tight Gabor frames.

Example B.4.2. Let Λ be a lattice in $\mathbb{R}^{2d}$ and $(\varphi_n)_n$ in $L^2(\mathbb{R}^d)$ a collection of windows such that (φ_n, Λ) is a tight frame with frame constant A for each n. Then if $(s_n)_n \in \ell^1$ is a convex combination so that

$$S = \sum_n s_n(\varphi_n \otimes \varphi_n)$$

is a positive trace-class operator, the pair (S, Λ) is a tight mixed-state Gabor frame with frame constant A. Indeed, we verify

$$\begin{aligned}\sum_{\lambda\in\Lambda} Q_S(\psi)(\lambda) &= \sum_{\lambda\in\Lambda}(\psi\otimes\psi)\star\left(\sum_n s_n(\varphi_n\otimes\varphi_n)\check{}\right)\\ &= \sum_n s_n \sum_{\lambda\in\Lambda}|V_{\varphi_n}\psi(\lambda)|^2 = \sum_n s_n A\|\psi\|^2 = A\|\psi\|^2,\end{aligned}$$

using the fact that $(\psi\otimes\psi)\star(\varphi\otimes\varphi)\check{} = |V_\varphi\psi|^2$, the frame equality and Fubini for the change of order of summation.

Our concept of mixed-state Gabor frames can also be seen as a special case of Gabor g-frames, introduced in [193, Section 5]. More specifically, (S, Λ) is a mixed-state Gabor frame if and only if $\sqrt{S}$ generates a Gabor g-frame with respect to Λ. As we will not need the full generality of Gabor g-frames in the following, we stick to the notation above when proving the promised reconstruction formula.

Proposition B.4.3. Let (S, Λ) be a tight mixed-state Gabor frame with frame constant A. We then have the reconstruction of the identity

$$\sum_{\lambda\in\Lambda}\pi(\lambda)S\pi(\lambda)^*\psi = A\psi \qquad \text{for all } \psi\in L^2(\mathbb{R}^d). \tag{B.4.6}$$

Proof. We expand S in its singular value decomposition as $S = \sum_n s_n(\varphi\otimes\varphi)$ with $s_n \geq 0$ for all n by the positivity of S. The tightness of the mixed state Gabor frame can then be written more explicitly as

$$A\|\psi\|^2 = \sum_{\lambda\in\Lambda}\sum_n s_n|\langle\psi, \pi(\lambda)\varphi_n\rangle|^2, \tag{B.4.7}$$

while (B.4.6) can be written as

$$A\psi = \sum_{\lambda\in\Lambda}\sum_n s_n\langle\psi, \pi(\lambda)\varphi_n\rangle\pi(\lambda)\varphi_n. \tag{B.4.8}$$

Consider the linear mapping $\Theta : L^2(\mathbb{R}^d) \to \ell^2(\Lambda\times\mathbb{N})$ given by

$$\Theta\psi = \frac{1}{\sqrt{A}}\left(\sqrt{s_n}\langle\psi, \pi(\lambda)\varphi_n\rangle\right)_{\lambda, n}.$$

By (B.4.7), Θ is an isometry and hence preserves inner products. We expand ψ in an orthonormal basis $(e_m)_m$ and use this as

$$\begin{aligned}
A\psi &= A\sum_m \langle \psi, e_m\rangle e_m \\
&= A\sum_m \langle \Theta\psi, \Theta e_m\rangle e_m \\
&= \sum_m \sum_{\lambda\in\Lambda} \sum_n s_n \langle \psi, \pi(\lambda)\varphi_n\rangle \langle \pi(\lambda)\varphi_n, e_m\rangle e_m \\
&= \sum_{\lambda\in\Lambda} \sum_n s_n \langle \psi, \pi(\lambda)\rangle \pi(\lambda)\varphi_n,
\end{aligned}$$

which is precisely (B.4.8). □

Remark B.4.4. In the language of [193], the above proposition states that the pair $(\sqrt{S}, \sqrt{S})$ generate dual Gabor g-frames.

We can also deduce a lattice analogue of Proposition B.2.1 (x) using the tight frame condition.

Proposition B.4.5. Let $T, S \in \mathcal{S}^1$ and Λ a lattice such that $\big(S, \Lambda\big)$ is a tight mixed-state Gabor frame with frame constant 1. Then

$$\sum_{\lambda\in\Lambda} T \star \check{S}(\lambda) = \operatorname{tr}(T).$$

Proof. Expand T in its singular value decomposition with orthonormal bases $(\xi_n)_n$ and $(\eta_n)_n$ of $L^2(\mathbb{R}^d)$ as $T = \sum_n t_n(\xi_n \otimes \eta_n)$ and note that

$$\sum_{\lambda\in\Lambda} T \star \check{S}(\lambda) = \sum_n t_n \sum_{\lambda\in\Lambda} (\xi_n \otimes \eta_n) \star \check{S}(\lambda) = \sum_n t_n \sum_{\lambda\in\Lambda} Q_S(\xi_n, \eta_n)(\lambda).$$

To proceed, we need an analogue of the frame identity (B.4.5) for $Q_S(\xi_n, \eta_n)$ which can be shown using a polarization argument. A straightforward calculation shows that

$$Q_S(\psi, \phi)(\lambda) = \frac{1}{4}\Big(Q_S(\psi+\phi)(\lambda) - Q_S(\psi-\phi)(\lambda) + iQ_S(\psi+i\phi)(\lambda) - iQ_S(\psi-i\phi)(\lambda)\Big).$$

Hence by summing over all $\lambda \in \Lambda$ on both sides, we obtain

$$\sum_{\lambda\in\Lambda} Q_S(\psi, \phi)(\lambda) = \frac{1}{4}\Big(\|\psi+\phi\|^2 - \|\psi-\phi\|^2 + i\|\psi+i\phi\|^2 - i\|\psi-i\phi\|^2\Big) = \langle \psi, \phi\rangle.$$

We can now finish the computation as

$$\sum_{\lambda\in\Lambda} T \star \check{S}(\lambda) = \sum_n t_n \sum_{\lambda\in\Lambda} Q_S(\xi_n, \eta_n)(\lambda) = \sum_n t_n \langle \xi_n, \eta_n\rangle = \operatorname{tr}(T).$$

□

Next, we show how the frame properties of S are changed when we convolve it with a well-behaved measure. This generalizes the continuous property discussed in [155, Section 7.2].

Proposition B.4.6. Let (S, Λ) be a mixed-state Gabor frame with frame bounds A, B and $\mu_c \in M(\mathbb{R}^{2d})$ a discrete measure of the form $\mu_c = \sum_{\lambda\in\Lambda} c(\lambda)\delta_\lambda$ where $c \in \ell^1(\Lambda)$. Then $(\check{\mu_c} \star S, \Lambda)$ is also a mixed-state Gabor frame with frame bounds $A\sum_{\lambda\in\Lambda} c(\lambda), B\sum_{\lambda\in\Lambda} c(\lambda)$.

Proof. We compute the sum

$$\begin{aligned}
\sum_{\lambda\in\Lambda} Q_{\check{\mu_c}\star S}(\psi)(\lambda) &= \sum_{\lambda\in\Lambda} (\psi\otimes\psi)\star(\mu_c\star\check{S})(\lambda)\\
&= \sum_{\lambda\in\Lambda} \mu_c * Q_S(\psi)(\lambda)\\
&= \sum_{\lambda\in\Lambda}\sum_{\lambda'\in\Lambda} Q_S(\psi)(\lambda-\lambda')c(\lambda')\\
&= \left(\sum_{\lambda\in\Lambda} c(\lambda)\right)\left(\sum_{\lambda\in\Lambda} Q_S(\psi)(\lambda)\right),
\end{aligned}$$

where we used Proposition B.3.6 (iii) for the first equality, Proposition B.3.6 (vi) for the second and Fubini for the final. The desired conclusion now follows upon substituting into the frame condition (B.4.5). □

B.4.3 Eigenvalues of mixed-state Gabor multipliers

Gabor multipliers or localization operators on lattices were first systematically investigated by Feichtinger and Nowak in [75] while the generalization of replacing the window by an operator was first discussed briefly in [192, Section 4.1] but never named nor investigated further. Following the terminology of [154], when we replace the window of a Gabor multiplier with a trace-class operator, we refer to the resulting operator as a *mixed-state* Gabor multiplier. We make this precise in the following definition.

Definition B.4.7. Given a lattice $\Lambda \subset \mathbb{R}^{2d}$, an operator $S \in \mathcal{L}(L^2(\mathbb{R}^d))$ and a compact subset $\Omega \subset \mathbb{R}^{2d}$, we refer to the operator

$$G^S_{\Omega,\Lambda} = \sum_{\lambda\in\Lambda} \chi_\Omega(\lambda)\delta_\lambda \star S =: \mu^\Lambda_\Omega \star S \tag{B.4.9}$$

as a *mixed-state Gabor multiplier* and write $\mu^\Lambda_\Omega = \sum_{\lambda\in\Lambda}\chi_\Omega(\lambda)\delta_\lambda$ for the measure.

In light of the reconstruction formula in Proposition B.4.3 above it makes sense that we will need to place similar conditions on S and Λ in order for mixed-state Gabor multipliers to be well behaved. We collect the properties we will need in the following definition.

Definition B.4.8. Given a lattice Λ, an operator $S \in \mathcal{L}(L^2(\mathbb{R}^d))$ is said to be a *density operator with respect to* Λ if S is positive, trace-class and such that (S, Λ) is a tight mixed-state Gabor frame with frame constant 1.

Now we are ready to state our main theorem on the eigenvalues of mixed-state Gabor multipliers. It generalizes a result in [75] by a method similar to the way how the corresponding result for localization operators in [74, 155] was derived.

Theorem B.4.9. *Let S be a density operator with respect to Λ, let $\Omega \subset \mathbb{R}^{2d}$ be compact and fix $\delta \in (0,1)$. If $\{\lambda_k^{R\Omega}\}_k$ are the eigenvalues of $G^S_{R\Omega,\Lambda}$, then*

$$\frac{\#\big\{k : \lambda_k^{R\Omega} > 1-\delta\big\}}{|R\Omega \cap \Lambda| \operatorname{tr}(S)} \to 1 \quad \textit{as } R \to \infty.$$

In the pathological case where $(R\Omega)_{R>0}$ does not exhaust $\mathbb{R}^{2d}$ (e.g. $\Omega = \{0\}$), the result is not meaningful and we from here on assume that this is not the case.

Much of the proof and notation is analogous to that in [155] and before moving to the proof we state and prove some important lemmata.

Lemma B.4.10. *Let S be a density operator with respect to Λ and $\Omega \subset \mathbb{R}^{2d}$ a compact domain. Then the eigenvalues $\big\{\lambda_k^{\Omega}\big\}_k$ of $G^S_{\Omega,\Lambda}$ satisfy $0 \le \lambda_k^\Omega \le 1$.*

Proof. By the positivity of $G^S_{\Omega,\Lambda}$ from Proposition B.3.6 (i), the eigenvalues are non-negative and real-valued. For the upper bound we, let h_k^Ω denote the eigenfunction associated to λ_k^Ω, then

$$\begin{aligned}\lambda_k^\Omega &= \big\langle G^S_{\Omega,\Lambda} h_k^\Omega, h_k^\Omega\big\rangle = \left\langle \left(\sum_{\lambda\in\Lambda} \chi_\Omega(\lambda)\delta_\lambda \star S\right) h_k^\Omega, h_k^\Omega \right\rangle \\ &\le \sum_{\lambda\in\Lambda} \big\langle \pi(\lambda) S \pi(\lambda)^* h_k^\Omega, h_k^\Omega \big\rangle = \sum_{\lambda\in\Lambda} Q_S(h_k^\Omega)(\lambda) = \big\|h_k^\Omega\big\|^2 = 1.\end{aligned}$$

□

In the following, the function $\tilde{S} = S \star \check{S} : \mathbb{R}^{2d} \to \mathbb{R}$ will play an important role. We start by proving some of its properties.

Lemma B.4.11. *Let $G^S_{\Omega,\Lambda}$ be a mixed-state Gabor multiplier. Then*

$$\operatorname{tr}\left(\big(G^S_{\Omega,\Lambda}\big)^2\right) = \sum_{\lambda\in\Lambda\cap\Omega} \sum_{\lambda'\in\Lambda\cap\Omega} \tilde{S}(\lambda'-\lambda).$$

Proof. Note first that $\operatorname{tr}\left((G^S_{\Omega,\Lambda})^2\right) = G^S_{\Omega,\Lambda} \star \check{G}^S_{\Omega,\Lambda}(0)$ by (B.2.9) and hence by Proposition B.3.6 (iii), Proposition B.3.6 (vi) and the symmetry of $\star$,

$$\begin{aligned}
\operatorname{tr}\left((G^S_{\Omega,\Lambda})^2\right) &= \left(\mu^\Lambda_\Omega \star S\right) \star \left(\check{\mu}^\Lambda_\Omega \star \check{S}\right)(0) \\
&= \mu^\Lambda_\Omega * \left(\check{\mu}^\Lambda_\Omega * (S \star \check{S})\right)(0) \\
&= \int_{\mathbb{R}^{2d}} \check{\mu}^\Lambda_\Omega * \tilde{S}(0 - z)\, d\mu^\Lambda_\Omega(z) \\
&= \sum_{\lambda \in \Omega\cap\Lambda} \check{\mu}^\Lambda_\Omega * \tilde{S}(-\lambda) \\
&= \sum_{\lambda \in \Omega\cap\Lambda} \int_{\mathbb{R}^{2d}} \tilde{S}(-z-\lambda)\, d\check{\mu}^\Lambda_\Omega(z) \\
&= \sum_{\lambda \in \Omega\cap\Lambda} \sum_{\lambda' \in \Omega\cap\Lambda} \tilde{S}(\lambda' - \lambda).
\end{aligned}$$

□

Lemma B.4.12. *Let S be a density operator with respect to Λ. Then $\tilde{S}$ is non-negative and*

$$\sum_{\lambda\in\Lambda} \tilde{S}(\lambda) = \operatorname{tr}(S).$$

Proof. Non-negativity follows by Proposition B.2.1 (ii) and the sum follows from (S, Λ) being a tight mixed-state Gabor frame with frame constant 1 by Proposition B.4.5. □

Lemma B.4.13. *Let S be a density operator with respect to Λ and $\{\lambda^\Omega_k\}_k$ the eigenvalues of $G^S_{\Omega,\Lambda}$ counted with multiplicity. Then*

$$\sum_k \lambda^\Omega_k = |\Omega \cap \Lambda| \operatorname{tr}(S).$$

Proof. By a theorem of Lidskii [190], the sum of the eigenvalues counted with multiplicity is equal to the trace of the operator and so the result follows from Proposition B.3.6 (iv). □

The next lemma is our replacement for the approximation of the identity argument used in the corresponding proofs in [155] and [8].

Lemma B.4.14. *Let S be a density operator with respect to Λ. Then*

$$\lim_{R\to\infty} \left| \frac{1}{|R\Omega \cap \Lambda|} \sum_{\lambda \in R\Omega\cap\Lambda} \sum_{\lambda' \in R\Omega\cap\Lambda} \tilde{S}(\lambda' - \lambda) - \operatorname{tr}(S) \right| = 0.$$

Proof. Fix $\varepsilon > 0$ and note that since $\sum_{\lambda\in\Lambda} \tilde{S}(\lambda) = \operatorname{tr}(S)$, we can find a sufficiently large R_0 so that

$$\sum_{\lambda\in R_0\Omega\cap\Lambda} \tilde{S}(\lambda) > \operatorname{tr}(S) - \varepsilon.$$

Note also that if $\lambda \in (R - \operatorname{diam}(R_0\Omega))\Omega \cap \Lambda$ then

$$R_0\Omega \cap \Lambda \subset \{\lambda - \lambda' : \lambda' \in R\Omega \cap \Lambda\}.$$

Now let $R' = R - \operatorname{diam}(R_0\Omega)$ and write

$$\begin{aligned}\frac{1}{|R\Omega\cap\Lambda|}\sum_{\lambda\in R\Omega\cap\Lambda}\sum_{\lambda'\in R\Omega\cap\Lambda}\tilde{S}(\lambda'-\lambda) &\geq \frac{1}{|R\Omega\cap\Lambda|}\sum_{\lambda\in R'\Omega\cap\Lambda}\sum_{\lambda'\in R\Omega\cap\Lambda}\tilde{S}(\lambda'-\lambda)\\ &\geq \frac{1}{|R\Omega\cap\Lambda|}\sum_{\lambda\in R'\Omega\cap\Lambda}\sum_{\lambda'\in R_0\Omega\cap\Lambda}\tilde{S}(\lambda')\\ &> \frac{|R'\Omega\cap\Lambda|}{|R\Omega\cap\Lambda|}\big(\operatorname{tr}(S)-\varepsilon\big).\end{aligned}$$

Since $\frac{R'}{R} \to 1$ as $R \to \infty$, the quantity can be made arbitrarily close to $\operatorname{tr}(S)$. Hence the desired conclusion follows. □

Lemma B.4.15. *Let S be a density operator with respect to Λ. Then for each $\delta \in (0,1)$,*

$$\begin{aligned}&\Big|\#\{k : \lambda_k^\Omega > 1-\delta\} - |\Omega\cap\Lambda|\operatorname{tr}(S)\Big|\\ &\qquad\leq \max\left\{\frac{1}{\delta}, \frac{1}{1-\delta}\right\}\left|\sum_{\lambda\in\Omega\cap\Lambda}\sum_{\lambda'\in\Omega\cap\Lambda}\tilde{S}(\lambda'-\lambda) - |\Omega\cap\Lambda|\operatorname{tr}(S)\right|.\end{aligned}$$

Proof. Define the operator $H(G^S_{\Omega,\Lambda})$ using the eigendecomposition $G^S_{\Omega,\Lambda} = \sum_k \lambda_k^\Omega (h_k^\Omega \otimes h_k^\Omega)$ as

$$H\big(G^S_{\Omega,\Lambda}\big) = \sum_k H(\lambda_k^\Omega)(h_k^\Omega \otimes h_k^\Omega), \qquad H(t) = \begin{cases} -t & \text{if } 0 \leq t \leq 1-\delta, \\ 1-t & \text{if } 1-\delta < t \leq 1, \end{cases}$$

which is well defined since $0 \leq \lambda_k^\Omega \leq 1$ by Lemma B.4.10. From this it follows that

$$\operatorname{tr}\big(H\big(G^S_{\Omega,\Lambda}\big)\big) = \sum_k H(\lambda_k^\Omega) = \#\{k : \lambda_k^\Omega > 1-\delta\} - |\Omega\cap\Lambda|\operatorname{tr}(S)$$

by Lemma B.4.13 and the definition of H. Hence

$$\Big|\#\{k : \lambda_k^\Omega > 1-\delta\} - |\Omega\cap\Lambda|\operatorname{tr}(S)\Big| = \big|\operatorname{tr}\big(H\big(G^S_{\Omega,\Lambda}\big)\big)\big| \leq \operatorname{tr}\big(|H|\big(G^S_{\Omega,\Lambda}\big)\big).$$

The function H can be bounded as

$$|H(t)| \leq \max\left\{\frac{1}{\delta}, \frac{1}{1-\delta}\right\}(t-t^2),$$

and so it follows that

$$\left|\#\left\{k : \lambda_k^{\Omega} > 1-\delta\right\} - |\Omega\cap\Lambda|\operatorname{tr}(S)\right| \leq \max\left\{\frac{1}{\delta}, \frac{1}{1-\delta}\right\}\operatorname{tr}\left(G^S_{\Omega,\Lambda} - \left(G^S_{\Omega,\Lambda}\right)^2\right),$$

which can be written as in the original statement by Lemma B.4.11. □

We are now ready to finish the proof.

Proof of Theorem B.4.9. Applying Lemma B.4.15 to $R\Omega$, we find

$$\begin{aligned}&\left|\#\{k : \lambda_k^{R\Omega} > 1-\delta\} - |R\Omega\cap\Lambda|\operatorname{tr}(S)\right| \\ &\qquad\leq \max\left\{\frac{1}{\delta}, \frac{1}{1-\delta}\right\}\left|\sum_{\lambda\in R\Omega\cap\Lambda}\sum_{\lambda'\in R\Omega\cap\Lambda}\tilde{S}(\lambda'-\lambda) - |R\Omega\cap\Lambda|\operatorname{tr}(S)\right|.\end{aligned}$$

The desired conclusion then follows upon dividing by $|R\Omega\cap\Lambda|$ and applying Lemma B.4.14. □

B.4.4 A Berezin-Lieb inequality on lattices

Berezin-Lieb inequalities have been investigated in [138, 152, 211] and recently used in [60] in the continuous setting. As discussed above, in applications we are more interested in lattice formulations which is why we present a version of the Berezin-Lieb inequality for measure-operator convolutions and operator-operator convolutions on lattices below.

Theorem B.4.16. *Let $S \in \mathcal{S}^1$ be positive with $\operatorname{tr}(S) = 1$ and Λ be a lattice such that (S, Λ) is a mixed-state Gabor frame with upper frame constant B. If $T \in \mathcal{S}^1$ is positive and Φ is a non-negative, convex and continuous function on a domain containing the spectrum of T and the range of $T \star \check{S}$, then*

$$\sum_{\lambda\in\Lambda}\Phi\circ(T\star\check{S})(\lambda) \leq B\operatorname{tr}\left(\Phi(T)\right).$$

Similarly, if $\mu_c = \sum_{\lambda\in\Lambda}c(\lambda)\delta_\lambda$ is a positive bounded measure and Φ a non-negative, convex and continuous function on a domain containing the spectrum of $\mu_c \star S$ and the range of Bc restricted to Λ, then

$$\operatorname{tr}(\Phi(\mu_c\star S)) \leq \sum_{\lambda\in\Lambda}\Phi(Bc(\lambda)),$$

where we also have to assume that Φ is non-decreasing if (S, Λ) is not tight.

Proof. Using the singular value decomposition $T = \sum_n t_n(\xi_n \otimes \xi_n)$, we find that

$$\begin{aligned} T \star \check{S}(\lambda) &= \sum_n t_n \big\langle \check{S}\pi(\lambda)^*\xi_n, \pi(\lambda)^*\xi_n \big\rangle = \sum_n t_n Q_S(\xi_n)(\lambda) \\ \implies \Phi \circ (T \star \check{S})(\lambda) &\leq \sum_n \Phi(t_n) Q_S(\xi_n)(\lambda), \end{aligned}$$

using Jensen's inequality. Thus by summing over Λ we get

$$\begin{aligned} \sum_{\lambda \in \Lambda} \Phi \circ (T \star \check{S})(\lambda) &\leq \sum_{\lambda \in \Lambda} \sum_n \Phi(t_n) Q_S(\xi_n)(\lambda) \\ &= \sum_n \Phi(t_n) \sum_{\lambda \in \Lambda} Q_S(\xi_n)(\lambda) \\ &= B \operatorname{tr}(\Phi(T)). \end{aligned}$$

Next for the measure-operator convolution statement, note that by the positivity of μ_c and S, $\mu_c \star S$ is positive via Proposition B.3.6 (i) and so we can write its singular value decomposition as $\mu_c \star S = \sum_n \lambda_n(\xi_n \otimes \xi_n)$. Hence

$$\begin{aligned} \operatorname{tr}(\Phi(\mu_c \star S)) &= \operatorname{tr}\left(\sum_n \Phi(\lambda_n)(\xi_n \otimes \xi_n) \right) \\ &= \sum_n \Phi\big(\langle (\mu_c \star S)\xi_n, \xi_n \rangle\big) \\ &= \sum_n \Phi\left(\sum_{\lambda \in \Lambda} c(\lambda) \langle S\pi(\lambda)^*\xi_n, \pi(\lambda)^*\xi_n \rangle \right), \end{aligned}$$

where we used that $\lambda_n = \langle (\mu_c \star S)\xi_n, \xi_n \rangle$. Next if (S, Λ) is not a tight frame, we use that Φ is non-decreasing to multiply the argument by $\frac{B}{\sum_{\lambda \in \Lambda} Q_{\check{S}}(\xi_n)(\lambda)}$ which yields

$$\begin{aligned} \operatorname{tr}(\Phi(\mu_c \star S)) &\leq \sum_n \Phi\left(\frac{\sum_{\lambda \in \lambda} Bc(\lambda) Q_{\check{S}}(\xi_n)(\lambda)}{\sum_{\lambda \in \Lambda} Q_{\check{S}}(\xi_n)(\lambda)} \right) \\ &\leq \sum_n \sum_{\lambda \in \Lambda} \Phi(Bc(\lambda)) Q_{\check{S}}(\xi_n)(\lambda), \end{aligned}$$

where the last step follows by Jensen's inequality. By switching the order of summation and using that $\{\pi(\lambda)^*\xi_n\}_n$ is an orthonormal basis of $L^2(\mathbb{R}^d)$ by the unitarity of π, we deduce that the last quantity can be written as $\sum_{\lambda \in \Lambda} \Phi(Bc(\lambda))$ which is what we wished to show. ◻

B.4.5 Convergence of sequences of Gabor multipliers

In this section, we focus our attention on deducing the consequences of Corollary B.3.8. We have already seen that measure-operator convolutions generalize both Gabor multipliers and localization operators and so our goal will be to apply the corollary to these settings. In Sections B.2 and B.3, we have used the BUPU framework for the bounded and tight nets of measures but we will now turn our attention to the more explicit setting of parameterized lattices. More specifically, we will consider rectangular lattices $\Lambda_{\alpha,\beta} = \alpha\mathbb{Z}^d \times \beta\mathbb{Z}^d \lhd \mathbb{R}^{2d}$ parameterized by $\alpha, \beta > 0$. This setting is prevalent in applications and while our results can easily be extended to larger classes of lattices we settle for this setting in the interest of clarity.

Given such a rectangular parameterized lattice, we can define the discretization in measure form of a function $m : \mathbb{R}^{2d} \to \mathbb{C}$ as

$$\mu^m_{\alpha,\beta} = \alpha^d \beta^d \sum_{\lambda \in \Lambda_{\alpha,\beta}} m(\lambda)\delta_\lambda. \tag{B.4.10}$$

In order to apply Corollary B.3.8, we will need to prove that the above measures are uniformly bounded in α and β. An important tool for this will be the following lemma.

Lemma B.4.17. *For any $m \in W(L^\infty, \ell^1)(\mathbb{R}^{2d})$, we can choose α_0, β_0 such that*

$$\left\|\mu^m_{\alpha,\beta}\right\|_M \leq C(\alpha_0, \beta_0)\|m\|_{W(L^\infty,\ell^1)}$$

for all $\alpha < \alpha_0$ and $\beta < \beta_0$.

Proof. Each point in $\Lambda_{\alpha,\beta}$ may be associated to a half-open square of the form $Q^{\alpha_0,\beta_0}_{n,k} = \big[\alpha_0 n, \alpha_0(n+1)\big) \times \big[\beta_0 k, \beta_0(k+1)\big)$ as these cover $\mathbb{R}^{2d}$. We label these collections of points as $P_{n,k} = \Lambda_{\alpha,\beta} \cap Q^{\alpha_0,\beta_0}_{n,k}$. It then holds that

$$\|\mu^m_{\alpha,\beta}\|_M = \alpha^d\beta^d \sum_{\lambda\in\Lambda_{\alpha,\beta}} |m(\lambda)| = \alpha^d\beta^d \sum_{n,k}\sum_{\lambda\in P_{n,k}} |m(\lambda)|.$$

Meanwhile, by the equivalence of norms on $W(L^\infty, \ell^1)(\mathbb{R}^{2d})$ for different BUPU, we have that $\|m\|_{W(L^\infty,\ell^1)} \geq C(\alpha_0,\beta_0)\sum_{n,k}\|m\|_{L^\infty(Q^{\alpha_0,\beta_0}_{n,k})}$ and so our desired result will follow if we can show that

$$\alpha^d\beta^d \sum_{\lambda\in P_{n,k}} |m(\lambda)| \leq C\|m\|_{L^\infty(Q^{\alpha_0,\beta_0}_{n,k})}$$

for some constant C. Indeed, each evaluation $|m(\lambda)|$ for $\lambda \in P_{n,k}$ is bounded from above by $\|m\|_{L^\infty(Q_{n,k}^{\alpha_0,\beta_0})}$ because $P_{n,k} \subset Q_{n,k}^{\alpha_0,\beta_0}$. Moreover, because $\alpha < \alpha_0, \beta < \beta_0$, the quantity

$$\sum_{\lambda \in P_{n,k}} \alpha^d \beta^d = \alpha^d \beta^d \left|\Lambda_{\alpha,\beta} \cap Q_{n,k}^{\alpha_0,\beta_0}\right|$$

is uniformly bounded, which finishes the proof. □

Approximating localization operators

As a first application, we will show that the discretization (B.4.10) essentially approaches the original mask for our purposes, when the lattice scaling parameters tends to zero.

Theorem B.4.18. *Let $m : \mathbb{R}^{2d} \to \mathbb{C}$ be a Riemann-integrable function in $W(L^\infty, \ell^1)(\mathbb{R}^{2d})$ and S a trace-class operator. Then we have the convergence*

$$\left\|\mu_{\alpha,\beta}^m \star S - m \star S\right\|_{\mathcal{S}^1} \to 0 \quad \text{as } \alpha, \beta \to 0.$$

Proof. As a first step we will show that m may be taken to have compact support without loss of generality. Fix $\varepsilon > 0$ and let $K \subset \mathbb{R}^{2d}$ be such that $\|m\chi_{K^c}\|_{W(L^\infty,\ell^1)} < \varepsilon$. We may then estimate

$$\begin{aligned}
\left\|\mu_{\alpha,\beta}^m \star S - m \star S\right\|_{\mathcal{S}^1} &\leq \left\|\mu_{\alpha,\beta}^m \star S - \mu_{\alpha,\beta}^{m\chi_K} \star S\right\|_{\mathcal{S}^1} + \left\|\mu_{\alpha,\beta}^{m\chi_K} \star S - m \star S\right\|_{\mathcal{S}^1} \\
&\leq \left\|\mu_{\alpha,\beta}^{m\chi_{K^c}}\right\|_M \|S\|_{\mathcal{S}^1} + \left\|\mu_{\alpha,\beta}^{m\chi_K} \star S - m \star S\right\|_{\mathcal{S}^1} \\
&\leq C\varepsilon \|S\|_{\mathcal{S}^1} + \left\|\mu_{\alpha,\beta}^{m\chi_K} \star S - m\chi_K \star S\right\|_{\mathcal{S}^1} + \left\|m\chi_K \star S - m \star S\right\|_{\mathcal{S}^1} \\
&\leq C\varepsilon \|S\|_{\mathcal{S}^1} + \left\|\mu_{\alpha,\beta}^{m\chi_K} \star S - m\chi_K \star S\right\|_{\mathcal{S}^1} + \|m\chi_{K^c}\|_M \|S\|_{\mathcal{S}^1} \\
&\leq 2C\varepsilon \|S\|_{\mathcal{S}^1} + \left\|\mu_{\alpha,\beta}^{m\chi_K} \star S - m\chi_K \star S\right\|_{\mathcal{S}^1},
\end{aligned}$$

where we used Lemma B.4.17 and the estimate of Proposition B.3.4 twice. Since ε was arbitrary, we may now replace $m\chi_K$ by m and assume that m has compact support.

To apply Corollary B.3.8 to $\|\mu_{\alpha,\beta}^m \star S - m \star S\|_{\mathcal{S}^1}$ and get the desired convergence, we need to show that $\mu_{\alpha,\beta}^m \to m$ in the weak-* sense and that the sequence $\left(\mu_{\alpha,\beta}^m\right)_{\alpha,\beta}$ is bounded and tight. Boundedness follows directly from Lemma B.4.17 while tightness follows from the compactness of the support of m established above.

The desired weak-* convergence can be formulated as that for any $f \in C_b(\mathbb{R}^{2d})$,

$$\mu_{\alpha,\beta}^m(f) \to \int_{\operatorname{supp}(m)} m(z) f(z)\, dz \qquad \text{as } \alpha, \beta \to 0.$$

We claim that the left-hand side can be realized as a Riemann sum approximating the right hand side. Indeed, from the definition of $\mu^m_{\alpha,\beta}$ we have that

$$\mu^m_{\alpha,\beta}(f) = \sum_{\lambda\in\Lambda_{\alpha,\beta}} m(\lambda)f(\lambda)\alpha^d\beta^d, \tag{B.4.11}$$

and this sum goes over the rectangles $\big[n\alpha, (n+1)\alpha\big)\times\big[k\beta, (k+1)\beta\big)$ which have area $\alpha^d\beta^d$. Now since m is Riemann integrable and f is continuous and bounded and therefore uniformly bounded on the support of m, $m\cdot f$ is also Riemann integrable and (B.4.11) converges to $\int_{\mathrm{supp}(m)} m(z)f(z)\,dz$ as promised. □

As discussed in the preliminaries section, $m \star S$ is a mixed-state localization operator. Moreover, convolving the discretized version of m with S essentially gives a mixed-state Gabor multiplier. To see this, we define Gabor multipliers $G^\varphi_{m,\alpha,\beta}$ based on the lattice parameters as

$$\begin{aligned} G^\varphi_{m,\alpha,\beta}\psi &= |\Lambda_{\alpha,\beta}| \sum_{n,k\in\mathbb{Z}} m(n\alpha, k\beta)\langle\psi, \pi(n\alpha, k\beta)\varphi\rangle\pi(n\alpha, k\beta)\varphi \\ &= \mu^m_{\alpha,\beta} \star (\varphi\otimes\varphi)(\psi). \end{aligned} \tag{B.4.12}$$

Note that this definition differs from (B.4.9) used earlier, even when the convolution is with an arbitrary trace-class operator. The reason is that we incorporate the lattice normalization as we will take the mask to be constant and the lattice parameters to be varying.

By taking m to be Riemann integrable and in $W(L^\infty,\ell^1)(\mathbb{R}^{2d})$, the rank-one case of Theorem B.4.18 allows us to use the definition (B.4.12) and deduce that

$$\big\|G^\varphi_{m,\alpha,\beta} - A^\varphi_m\big\|_{\mathcal{S}^1} \to 0 \qquad \text{as } \alpha,\beta\to 0,$$

since the localization operator A^φ_m, from (B.2.3), can be written as $A^\varphi_m = m\star(\varphi\otimes\varphi)$. Even more explicitly, we can formulate the following corollary when the mask is simply an indicator function.

Corollary B.4.19. Let $\Omega\subset\mathbb{R}^{2d}$ a compact Jordan measurable subset, $\varphi\in L^2(\mathbb{R}^d)$ and $G^\varphi_{m,\alpha,\beta}$ as in (B.4.12). Then

$$\big\|G^\varphi_{\Omega,\alpha,\beta} - A^\varphi_\Omega\big\|_{\mathcal{S}^1} \to 0 \qquad \text{as } \alpha,\beta\to 0.$$

The proof is essentially immediate upon noting that Jordan measurability of Ω is equivalent to Riemann integrability of χ_Ω and that compact support implies $\chi_\Omega\in W(L^\infty,\ell^1)(\mathbb{R}^{2d})$.

These results should be contrasted with those [75, Section 5.9] where localization operators are approximated by Gabor multipliers in the trace-class norm in a similar way. The largest difference is that our result is valid for arbitrary L^2 windows (and of course also operator windows).

One consequence of this result is that (mixed-state) localization operators are dense in the space of localization operators. In [24], it was shown that localization operators with varying masks are dense in $\mathcal{S}^1$ as long as $V_\varphi \varphi \neq 0$. The generalization of this to mixed-state localization operators was established in [136, 153] where the condition was replaced by $\mathcal{F}_W(S) \neq 0$. Combining these results, we can approximate arbitrary $\mathcal{S}^1$ operators as follows.

Corollary B.4.20. For any trace-class operator S with $\mathcal{F}_W(S)$ free of zeros, the set

$$\left\{\mu^m_{\alpha,\beta} \star S : \alpha, \beta > 0, m \in L^1(\mathbb{R}^{2d})\right\}$$

is dense in $\mathcal{S}^1$.

Proof. The result follows from [153, Theorem 7.6] and Theorem B.4.18 upon noting that $W(L^\infty, \ell^1)(\mathbb{R}^{2d})$ is dense in $L^1(\mathbb{R}^{2d})$. □

Mask, lattice and window continuity

Next, we turn our attention to investigating how the mixed-state Gabor multipliers $\mu^m_{\alpha,\beta} \star S$ behave when varying the parameters. Results of this nature have been proved in [71, 75] but we will be able to prove stronger convergence under stricter conditions on the mask but weaker conditions on the window.

Theorem B.4.21. *Let $(\alpha_n)_n$ and $(\beta_n)_n$ be sequences converging to $\alpha, \beta > 0$ respectively, $(m_n)_n$ a sequence in the Wiener amalgam space $W(C_0, \ell^1)(\mathbb{R}^d)$ converging to m in the corresponding norm and $(S_n)_n$ a sequence of trace-class operators converging to S in $\mathcal{S}^1$. Then for the associated mixed-state Gabor multipliers we have the convergence*

$$G^{S_n}_{m_n,\alpha_n,\beta_n} \to G^S_{m,\alpha,\beta} \quad \text{in } \mathcal{S}^1 \text{ as } n \to \infty.$$

Proof. To begin, we will need to prove that $\left(\mu^{m_n}_{\alpha_n,\beta_n}\right)_n$ is a tight and bounded sequence of measures. Boundedness follows from Lemma B.4.17 as functions in $W(C_0, \ell^1)(\mathbb{R}^{2d})$ clearly are Riemann integrable and in $W(L^\infty, \ell^1)(\mathbb{R}^{2d})$ and we can bound $\|m_n\|_{W(L^\infty,\ell^1)}$ from above uniformly since $(m_n)_n$ converges in this space. For tightness, we again use Lemma B.4.17 to relate the $M(\mathbb{R}^{2d})$ and $W(C_0, \ell^1)(\mathbb{R}^{2d})$ norms and use that for any set K, the sequence $\left(m_n\big|_{K^c}\right)_n$ converges in $W(C_0, \ell^1)(\mathbb{R}^{2d})$ which can be made arbitrarily small by letting K be large due to the finiteness of $\|m\|_{W(C_0,\ell^1)}$.

With verification out of the way, we next show that we can treat the convergence of $(S_n)_n$ separately. Indeed, by the triangle inequality

$$\begin{aligned} \big\|G^{S_n}_{m_n,\alpha_n,\beta_n} - G^S_{m,\alpha,\beta}\big\|_{\mathcal{S}^1} \leq \big\|G^{S_n}_{m_n,\alpha_n,\beta_n} - G^S_{m_n,\alpha_n,\beta_n}\big\|_{\mathcal{S}^1} \\ + \big\|G^S_{m_n,\alpha_n,\beta_n} - G^S_{m,\alpha,\beta}\big\|_{\mathcal{S}^1}. \end{aligned} \tag{B.4.13}$$

Expanding the first term as $\big\|\mu^{m_n}_{\alpha_n,\beta_n} \star (S_n - S)\big\|_{\mathcal{S}^1}$ and applying the estimate of Corollary B.2.8, we see that this term can be made arbitrarily small uniformly by the boundedness of $\big(\mu^{m_n}_{\alpha_n,\beta_n}\big)_n$. We may hence focus our attention on showing convergence of the last term of (B.4.13). As is hopefully expected, this is done by applying Corollary B.3.8 to

$$\big\|G^S_{m_n,\alpha_n,\beta_n} - G^S_{m,\alpha,\beta}\big\|_{\mathcal{S}^1} = \big\|\mu^{m_n}_{\alpha_n,\beta_n} \star S - \mu^m_{\alpha,\beta} \star S\big\|_{\mathcal{S}^1}.$$

Since we have already shown that $\big(\mu^{m_n}_{\alpha_n,\beta_n}\big)_n$ is a tight and bounded sequence, it only remains to show that for all $f \in C_b(\mathbb{R}^{2d})$,

$$\mu^{m_n}_{\alpha_n,\beta_n}(f) \to \mu^m_{\alpha,\beta}(f) \qquad \text{as } n \to \infty.$$

We treat m first by estimating

$$\begin{aligned} \big|\mu^{m_n}_{\alpha_n,\beta_n}(f) - \mu^m_{\alpha,\beta}(f)\big| &\leq \big|\mu^{m_n}_{\alpha_n,\beta_n}(f) - \mu^m_{\alpha_n,\beta_n}(f)\big| + \big|\mu^m_{\alpha_n,\beta_n}(f) - \mu^m_{\alpha,\beta}(f)\big| \\ &\leq \big|\mu^{m_n-m}_{\alpha_n,\beta_n}(f)\big| + \big|\mu^m_{\alpha_n,\beta_n}(f) - \mu^m_{\alpha,\beta}(f)\big| \\ &\leq C(2\alpha, 2\beta)\|m - m_n\|_{W(C_0,\ell^1)}\|f\|_{C_b} + \big|\mu^m_{\alpha_n,\beta_n}(f) - \mu^m_{\alpha,\beta}(f)\big|, \end{aligned} \tag{B.4.14}$$

where we for the last step used Lemma B.4.17 with the fact that for large enough n, $\alpha_n \leq 2\alpha$ and $\beta_n \leq 2\beta$ since $\alpha_n \to \alpha$ and $\beta_n \to \beta$ as $n \to \infty$. As the first term can be made arbitrarily small by considering sufficiently large n, we may restrict our attention to the second term. Here we must exchange m for a function with compact support. This is done by first estimating that $\big|\mu^m_{\alpha_n,\beta_n}(f) - \mu^m_{\alpha,\beta}(f)\big|$ is less than

$$\big|\mu^m_{\alpha_n,\beta_n}(f) - \mu^{m\chi_K}_{\alpha_n,\beta_n}(f)\big| + \big|\mu^{m\chi_K}_{\alpha_n,\beta_n}(f) - \mu^{m\chi_K}_{\alpha,\beta}(f)\big| + \big|\mu^{m\chi_K}_{\alpha,\beta}(f) - \mu^m_{\alpha,\beta}(f)\big| \tag{B.4.15}$$

for any set $K \subset \mathbb{R}^{2d}$ by the triangle inequality. The first and last terms may be estimated in the same way as in (B.4.14) and upon choosing K large enough so that $\|m\chi_{K^c}\|_{W(C_0,\ell^1)}$ is sufficiently small, we see that we only need to estimate the middle term of (B.4.15) where we can replace $m\chi_K$ by a new m with compact support.

Now since m has compact support, the sum

$$\begin{aligned}&\left|\mu^m_{\alpha_n,\beta_n}(f)-\mu^m_{\alpha,\beta}(f)\right|\\&\quad\leq\sum_{k,j\in\mathbb{Z}^d}\left|m(k\alpha_n,j\beta_n)f(k\alpha_n,j\beta_n)\alpha_n^d\beta_n^d-m(k\alpha,j\beta)f(k\alpha,j\beta)\alpha^d\beta^d\right|\end{aligned}$$

has a finite number of terms and so to show that it goes to zero as $n\to\infty$, it suffices to show that each term goes to zero as $n\to\infty$. This is however easy to see by the continuity of m and f and consequently, we are done. □

For a comparable earlier result in [71, Theorem 3.3], we had to assume $m\in W(C_0,\ell^2)(\mathbb{R}^{2d})$ and as a result obtained $\mathcal{S}^2$ convergence. In our setting, requiring $m\in W(C_0,\ell^1)(\mathbb{R}^{2d})$ is more natural as Corollary B.3.8 only works for $\mathcal{S}^1$ convergence which requires the mask to be a bounded measure or integrable. It should also be noted that the continuity assumption also is natural as the lattice setting means that any discontinuities in the mask may be picked up for some parameters but not all which can change the operator norm. In [75, Theorem 5.6.7] trace-class convergence was proved but the mask was assumed to be in $S_0(\mathbb{R}^{2d})$ and then windows were not allowed to vary. The relaxing of conditions compared to the earlier results should largely be attributed to the alternative path: instead of performing the proof at the level of operator symbol we use methods from quantum harmonic analysis and the framework laid out in [73].

In the interest of clarity, we again highlight the rank-one case of Gabor multipliers.

Corollary B.4.22. Let $(\alpha_n)_n$ and $(\beta_n)_n$ be sequences converging to $\alpha,\beta>0$ respectively, $(m_n)_n$ a sequence in the Wiener amalgam space $W(C_0,\ell^1)(\mathbb{R}^d)$ converging to m and $(\varphi_n)_n$ a sequence of $L^2(\mathbb{R}^d)$ windows converging to $\varphi\in L^2(\mathbb{R}^d)$. Then for the associated Gabor multipliers we have the convergence

$$G^{\varphi_n}_{m_n,\alpha_n,\beta_n}\to G^{\varphi}_{m,\alpha,\beta}\quad\text{in }\mathcal{S}^1\text{ as }n\to\infty.$$

Paper C

Weyl Quantization of Exponential Lie Groups for Square Integrable Representations

Stine Marie Berge and Simon Halvdansson

Preprint

Paper C

Weyl Quantization of Exponential Lie Groups for Square Integrable Representations

Abstract

We construct a general quantization procedure for square integrable functions on well-behaved connected exponential Lie groups. The Lie groups in question should admit at least one co-adjoint orbit of maximal possible dimension. The construction is based on composing the Fourier-Wigner transform with another Fourier transform we call the Fourier-Kirillov transform. This quantization has many desirable properties including respecting function translations and inducing a well-behaved Wigner distribution.

Moreover, we investigate the connection to the operator convolutions of quantum harmonic analysis. This is intricately connected to Weyl quantization in the Weyl-Heisenberg setting. We find that convolution relations in quantum harmonic analysis can be written as group convolutions of Weyl quantizations. This implies that the squared modulus of the wavelet transform of the representation can be written as a convolution between two Wigner distributions. Lastly, we look at how we can extend known results based on Weyl quantization to wider classes of groups using our quantization procedure.

C.1 Introduction

The Weyl quantization mapping, originally studied by H. Weyl in [212, 213], is an object of fundamental importance in analysis and mathematical physics. One view of the mapping is as a correspondence rule between functions on phase space, $L^2(\mathbb{R}^{2d})$, and observables in the form of Hilbert-Schmidt operators on the Hilbert space $L^2(\mathbb{R}^d)$. There is a large body of influential work studying this mapping, including but not limited to [104, 173, 189]. We refer to [218] for a comprehensive overview.

Various generalizations of the Weyl quantization to different phase spaces and function spaces have been proposed and studied over the last decades. Some generalizations focus on the dequantization of rank-one operators, that is, dequantization of pure states. This perspective has been applied to a wide variety of contexts, see e.g. [12, 36, 100, 172, 176]. Following the work in [31, 94] we set out to formulate Weyl quantization in the general context in a more systematic way. This formulation is based on the composition of the Fourier-Wigner transform and a *Fourier-Kirillov transform*, with the Fourier-Kirillov transform taking the role of the symplectic Fourier transform. In this way, we are able to formulate our results in a general setting without reference to the specific group structure. It should be mentioned that the proposed Weyl quantization given is similar in spirit to the one given in [161], and enjoys many of the same properties.

Main construction

The goal of this article is to set up a quantization scheme for functions on a exponential Lie group G. The proposed scheme generalizes the Weyl quantization for the affine group, see [31, 32, 94, 96, 97]. Integral to the quantization is the existence of square integrable representations. From Kirillov orbit theory, it is well known how to construct the irreducible unitary representations $\pi : G \to \mathcal{U}(\mathcal{H})$ from the co-adjoint orbits. When the dimension of the co-adjoint orbit coincides with the dimension of the group, then the corresponding representation is square integrable. Hence in our construction we will always assume that the orbits are of the same dimension as the group. Notably, this includes the affine group [32] and the shearlet group [144].

As mentioned above, our construction of Weyl quantization is based on generalizing the relation $a_S = \mathcal{F}_\sigma(\mathcal{F}_W(S))$ where $\mathcal{F}_\sigma$ is the symplectic Fourier transform and $\mathcal{F}_W$ is the Fourier-Wigner transform. The Fourier-Wigner transform is a standard object in this setup and it and its inverse are defined as

$$\mathcal{F}_W(A)(x) = \mathrm{tr}(A\mathcal{D}\pi(x)), \qquad \mathcal{F}_W^{-1}(f) = \pi(\check{f}) \circ \mathcal{D}.$$

where $\check{f}(x) = f(x^{-1})$, $\mathcal{D}$ is the inverse Duflo-Moore operator and $\pi(f)$ is the (left) integrated representation of $f : G \to \mathbb{C}$. We will properly define and give background on these objects in the preliminaries section. Section C.3.1 is devoted to working out all of the properties of the Fourier-Wigner transform which we will need, including that it can be extended to a unitary isometry from the Hilbert-Schmidt operators $\mathcal{S}^2(\mathcal{H})$ to $L^2_r(G)$, the space of square integrable functions on G with respect to the right Haar measure.

In place of the symplectic Fourier transform, we will set up a *Fourier-Kirillov transform* $\mathcal{F}_{\mathrm{KO}}$. This transform is coherently linked to a chosen co-adjoint orbit. Let $K(x)$ denote the co-adjoint map and $\mathcal{O}_F$ the co-adjoint orbit $\{K(x)F : x \in G\}$. Then $\mathcal{F}_{\mathrm{KO}} : L^2_r(G) \to L^2_r(G)$ is defined as

$$\mathcal{F}_{\mathrm{KO}}(f)(x) = \frac{1}{\sqrt{|\mathrm{Pf}_F| \cdot \Delta(x)}} \int_G f(y) e^{2\pi i \langle K(x^{-1})F, \log(y)\rangle} \frac{1}{\sqrt{\Theta(\log(y))}} \, \mathrm{d}\mu_r(y)$$

where Pf_F, Δ and Θ are functions which ensure proper normalization across spaces to be further detailed later. The Fourier-Kirillov transform can be viewed, up to some normalizing factors, as the Fourier transform on $L^2(\mathfrak{g}) \to L^2(\mathfrak{g}^*)$ restricted to the orbit corresponding to the element $F \in \mathcal{O}_F$.

With the two Fourier transforms defined, we define the quantization A and dequantization $a = A^{-1}$ by composing them as

$$A_f = \mathcal{F}_{\mathrm{W}}^{-1}(\mathcal{F}_{\mathrm{KO}}^{-1}(f)), \qquad a_S = \mathcal{F}_{\mathrm{KO}}(\mathcal{F}_{\mathrm{W}}(S)).$$

Using the properties of the transforms, we are able to deduce that the full quantization map $A : L^2_r(G) \to \mathcal{S}^2(\mathcal{H})$ is a unitary isometry. This quantization scheme is similar to the Wigner map outlined in [12, 13], however the inclusion of the co-adjoint map creates an algebraic structure on the phase space, namely the original group structure.

Quantization structure

Having proved that the quantization is a unitary isometry, we move on to showing that quantization respects translations and complex conjugation in an appropriate manner. Specifically, we show the two identities

$$\pi(x)^* A_f \pi(x) = A_{R_{x^{-1}} f}, \qquad A_f^* = A_{\bar{f}}, \tag{C.1.1}$$

where $R_{x^{-1}} f(y) = f(yx^{-1})$. These properties hold true for classical Weyl quantization but are not given for quantization schemes based on the Wigner distribution. Moreover, classical Weyl quantization can be given the additional structure of being an isometric $*$-isomorphism between H^* algebras as was shown by Pool [173]. This means that taking the adjoint and composing operators in $\mathcal{S}^2$ corresponds to certain actions on functions in a isomorphic way. We show that the same is true for our quantization which lead to notions of *twisted convolutions* and *twisted multiplication*. Specifically, we show that all maps are isometric $*$-isomorphisms between H^* algebras in the commutative diagram

$$\begin{tikzcd}
\big(\mathcal{S}^2(\mathcal{H}), \circ, {}^*\big) \arrow[d, "\mathcal{F}_\mathrm{W}"] \arrow[drr, dotted, "a"] & & \\
\big(\mathcal{F}_\mathrm{W}(\mathcal{S}^2), \natural, \sqrt{\Delta(\cdot)}\,\bar{\check{\ }}\big) \arrow[rr, "\mathcal{F}_\mathrm{KO}"] & & \big(L^2_r(G), \sharp, {}^{-}\big)
\end{tikzcd}$$

where the intermediate space $\mathcal{F}_\mathrm{W}(\mathcal{S}^2) = \mathcal{F}_\mathrm{KO}^{-1}(L^2_r(G))$ is a subset of $L^2_r(G)$. This construction is detailed in Section C.4.2.

Wigner distributions

Wigner distributions are most commonly related to Weyl quantization through the weak relation

$$\big\langle f, W(\phi,\psi)\big\rangle_{L^2_r} = \big\langle A_f\psi, \phi\big\rangle_{\mathcal{H}}$$

for all $f \in L^2_r(G)$ and $\psi, \phi \in \mathcal{H}$. This relation is equivalent to the Wigner distribution $W(\psi,\phi)$ being the dequantization of the rank-one operator $(\psi \otimes \phi) : \xi \mapsto \langle \xi, \phi\rangle\psi$. For this reason, we set $W(\psi,\phi) = a_{\psi\otimes\phi}$ and get the properties that the mapping $(\psi,\phi) \mapsto W(\psi,\phi)$ is sesquilinear, that $\overline{W(\psi,\phi)} = W(\phi,\psi)$ and $R_x W(\psi,\phi) = W(\pi(x)\psi, \pi(x)\phi)$ from general quantization properties such as (C.1.1) for free. In the case that the group is unimodular, we show that

$$\int_G W(\psi,\phi)(x)\,\mathrm{d}\mu(x) = \langle \psi, \phi\rangle.$$

Quantum harmonic analysis

The framework of quantum harmonic analysis is concerned with convolutions, translations and Fourier transforms of operators and their interactions. One of several ways to define these operations is through their interaction with Weyl quantization. In recent years, quantum harmonic analysis has been developed for the affine group [32] in conjunction with a Weyl quantization procedure [31] as well as for general locally compact groups [115] without any connection to quantization. We will show that our quantization procedure is compatible with the quantum harmonic analysis operations in [32] and [115], thus imbuing them with additional structure. Specifically, we will show that the operator convolutions from quantum harmonic analysis can be realized as group convolutions of Weyl symbols as

$$\begin{aligned} f \star S &= A_{f * a_S}, \\ T \star S &= a_T * \check{a_S}. \end{aligned}$$

As a consequence, we can generalize relations such as the convolution of two Wigner distributions being a spectrogram,

$$W(\psi) * \widetilde{W(\phi)} = |\mathcal{W}_\phi \psi|^2 \tag{C.1.2}$$

where $\mathcal{W}_\phi \psi$ is the wavelet transform of ψ with respect to ϕ.

Applications

Since our construction is a generalization of that for the Weyl-Heisenberg and affine groups, for applications our hope is to generalize some of the properties of those Weyl quantization. The first application in Section C.5 has not been explored for the affine group but uses the relation (C.1.2) to develop a criterion for *phase retrieval*, meaning inversion of the map $\mathcal{W}_\phi \psi \mapsto |\mathcal{W}_\phi \psi|$.

Having developed a Wigner distribution, we treat the *Wigner approximation problem* which asks how close a given $L^2_r(G)$ function is to being a Wigner distribution using the same approach as has been used for the Weyl-Heisenberg and affine groups earlier. Similarly, we are able to extend a proof technique which shows that for both Wavelet spaces $\mathcal{W}_\phi(\mathcal{H})$ and Wigner spaces $W(\mathcal{H}, \phi)$, spaces induced by different ϕ_1, ϕ_2 have trivial intersection as long as ϕ_1 and ϕ_2 are linearly independent. In both of these cases, we have show how general Weyl quantization allows us to apply the proof techniques from standard Weyl quantization to more general contexts.

These examples of applications are not meant to be exhaustive but rather to illustrate the value of the tools developed in the article.

C.2 Preliminaries

C.2.1 Weyl quantization on the Weyl-Heisenberg Group

To set the stage, we will review how Weyl quantization is set up in the well-known case of the Weyl-Heisenberg group. Specifically, we will outline how we view Weyl quantization as being induced by a pair of Fourier transforms instead of the Wigner distribution. This viewpoint is more amenable to generalizations. For a more physics-oriented perspective, see e.g. [114, Chap. 13].

Via Wigner distribution

The original construction by Weyl is most easily formulated using the *cross-Wigner distribution* [215], also called the *Wigner transform*, given by

$$W(\psi, \phi)(x, \omega) = \int_{\mathbb{R}^d} \psi(x + t/2)\overline{\phi(x - t/2)} e^{-2\pi i \omega \cdot t} \, \mathrm{d}t, \qquad \psi, \phi \in L^2(\mathbb{R}^d).$$

We write $W(\phi) = W(\phi, \phi)$ for convenience. Using the Wigner transform, we can define the Weyl quantization A_f of a function $f \in L^2(\mathbb{R}^{2d})$ weakly via the relation

$$\langle A_f \psi, \phi \rangle_{L^2(\mathbb{R}^d)} = \big\langle f, W(\phi, \psi) \big\rangle_{L^2(\mathbb{R}^{2d})}. \tag{C.2.1}$$

We say that the map $f \mapsto A_f$ is the *quantization map* or *Weyl transform* and write $S \mapsto a_S$ for the inverse, called the *dequantization map*. The dequantization map is a bijective isometry from the space of Hilbert-Schmidt operators $\mathcal{S}^2$ to $L^2(\mathbb{R}^{2d})$, see [173]. Hence $A_{a_S} = S$ and $f = a_{A_f}$. Through an elementary computation, one can show that the following *Moyal identity* holds for the Wigner transform.

Theorem C.2.1. *For all $\psi_1, \psi_2, \phi_1, \phi_2 \in L^2(\mathbb{R}^d)$, it holds that*

$$\big\langle W(\psi_1, \phi_1), W(\psi_2, \phi_2) \big\rangle_{L^2(\mathbb{R}^{2d})} = \langle \psi_1, \psi_2 \rangle_{L^2(\mathbb{R}^d)} \overline{\langle \phi_1, \phi_2 \rangle}_{L^2(\mathbb{R}^d)}. \tag{C.2.2}$$

In view of this result, the quantization rule (C.2.1) implies that the dequantization of the rank-one operator $\psi \otimes \phi$ is the Wigner transform $W(\psi, \phi)$. Consequently, one can view quantization as being induced by the Wigner transform or vice versa.

Define the (projective) *Schrödinger representation* by

$$\pi(z) f(t) = \pi(x, \omega) f(t) = M_\omega T_x f(t) = e^{2\pi i \omega t} f(t - x)$$

where we have made the identification $z = (x, \omega) \in \mathbb{R}^{2d}$. We will discuss the time and frequency-shifts T_x, M_Ω further in Section C.2.5. The key relation between the Schrödinger representation and the Weyl quantization is included in the following proposition:

Proposition C.2.2. Let $f \in L^2(\mathbb{R}^{2d})$. Then $A_{\bar{f}} = A_f^*$ and

$$A_{f(\cdot - z)} = \pi(z) A_f \pi(z)^*.$$

Via Fourier transforms

When generalizing Weyl quantization beyond the standard $\mathbb{R}^{2d}$ setting, a common practice is to generalize the definition of the Wigner distribution and use (C.2.1) to induce a quantization map. In this paper, we argue that while the relation (C.2.1) indeed is fundamental to quantization, directly defining a different version of the Wigner distribution is not. In fact, there is another way to define Weyl quantization which itself induces a Wigner distribution which is what we will detail in this section.

Recall that phase space $\mathbb{R}^{2d}$ is a symplectic space when equipped with the symplectic form

$$\sigma(x_1, \omega_1, x_2, \omega_2) = \omega_1 \cdot x_2 - \omega_2 \cdot x_1, \qquad (x_1, \omega_1),\ (x_2, \omega_2) \in \mathbb{R}^{2d}.$$

Here the appropriate Fourier transform is the *symplectic Fourier transform* $\mathcal{F}_\sigma$, defined for $f \in L^1(\mathbb{R}^{2d})$ as

$$\mathcal{F}_\sigma(f)(z) = \int_{\mathbb{R}^{2d}} f(z')e^{-2\pi i\sigma(z,z')}\,\mathrm{d}z'.$$

As is the case with the standard Fourier transform, the symplectic Fourier transform can be extended to square integrable functions as well. On the space of operators, we will instead use the Fourier-Wigner (sometimes called Fourier-Weyl) transform. The *Fourier-Wigner transform* $\mathcal{F}_W$ is a map from the trace-class operators on $L^2(\mathbb{R}^d)$, denoted by $\mathcal{S}^1$, to a subset of $L^2(\mathbb{R}^{2d})$, given by

$$\mathcal{F}_W(S)(z) = e^{-\pi i x\cdot\omega}\,\mathrm{tr}(\pi(-z)S).$$

It follows from [84, 153, 218] that the Weyl quantization can be written as

$$a_S = \mathcal{F}_\sigma(\mathcal{F}_W(S)).$$

C.2.2 Representation theory on locally compact groups

In this section, let G denote a locally compact group of *type I*, see e.g. [85, p. 229]. In the context of Lie groups, "type I" is often referred to as *tame groups*, in contrast with *wild groups*. Tame groups are a broad class of Lie groups that include all exponential Lie groups, see [105, 200]. Given such a group we will denote the *left Haar measure* on the group G by μ_l^G and the *right Haar measure* by μ_r^G, see [84] for the definition. There exists a multiplicative function $\Delta_G : G \to \mathbb{R}^+$ called the *modular function* defined by the relationship

$$\mu_l^G(x) = \Delta_G(x)\mu_r^G(x), \qquad x \in G.$$

Whenever it is clear from the context, we will suppress the group G in the notation μ_l^G, μ_r^G and Δ_G. The right versus left Haar measure on an L^p-space will be indicated by either an r or l subscript.

Given functions $f_1, f_2 \in L_r^1(G)$ we define their *(right-)convolution* by

$$f_1 * f_2(x) = \int_G f_1(y)f_2(xy^{-1})\,\mathrm{d}\mu_r(y).$$

Define *right*- and *left-translation* of a function $f : G \to \mathbb{C}$ by $(R_y f)(x) = f(xy)$ and $(L_y f)(x) = f(y^{-1}x)$, respectively. The *involution* of a function $f \in L_l^1(G)$ is defined by

$$\check{f}(x) = f(x^{-1}), \qquad \text{for } x \in G.$$

Notice that the involution maps $L^1_l(G)$ to $L^1_r(G)$. The inverse of the involution will also be denoted by $\check{\cdot}$. It should also be noted that different definitions of function involution exists in the literature, see e.g. [85].

We will denote by $(\pi, \mathcal{H}_\pi)$ a *irreducible unitary representation* π of the group G acting on the separable Hilbert space $\mathcal{H}_\pi$. When it is clear from the context, we will write $\mathcal{H}$ instead of $\mathcal{H}_\pi$. The representation $(\pi, \mathcal{H})$ induces an *integrated (left) representation* acting on $f \in L^1_l(G)$ by

$$\pi(f)\phi = \int_G f(x)\pi(x)\phi \, \mathrm{d}\mu_l(x), \qquad \text{for } \phi \in \mathcal{H}.$$

Given an irreducible unitary representation $(\pi, \mathcal{H})$, we denote the *wavelet transform* by

$$\mathcal{W}_\phi\psi(x) = \langle \psi, \pi(x)^*\phi \rangle_{\mathcal{H}}. \tag{C.2.3}$$

The convention of using $\pi(x)^*$ instead of $\pi(x)$ for the wavelet transform is non-standard but more compatible with the right Haar measure which we will prefer. We say that $(\pi, \mathcal{H})$ is *square integrable* if there exists a non-zero $\phi \in \mathcal{H}$ such that $\mathcal{W}_\phi\phi \in L^2_r(G)$. For square integrable representations, we have the following classical orthogonality relation from [64].

Theorem C.2.3 (Duflo-Moore Theorem)**.** *Let $(\pi, \mathcal{H})$ be a unitary square integrable representation. There exists a unique positive densely defined operator $\mathcal{D}^{-1} : \mathrm{Dom}(\mathcal{D}^{-1}) \subset \mathcal{H} \to \mathcal{H}$ such that*

$$\langle \mathcal{W}_{\phi_1}\psi_1, \mathcal{W}_{\phi_2}\psi_2 \rangle_{L^2_r} = \langle \psi_1, \psi_2 \rangle_{\mathcal{H}} \overline{\langle \mathcal{D}^{-1}\phi_1, \mathcal{D}^{-1}\phi_2 \rangle_{\mathcal{H}}}, \tag{C.2.4}$$

for $\phi_1, \phi_2 \in \mathrm{Dom}(\mathcal{D}^{-1})$ and $\psi_1, \psi_2 \in \mathcal{H}$.

We will refer to the operator $\mathcal{D}^{-1}$ in the theorem as the *Duflo-Moore operator*. Functions $\phi \in \mathrm{Dom}(\mathcal{D}^{-1})$ are said to be *admissible*. Below we collect some standard results on how the modular function Δ, representation π, and Duflo-Moore operator $\mathcal{D}^{-1}$ interact.

Lemma C.2.4. *The following properties hold:*

(i) $d\mu_l(xy) = \Delta(y) \, \mathrm{d}\mu_l(x)$,

(ii) $d\mu_r(x) = \Delta(x^{-1}) \, \mathrm{d}\mu_l(x)$,

(iii) $d\mu_r(x^{-1}) = \Delta(x) \, \mathrm{d}\mu_r(x)$,

(iv) $\mathcal{D}\pi(x) = \sqrt{\Delta(x)}\pi(x)\mathcal{D}$.

C.2.3 Lie groups

In this section, we will assume that G is a connected Lie group with Lie algebra $\mathfrak{g}$. We will denote the (Lie group) *exponential map* by $\exp : \mathfrak{g} \to G$. In the case that the exponential map is a diffeomorphism, we say that the Lie group G is *exponential*. The inverse of the exponential map is called the logarithm and is denoted by log. For a Lie group homeomorphism $\Phi : G \to H$ the following diagram commutes

$$\begin{array}{ccc} \mathfrak{g} & \xrightarrow{\Phi_*} & \mathfrak{h} \\ {\scriptstyle \exp}\downarrow & & \downarrow{\scriptstyle \exp} \\ G & \xrightarrow{\Phi} & H \end{array}, \tag{C.2.5}$$

where $\mathfrak{g}$ and $\mathfrak{h}$ are the Lie algebras of G and H, respectively.

Let $\mathrm{d}X$ denote the Euclidean measure on $\mathfrak{g}$. There exists a function $\Theta : \mathfrak{g} \to \mathbb{R}^+$ such that

$$\mathrm{d}\mu_r(\exp(X)) = \Theta(X)\,\mathrm{d}X$$

is the right Haar measure in exponential coordinates. Additionally,

$$\mathrm{d}\mu_l(\exp(X)) = \Theta(-X)\,\mathrm{d}X.$$

Hence

$$\frac{\Theta(-X)}{\Theta(X)} = \Delta_G(\exp(X)). \tag{C.2.6}$$

We can compute Θ by using the formula

$$\Theta(X) = \left|\det\left(\frac{e^{\mathrm{ad}(X)} - 1}{\mathrm{ad}(X)}\right)\right|. \tag{C.2.7}$$

In the case that G is a nilpotent group, we have that $\Theta(X) = 1$ for all $X \in \mathfrak{g}$.

A Lie group can act on itself via conjugation $\Psi : G \to \mathrm{Aut}(G)$ given by $\Psi_x(y) = xyx^{-1}$. Fixing $x \in G$, the *adjoint representation* is given by

$$\mathrm{Ad}_x = (\Psi_x)_* : \mathfrak{g} \to \mathfrak{g}.$$

The modular function can be written by using the adjoint map as

$$\det(\mathrm{Ad}_x) = \Delta(x). \tag{C.2.8}$$

By using (C.2.5) we get that $\exp(\mathrm{Ad}_x X) = \Psi_x(\exp(X))$. Fixing $Y \in \mathfrak{g}$ and taking the derivative in the other variable, we get that $\mathrm{ad}_Y : \mathfrak{g} \to \mathfrak{g}$ can be computed as $\mathrm{ad}_Y X = [Y, X]$.

Group examples

Before going further, it is instructive to look at a couple of examples of exponential Lie groups where the representations are known.

Example C.2.5 (Heisenberg Group)**.** The Heisenberg group is the underlying group for the standard Wigner transform outlined in Section C.2.5. It should however be noted that the representation given by co-adjoint orbit theory is not square integrable.

Let $\mathbb{H}$ denote the Heisenberg group, which when written in matrix form is the group

$$\mathbb{H} = \left\{ (x, y, z) = \begin{pmatrix} 1 & x & z \\ 0 & 1 & y \\ 0 & 0 & 1 \end{pmatrix} : x, y, z \in \mathbb{R} \right\}.$$

The corresponding Lie algebra $\mathfrak{h}$ is the span of

$$X = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad Y = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \quad Z = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

It is endowed with the bracket relations

$$[X, Z] = [Y, Z] = 0 \text{ and } [X, Y] = Z.$$

The exponential map is given by

$$\exp(xX + yY + zZ) = (x, y, z + xy/2).$$

It is well-known that both the right and the left Haar-measure is given by

$$\mathrm{d}\mu^{\mathbb{H}}(x, y, z) = \mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z.$$

Hence the modular function $\Delta_{\mathbb{H}}(x, y, z) = 1$ and $\Theta(xX + yY + zZ) = 1$. Except the characters, the irreducible representations on $\mathbb{H}$ is, up to equivalence, given by

$$\pi_\hbar(x, y, z) f(w) = \exp(2\pi i\hbar(z + yw)) f(w + x),$$

where $\hbar \in \mathbb{R} \setminus \{0\}$ and $f \in L^2(\mathbb{R})$.

Example C.2.6 (Affine Group)**.** The *(reduced) affine group* $(\mathrm{Aff}, \cdot_{\mathrm{Aff}})$ is the Lie group whose underlying set is the upper half plane $\mathrm{Aff} := \mathbb{R}^+ \times \mathbb{R} := (0, \infty) \times \mathbb{R}$, while the group operation is given by

$$(a, x) \cdot_{\mathrm{Aff}} (b, y) := (ab, ay + x), \qquad (a, x), (b, y) \in \mathrm{Aff}.$$

We can represent the affine group Aff and its Lie algebra $\mathfrak{aff}$ in matrix form

$$\mathrm{Aff} = \left\{ \begin{pmatrix} a & x \\ 0 & 1 \end{pmatrix} : a > 0,\, x \in \mathbb{R} \right\}, \quad \mathfrak{aff} = \{uU + vV : u, v \in \mathbb{R}\},$$

where

$$U = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \text{ and } V = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}.$$

The Lie algebra structure of $\mathfrak{aff}$ is completely determined by

$$[U, V] = V.$$

The right and left Haar measure are given by

$$\mathrm{d}\mu_r(a, x) = \frac{\mathrm{d}a\,\mathrm{d}x}{a}, \qquad \mathrm{d}\mu_l(a, x) = \frac{\mathrm{d}a\,\mathrm{d}x}{a^2},$$

hence the modular function is $\Delta(a, x) = \frac{1}{a}$. Computing the exponential map gives

$$\exp(uU + vV) = (e^u, v\lambda(u)),$$

where $\lambda(u) = \frac{e^u - 1}{u}$. In the case of the exponential map Θ is given by

$$\Theta(uU + vV) = e^{-u} \cdot \lambda(u).$$

Except the characters, there are only two non-equivalent irreducible representations of the affine group. These are given by

$$\pi_+(a, x)\psi(r) := e^{-2\pi i x r}\psi(ar), \qquad \psi \in L^2(\mathbb{R}^+, r^{-1}\, dr)$$

and

$$\pi_-(a, x)\psi(r) := e^{2\pi i x r}\psi(ar), \qquad \psi \in L^2(\mathbb{R}^+, r^{-1}\, dr).$$

Example C.2.7 (Shearlet Group)**.** A matrix representation of the Shearlet group $\mathbb{S}$ is given by

$$(a, s, x_1, x_2) = (a, s, x) = \begin{pmatrix} a & \sqrt{a}s & x_1 \\ 0 & \sqrt{a} & x_2 \\ 0 & 0 & 1 \end{pmatrix}, \quad a > 0,\ s \in \mathbb{R}, x \in \mathbb{R}^2.$$

The multiplication of two elements is defined as

$$\begin{aligned}(a, s, x_1, x_2)(b, t, y_1, y_2) &= (ab, s + \sqrt{a}t, x + S_s A_a y) \\ &= (ab, s + \sqrt{a}t, x_1 + ay_1 + \sqrt{a}sy_2, x_2 + \sqrt{a}y_2)\end{aligned}$$

where

$$A_a = \begin{pmatrix} a & 0 \\ 0 & \sqrt{a} \end{pmatrix}, \qquad S_s = \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix}.$$

The left- and right-Haar measure are given by

$$\mathrm{d}\mu_l(a,s,x_1,x_2) = \frac{\mathrm{d}a\,\mathrm{d}s\,\mathrm{d}x_1\,\mathrm{d}x_2}{a^3}, \qquad \mathrm{d}\mu_r(a,s,x_1,x_2) = \frac{\mathrm{d}a\,\mathrm{d}s\,\mathrm{d}x_1\,\mathrm{d}x_2}{a}.$$

Hence the modular function is

$$\Delta_{\mathbb{S}}(a,s,x) = \frac{1}{a^2}.$$

The Lie algebra in matrix form is given by

$$\mathfrak{s} = \left\{ \begin{pmatrix} \alpha & \sigma & \xi_1 \\ 0 & \alpha/2 & \xi_2 \\ 0 & 0 & 0 \end{pmatrix} : \alpha, \sigma, \xi_1, \xi_2 \in \mathbb{R} \right\}.$$

Let A, B, C and D be a basis for the Lie algebra consisting of the matrices

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/2 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \; B = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \; C = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \text{ and } D = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}.$$

The Lie bracket is given by

$$\begin{aligned}[][\alpha A + \sigma B + \xi_1 C + \xi_2 D, \tilde{\alpha} A + \tilde{\sigma} B + \tilde{\xi}_1 C + \tilde{\xi}_2 D] \\ = \frac{\alpha\tilde{\sigma} - \tilde{\alpha}\sigma}{2} B + (\alpha\tilde{\xi}_1 - \tilde{\alpha}\xi_1 + \sigma\tilde{\xi}_2 - \tilde{\sigma}\xi_2) C + \frac{\alpha\tilde{\xi}_2 - \tilde{\alpha}\xi_2}{2} D.\end{aligned}$$

In this case we have the exponential map

$$\exp(\alpha A + \sigma B + \xi_1 C + \xi_2 D) = (e^\alpha, \sigma\lambda(\alpha/2), \xi_1\lambda(\alpha) + \sigma\xi_2\lambda(\alpha/2)^2/2, \xi_2\lambda(\alpha/2)),$$

where

$$\lambda(\alpha) = \frac{e^\alpha - 1}{\alpha}.$$

The function Θ giving the measures in exponential coordinates is given by

$$\Theta(\alpha A + \sigma B + \xi_1 C + \xi_2 D) = \lambda(\alpha)\lambda(\alpha/2)^2.$$

The irreducible representations of central importance are

$$\pi_+(a,s,x_1,x_2)\phi(b,t) = e^{-2\pi i(bx_1 + \sqrt{b}tx_2)}\phi\left(ba, t + s\sqrt{b}\right)$$

and

$$\pi_-(a,s,x_1,x_2)\phi(b,t) = e^{2\pi i(bx_1 + \sqrt{b}tx_2)}\phi\left(ba, t + s\sqrt{b}\right)$$

where $\phi \in L^2(\mathbb{R}^+ \times \mathbb{R}, \frac{\mathrm{d}b\,\mathrm{d}t}{b})$.

C.2.4 Representation theory of exponential Lie groups

Induced representations on Lie groups

In this section, let G be a Lie group and H a closed subgroup of G. Then a *right-coset* is the set $Hg = \{hg : h \in H\}$ for all $g \in G$. The collection of all right cosets will be denoted by $R = H \setminus G$. There exists a quotient map $q : G \to R$, sending $g \mapsto Hg$. It is well known that R is a homogeneous manifold equipped with a canonical measure inherited from the Lie group. The measure on R has the property that for every continuous compactly supported function f on G we have

$$\int_G f(g)\,\mathrm{d}\mu_r^G(g) = \int_R \int_H f(hg)\frac{\Delta_G(h)}{\Delta_H(h)}\,\mathrm{d}\mu_r^H(h)\,\mathrm{d}\mu^R(g).$$

Using this measure we can define the space $L^2(R)$ of all square integrable functions. Whenever we are given a function $f \in L_r^1(G)$ we have that

$$\tilde{f}(Hx) = \int_H f(hx)\frac{\Delta_G(h)}{\Delta_H(h)}\,\mathrm{d}\mu_r^H(h)$$

is in $L^1(R)$.

Given a section $s : U \subset R \to G$ define the function $h_s : R \times G \to H$ by

$$s(x)g = h_s(x,g)s(xg).$$

The function h_s has the additional property that

$$h_s(x, g_1g_2) = h_s(x, g_1)h_s(xg_1, g_2).$$

Let $(\pi, \mathcal{H})$ be a representation of the group H. Then we define the induced representation acting on $L^2(R, \mathcal{H})$ by

$$(\mathrm{ind}_H^G\pi(g)f)(x) = \sqrt{\frac{\Delta_H(h_s(x,g))}{\Delta_G(h_s(x,g))}}\pi(h_s(x,g))f(x\cdot g).$$

An important special case is when π is a character, i.e., a representation acting on the Hilbert space $\mathbb{C}$. Let $F \in \mathfrak{h}^*$. Then when H is an exponential group we have that

$$\pi(h) = e^{-2\pi i\langle F, \log(h)\rangle}$$

is a character. Then the induced representation acts on $L^2(R)$ by

$$(\mathrm{ind}_H^G\pi(g)f)(x) = \sqrt{\frac{\Delta_H(h_s(x,g))}{\Delta_G(h_s(x,g))}}e^{-2\pi i\langle F, \log_H(h_s(x,g))\rangle}f(x\cdot g).$$

Semi-direct product

In the case that G can be written as the semi-direct product of R and H, the formulas for induced representation simplify. The definition of semi-direct product relies on the existence of a group homomorphism $\phi : R \to \mathrm{Aut}(H)$ such that the product on $G = R \times H$ is given by

$$(r, h) \cdot (s, j) = (rs, h\phi(r)j).$$

We will denote the group endowed with the product given by the semi-direct product with respect to ϕ by $R \rtimes_\phi H$. Then R can be identified with the right cosets $H \setminus G$. Notice that

$$(e_R, \phi(a)y) = (a, e_H)(e_R, y)(a, e_H)^{-1} = \Psi_{(a,e_H)}(e_R, y).$$

Since H is a normal subgroup, we have that $\Delta_H = \Delta_G$. The semi-direct product have the structure of a quotient group with the right Haar-measure given by

$$\mathrm{d}\mu_r^G = \mathrm{d}\mu_r^R \, \mathrm{d}\mu_r^H.$$

Set $s : R \to G$ to be $a \mapsto (a, e_H)$. Then $h_s(a, (b, y)) = (e_R, \phi(a)y)$. We have that $\Delta_H = \Delta_G$, hence the induced representation simplifies to

$$(\mathrm{ind}_H^G \pi((b, y)) f)(a) = \pi(\phi(a)y) f(a \cdot b).$$

If $(\pi, \mathbb{C})$ is the character

$$\pi(h) = e^{-2\pi i F(\log(h))}$$

defined on H, then the induced representation is given by

$$\begin{aligned}(\mathrm{ind}_H^G \pi((b, y)) f)(a) &= e^{-2\pi i F \log_G(\phi(a)y)} f(a \cdot b) \\ &= e^{-2\pi i F \log_H(\phi(a)y)} f(a \cdot b) \qquad \text{(C.2.9)}\end{aligned}$$

acting on $L^2(R, d\mu_r^R)$. The modular function is given by

$$\Delta_G(a, x) = \frac{1}{|\det_H(\phi(a)_*)|} \Delta_R(a) \Delta_H(x) = \frac{1}{|\det_H(\mathrm{Ad}_a)|} \Delta_R(a) \Delta_H(x).$$

In the case that H is unimodular, we have that $\Delta_G(a, x) = \Delta_G(a, e_H)$. In this case, we set

$$(\mathcal{D}^{-1}\phi)(r) = \sqrt{\Delta_G(r, e_H)} \phi(r) = \frac{\sqrt{\Delta_R(r)}}{\sqrt{|\det_H(\mathrm{Ad}_r)|}} \phi(r).$$

Kirillov co-adjoint orbit theory

The objective of Kirillov orbit theory is to show how the co-adjoint orbits and the representations are linked. The Kirillov orbit method gives us an explicit way to construct all irreducible representation on certain Lie groups. As the name implies, the method states that for every co-adjoint orbit there exists a unique, up to equivalence, irreducible representation associated to it. Moreover, the theory states that all irreducible representations can be constructed in such a way. We will now give a short introduction of how the irreducible representations are constructed from the orbits. For a complete overview, see [18, 134].

Associated to the adjoint map we can define the *co-adjoint map* $K(g) : \mathfrak{g} \to \mathfrak{g}$ by the equation

$$\langle K(g)F, X\rangle = \langle F, \mathrm{Ad}_{g^{-1}}X\rangle.$$

We will denote the stabilizer of this map by $\mathrm{Stab}(F) = \{g \in G : K(g)F = F\}$. The derivative of the co-adjoint map is given by the equation

$$\langle K_*(X)F, Y\rangle = \langle F, -\mathrm{ad}_X Y\rangle.$$

Given a F we denote the *co-adjoint orbit* by $O_F = K(G)F \subset \mathfrak{g}$. We can identify the orbit with the quotient $\mathrm{Stab}(F) \setminus G = \{\mathrm{Stab}(F)g : g \in G\}$ by using the map $\kappa : \mathrm{Stab}(F) \setminus G \to O_F$ defined as $\kappa(x) = K(x^{-1})F$, where $x \in \mathrm{Stab}(F) \setminus G$. In the case that $\mathrm{Stab}(F) = \{e\}$, the orbit can be identified with the original group G. In this case, we will denote the inverse of κ by $\kappa^{-1} : O_F \to G$.

Symplectic structure of the orbit

We will take a *symplectic space* to mean a (smooth) manifold M equipped with a closed non-degenerate 2-form $\omega : \wedge^2(T^*M) \to \mathbb{R}$. The form ω is referred to as the *symplectic form*. All symplectic manifolds are even dimensional, hence we will denote the dimension of M by $2d$. The simplest example of a symplectic manifold is $\mathbb{R}^{2d}$ endowed with the standard symplectic form

$$\omega_{(x_1,\ldots,x_d,y_1,\ldots,y_d)} = \sum_{i=1}^{d} \mathrm{d}x_i \wedge \mathrm{d}y_i,$$

where $(x_1, \ldots, x_d, y_1, \ldots, y_d) \in \mathbb{R}^{2d}$ are the standard coordinates. The *volume form* on a symplectic manifold is given by

$$\mathrm{d}\omega = \frac{\omega^d}{d!}.$$

A diffeomorphism $f : M_1 \to M_2$ between two symplectic manifolds (M_j, ω_j) is called a *symplectomorphism* if $f^*(\omega_2) = \omega_1$, where f^* denotes the pullback of the function f. If such a map f exists between two symplectic manifolds, we say that the manifolds are *symplectomorphic*. Darboux's theorem states that every symplectic manifold of dimension $2d$ is locally symplectomorphic to $\mathbb{R}^{2d}$ endowed with the standard symplectic form.

Let $N \subset M$ be a submanifold of the symplectic manifold (M, ω). Then we say that N is *isotropic* if $\omega|_{TN} = 0$. A *Lagrangian manifold* is an isotropic manifold of maximal dimension $\dim(N) = \frac{1}{2}\dim(M)$. Alternatively, we can define a *Lagrangian manifold* to be an isotropic manifold where $\omega_{TN^\perp} = 0$. We define a *Lagrangian fibration* to be a fibration where all the fibers are Lagrangian manifolds. Locally, any Lagrangian filtration can be written as $(x_1, \ldots, x_n, y_1, \ldots, y_n)$ where the symplectic form can be written as

$$\omega_{(x_1,\ldots,x_n,y_1,\ldots,y_n)} = \sum_{j=1}^{n} \mathrm{d}x_j \wedge \mathrm{d}y_j .$$

One can induce a symplectic form on the co-adjoint orbit $\mathcal{O}_F \sim \mathrm{Stab}(F) \setminus G$ as follows: Define the symplectic form at the point $F \in \mathcal{O}_F$ by

$$\omega_F(K_*(F)X, K_*(F)Y) = \langle F, [X, Y] \rangle,$$

where $X, Y \in \mathfrak{g}/\mathrm{stab}(F)$ that can be identified with $T_F^*\mathcal{O}_F$. We can define the symplectic form on the rest of the $T^*\mathcal{O}_F$ by using the group action on the orbit. Then the symplectic form is defined by $\omega_{\kappa(x)} = \kappa(x)^*\omega_F$. Hence by definition, the symplectic form becomes right invariant with respect to the group action.

A *invariant complex-polarization* with respect to ω_F is a subalgebra $\mathfrak{h}$ of $\mathfrak{g} \otimes \mathbb{C}$ such that

- $\mathfrak{h}$ is a Lagrangian subspace of ω_F.
- $\mathfrak{h} + \overline{\mathfrak{h}}$ is a subalgebra of $\mathfrak{g} \otimes \mathbb{C}$.
- $\mathrm{Ad}(s)\mathfrak{h} = \mathfrak{h}$ for every $s \in \mathrm{Stab}(F)$.

Recall that we say that a subalgebra $\mathfrak{k}$ of $\mathfrak{g} \otimes \mathbb{C}$ is real if

$$\mathfrak{k} = \overline{\mathfrak{k}}.$$

We let $\mathfrak{e}$ denote the largest real subalgebra of $\mathfrak{g}$, i.e.

$$\mathfrak{e} = (\mathfrak{h} + \overline{\mathfrak{h}}) \cap \mathfrak{g}.$$

Additionally, we define the subalgebra $\mathfrak{d}$ to be the

$$\mathfrak{d} = (\mathfrak{h} \cap \overline{\mathfrak{h}}) \cap \mathfrak{g}.$$

Notice that in the case that $\mathfrak{h}$ is a real subalgebra, we have that

$$\mathfrak{e} = \mathfrak{d}.$$

Definition C.2.8. We say that an invariant complex polarization satisfies the Pukanszky condition if

$$K(\exp(\mathfrak{d}))F = F + \mathrm{An}(\mathfrak{e}),$$

where $\mathrm{An}(\mathfrak{e})$ denotes the annihilator of $\mathfrak{e}$.

Representations from co-adjoint orbits

In this section we will let G be a exponential Lie group. Let $\mathcal{O}$ be a co-adjoint orbit containing a point $F \in \mathcal{O}$. Find a subgroup of maximal dimension H with Lie algebra $\mathfrak{h}$ such that $\langle F, [\mathfrak{h}, \mathfrak{h}]\rangle = 0$. It can be shown that $\dim(H \setminus G) = \frac{1}{2}\dim(\mathcal{O})$. Define the character on H by $\chi_F(h) = e^{2\pi i\langle F, \log(h)\rangle}$. Then the induced representation

$$\pi_F = \mathrm{ind}_H^G \chi_F$$

is an irreducible representation on G. In general, these representations are only square integrable when $\mathrm{Stab}(F)$ consists of only the identity element.

Example C.2.9 (Heisenberg Group)**.** Continuing with the notation from Example C.2.5, let X^*, Y^* and Z^* be the dual basis of X, Y and Z. Then we have that the co-adjoint map satisfies

$$K(x, y, z)(aX^* + bY^* + cZ^*) = (a + yc)X^* + (b - xc)Y * +cZ^*.$$

From the co-adjoint map, we see that we have two kinds of orbits. Setting $F = \hbar Z^*$ for $\hbar \neq 0$ gives the two-dimensional orbits

$$\mathcal{O}_\hbar = \{aX^* + bY^* + \hbar Z^* : a, b \in \mathbb{R}\}.$$

When $F = aX^* + bY^*$ we get the zero-dimensional orbits

$$\mathcal{O}_{(a,b)} = \{aX^* + bY^* : a, b \in \mathbb{R}\}.$$

The representations corresponding to the orbits $\mathcal{O}_\hbar$ is $\pi_{-\hbar}$. Since the $\mathrm{Stab}(\hbar Z^*) = (0, 0, z)$, this representation is not square integrable.

Example C.2.10 (Affine Group)**.** The affine group has two two-dimensional orbits. Continuing with the notation from Example C.2.6, let U^* and V^* be the dual basis to the Lie algebra consisting of U and V. The co-adjoint map is given by

$$K(a, x)(uU^* + vV^*) = (u + a^{-1}xv)U^* + a^{-1}vV^*.$$

This means that the orbits are given by

$$\mathcal{O}_+ = \{K(a, x)V^* : (a, x) \in \mathrm{Aff}\} = \{uU^* + vV^* : v \in \mathbb{R},\ u > 0\}$$

and

$$\mathcal{O}_- = \{K(a,x)(-V^*) : (a,x) \in \mathrm{Aff}\} = \{uU^* + vV^* : v \in \mathbb{R},\ u < 0\}.$$

The affine group has the structure of semi-direct product between $R = (\mathbb{R}^+, \cdot)$ and $H = (\mathbb{R}, +)$ with $\phi(a)x = a \cdot x$. Using Equation C.2.9, we see that the orbit $\mathcal{O}_\pm$ with $F = \pm V^*$ gives the representations $\pi_\pm$.

Example C.2.11 (Shearlet Group)**.** The shearlet group has two four-dimensional orbits. The dual basis is then denoted by A^*, B^*, C^* and D^*. The co-adjoint map is given by

$$\begin{aligned} K(a,s,x_1,x_2)(\alpha A^* + \beta B^* + \gamma C^* + \delta D^*) = (\alpha + \frac{\beta s}{2\sqrt{a}} + \gamma\frac{2x_1 - sx_2}{2a} + \frac{\delta x_2}{2\sqrt{a}})A^* \\ + (\frac{\beta}{\sqrt{a}} + \frac{\gamma x_2}{a})B^* + \frac{\gamma}{a}C^* - (\frac{\gamma s}{a} - \frac{\delta}{\sqrt{a}})D^*. \end{aligned}$$

This means that the orbits are given by

$$\mathcal{O}_+ = \{K(x)C^* : x \in \mathbb{S}\} = \{\alpha A^* + \beta B^* + \gamma C^* + \delta D^* : \alpha,\beta,\delta \in \mathbb{R},\ \gamma > 0\}$$

and

$$\mathcal{O}_- = \{K(x)(-C^*) : x \in \mathbb{S}\} = \{\alpha A^* + \beta B^* + \gamma C^* + \delta D^* : \alpha,\beta,\delta \in \mathbb{R},\ \gamma < 0\}.$$

Let $H = \exp(\mathrm{span}(D,C))$. We can describe the group as a semi-direct product: Define $\phi.: (\mathbb{R}^+ \times \mathbb{R}) \to \mathrm{Aut}(\mathbb{R}^2)$ defined by $\phi_{(a,s)} = S_s A_a$. Then we can define $(\mathbb{R}^+ \times \mathbb{R}) \ltimes_\phi \mathbb{R}^2$. Choosing $F = \pm C^*$ then gives the induced representations $\pi_\pm$ from Example C.2.7.

C.2.5 Time-frequency analysis

Time-frequency analysis and its subfields are central pieces of applied harmonic analysis and have a strong connection to Weyl quantization. The (projective) representation $\pi(z)\psi(t) = \pi(x,\omega)\psi(t) = e^{2\pi i \omega t}\psi(t-x)$ used in Section C.2.1 induces the main object of time-frequency analysis, the *short-time Fourier transform*, defined as

$$V_\phi\psi(z) = \langle \psi, \pi(z)\phi \rangle. \tag{C.2.10}$$

Note that this quantity is a special case of a wavelet transform (C.2.3) if we disregard the $\pi(z)^*$ convention. In applications, it is often the squared modulus $|V_\phi\psi|^2$ of the short-time Fourier transform, the *spectrogram*, that is used because it is non-negative. In view of Theorem C.2.3 it also integrates to 1 and so can be seen as a probability distribution on $\mathbb{R}^{2d}$.

As time-frequency analysis is only tangentially related to the subject at hand, we only give a short primer on those topics which will be needed later on. A standard reference for time-frequency analysis in which much of the content of this section can be found is [107].

We already saw the cross-Wigner distribution $W(\psi, \phi)$ in Section C.2.1 as the dequantization of a rank-one operator. In time-frequency analysis, the cross-Wigner distribution is known as a time-frequency distribution related to short-time Fourier transform by

$$W(\psi, \phi)(x, \omega) = 2^d e^{4\pi i x\cdot\omega} V_{\check{\phi}}\psi(2x, 2\omega).$$

The Wigner distribution also has an inversion formula and is uniquely determined by the window up to a constant, i.e., if $W(\psi) = W(\phi)$ then $\psi = c\phi$ where $|c| = 1$.

The more general form of the short-time Fourier transform, the wavelet transform $\mathcal{W}_\phi$ mentioned in (C.2.3), is the typical generalization of time-frequency analysis. To define is, we need a locally compact group G, a Hilbert space $\mathcal{H}$ and a square integrable irreducible representation $\pi : G \to \mathcal{U}(\mathcal{H})$. The most classical incarnation of this setup, apart from time-frequency analysis, is *time-scale analysis* on the *affine group*, discussed in Section C.2.3. In time-scale analysis, the squared modulus of the wavelet transform is called the *scalogram* and is used similarly to the spectrogram. Other classical generalizations include the shearlet transform [111, 144] and various two-dimensional wavelet transforms such as the similitude transform [14, 16].

C.2.6 Quantum harmonic analysis

The theory of *quantum harmonic analysis* (QHA), originally developed by R. Werner in 1984 [211], lays out a framework in which many of the classical operations and results of harmonic analysis is generalized to operators. Specifically, convolutions are defined between functions on $\mathbb{R}^{2d}$ and operators on $L^2(\mathbb{R}^d)$, as well as pairs of operators on $L^2(\mathbb{R}^d)$, as

$$f \star S = \int_{\mathbb{R}^{2d}} f(z)\pi(z)S\pi(z)^*\,dz, \quad T \star S(z) = \operatorname{tr}\left(T\pi(z)PSP\pi(z)^*\right) \tag{C.2.11}$$

where $\pi(z)f(t) = \pi(x, \omega)f(t) = e^{2\pi i\omega\cdot t}f(t - x)$ is the projective representation of the Weyl-Heisenberg group from Section C.2.5 and $P : f \mapsto \check{f}$ is the parity operator. Together with the Fourier-Wigner transform $\mathcal{F}_W : \mathcal{S}^1 \to C_0(\mathbb{R}^{2d})$, from Section C.2.1, defined as $\mathcal{F}_W(S)(z) = e^{-i\pi x\cdot\omega}\operatorname{tr}(\pi(-z)S)$, they make up the main tools of QHA. Both convolution definitions in (C.2.11) conjugate the operator S by

$\pi(z)$, this operation is called an *operator translation* and is commonly denoted by $\alpha_z(S) = \pi(z)S\pi(z)^*$.

Remark C.2.12. Note that the above definitions are not consistent with the rest of this article but rather agree with those commonly found in the greater quantum harmonic analysis literature. When generalizing beyond the Weyl-Heisenberg group it is beneficial to adopt some alternative formulations. We will detail these and their motivation in Section C.2.6 below.

Standard properties

Below we collect some of the main properties of operator convolutions, all of which have counterparts in harmonic analysis. For proofs and more details, see [153].

Proposition C.2.13. Let $f, g \in L^1(\mathbb{R}^{2d})$, $S \in \mathcal{S}^p$, $T \in \mathcal{S}^q$ for $1 \leq p, q, r \leq \infty$ with $\frac{1}{p} + \frac{1}{q} = 1 + \frac{1}{r}$ and $R \in B(L^2)$ be compact, then

(i) $f \star S$ is positive if f is non-negative and S is positive,

(ii) $T \star S$ is non-negative if T and S are positive,

(iii) $(f \star S)^* = \bar{f} \star S^*$,

(iv) $(f \star S) \star T = f * (S \star T)$,

(v) $(f * g) \star S = f \star (g \star S)$,

(vi) $f \star R$ is compact,

(vii) $\|f \star S\|_{\mathcal{S}^p} \leq \|f\|_{L^1} \|S\|_{\mathcal{S}^p}$,

(viii) $\|T \star S\|_{L^r} \leq \|S\|_{\mathcal{S}^p} \|T\|_{\mathcal{S}^q}$.

To formulate Fourier-analytic results we need a Fourier transform which works on the phase space $\mathbb{R}^{2d}$. This will be the same symplectic Fourier transform as in Section C.2.1, namely

$$\mathcal{F}_\sigma(f)(z) = \int_{\mathbb{R}^{2d}} f(z')e^{-2\pi i \sigma(z,z')}\, dz', \tag{C.2.12}$$

where $z = (x, \omega)$, $z' = (x', \omega')$ and $\sigma(z, z') = \omega x' - \omega' x$ is the standard symplectic form. The correct setting to investigate Fourier-related properties is requiring all operators to be of trace-class and here we have some more noteworthy results.

Proposition C.2.14. Let $f \in L^1(\mathbb{R}^{2d})$ and $T, S \in \mathcal{S}^1$, then

(i) $\mathcal{F}_W(f \star S) = \mathcal{F}_\sigma(f)\mathcal{F}_W(S)$,

(ii) $\mathcal{F}_\sigma(T \star S) = \mathcal{F}_W(T)\mathcal{F}_W(S)$,

(iii) $\operatorname{tr}(f \star S) = \operatorname{tr}(S) \int_{\mathbb{R}^{2d}} f(z)\, dz$,

(iv) $\int_{\mathbb{R}^{2d}} T \star S(z)\, dz = \operatorname{tr}(T)\operatorname{tr}(S)$.

These results should be seen as generalizations and consequences of the standard convolution theorem.

Part of the power of quantum harmonic analysis is that various seemingly unrelated objects can be realized as operator convolutions where their properties have clear explanations. Examples of this include localization operators and Cohen's class distributions from time-frequency analysis [153, 154] (Section C.2.5) and Bergman-Fock Toeplitz operators [92, 156].

Relation to Weyl quantization

As quantum harmonic analysis is concerned with the interplay between functions and operators, it should come as no surprise that the theory is intimately connected with that of Weyl quantization. In fact, one can use Weyl quantization to induce all of the main operations of QHA by noting that

$$T \star S(z) = a_T * a_S(z), \qquad f \star S = A_{f * a_S}. \tag{C.2.13}$$

Moreover, Weyl quantization can be written explicitly as

$$a_S = \mathcal{F}_\sigma(\mathcal{F}_W(S)) \tag{C.2.14}$$

when $S \in \mathcal{S}^2$, so the two central Fourier transforms of QHA make up the core of the Weyl quantization procedure. From (C.2.13) and (C.2.14), all of Proposition C.2.14 follow once we note that

$$\int_{\mathbb{R}^{2d}} a_S(z)\, dz = \mathcal{F}_\sigma(a_S)(0) = \mathcal{F}_\sigma(\mathcal{F}_\sigma(\mathcal{F}_W(S)))(0) = \mathcal{F}_W(S)(0) = \operatorname{tr}(S).$$

The action of translating an operator using the operator translation $\alpha_z : S \mapsto \pi(z)S\pi(z)^*$ which is used in quantum harmonic analysis can also be seen as being induced through Weyl quantization via the relation

$$\alpha_z(S) = A_{T_z a_S}$$

from Proposition C.2.2, where $T_z f(x) = f(x - z)$ is the translation operator.

Generalizing beyond the Weyl-Heisenberg group

The setting of $L^2(\mathbb{R}^d)$ as the Hilbert space on which operators act and $\pi : \mathbb{R}^{2d} \to \mathcal{U}(L^2(\mathbb{R}^d))$ as the associated representation has underpinned our discussion of QHA so far. The work on generalizing this beyond the Weyl-Heisenberg group started in [32] where $\mathbb{R}^{2d}$ was replaced by the affine group Aff, the Hilbert space $L^2(\mathbb{R}^d)$ specialized to $L^2(\mathbb{R}^+)$ and the irreducible representation set to

$$U : \mathrm{Aff} \to \mathcal{U}(L^2(\mathbb{R}^+)), \qquad U(x,a)\psi(t) = e^{2\pi i x\cdot t}\psi(at).$$

In this setting, a suitable Wigner transform was also developed using the work in [31] which required the introduction of a suitable replacement of the symplectic Fourier transform, namely the *Fourier-Kirillov transform* which acts on Aff. Later on in [115], using similar techniques, the setup was further generalized to allow for irreducible representations π of (separable) locally compact groups G. In this setting and the affine, the function-operator and operator-operator convolutions take the following form with the right Haar measure convention:

$$f \star S = \int_G f(x)\pi(x)^* S\pi(x)\,\mathrm{d}\mu_r(x), \qquad T \star S(x) = \mathrm{tr}\left(T\pi(x)^* S\pi(x)\right).$$

For these convolutions, variants of Proposition C.2.13 and C.2.14 hold with the exception of the two Fourier properties. QHA was also generalized to the non-separable abelian setting in [91].

It turns out that most of the difficulties related to generalizing QHA stem from the possible non-unimodularity of the underlying group. Specifically, this leads to the introduction of the concept of *admissible* operators which are operators S on $\mathcal{H}$ such that $\mathcal{D}^{-1}S\mathcal{D}^{-1}$ is trace-class. One central result is that an operator-operator convolution only is integrable if one of the operators is trace-class and the other is admissible as is clear from the formula

$$\int_G T \star S(x)\,\mathrm{d}\mu_r(x) = \mathrm{tr}(T)\,\mathrm{tr}(\mathcal{D}^{-1}S\mathcal{D}^{-1}). \tag{C.2.15}$$

C.3 Two Fourier Transforms

We are now ready to set up the Fourier-Wigner and Fourier-Kirillov transform alluded to in the introduction. Throughout the section we will assume that $\mathrm{Stab}(F) = \{e\}$ of G, and hence the co-adjoint orbit have the same dimension as the group. We will also assume that H is a closed subgroup inducing the representation described in Section C.2.4.

C.3.1 Fourier-Wigner transform

The Fourier-Wigner transform defined in this section is the Fourier transform in the Plancherel theorem, see e.g. [87, Thm. 3.48], restricted to the representation π. For convenience, we give the proofs of properties we will need in later sections.

Let $\pi_F : G \to \mathcal{H}$ be the representation associated to the point $F \in \mathfrak{g}^*$ discussed in Section C.2.4, where $\dim(O_F) = \dim(G)$. From now on, we will drop the F subscript.

Definition C.3.1 (Fourier-Wigner transform)**.** Let $A \in \mathcal{S}^2$ be such that $A\mathcal{D}$ extends to a trace class operator. We then define the *Fourier-Wigner transform* $\mathcal{F}_{\mathrm{W}}(A)$ as

$$\mathcal{F}_{\mathrm{W}}(A)(x) = \mathrm{tr}(A\mathcal{D}\pi(x)).$$

If $f \in L^1_r(G)$ and $\phi \in \mathrm{Dom}(\mathcal{D})$, then we define the *inverse Fourier-Wigner transform* by

$$\mathcal{F}_{\mathrm{W}}^{-1}(f)(\phi) = (\pi(\check{f}) \circ \mathcal{D})\phi = \int_G f(x)\pi(x^{-1})\mathcal{D}\phi \, \mathrm{d}\mu_r(x).$$

We will now show that the Fourier-Wigner transform extends to an injective isometry from $\mathcal{S}^2(\mathcal{H}) \to L^2_r(G)$, with the additional property that the inverse Fourier-Wigner transform is a left inverse. To do so, we first need a preliminary result.

Lemma C.3.2. *The space $\mathcal{S}^1\mathcal{D}^{-1} \cap \mathcal{S}^2$ is a dense subspace of $\mathcal{S}^2$.*

Proof. Let $A = \sum_n a_n(\psi_n \otimes \phi_n)$ be an arbitrary operator in $\mathcal{S}^2$, fix $\varepsilon > 0$ and define

$$A_N = \sum_{n=1}^N a_n(\psi_n \otimes \phi_n), \qquad X_N = \sum_{n=1}^N a_n(\psi_n \otimes \mathcal{D}^{-1}\xi_n)$$

where $\xi_n = \mathcal{D}\phi_n'$ and ϕ_n' are functions in the domain of $\mathcal{D}$ which satisfy $\|\phi_n - \phi_n'\| < \varepsilon$ for each n. This is possible because the domain of $\mathcal{D}$ is dense in $\mathcal{H}$ by Theorem C.2.3. Since $A \in \mathcal{S}^2$, there exists a N such that $\|A - A_N\|_{\mathcal{S}^2} < \varepsilon$. We now claim that $\|A - X_N\|_{\mathcal{S}^2} < \varepsilon(1 + \|A\|_{\mathcal{S}^2})$, which can be made arbitrarily small, and that $X_N\mathcal{D} \in \mathcal{S}^1$. Indeed, $X_N\mathcal{D} \in \mathcal{S}^1$ since

$$X_N\mathcal{D} = \sum_{n=1}^N a_n(\psi_n \otimes \xi_n)$$

is a finite sum of bounded rank-one operators and $\psi_n, \xi_n \in \mathcal{H}$. Moreover, $\|X_N - A_N\|_{\mathcal{S}^2}$ can be bounded as

$$\begin{aligned}\|X_N - A_N\|_{\mathcal{S}^2} &\leq \left(\sum_{n=1}^{N} |a_n|^2 \|\psi_n\| \big\|\mathcal{D}^{-1}\xi_n - \phi_n\big\|\right)^{1/2} = \left(\sum_{n=1}^{N} |a_n|^2 \big\|\phi_n' - \phi_n\big\|\right)^{1/2} \\ &\leq \varepsilon \|A_N\|_{\mathcal{S}^2} \leq \varepsilon \|A\|_{\mathcal{S}^2},\end{aligned}$$

finishing the proof. □

Proposition C.3.3. Let $\psi \in \mathcal{H}$ and $\phi \in \mathrm{Dom}(\mathcal{D}^{-1})$. Then

$$\mathcal{F}_{\mathrm{W}}(\psi \otimes \mathcal{D}^{-1}\phi)(x) = \mathcal{W}_\phi \psi(x).$$

Proof. Plugging in the definition of $\mathcal{F}_{\mathrm{W}}$, we find that

$$\begin{aligned}\mathcal{F}_{\mathrm{W}}(\psi \otimes \mathcal{D}^{-1}\phi)(x) &= \mathcal{F}_{\mathrm{W}}((\psi \otimes \phi)\mathcal{D}^{-1})(x) \\ &= \operatorname{tr}\big((\psi \otimes \phi)\pi(x)\big) \\ &= \langle \psi, \pi(x)^* \phi \rangle.\end{aligned}$$

□

Using linearity, we can use this last result to define $\mathcal{F}_{\mathrm{W}}$ on all of $\mathcal{S}^1\mathcal{D}^{-1}$ using the singular value decomposition. In the same way, we will now show how $\mathcal{F}_{\mathrm{W}}$ can be further extended to all of $\mathcal{S}^2$ using Lemma C.3.2.

Proposition C.3.4. The Fourier-Wigner transform $\mathcal{F}_{\mathrm{W}} : \mathcal{S}^1\mathcal{D}^{-1} \to L^2_r(G)$ can be continuously extended to all of $\mathcal{S}^2$ as an injective isometry meaning that

$$\|A\|_{\mathcal{S}^2} = \|\mathcal{F}_{\mathrm{W}}(A)\|_{L^2_r}$$

for all $A \in \mathcal{S}^2$.

Proof. We will show that $\mathcal{F}_{\mathrm{W}}$ is an isometry on $\mathcal{S}^1\mathcal{D}^{-1} \cap \mathcal{S}^2$ which is dense in $\mathcal{S}^2$ by Lemma C.3.2. Let $A \in \mathcal{S}^1\mathcal{D}^{-1}$ so that we can write

$$A = \sum_n s_n(\psi_n \otimes \phi_n)$$

where $\phi_n \in \mathrm{Dom}(\mathcal{D})$ and both $(\psi_n)_n, (\phi_n)_n$ are orthonormal sequences in $\mathcal{H}$. We can now use Proposition C.3.3 to conclude that,

$$\begin{aligned}\|\mathcal{F}_{\mathrm{W}}(A)\|_{L_r^2}^2 = \langle \mathcal{F}_{\mathrm{W}}(A), \mathcal{F}_{\mathrm{W}}(A)\rangle &= \sum_n \sum_m s_n \overline{s_m} \langle \mathcal{F}_{\mathrm{W}}(\psi_n \otimes \phi_n), \mathcal{F}_{\mathrm{W}}(\psi_m \otimes \phi_m)\rangle \\ &= \sum_n \sum_m s_n \overline{s_m} \langle \mathcal{W}_{\mathcal{D}\phi_n}\psi_n, \mathcal{W}_{\mathcal{D}\phi_m}\psi_m\rangle \\ &= \sum_n \sum_m s_n \overline{s_m} \langle \psi_n, \psi_m\rangle \overline{\langle \phi_n, \phi_m\rangle} \\ &= \sum_n |s_n|^2 \|\phi_n\|^2 \|\psi_n\|^2 = \sum_n |s_n|^2 = \|A\|_{\mathcal{S}^2}^2\end{aligned}$$

where we made use of the Duflo-Moore theorem (Theorem C.2.3). □

The inverse Fourier-Wigner transform $\mathcal{F}_{\mathrm{W}}^{-1}$ is a left inverse to $\mathcal{F}_{\mathrm{W}}$ as we show in this next proposition.

Proposition C.3.5. The composition $\mathcal{F}_{\mathrm{W}}^{-1} \circ \mathcal{F}_{\mathrm{W}}$ is the identity operator on $\mathcal{S}^2$.

Proof. We will prove the proposition for $\mathcal{S}^1\mathcal{D}^{-1}$ and conclude the full case by the density from Lemma C.3.2. Let $A, B \in \mathcal{S}^1\mathcal{D}^{-1}$ be arbitrary, then with $A = \sum_n a_n(\psi_n \otimes \mathcal{D}^{-1}\phi_n)$ and $B = \sum_n b_n(\xi_n \otimes \mathcal{D}^{-1}\eta_n)$, we have that

$$\langle \mathcal{F}_{\mathrm{W}}^{-1}(\mathcal{F}_{\mathrm{W}}(A)), B\rangle_{\mathcal{S}^2} = \sum_n \sum_m a_n \overline{b_m} \langle \mathcal{F}_{\mathrm{W}}^{-1}(\mathcal{W}_{\phi_n}\psi_n), (\xi_m \otimes \mathcal{D}^{-1}\eta_m)\rangle_{\mathcal{S}^2} \tag{C.3.1}$$

by Proposition C.3.3. We can now compute each term as

$$\begin{aligned}\langle \mathcal{F}_{\mathrm{W}}^{-1}(\mathcal{W}_\phi\psi), (\xi \otimes \mathcal{D}^{-1}\eta)\rangle_{\mathcal{S}^2} &= \mathrm{tr}\left(\mathcal{F}_{\mathrm{W}}^{-1}(\mathcal{W}_\phi\psi)(\mathcal{D}^{-1}\eta \otimes \xi)\right) \\ &= \sum_n \langle \mathcal{F}_{\mathrm{W}}^{-1}(\mathcal{W}_\phi\psi)\langle e_n, \xi\rangle \mathcal{D}^{-1}\eta, e_n\rangle \\ &= \langle \mathcal{F}_{\mathrm{W}}^{-1}(\mathcal{W}_\phi\psi)\mathcal{D}^{-1}\eta, \xi\rangle \\ &= \int_G \langle \psi, \pi(x)\phi\rangle \langle \pi(x)\mathcal{D}\mathcal{D}^{-1}\eta, \xi\rangle \,\mathrm{d}\mu_l(x) \\ &= \int_G \mathcal{W}_\phi\psi(x^{-1})\overline{\mathcal{W}_\eta\xi(x^{-1})} \,\mathrm{d}\mu_l(x) \\ &= \int_G \mathcal{W}_\phi\psi(x)\overline{\mathcal{W}_\eta\xi(x)} \,\mathrm{d}\mu_r(x) \\ &= \langle \psi, \xi\rangle \overline{\langle \mathcal{D}^{-1}\phi, \mathcal{D}^{-1}\eta\rangle}\end{aligned}$$

using Theorem C.2.3. Plugging this result into each term of (C.3.1), we conclude that

$$\left\langle \mathcal{F}_{\mathrm{W}}^{-1}(\mathcal{F}_{\mathrm{W}}(A)), B\right\rangle_{\mathcal{S}^2} = \sum_n \sum_m a_n \overline{b_m} \langle \psi_n, \eta_m\rangle \overline{\langle \mathcal{D}^{-1}\phi_n, \mathcal{D}^{-1}\eta_m\rangle} = \langle A, B\rangle_{\mathcal{S}^2}$$

which implies that $\mathcal{F}_{\mathrm{W}}^{-1} \circ \mathcal{F}_{\mathrm{W}}$ is the identity on $\mathcal{S}^1\mathcal{D}^{-1}$. □

The following is as close as we are able to get to a Riemann-Lebesgue lemma for $\mathcal{F}_{\mathrm{W}}$ as decay at infinity is not always possible to define for a group.

Proposition C.3.6. If $A \in \mathcal{S}^1\mathcal{D}^{-1}$, then $\mathcal{F}_{\mathrm{W}}(A)$ is continuous. Moreover, if for all $\varepsilon > 0$ and $\psi, \phi \in \mathcal{H}$ there exist sets $E(\varepsilon, \psi, \phi)$ such that

$$x \in E(\varepsilon, \psi, \phi) \implies |\mathcal{W}_\phi \psi(x)| < \varepsilon,$$

then given an $\varepsilon > 0$, there exists a set $E(\varepsilon, A)$ which is a finite intersection of sets of the form $E(\varepsilon', \psi, \phi)$ such that

$$x \in E(\varepsilon, A) \implies |\mathcal{F}_{\mathrm{W}}(A)(x)| < \varepsilon.$$

Proof. For continuity, let $x \to x_0$ and estimate

$$\begin{aligned}\left|\mathcal{F}_{\mathrm{W}}(A)(x) - \mathcal{F}_{\mathrm{W}}(A)(x_0)\right| &= \left|\operatorname{tr}\left(A\mathcal{D}\pi(x) - A\mathcal{D}\pi(x_0)\right)\right| \\ &\leq \|A\mathcal{D}\|_{\mathcal{S}^1}\|\pi(x) - \pi(x_0)\|_{\mathcal{S}^\infty}\end{aligned}$$

using $\|AB\|_{\mathcal{S}^1} \leq \|A\|_{\mathcal{S}^1}\|B\|_{\mathcal{S}^\infty}$. The last factor goes to zero by the strong continuity of the representation.

For the decay, fix $\varepsilon > 0$ and decompose $A\mathcal{D}$ as $A\mathcal{D} = \sum_n a_n(\psi_n \otimes \phi_n)$ where $(\psi_n)_n$ and $(\phi_n)_n$ are orthonormal sequences. Since $(a_n)_n \in \ell^1$, we can find an integer $N > 0$ such that $\sum_{n=N+1}^\infty |a_n| < \varepsilon/2$ and therefore,

$$\left|\mathcal{F}_{\mathrm{W}}(A)(x)\right| \leq \sum_{n=1}^N |a_n|\left|\mathcal{W}_{\phi_n}\psi_n(x)\right| + \underbrace{\sum_{n=N+1}^\infty |a_n|\left|\mathcal{W}_{\phi_n}\psi_n(x)\right|}_{<\varepsilon/2}$$

since $|\mathcal{W}_{\phi_n}\psi_n(x)| \leq 1$ by Cauchy-Schwarz. Now let $E(\varepsilon, A) = \bigcap_{n=1}^N E\left(\frac{\varepsilon}{2\|A\|_{\mathcal{S}^1}}, \psi_n, \phi_n\right)$ and the result follows. □

Remark C.3.7. The decay condition on the wavelet transform $\mathcal{W}_\phi\psi$ is rather natural. In the time-frequency literature, one often places assumptions on the short-time Fourier transform of the type

$$|V_\phi\psi(x)| \leq C(1 + |x|)^{-s}$$

for some positive s dependent on the dimension of the ambient space.

In computations involving representations on semi-direct products, we can compute the Fourier-Wigner transform using the integral kernel of an operator. Denote by $R = H/G$. Recall that if $A \in \mathcal{S}^2(L^2_r(R))$ then the integral kernel $K_A \in L^2(R \times R)$ is defined by

$$A\phi = \int_R K_A(\cdot, t)\phi(t)\mathrm{d}\mu^R(t).$$

Proposition C.3.8. Assume that $G = R \rtimes_\phi H$ and $A = \psi_1 \otimes \psi_2 \in \mathcal{S}^2(L^2_r(R))$, where $\psi_2 \in \mathrm{Dom}(\mathcal{D})$. Then the kernel of $A\mathcal{D}\pi((r,h))$ is given by

$$\psi_1(s)\overline{\psi_2(tr^{-1})}\frac{\sqrt{|\det_H(\phi_*(tr^{-1}))|}}{\sqrt{\Delta_R(tr^{-1})}}e^{-2\pi i F \log_H(\phi(tr^{-1})h)}.$$

Taking the trace gives

$$\mathcal{F}_W(\psi_1 \otimes \psi_2)(r,h) = \int_R \psi_1(t)\overline{\psi_2(tr^{-1})}e^{-2\pi i F \log_H(\phi(tr^{-1})h)}\frac{\sqrt{|\det_H(\phi_*(tr^{-1}))|}}{\sqrt{\Delta_R(tr^{-1})}}\mathrm{d}\mu^R(t).$$

Proof. Using (C.2.9) we have

$$(\mathcal{D}\pi(r,h)\xi)(t) = \frac{\sqrt{|\det_H(\phi_*(t))|}}{\sqrt{\Delta_R(t)}}e^{-2\pi i F \log_H(\phi(t)h)}\xi(tr).$$

The definition of the Fourier-Wigner transform now implies that

$$\mathcal{F}_{\mathrm{W}}(\psi_1 \otimes \psi_2)(r) = \int_R \psi_1(s)\overline{\psi_2(tr^{-1})}\frac{\sqrt{|\det_H(\phi_*(tr^{-1}))|}}{\sqrt{\Delta_R(tr^{-1})}}e^{-2\pi i F \log_H(\phi(tr^{-1})h)}\xi(t)\,\mathrm{d}\mu^R(t)$$

Hence the kernel is

$$\psi_1(s)\overline{\psi_2(tr^{-1})}\frac{\sqrt{|\det_H(\phi_*(tr^{-1}))|}}{\sqrt{\Delta(tr^{-1})}}e^{-2\pi i F \log_H(\phi(t^{-1}rtr^{-1})h)}.$$

Using that the trace of a rank-one operator $\psi_1 \otimes \psi_2$ is given by

$$\mathrm{tr}(\psi_1 \otimes \psi_2) = \int_R \psi_1(t)\overline{\psi_2(t)}\,\mathrm{d}\mu^R(t)$$

gives the result. □

Lastly we collect some fundamental results on how the inverse Fourier-Wigner transform interacts with translations, adjoints and convolutions.

Lemma C.3.9. *Let $f, g \in L^1_r(G)$ and $x \in G$, then*

(i) $\mathcal{F}_{\mathrm{W}}^{-1}(f)\pi(x) = \mathcal{F}_{\mathrm{W}}^{-1}\left(\frac{1}{\sqrt{\Delta(x)}} L_{x^{-1}} f\right)$,

(ii) $\pi(x)\mathcal{F}_{\mathrm{W}}^{-1}(f) = \mathcal{F}_{\mathrm{W}}^{-1}(R_x f)$,

Proof. Properties (i) and (ii) are essentially reformulations of [85, Thm. 3.9]. For (i), we can compute

$$\begin{aligned}
\mathcal{F}_{\mathrm{W}}^{-1}(f)\pi(x) &= \int_G \check{f}(y)\pi(y)\mathcal{D}\pi(x)\,\mathrm{d}\mu_l(y) \\
&= \int_G \check{f}(y)\pi(yx)\sqrt{\Delta(x)}\mathcal{D}\,\mathrm{d}\mu_l(y) && \left(\mathcal{D}\pi(x) = \sqrt{\Delta(x)}\pi(x)\mathcal{D}\right) \\
&= \int_G \check{f}(zx^{-1})\pi(z)\sqrt{\Delta(x)}\mathcal{D}\,\mathrm{d}\mu_l(zx^{-1}) && \left(z = yx \implies y = zx^{-1}\right) \\
&= \int_G \check{f}(zx^{-1})\pi(z)\frac{1}{\sqrt{\Delta(x)}}\mathcal{D}\,\mathrm{d}\mu_l(z) && \left(d\mu_l(zx^{-1}) = \Delta(x^{-1})d\mu_l(z)\right) \\
&= \int_G \widetilde{L_x f}(z)\pi(z)\frac{1}{\sqrt{\Delta(x)}}\mathcal{D}\,\mathrm{d}\mu_l(z) && \left(\check{f}(zx^{-1}) = \widetilde{L_x f}(z)\right) \\
&= \mathcal{F}_{\mathrm{W}}^{-1}\left(\frac{1}{\sqrt{\Delta(x)}} L_{x^{-1}} f\right).
\end{aligned}$$

To see that $\check{f}(zx^{-1}) = \widetilde{L_{x^{-1}} f}(z)$, note that $\check{f}(zx^{-1}) = f(xz^{-1}) = (L_{x^{-1}} f)(z^{-1})$.

Meanwhile for (ii), we have that

$$\begin{aligned}
\pi(x)\mathcal{F}_{\mathrm{W}}^{-1}(f) &= \int_G \check{f}(y)\pi(xy)\mathcal{D}\,\mathrm{d}\mu_l(y) \\
&= \int_G \check{f}(x^{-1}z)\pi(z)\mathcal{D}\,\mathrm{d}\mu_l(x^{-1}z) && \left(z = xy \implies y = x^{-1}z\right) \\
&= \int_G \widetilde{R_x f}(z)\pi(z)\mathcal{D}\,\mathrm{d}\mu_l(x^{-1}z) && \left(\check{f}(x^{-1}z) = \widetilde{R_x f}(z)\right) \\
&= \int_G \widetilde{R_x f}(z)\pi(z)\mathcal{D}\,\mathrm{d}\mu_l(z) && \left(d\mu_l(yz) = d\mu_l(z)\right) \\
&= \mathcal{F}_{\mathrm{W}}^{-1}(R_x f).
\end{aligned}$$

□

C.3.2 Fourier-Kirillov transform

We will now construct a function Fourier transform. Notice in Section C.2.4 the symplectic form is defined via group action. Hence map κ induces, ut to a constant, an isomorphism from $L_r^2(G)$ to $L^2(O_F, \mathrm{d}\omega)$ by mapping $f \mapsto f \circ \kappa^{-1}$. Since the right Haar measure is unique up to a constant, we will choose the right Haar measure of the group G that makes this map into an isomorphism.

Lemma C.3.10. *Let $X_1, \ldots, X_{2n}$ be a basis for $\mathfrak{g}$, and $\mathrm{d}Y_1, \ldots, \mathrm{d}Y_{2n}$ be the dual basis. Set Pf_F to be the Pfaffian*

$$\mathrm{Pf}_F = \mathrm{Pf}(F([X_i, X_j])) = \frac{1}{2^n n!} \sum_{\sigma \in \mathrm{S}_{2n}} \mathrm{sgn}(\sigma) \prod_{j=1}^{n} F([X_{\sigma(2j-1)}, X_{\sigma(2j)}]),$$

where S_{2n} is the symmetric group. Then

$$\mathrm{Pf}_F \cdot \Delta(\kappa^{-1}(Y))\mathrm{d}Y_1 \wedge \cdots \wedge \mathrm{d}Y_{2n} = \mathrm{Pf}_F \cdot \Delta(\kappa^{-1}(Y))\, \mathrm{d}Y = \mathrm{d}\omega(Y) = \frac{\omega^n(Y)}{n!}.$$

Proof. By the definition of the symplectic form, see [18, Lem. 4.1.3], we have that

$$\mathrm{Pf}(Y[X_{\sigma(2j-1)}, X_{\sigma(2i)}])\mathrm{d}Y_1 \wedge \cdots \wedge \mathrm{d}Y_{2n} = \mathrm{d}\omega(Y).$$

By the definition of κ, we can write

$$Y = K(\kappa^{-1}(Y)^{-1})F,$$

hence

$$\mathrm{Pf}(Y[X_{\sigma(2i-1)}, X_{\sigma(2i)}])\, \mathrm{d}Y = \mathrm{Pf}(F([\mathrm{Ad}_{\kappa^{-1}(Y)} X_{\sigma(2i-1)}, \mathrm{Ad}_{\kappa^{-1}(Y)} X_{\sigma(2i)}]))\, \mathrm{d}Y.$$

The Pfaffian, see e.g. [65], has the property that

$$\mathrm{Pf}(BAB^*) = \det(B)\mathrm{Pf}(A).$$

Using this together with (C.2.8) we get that

$$\mathrm{Pf}_F \cdot \Delta(\kappa^{-1}(Y))\mathrm{d}Y_1 \wedge \cdots \wedge \mathrm{d}Y_{2n} = \mathrm{d}\omega(Y). \qquad \square$$

Define the *Fourier-Kirillov transform* to be a mapping $\mathcal{F}_{\mathrm{KO}} : L_r^2(G) \to L_r^2(G)$ by

$$\mathcal{F}_{\mathrm{KO}}(f)(x) = \frac{1}{\sqrt{|\mathrm{Pf}_F| \cdot \Delta(x)}} \int_G f(y) e^{2\pi i \langle \kappa(x), \log(y) \rangle} \frac{1}{\sqrt{\Theta(\log(y))}}\, \mathrm{d}\mu_r(y) \tag{C.3.2}$$

$$= \frac{1}{\sqrt{|\mathrm{Pf}_F| \cdot \Delta(x)}} \int_{\mathfrak{g}} f(\exp(X)) e^{2\pi i \langle \kappa(x), X \rangle} \sqrt{\Theta(X)}\, \mathrm{d}X.$$

The *inverse Fourier-Kirillov transform* $\mathcal{F}_{\mathrm{KO}}^{-1} : L_r^2(G) \to L_r^2(G)$ is given by

$$\begin{aligned}\mathcal{F}_{\mathrm{KO}}^{-1}(f)(x) &= \sqrt{\Theta(\log(x))} \int_{\mathfrak{g}^*} \chi_{O_F}(Y) f(\kappa^{-1}(Y)) e^{-2\pi i \langle Y, \log(x)\rangle} \sqrt{|\mathrm{Pf}_F| \cdot \Delta(\kappa^{-1}(Y))}\, \mathrm{d}Y \\ &= \sqrt{\Theta(\log(x))} \int_G f(y) e^{-2\pi i \langle \kappa(y), \log(x)\rangle} \frac{1}{\sqrt{|\mathrm{Pf}_F| \cdot \Delta(y)}}\, \mathrm{d}\mu_r(y).\end{aligned} \tag{C.3.3}$$

Proposition C.3.11. The Fourier-Kirillov and inverse Fourier-Kirillov transforms are well-defined and have the property that

$$\mathcal{F}_{\mathrm{KO}} \mathcal{F}_{\mathrm{KO}}^{-1} = \mathrm{id}.$$

Moreover, $\mathcal{F}_{\mathrm{KO}}^{-1}$ is unitary, i.e.,

$$\left\langle \mathcal{F}_{\mathrm{KO}}^{-1}(f), \mathcal{F}_{\mathrm{KO}}^{-1}(g) \right\rangle_{L_r^2} = \langle f, g \rangle_{L_r^2},$$

and $\mathcal{F}_{\mathrm{KO}}^{-1} \mathcal{F}_{\mathrm{KO}}$ is a projection onto $\mathcal{F}_{\mathrm{KO}}^{-1}(L_r^2(G))$.

Proof. Let us start by showing well-definedness of the transforms. It is clear that the following maps are surjective isomorphisms:

- $\mathrm{Mul}_{\sqrt{\Theta}} : L_r^2(G) \to L^2(\mathfrak{g}, \mathrm{d}X)$ defined by

$$\mathrm{Mul}_{\sqrt{\Theta}}(f)(X) = \sqrt{\Theta(X)} f(e^X).$$

- $\mathcal{F} : L^2(\mathfrak{g}, \mathrm{d}X) \to L^2(\mathfrak{g}^*, \mathrm{d}Y)$ defined by

$$\mathcal{F}(f) = \int_{\mathfrak{g}} f(X) e^{2\pi i \langle X, Y\rangle}\, \mathrm{d}X.$$

- $\mathrm{Mul}_{1/\sqrt{|\mathrm{Pf}_F|\Delta}} : L^2(O_F, \mathrm{d}Y) \to L^2(O_F, \mathrm{d}\omega)$ defined by $\mathrm{Mul}_{1/\sqrt{|\mathrm{Pf}_F|\Delta}}(f)(Y) = \frac{f(Y)}{\sqrt{|\mathrm{Pf}_F| \cdot \Delta(\kappa^{-1}(Y))}}$. See Lemma C.3.10.

Additionally, we have the restriction operator $\mathrm{Res}_{O_F} : L^2(\mathfrak{g}^*, \mathrm{d}Y) \to L^2(O_F, \mathrm{d}Y)$ defined by

$$\mathrm{Res}_{O_F}(f)(Y) = f|_{O_F}(Y)$$

and the inclusion operator $\mathrm{Inj}_{O_F} : L^2(O_F, \mathrm{d}Y) \to L^2(\mathfrak{g}^*, \mathrm{d}Y)$ defined by

$$\mathrm{Inj}_{O_F}(f)(Y) = \begin{cases} f(Y) & Y \in O_F, \\ 0 & \text{otherwise.} \end{cases}$$

Now we can write

$$(\mathcal{F}_{\mathrm{KO}} f)(x) = \mathrm{Mul}_{1/\sqrt{|\mathrm{Pf}_F|\Delta}} \circ \mathrm{Res}_{O_F} \circ \mathcal{F} \circ \mathrm{Mul}_{\sqrt{\Theta}}(f)(\kappa(x))$$

and

$$(\mathcal{F}_{\mathrm{KO}}^{-1} f)(x) = \mathrm{Mul}_{\sqrt{\Theta}}^{-1} \circ \mathcal{F}^{-1} \circ \mathrm{Inj}_{O_F} \circ \mathrm{Mul}_{1/\sqrt{|\mathrm{Pf}_F|\Delta}}^{-1}(f)(x).$$

Notice that $\mathrm{Inj}_{O_F} \circ \mathrm{Res}_{O_F}$ is a projection and $\mathrm{Res}_{O_F} \circ \mathrm{Inj}_{O_F} = \mathrm{id}$ which implies that $\mathcal{F}_{\mathrm{KO}}^{-1}\mathcal{F}_{\mathrm{KO}}$ is a projection and $\mathcal{F}_{\mathrm{KO}}\mathcal{F}_{\mathrm{KO}}^{-1} = \mathrm{id}$.

That the operator $\mathcal{F}_W^{-1}$ is unitary follows from the computation

$$\begin{aligned}
\left\langle \mathcal{F}_{\mathrm{KO}}^{-1}(f), \mathcal{F}_{\mathrm{KO}}^{-1}(g) \right\rangle_{L_r^2} &= \left\langle \mathrm{Inj}_{O_F}(\mathrm{Mul}_{1/\sqrt{|\mathrm{Pf}_F|\Delta}}^{-1}(f)), \mathrm{Inj}_{O_F}(\mathrm{Mul}_{1/\sqrt{|\mathrm{Pf}_F|\Delta}}^{-1}(g)) \right\rangle_{L^2(\mathfrak{g}^*, \mathrm{d}Y)} \\
&= \left\langle \mathrm{Mul}_{1/\sqrt{|\mathrm{Pf}_F|\Delta}}^{-1}(f), \mathrm{Mul}_{1/\sqrt{|\mathrm{Pf}_F|\Delta}}^{-1}(g) \right\rangle_{L^2(O_F, \mathrm{d}Y)} \\
&= \langle f, g \rangle_{L_r^2},
\end{aligned}$$

where we have used that the inclusion operator is an injective isomorphism. □

Lemma C.3.12. *Let $f \in L_r^2(G)$, then $\mathcal{F}_{\mathrm{KO}}(\sqrt{\Delta_G(\cdot)}\check{f}(\cdot)) = \overline{\mathcal{F}_{\mathrm{KO}}(\overline{f(\cdot)})}$.*

Proof. Using (C.2.6) we have that

$$\begin{aligned}
\overline{\mathcal{F}_{\mathrm{KO}}(\overline{f})(x)} &= \frac{1}{\sqrt{|\mathrm{Pf}_F|\Delta(x)}} \int_{\mathfrak{g}} f(\exp(X)) e^{-2\pi i \langle \kappa(x), X \rangle} \sqrt{\Theta(X)}\, \mathrm{d}X \\
&= \frac{1}{\sqrt{|\mathrm{Pf}_F|\Delta(x)}} \int_{\mathfrak{g}} f(\exp(-X)) e^{2\pi i \langle \kappa(x), X \rangle} \sqrt{\Theta(-X)}\, \mathrm{d}X \\
&= \frac{1}{\sqrt{|\mathrm{Pf}_F|\Delta(x)}} \int_{\mathfrak{g}} f((\exp(X))^{-1}) e^{2\pi i \langle \kappa(x), X \rangle} \sqrt{\Theta(-X)}\, \mathrm{d}X \\
&= \frac{1}{\sqrt{|\mathrm{Pf}_F|\Delta(x)}} \int_{\mathfrak{g}} \sqrt{\Delta_G(\exp(X))}\check{f}(\exp(X)) e^{2\pi i \langle \kappa(x), X \rangle} \sqrt{\Theta(X)}\, \mathrm{d}X,
\end{aligned}$$

which completes the proof. □

Proposition C.3.13. Let $f \in L_r^2(G)$ and define $L_x f(y) = f(x^{-1}y)$, $R_x f(y) = f(yx)$ be the right and left translation operators, and $\Psi_{x^{-1}} = L_x R_x$. Then,

$$\mathcal{F}_{\mathrm{KO}}^{-1}(R_x f) = \sqrt{\Delta(x)}\Psi_{x^{-1}}\mathcal{F}_{\mathrm{KO}}^{-1}(f) \quad \text{and} \quad R_x \mathcal{F}_{\mathrm{KO}}(f) = \sqrt{\Delta(x)}\mathcal{F}_{\mathrm{KO}}(\Psi_x f).$$

Proof. We have that

$$\begin{aligned}
\left\langle K((yx^{-1})^{-1})F, \log(z) \right\rangle &= \left\langle F, \mathrm{Ad}_{yx^{-1}} \log(z) \right\rangle \\
&= \left\langle K(y^{-1})F, \mathrm{Ad}_{x^{-1}} \log(z) \right\rangle = \left\langle K(y^{-1})F, \log(\Psi_{x^{-1}} z) \right\rangle.
\end{aligned}$$

Hence

$$\begin{aligned}
\mathcal{F}_{\mathrm{KO}}^{-1}(f)(x^{-1}zx) &= \sqrt{\Theta(\log(x^{-1}zx))}\int_G f(y)e^{-2\pi i\langle\kappa(y),\log(\Psi_{x^{-1}}(z))\rangle}\frac{1}{\sqrt{|\mathrm{Pf}_F|\cdot\Delta(y)}}\,\mathrm{d}\mu_r(y)\\
&= \sqrt{\Theta(\mathrm{Ad}_{x^{-1}}\log(z))}\int_G f(y)e^{-2\pi i\langle\kappa(yx^{-1}),\log(z)\rangle}\frac{1}{\sqrt{|\mathrm{Pf}_F|\cdot\Delta(y)}}\,\mathrm{d}\mu_r(y)\\
&= \sqrt{\Theta(\mathrm{Ad}_{x^{-1}}\log(z))}\int_G f(yx)e^{-2\pi i\langle\kappa(y),\log(z)\rangle}\frac{1}{\sqrt{|\mathrm{Pf}_F|\cdot\Delta(yx)}}\,\mathrm{d}\mu_r(y)\\
&= \frac{\sqrt{\Theta(\log(z))}}{\sqrt{\Delta(x)}}\int_G f(yx)e^{-2\pi i\langle\kappa(y),\log(z)\rangle}\frac{1}{\sqrt{|\mathrm{Pf}_F|\cdot\Delta(y)}}\,\mathrm{d}\mu_r(y).
\end{aligned}$$

The final step follows from (C.2.7) together with the determinant being invariant under automorphisms. The identity $R_x\mathcal{F}_{\mathrm{KO}}(f) = \sqrt{\Delta(x)}\mathcal{F}_{\mathrm{KO}}(\Psi_x f)$ is proved similarly. □

For semi-direct products we have a slight simplification.

Proposition C.3.14. Let $G = R \rtimes_\phi H$. Then

$$\mathcal{F}_{\mathrm{KO}}(f)(r_1,h_1) = \frac{1}{\sqrt{|\mathrm{Pf}_F|\cdot\Delta_G(r_1,h_1)}}\int_G f(r_2,h_2)e^{2\pi i\langle F,\log_G(r_1r_2r_1^{-1},h_1\phi(r_1^{-1}r_2^{-1})(h_2)h_1^{-1})\rangle}\frac{\mathrm{d}\mu_r(r_2,h_2)}{\sqrt{\Theta(\log(r_2,h_2))}}.$$

Proof. Let $(r_1,h_1),\ (r_2,h_2)\in R\rtimes_\phi H$. We have that

$$\begin{aligned}
\langle K((r_1,h_1)^{-1})F,\log_G(r_2,h_2)\rangle &= \langle F,\mathrm{Ad}_{(r_1,h_1)}\log_G(r_2,h_2)\rangle\\
&= \langle F,\log_G((r_1,h_1)(r_2,h_2)(r_1,h_1)^{-1})\rangle\\
&= \langle F,\log_G(r_1r_2r_1^{-1},h_1\phi(r_1^{-1}r_2^{-1})(h_2)h_1^{-1}))\rangle.
\end{aligned}$$

□

C.3.3 Combining the transforms

As described in the introduction our goal is to define quantization by combining the Fourier transforms. Hence this section will be devoted to showing the connections between the range of the Fourier-Kirillov transform and the inverse Fourier-Wigner transform.

Lemma C.3.15. *Let f be continuous on G with compact support. Then*

$$\mathcal{F}_{\mathrm{KO}}^{-1}(\mathcal{F}_{\mathrm{KO}}(f))(e) = \mathcal{F}_{\mathrm{W}}(\mathcal{F}_{\mathrm{W}}^{-1}(f))(e),$$

where $e\in G$ is the group identity element.

Proof. This is simply a reformulation of [18, Prop. 6.3.1] by writing

$$\mathcal{F}_{\mathrm{W}}(\mathcal{F}_{\mathrm{W}}^{-1}(f))(e) = \operatorname{tr}(\pi(\check{f}) \circ \mathcal{D} \circ \mathcal{D}) = \operatorname{tr}(\mathcal{D} \circ \pi(\check{f}) \circ \mathcal{D})$$

and

$$\mathcal{F}_{\mathrm{KO}}^{-1}(\mathcal{F}_{\mathrm{KO}}(f))(e) = \int_{O_F} \int_{\mathfrak{g}} \check{f}(\exp(X)) e^{-2\pi i \langle Y, X\rangle} \sqrt{\Theta(-X)}\, \mathrm{d}X\, \mathrm{d}Y. \qquad \square$$

Theorem C.3.16. *Let $f \in L^2_r(G)$. Then*

$$\mathcal{F}_{\mathrm{KO}}^{-1}(\mathcal{F}_{\mathrm{KO}}(f)) = \mathcal{F}_{\mathrm{W}}(\mathcal{F}_{\mathrm{W}}^{-1}(f)). \tag{C.3.4}$$

Consequently, $\mathcal{F}_{\mathrm{W}}\mathcal{F}_{\mathrm{W}}^{-1}$ is a projection onto $\mathcal{F}_{\mathrm{KO}}^{-1}(L^2_r(G))$ and $\mathcal{F}_{\mathrm{KO}}^{-1}(L^2_r(G)) = \mathcal{F}_{\mathrm{W}}(\mathcal{S}^2(\mathcal{H}))$.

Proof. We will start by assuming that f is continuous with compact support. Then from Lemma C.3.15 we have that

$$\mathcal{F}_{\mathrm{KO}}^{-1}(\mathcal{F}_{\mathrm{KO}}(f))(e) = \mathcal{F}_{\mathrm{W}}(\mathcal{F}_{\mathrm{W}}^{-1}(f))(e).$$

Using Proposition C.3.13 we have that

$$\begin{aligned}
\mathcal{F}_{\mathrm{KO}}^{-1}(\mathcal{F}_{\mathrm{KO}}(f))(x) &= R_x \mathcal{F}_{\mathrm{KO}}^{-1}(\mathcal{F}_{\mathrm{KO}}(f))(e) \\
&= \mathcal{F}_{\mathrm{KO}}^{-1}(\sqrt{\Delta(x)}\Psi_x \mathcal{F}_{\mathrm{KO}}(f))(e) \\
&= \mathcal{F}_{\mathrm{KO}}^{-1}(\mathcal{F}_{\mathrm{KO}}(R_x f))(e).
\end{aligned}$$

Additionally, using that $R_x \mathcal{F}_{\mathrm{W}}(A) = \mathcal{F}_{\mathrm{W}}(\pi(x)A)$ (from the definition of $\mathcal{F}_{\mathrm{W}}$) and Lemma C.3.9 (ii) we can deduce that

$$\begin{aligned}
\mathcal{F}_{\mathrm{W}}(\mathcal{F}_{\mathrm{W}}^{-1}(f))(x) &= R_x \mathcal{F}_{\mathrm{W}}(\mathcal{F}_{\mathrm{W}}^{-1}(f))(e) \\
&= \mathcal{F}_{\mathrm{W}}(\pi(x)\mathcal{F}_{\mathrm{W}}^{-1}(f))(e) \\
&= \mathcal{F}_{\mathrm{W}}(\mathcal{F}_{\mathrm{W}}^{-1}(R_x f))(e).
\end{aligned}$$

Hence we have that

$$\mathcal{F}_{\mathrm{KO}}^{-1}(\mathcal{F}_{\mathrm{KO}}(f)) = \mathcal{F}_{\mathrm{W}}(\mathcal{F}_{\mathrm{W}}^{-1}(f))$$

for continuous f with compact support. Since continuous functions with compact support are dense in $L^2_r(G)$, the result follows.

To see that $\mathcal{F}_{\mathrm{W}}\mathcal{F}_{\mathrm{W}}^{-1}$ is a projection onto $\mathcal{F}_{\mathrm{KO}}^{-1}(L^2_r(G))$, let $F = \mathcal{F}_{\mathrm{KO}}^{-1}(f)$ be an arbitrary elements of $\mathcal{F}_{\mathrm{KO}}^{-1}(L^2_r(G))$. Then

$$\mathcal{F}_{\mathrm{W}}(\mathcal{F}_{\mathrm{W}}^{-1}(F)) = \mathcal{F}_{\mathrm{KO}}^{-1}(\mathcal{F}_{\mathrm{KO}}(F)) = \mathcal{F}_{\mathrm{KO}}^{-1}(\mathcal{F}_{\mathrm{KO}}(\mathcal{F}_{\mathrm{KO}}^{-1}(f))) = \mathcal{F}_{\mathrm{KO}}^{-1}(f) = F$$

by (C.3.4) and Proposition C.3.11.

Lastly for $\mathcal{F}_{\mathrm{KO}}^{-1}(L_r^2(G)) = \mathcal{F}_{\mathrm{W}}(\mathcal{S}^2(\mathcal{H}))$, we first let $g \in \mathcal{F}_{\mathrm{KO}}^{-1}(L_r^2(G))$ with $g = \mathcal{F}_{\mathrm{KO}}^{-1}(f)$ for some $f \in L_r^2(G)$. Then

$$g = \mathcal{F}_{\mathrm{KO}}^{-1}(f) = \mathcal{F}_{\mathrm{KO}}^{-1}(\mathcal{F}_{\mathrm{KO}}(\mathcal{F}_{\mathrm{KO}}^{-1}(f))) = \mathcal{F}_{\mathrm{W}}(\mathcal{F}_{\mathrm{W}}^{-1}(g)) \in \mathcal{F}_{\mathrm{W}}(\mathcal{S}^2)$$

where we used Proposition C.3.11 for the second step. Similarly, for the other direction we let $g \in \mathcal{F}_{\mathrm{W}}(\mathcal{S}^2)$ with $g = \mathcal{F}_{\mathrm{W}}(A)$ for some $A \in \mathcal{S}^2$. It then holds that

$$g = \mathcal{F}_{\mathrm{W}}(A) = \mathcal{F}_{\mathrm{W}}(\mathcal{F}_{\mathrm{W}}^{-1}(\mathcal{F}_{\mathrm{W}}(A))) = \mathcal{F}_{\mathrm{KO}}^{-1}(\mathcal{F}_{\mathrm{KO}}(g)) \in \mathcal{F}_{\mathrm{KO}}^{-1}(L_r^2(G))$$

and consequently the sets are identical. $\square$

Example C.3.17 (Affine Group)**.** Denote by $X_1 = U$ and $X_2 = V$. We have for $F = V^*$ that

$$\mathrm{Pf}_{V^*} = \frac{1}{2} \sum_{\sigma \in S_2} \mathrm{sgn}(\sigma) V^*([X_{\sigma(1)}, X_{\sigma(2)}]) = 1.$$

Additionally, we have that $\kappa(a,x) = aV^* - xU^*$, since co-adjoint map is given by

$$K(a,x)(V^*) = a^{-1}V^* + a^{-1}xU^*.$$

This means that

$$\langle \kappa(a,x), uU + vV \rangle = av - xu,$$

which is the standard symplectic form. Hence

$$\begin{aligned}
\mathcal{F}_{\mathrm{KO}}(f)(a,x) &= \sqrt{a} \int_{\mathrm{Aff}} f(b,y) e^{2\pi i (ay/\lambda(\log(b)) - x\log(b))} \frac{\mathrm{d}\mu_r(b,y)}{\sqrt{b\lambda(\log(b))}} \\
&= \sqrt{a} \int_{\mathbb{R}^2} f\left(e^u, v \cdot \lambda(u)\right) e^{2\pi i (av - xu)} \sqrt{e^u \lambda(u)}\, \mathrm{d}u\, \mathrm{d}v.
\end{aligned}$$

Notice that if we had chosen the other orbit corresponding to $F = -V^*$ we would get the Fourier transformation

$$\begin{aligned}
\mathcal{F}_{\mathrm{KO}}^{-}(f)(a,x) &= \sqrt{a} \int_{\mathrm{Aff}} f(b,y) e^{2\pi i (x\log(b) - ay/\lambda(\log(b)))} \frac{\mathrm{d}\mu_r(b,y)}{\sqrt{b\lambda(\log(b))}} \\
&= \sqrt{a} \int_{\mathbb{R}^2} f\left(e^u, v\lambda(u)\right) e^{2\pi i (xu - av)} \sqrt{e^u \lambda(u)}\, \mathrm{d}u\, \mathrm{d}v.
\end{aligned}$$

Example C.3.18 (Shearlet Group)**.** For the Shearlet group we have that $\mathcal{D}\phi(a,t) = a\phi(a,t)$. Let

$$Af(c,r) = \int_{\mathbb{R}^+ \times \mathbb{R}} K_A((c,r),(b,t)) f(b,t) \frac{\mathrm{d}b\, \mathrm{d}t}{b}.$$

Hence we have that

$$\mathcal{F}_{\mathrm{W}}(A)(a,s,x_1,x_2) = \int_{\mathbb{R}^+\times\mathbb{R}} K_A((b,t),(ab,t+s\sqrt{b}))e^{-2\pi i(bx_1+\sqrt{b}tx_2)}\,\mathrm{d}b\,\mathrm{d}t.$$

Denote the basis by $X_1 = A$, $X_2 = B$, $X_3 = C$, and $X_4 = D$, and Y_i the dual basis. Computing Pfaffian for $F = \pm Y_3$ we get that $|\mathrm{Pf}_{\pm Y_3^*}| = 1$. Computing for $F = Y_3^*$ we get that

$$\langle \kappa(a,s,x_1,x_2), \alpha A + \sigma B + \xi_1 C + \xi_2 D\rangle\rangle = \frac{sx_2}{2}\alpha + \sqrt{a}(s\xi_2 - x_2\sigma) + a\xi_1 - x_1\alpha.$$

The Fourier-Kirillov transform for $F = Y_3^*$ hence becomes

$$\begin{aligned}&\mathcal{F}_{\mathrm{KO}}(f)(a,s,x_1,x_2)\\&= a\int_{\mathfrak{g}} f(\exp(\alpha,\beta,\gamma,\delta))e^{2\pi i\left(\frac{sx_2}{2}\alpha+\sqrt{a}(s\xi_2-x_2\sigma)+a\xi_1-x_1\alpha\right)}\sqrt{\lambda(\alpha)}\lambda(\alpha/2)\,\mathrm{d}\alpha\,\mathrm{d}\sigma\,\mathrm{d}\xi_1\,\mathrm{d}\xi_2,\end{aligned}$$

where

$$\exp(\alpha,\beta,\gamma,\delta) = (e^{\alpha}, \sigma\lambda(\alpha/2), \xi_1\lambda(\alpha) + \sigma\xi_2\lambda(\alpha/2)^2/2, \xi_2\lambda(\alpha/2)).$$

C.4 Weyl Quantization

So far, we have extensively discussed the Fourier-Wigner (Section C.3.1) and Fourier-Kirillov (Section C.3.2) transforms. We will now compose these to define our version of Weyl quantization in the same way as was done in Section C.2.1. This means that we define the quantization mapping as

$$A : f \mapsto A_f = \mathcal{F}_{\mathrm{W}}^{-1}(\mathcal{F}_{\mathrm{KO}}^{-1}(f)). \tag{C.4.1}$$

From the results about the Fourier-Wigner and Fourier-Kirillov transforms of Section C.3, we can deduce the following important property.

Theorem C.4.1. *The quantization mapping $A : L_r^2(G) \to \mathcal{S}^2$ and its inverse $a : \mathcal{S}^2 \to L_r^2(G)$ are both linear unitary isometries.*

Proof. Linearity of A and a follow from the definitions of $\mathcal{F}_{\mathrm{W}}$ and $\mathcal{F}_{\mathrm{KO}}$. For the unitary isometry part, note first that for $A, B \in \mathcal{S}^2$,

$$\left\langle \mathcal{F}_{\mathrm{KO}}(\mathcal{F}_{\mathrm{W}}(A)), \mathcal{F}_{\mathrm{KO}}(\mathcal{F}_{\mathrm{W}}(B))\right\rangle_{L_r^2} = \left\langle \mathcal{F}_{\mathrm{KO}}^{-1}(\mathcal{F}_{\mathrm{KO}}(\mathcal{F}_{\mathrm{W}}(A))), \mathcal{F}_{\mathrm{KO}}^{-1}(\mathcal{F}_{\mathrm{KO}}(\mathcal{F}_{\mathrm{W}}(B)))\right\rangle_{L_r^2}$$

since $\mathcal{F}_{\mathrm{KO}}^{-1}$ is unitary by Proposition C.3.11. The same proposition also states that $\mathcal{F}_{\mathrm{KO}}^{-1}\mathcal{F}_{\mathrm{KO}}$ is a projection onto $\mathcal{F}_{\mathrm{KO}}^{-1}(L_r^2)$ which is equal to $\mathcal{F}_{\mathrm{W}}(\mathcal{S}^2)$ by Theorem C.3.16. Consequently,

$$\left\langle \mathcal{F}_{\mathrm{KO}}^{-1}(\mathcal{F}_{\mathrm{KO}}(\mathcal{F}_{\mathrm{W}}(A))), \mathcal{F}_{\mathrm{KO}}^{-1}(\mathcal{F}_{\mathrm{KO}}(\mathcal{F}_{\mathrm{W}}(B)))\right\rangle_{L_r^2} = \left\langle \mathcal{F}_{\mathrm{W}}(A), \mathcal{F}_{\mathrm{W}}(B)\right\rangle_{L_r^2}$$

and the result now follows from $\mathcal{F}_{\mathrm{W}}$ being unitary by Proposition C.3.4.

The other direction of the quantization follows from a similar argument. Let $f, g \in L_r^2(G)$ and note that by $\mathcal{F}_{\mathrm{W}}$ being unitary (Proposition C.3.4) and by $\mathcal{F}_{\mathrm{W}}\mathcal{F}_{\mathrm{W}}^{-1}$ being a projection onto $\mathcal{F}_{\mathrm{KO}}^{-1}(L_r^2(G))$ (Theorem C.3.16),

$$\begin{aligned}
\left\langle \mathcal{F}_{\mathrm{W}}^{-1}(\mathcal{F}_{\mathrm{KO}}^{-1}(f)), \mathcal{F}_{\mathrm{W}}^{-1}(\mathcal{F}_{\mathrm{KO}}^{-1}(g))\right\rangle_{\mathcal{S}^2} &= \left\langle \mathcal{F}_{\mathrm{W}}(\mathcal{F}_{\mathrm{W}}^{-1}(\mathcal{F}_{\mathrm{KO}}^{-1}(f))), \mathcal{F}_{\mathrm{W}}(\mathcal{F}_{\mathrm{W}}^{-1}(\mathcal{F}_{\mathrm{KO}}^{-1}(g)))\right\rangle_{L_r^2} \\
&= \left\langle \mathcal{F}_{\mathrm{KO}}^{-1}(f), \mathcal{F}_{\mathrm{KO}}^{-1}(g)\right\rangle_{L_r^2} \\
&= \langle f, g\rangle_{L_r^2}
\end{aligned}$$

where we in the last step used that $\mathcal{F}_{\mathrm{KO}}^{-1}$ is unitary by Proposition C.3.11. □

We now move to showing more detailed properties of the quantization mapping, starting with the actions of translation and conjugation.

Example C.4.2 (Affine group)**.** For the representation given by $F = V^*$ we get the quantization outlined in [94]. The dequantization of an operator A is explicitly given by

$$f_A(a,x) = \int_{-\infty}^{\infty} K_A\left(\frac{aue^u}{e^u-1}, \frac{au}{e^u-1}\right) e^{-2\pi i x u}\, du,$$

where $K_A \colon \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{C}$ is the integral kernel of A defined by

$$A\psi(r) = \int_0^{\infty} K_A(r,s)\psi(s)\,\frac{ds}{s}, \qquad \psi \in L^2(\mathbb{R}_+).$$

Example C.4.3 (Shearlet group)**.** Let

$$Af(c,r) = \int_{\mathbb{R}^+\times\mathbb{R}} K_A((c,r),(b,t))f(b,t)\,\frac{\mathrm{d}b\,\mathrm{d}t}{b}.$$

Using the formula for the Fourier-Kirillov and Fourier-Wigner transform for the Shearlet group we get that the dequantization $f_A = \mathcal{F}_{\mathrm{KO}}(\mathcal{F}_W(A))$ of A at the point $(a,b,c,d) \in \mathbb{S}$ has the explicit expression

$$\int_{\mathbb{R}^2} K_A\left(\left(\frac{ae^{\alpha}}{\lambda(\alpha)}, \frac{2\lambda(\alpha)s + \sigma\lambda(\alpha/2)^2}{2\sqrt{\lambda(\alpha)}\lambda(\alpha/2)}\right), \left(\frac{a}{\lambda(\alpha)}, \frac{2\lambda(\alpha)s - \sigma\lambda(\alpha/2)^2}{2\sqrt{\lambda(\alpha)}\lambda(\alpha/2)}\right)\right) \times e^{2\pi i\left(\frac{sx_2-2x_1}{2}\alpha - x_2\sigma\right)}\,\mathrm{d}\alpha\,\mathrm{d}\sigma.$$

C.4.1 Translation and conjugation

In this section we investigate how quantization behaves with respect to translation and complex conjugation. In the proofs, we will have to follow the quantization through the two Fourier transforms $\mathcal{F}_{\mathrm{W}}$ and $\mathcal{F}_{\mathrm{KO}}$ in (C.4.1). For translation, most of the work was done in Lemma C.3.9 and Proposition C.3.13. Note that the following relation is the same as was shown for the affine group in [32, p. 47].

Proposition C.4.4. Let $f \in L^2_r(G)$ and $x \in G$, then

$$\pi(x)^* A_f \pi(x) = A_{R_{x^{-1}} f}$$

where $R_{x^{-1}} f(y) = f(yx^{-1})$.

Proof. Assume that $f \in L^2_r(G) \cap L^1_r(G)$ so that we may apply Lemma C.3.9. The full result then follows by density and the continuity of A from $L^2_r(G)$ to $\mathcal{S}^2$. We can then compute

$$\begin{aligned}
\pi(x)^* A_f \pi(x) &= \pi(x^{-1}) \mathcal{F}_{\mathrm{W}}^{-1}(\mathcal{F}_{\mathrm{KO}}^{-1}(f))\pi(x) \\
&= \mathcal{F}_{\mathrm{W}}^{-1}(R_{x^{-1}} \mathcal{F}_{\mathrm{KO}}^{-1}(f))\pi(x) \\
&= \mathcal{F}_{\mathrm{W}}^{-1}\left(\frac{1}{\sqrt{\Delta(x)}} L_{x^{-1}} R_{x^{-1}} \mathcal{F}_{\mathrm{KO}}^{-1}(f) \right)
\end{aligned}$$

where we in the last step used Proposition C.3.13 as $\frac{1}{\sqrt{\Delta(x)}} L_{x^{-1}} R_{x^{-1}} \mathcal{F}_{\mathrm{KO}}^{-1}(f) = \mathcal{F}_{\mathrm{KO}}^{-1}(R_{x^{-1}} f)$. □

We now move on to the quantization of the complex conjugates.

Proposition C.4.5. Let $f \in L^2_r(G)$, then

$$A_f^* = A_{\bar{f}}.$$

Proof. To start, we assume that f is such that $A_f \mathcal{D} \in \mathcal{S}^1$ so that we can apply the Fourier-Wigner transform. Recall that $A_f = \mathcal{F}_{\mathrm{W}}^{-1}(\mathcal{F}_{\mathrm{KO}}^{-1}(f))$. We will compute $\mathcal{F}_{\mathrm{W}}(A_f^*)$ and show that it coincides with $\mathcal{F}_{\mathrm{W}}(A_{\bar{f}}) = \mathcal{F}_{\mathrm{KO}}^{-1}(\bar{f})$ which yields the desired conclusion by the injectivity of $\mathcal{F}_{\mathrm{W}}$ guaranteed by Proposition C.3.4.

Let $A_f = \sum_m s_m(\psi_m \otimes \phi_m)$ be the singular value decomposition of A_f and

$A = A_f \mathcal{D}$ so that we may conclude that

$$\begin{aligned}
\mathcal{F}_{\mathrm{W}}(A_f)(x) &= \mathcal{F}_{\mathrm{W}}(A\mathcal{D}^{-1})(x) = \operatorname{tr}\left(A\pi(x)\right) = \operatorname{tr}\left(A_f\mathcal{D}\pi(x)\right) \\
&= \sum_n \langle A_f \mathcal{D}\pi(x)\xi_n, \xi_n\rangle \\
&= \sum_n \left\langle \sum_m s_m(\psi_m \otimes \phi_m)\mathcal{D}\pi(x)\xi_n, \xi_n \right\rangle \\
&= \sum_n \sum_m s_m \langle \mathcal{D}\pi(x)\xi_n, \phi_m\rangle\langle\psi_m, \xi_n\rangle \\
&= \sum_m s_m\langle\psi_m, \pi(x^{-1})\mathcal{D}\phi_m\rangle.
\end{aligned}$$

Meanwhile, using that $A_f^* = \mathcal{D}^{-1}A^*$ we find that

$$\begin{aligned}
\mathcal{F}_{\mathrm{W}}(A_f^*)(x) &= \mathcal{F}_{\mathrm{W}}(\mathcal{D}^{-1}A^*)(x) = \mathcal{F}_{\mathrm{W}}(\mathcal{D}^{-1}A^*\mathcal{D}\mathcal{D}^{-1}) \\
&= \operatorname{tr}\left(\mathcal{D}^{-1}A^*\mathcal{D}\pi(x)\right) = \operatorname{tr}\left(A_f^*\mathcal{D}\pi(x)\right) \\
&= \sum_n \langle A_f^*\mathcal{D}\pi(x)\xi_n, \xi_n\rangle \\
&= \overline{\sum_m s_m\langle \pi(x)^*\mathcal{D}\psi_m, \phi_m\rangle}
\end{aligned}$$

where we in the last step used the same argument as for $\mathcal{F}_{\mathrm{W}}(A_f)$. Continuing, we find that

$$\begin{aligned}
\mathcal{F}_{\mathrm{W}}(A_f^*)(x) &= \overline{\sum_m s_m\langle\psi_m, \sqrt{\Delta(x)}\pi(x)\mathcal{D}\phi_m\rangle} \\
&= \sqrt{\Delta(x)}\,\overline{\sum_m s_m\langle\psi_m, \pi(x)\mathcal{D}\phi_m\rangle} \\
&= \sqrt{\Delta(x)}\,\overline{\mathcal{F}_{\mathrm{W}}(A_f)(x^{-1})}.
\end{aligned}$$

From Lemma C.3.12 we have that

$$\mathcal{F}_{\mathrm{KO}}^{-1}(\bar{f})(x) = \sqrt{\Delta(x)}\overline{\widecheck{\mathcal{F}_{\mathrm{KO}}^{-1}(f)}(x)}$$

and plugging this into the above yields

$$\mathcal{F}_{\mathrm{W}}(A_f^*)(x) = \sqrt{\Delta(x)}\frac{1}{\sqrt{\Delta(x)}}\mathcal{F}_{\mathrm{KO}}^{-1}(\bar{f})(x)$$

from which it follows that $A_f^* = A_{\bar{f}}$ using the injectivity of $\mathcal{F}_{\mathrm{W}}$.

For the full case, for any $\varepsilon > 0$ we can by Lemma C.3.2 find an $A_g \in \mathcal{S}^2$ with $A_g\mathcal{D} \in \mathcal{S}^1$ such that

$$\|A_f - A_g\|_{\mathcal{S}^2} < \varepsilon.$$

It then holds that $A_g^* = A_{\bar{g}}$ and so we have

$$\|A_f^* - A_{\bar{f}}\|_{\mathcal{S}^2} \leq \|A_f^* - A_g^*\|_{\mathcal{S}^2} + \|A_{\bar{g}} - A_{\bar{f}}\|_{\mathcal{S}^2} < 2\varepsilon.$$

Since ε was arbitrary, the conclusion follows. □

C.4.2 Algebraic structure

With the basic properties of the quantization mapping established, we now move on to endowing each step of the quantization map with algebraic structure. First we show that $A : L^2_r(G) \to \mathcal{S}^2$ is a form of isomorphism and then move on to the individual Fourier transforms.

An H^*-algebra $*$-isomorphism

In his seminal 1966 paper [173], Pool showed that Weyl quantization on the Weyl-Heisenberg group can be realized as an isometric $*$-isomorphism between H^*-algebras. In this section, we set out to do the same in our more general setup. As a first step, we recall the definition of an H^*-algebra [218]. Note that the Hilbert space $\mathcal{H}$ in the definition below is not the same as the one associated with our representation π.

Definition C.4.6. Let $\mathcal{H}$ be a complex and separable Hilbert space and $\cdot : a, b \mapsto a \cdot b, {}^* : a \mapsto a^*$ two operations satisfying the properties

(i) $a \cdot (b + c) = a \cdot b + a \cdot c, \quad (a + b) \cdot c = a \cdot c + b \cdot c,$

(ii) $\lambda(a \cdot b) = (\lambda a) \cdot b = a \cdot (\lambda b),$

(iii) $a \cdot (b \cdot c) = (a \cdot b) \cdot c,$

(iv) $a^{**} = a,$

(v) $(a + b)^* = a^* + b^*,$

(vi) $(a \cdot b)^* = b^* \cdot a^*,$

(vii) $(\lambda a)^* = \bar{\lambda} a^*,$

(viii) $\|a^*\| = \|a\|,$

(ix) $\|a \cdot b\| \leq \|a\|\|b\|$,

(x) $\langle a \cdot b, c\rangle = \langle b, a^* \cdot c\rangle$,

for $a, b, c \in \mathcal{H}$ and $\lambda \in \mathbb{C}$. Then we call $(\mathcal{H}, \cdot, ^*)$ an H^*-algebra with respect to the given operations.

We want to show that the quantization mapping $A : f \mapsto A_f$ (C.4.1) provides a $*$-isomorphism between the function and operator spaces. Since we have already showed that complex conjugation corresponds to taking the adjoint in Proposition C.4.5, the main obstacle to this is defining a mapping on $L^2_r(G)$ which should correspond to composing operators through quantization. In the Weyl-Heisenberg case, this need is fulfilled by *twisted multiplication* [84, Chap. 2.3], sometimes also referred to as *Moyal products* [114, Sec. 13.3.3], which have an explicit definition. In the general setting, such an explicit expression is too much to hope for but it is possible to set up a suitable mapping in a backwards manner as

$$f \sharp g := a_{A_f A_g} \iff A_{f\sharp g} = A_f A_g \tag{C.4.2}$$

and we refer to it as twisted multiplication as it coincides with the Weyl-Heisenberg definition. Note that the twisted multiplication of two $L^2_r(G)$ functions is another $L^2_r(G)$ function since

$$\|A_{f\sharp g}\|_{\mathcal{S}^1} = \|A_f A_g\|_{\mathcal{S}^1} \leq \|A_f\|_{\mathcal{S}^2}\|A_g\|_{\mathcal{S}^2}$$

which implies that $A_{f\sharp g} \in \mathcal{S}^2$ since $\mathcal{S}^1 \subset \mathcal{S}^2$ and in turn $f\sharp g \in L^2_r(G)$ by Theorem C.4.1.

This manner of setting up the multiplication by tracing the quantization is not unique, see e.g. [34] for a similar construction for a Kohn-Nirenberg type quantization. Combining twisted convolutions with the result of Proposition C.4.5, we have all the ingredients we need to set up the desired isomorphism. However, we must first verify that the two spaces indeed are H^*-algebras in the first place.

Lemma C.4.7. *The Hilbert space $\mathcal{S}^2$ equipped with the operations of composition and taking the adjoint,*

$$A, B \mapsto A \circ B, \qquad A \mapsto A^*,$$

is an H^*-algebra.

Proof. All of the conditions of Definition C.4.6 are standard properties of Hilbert-Schmidt operators with the possible exception of submultiplicativity, item (ix), which follows from $\|TS\|_{\mathcal{S}^2} \leq \|T\|_{\mathcal{S}^2}\|S\|_\infty$ and $\|S\|_\infty \leq \|S\|_{\mathcal{S}^2}$, and adjoints in inner products which is equivalent to $\operatorname{tr}(ABC^*) = \operatorname{tr}(BC^*A)$ and follows from trace cyclicity. □

Lemma C.4.8. *The Hilbert space $L^2_r(G)$ equipped with the operations of twisted multiplication and complex conjugation,*

$$f, g \;\mapsto\; f\sharp g, \qquad f \;\mapsto\; \overline{f},$$

is an H^*-algebra.

Proof. We verify the associativity property (iii) for the twisted multiplication as

$$f\sharp(g\sharp h) = a_{A_f A_{g\sharp h}} = a_{A_f A_g A_h} = a_{A_{f\sharp g} A_h} = (f\sharp g)\sharp h.$$

Meanwhile submultiplicativity (ix) follows from

$$\|f\sharp g\|_{L^2_r} = \|A_{f\sharp g}\|_{\mathcal{S}^2} = \|A_f A_g\|_{\mathcal{S}^2} \le \|A_f\|_{\mathcal{S}^2}\|A_g\|_{\mathcal{S}^2} = \|f\|_{L^2_r}\|g\|_{L^2_r}.$$

Lastly for property (x) we use that the quantization mapping is unitary, that $\mathcal{S}^2$ is an H^*-algebra, Proposition C.4.5 and equation (C.4.2) to verify that

$$\begin{aligned}\langle f\sharp g, h\rangle_{L^2_r} = \langle A_f A_g, A_h\rangle_{\mathcal{S}^2} &= \langle A_g, A_f^* A_h\rangle_{\mathcal{S}^2} \\ &= \langle A_g, A_{\bar{f}} A_h\rangle_{\mathcal{S}^2} = \langle g, \bar{f}\sharp h\rangle_{L^2_r}.\end{aligned}$$

The remaining conditions are easy to verify and are skipped in the interest of brevity. □

We are now essentially done with the work of showing that the mapping is a $*$-isomorphism.

Theorem C.4.9. *The quantization mapping $A : L^2_r(G) \to \mathcal{S}^2$ as defined by* (C.4.1) *is an isometric $*$-isomorphism of the H^*-algebra $(L^2_r(G), \sharp, \overline{})$ onto the H^*-algebra $(\mathcal{S}^2, \circ, {}^*)$.*

Proof. The mapping being an isometry follows from Proposition C.3.4 and Proposition C.3.11. That $A_f^* = A_{\overline{f}}$ was established in Proposition C.4.5 and $A_{f\sharp g} = A_f A_g$ is the contents of (C.4.2). □

Intermediate twisted $*$-isomorphisms

While the twisted multiplication in (C.4.2) was crucial to setting up the H^*-algebra $*$-isomorphism, there is more structure induced by it available to uncover. Specifically, just as one can use the classical Fourier convolution theorem to define convolutions from multiplications on $L^2(\mathbb{R})$ as $f * g = \mathcal{F}^{-1}(\mathcal{F}(f)\cdot\mathcal{F}(g))$, we can also define an associated *twisted convolution*. The same construction has also been made in the Weyl-Heisenberg setting and again there is an explicit expression for the operation available which we will be unable to obtain in this more general setting. Denoting the twisted convolution by $\natural$, generalizing the convolution theorem means that we need to satisfy

$$\mathcal{F}_{\mathrm{KO}}(f \natural g) = \mathcal{F}_{\mathrm{KO}}(f) \sharp \mathcal{F}_{\mathrm{KO}}(g) \implies f \natural g = \mathcal{F}_{\mathrm{KO}}^{-1}\big(\mathcal{F}_{\mathrm{KO}}(f) \sharp \mathcal{F}_{\mathrm{KO}}(g)\big) \qquad \text{(C.4.3)}$$

since $\mathcal{F}_{\mathrm{KO}}$ is the Fourier transform in our setting and we use $\sharp$ as our multiplication. More explicitly, using $A_f = \mathcal{F}_{\mathrm{W}}^{-1}(\mathcal{F}_{\mathrm{KO}}^{-1}(f))$ and the definition of twisted convolution, we get that

$$\begin{aligned}
f \natural g &= \mathcal{F}_{\mathrm{KO}}^{-1}\left[a_{A_{\mathcal{F}_{\mathrm{KO}}(f)} A_{\mathcal{F}_{\mathrm{KO}}(g)}}\right] \\
&= \mathcal{F}_{\mathrm{KO}}^{-1}\left[a_{\mathcal{F}_{\mathrm{W}}^{-1}(f) \mathcal{F}_{\mathrm{W}}^{-1}(g)}\right] \\
&= \mathcal{F}_{\mathrm{KO}}^{-1}\left[\mathcal{F}_{\mathrm{KO}}(\mathcal{F}_{\mathrm{W}}(\mathcal{F}_{\mathrm{W}}^{-1}(f) \circ \mathcal{F}_{\mathrm{W}}^{-1}(g)))\right] \\
&= \mathcal{F}_{\mathrm{W}}\big(\mathcal{F}_{\mathrm{W}}^{-1}(f) \circ \mathcal{F}_{\mathrm{W}}^{-1}(g)\big). \qquad \text{(C.4.4)}
\end{aligned}$$

In view of the implication $\mathcal{F}_{\mathrm{W}}^{-1}(f \natural g) = \mathcal{F}_{\mathrm{W}}^{-1}(f) \circ \mathcal{F}_{\mathrm{W}}^{-1}(g)$, we could have obtained the same expression for $\natural$ by simply requiring that the triple $(\mathcal{F}_{\mathrm{W}}^{-1}, \natural, \circ)$ should have the same relation as the standard Fourier triple $(\mathcal{F}, *, \cdot)$ gets from the convolution theorem. From this point of view, it becomes clear that our original definition of $\sharp$ was induced by a similar relation but with $(A, \sharp, \circ)$ as the triple where $A : f \mapsto A_f$ is the quantization mapping.

So far we been endowing our quantization procedure A with additional structure. In this section we finish this task by showing that the diagram in Figure C.1 commutes.

$$\begin{tikzcd}
\big(\mathcal{S}^2(\mathcal{H}), \circ, {}^*\big) \arrow[d, "\mathcal{F}_{\mathrm{W}}"] \arrow[dr, dotted, "a"] & \\
\big(\mathcal{F}_{\mathrm{W}}(\mathcal{S}^2), \natural, \sqrt{\Delta(\cdot)}\,\check{\ }\big) \arrow[r, "\mathcal{F}_{\mathrm{KO}}"] & \big(L_r^2(G), \sharp, {}^-\big)
\end{tikzcd}$$

Figure C.1: Commutative diagram showing the twisted multiplication and twisted convolution which correspond to operator compositions.

Note that for the intermediate space we need to choose

$$\mathcal{F}_{\mathrm{W}}(\mathcal{S}^2) = \mathcal{F}_{\mathrm{KO}}^{-1}(L_r^2(G)) \subset L_r^2(G)$$

since this is the space that both $\mathcal{F}_{\mathrm{W}}$ and $\mathcal{F}_{\mathrm{KO}}^{-1}$ map to, not the full $L_r^2(G)$ space. Before showing that the mappings really respect involution and composition, we show that the intermediate object actually is an H^*-algebra.

Lemma C.4.10. *The triple* $\left(\mathcal{F}_{\mathrm{W}}(\mathcal{S}^2),\ \natural,\ \sqrt{\Delta(\cdot)}\,\overline{\check{\ }}\right)$ *is an* H^**-algebra.*

Proof. Linearity is clear while associativity of $\natural$ can be derived from Lemma C.4.7 or C.4.8 via (C.4.4) or (C.4.3). That $\sqrt{\Delta(\cdot)}\,\overline{\check{\ }}$ is an involution can be verified manually as

$$(f^*)^*(x) = \sqrt{\Delta(x)}\,\overline{\widecheck{\sqrt{\Delta(\cdot)}\,\overline{f(\cdot^{-1})}}} = \sqrt{\Delta(x)}\,\overline{\sqrt{\Delta(x^{-1})}\,\overline{f((x^{-1})^{-1})}} = f(x).$$

Meanwhile for property (vi), we can use Lemma C.3.12 to verify that, with $f^* = \sqrt{\Delta(\cdot)}\,\overline{\check{f}}$,

$$\begin{aligned}
f^*\natural g^* &= \mathcal{F}_{\mathrm{KO}}^{-1}\left(\mathcal{F}_{\mathrm{KO}}(f^*)\sharp\mathcal{F}_{\mathrm{KO}}(g^*)\right)\\
&= \mathcal{F}_{\mathrm{KO}}^{-1}\left(\overline{\mathcal{F}_{\mathrm{KO}}(f)}\sharp\overline{\mathcal{F}_{\mathrm{KO}}(g)}\right)\\
&= \mathcal{F}_{\mathrm{KO}}^{-1}\left(\overline{\mathcal{F}_{\mathrm{KO}}(g)\sharp\mathcal{F}_{\mathrm{KO}}(f)}\right)\\
&= \sqrt{\Delta(\cdot)}\,\overline{\mathcal{F}_{\mathrm{KO}}^{-1}(\widecheck{\mathcal{F}_{\mathrm{KO}}(g)\sharp}\mathcal{F}_{\mathrm{KO}}(f))}\\
&= (g\natural f)^*
\end{aligned}$$

as desired. That the involution is an isometry follows from the standard properties of the Haar measure summarized in Lemma C.2.4 while submultiplicativity of the twisted convolution can be verified by e.g. pulling back to $\mathcal{S}^2$ using (C.4.4). The same holds true for property (x) by the unitarity of $\mathcal{F}_{\mathrm{W}}$. □

All that remains now is to show that each of $\mathcal{F}_{\mathrm{W}}$ and $\mathcal{F}_{\mathrm{KO}}$ are isometric $*$-isomorphisms.

Proposition C.4.11. The mapping $\mathcal{F}_{\mathrm{W}} : \left(\mathcal{S}^2,\ \circ,\ ^*\right) \to \left(\mathcal{F}_{\mathrm{W}}(\mathcal{S}^2),\ \natural,\ \sqrt{\Delta(\cdot)}\,\overline{\check{\ }}\right)$ is an isometric $*$-isomorphism between H^* algebras.

Proof. The mapping being an isometry has already been shown in Proposition C.3.4 and the $*$-isomorphism property of multiplication was verified in (C.4.4). That the mapping respects the respective involutions can be verified by tracing out the diagram to yield

$$\mathcal{F}_{\mathrm{W}}(S^*) = \mathcal{F}_{\mathrm{KO}}^{-1}(\overline{a_S})$$

from which we get the desired conclusion through Lemma C.3.12. □

Proposition C.4.12. The mapping $\mathcal{F}_{\mathrm{KO}} : \left(\mathcal{F}_{\mathrm{W}}(\mathcal{S}^2),\ \natural,\ \sqrt{\Delta(\cdot)}\,\overline{\check{\ }}\right) \to \left(L_r^2(G),\ \sharp,\ ^{-}\right)$ is an isometric $*$-isomorphism between H^* algebras.

Proof. The mapping being an isometry follows from Proposition C.3.11 and the $*$-isometry properties follow from Lemma C.3.12 and (C.4.3). □

Combining these two propositions and Lemma C.4.10 with the results of Section C.4.2 allows us to conclude that the diagram in Figure C.1 indeed does commute.

C.4.3 Wigner distributions

As we saw in Section C.2.1, defining a Wigner distribution and setting up a quantization scheme are really two sides of the same coin. In this article, we take the view that it is the quantization that induces the Wigner distribution, meaning that we will define

$$W(\psi, \phi)(x) := a_{\psi\otimes\phi}(x) = \mathcal{F}_{\mathrm{KO}}(\mathcal{F}_{\mathrm{W}}(\psi \otimes \phi))(x)$$

for $\psi, \phi \in \mathcal{H}$ and $x \in G$. As we will see shortly, this definition is equivalent to $\langle f, W(\phi, \psi)\rangle_{L^2_r} = \langle A_f \psi, \phi\rangle_{\mathcal{H}}$ and so should be expected. Before moving on to properties of this function, we expand this definition and the associated inverse to get as close as possible to a computable formula. Using the result of Proposition C.3.3, the fact that $\mathcal{F}_{\mathrm{W}}$ can be extended to all of $\mathcal{S}^2$ using Proposition C.3.4 and the definition of the Fourier-Kirillov transform (C.3.2), we get

$$\begin{aligned} W(\psi, \phi)(x) &= \mathcal{F}_{\mathrm{KO}}(\mathcal{F}_{\mathrm{W}}(\psi \otimes \phi))(x) \\ &= \frac{1}{\sqrt{|\mathrm{Pf}_F| \cdot \Delta(x)}} \int_G \mathcal{W}_{\mathcal{D}\phi}\psi(y) e^{2\pi i \langle K(x^{-1})F, \log(y)\rangle} \frac{1}{\sqrt{\Theta(\log(y))}} \,\mathrm{d}\mu_r(y). \end{aligned}$$

The inverse operation, going from the Wigner distribution to the rank-one operator $\psi \otimes \phi$, can also be written out explicitly as

$$\begin{aligned} \psi \otimes \phi &= \mathcal{F}_{\mathrm{W}}^{-1}(\mathcal{F}_{\mathrm{KO}}^{-1}(W(\psi, \phi))) \\ &= \mathcal{F}_{\mathrm{W}}^{-1}\left(\sqrt{\Theta(\log(\cdot))} \int_G W(\psi, \phi)(x) e^{-2\pi i \langle K(x^{-1})F, \log(\cdot)\rangle} \frac{1}{\sqrt{|\mathrm{Pf}_F| \cdot \Delta(x)}} \,\mathrm{d}\mu_r(x) \right) \\ &= \int_G \int_G \sqrt{\Theta(\log(y^{-1}))} W(\psi, \phi)(x) e^{-2\pi i \langle K(x^{-1})F, \log(y^{-1})\rangle} \\ &\qquad \times \frac{1}{\sqrt{|\mathrm{Pf}_F| \cdot \Delta(x)}} \pi(y)\mathcal{D} \,\mathrm{d}\mu_r(x) \,\mathrm{d}\mu_r(y). \end{aligned} \tag{C.4.5}$$

Next we establish some of the basic properties of this function.

Proposition C.4.13. The mapping $(\psi, \phi) \mapsto W(\psi, \phi)$ is linear in the first argument and anti-linear in the second argument.

Proof. Using that $W(\psi, \phi) = a_{\psi \otimes \phi}$, we can see that this holds because $(\psi, \phi) \mapsto \psi \otimes \phi$ is linear in the first argument and antilinear in the second argument while $a : \mathcal{S}^2 \to L_r^2(G)$ is linear by Theorem C.4.1. □

Proposition C.4.14. Let $\psi, \phi \in \mathcal{H}$, then the Wigner distribution $W(\psi, \phi)$ belongs to $L_r^2(G)$ and satisfies the weak relation

$$\left\langle f, W(\phi, \psi)\right\rangle_{L_r^2} = \langle A_f \psi, \phi\rangle_{\mathcal{H}} \tag{C.4.6}$$

for $f \in L_r^2(G)$, as well as the identities

$$\begin{aligned} \overline{W(\psi, \phi)(x)} &= W(\phi, \psi)(x), \\ R_x W(\psi, \phi) &= W(\pi(x)\psi, \pi(x)\phi). \end{aligned}$$

Proof. We apply quantization to both sides of (C.4.6) and use the definition of inner products in the space of Hilbert-Schmidt operators and the definition of rank-one operators to get

$$\begin{aligned} \langle f, W(\phi, \psi)\rangle_{L_r^2} &= \langle A_f, \phi \otimes \psi\rangle_{\mathcal{S}^2} \\ &= \sum_n \langle A_f(\psi \otimes \phi)\xi_n, \xi_n\rangle_{\mathcal{H}} \\ &= \sum_n \langle \xi_n, \phi\rangle\langle A_f\psi, \xi_n\rangle_{\mathcal{H}} \\ &= \langle A_f\psi, \phi\rangle_{\mathcal{H}}. \end{aligned}$$

For the first of the two identities, we let $f \in L_r^2(G)$ be arbitrary and look at the weak action of $W(\psi, \phi)$ to conclude

$$\begin{aligned} \langle f, W(\psi, \phi)\rangle_{L_r^2} &= \langle A_f\phi, \psi\rangle_{\mathcal{H}} \\ &= \overline{\langle A_{\bar{f}}\psi, \phi\rangle_{\mathcal{H}}} \\ &= \left\langle f, \overline{W(\psi, \phi)}\right\rangle_{L_r^2}. \end{aligned}$$

Since f was arbitrary, it follows that $W(\psi, \phi) = \overline{W(\phi, \psi)}$.

Lastly, we compute the quantization of $R_x W(\psi, \phi)$ using Proposition C.4.4 as

$$R_x W(\psi, \phi) = R_x a_{\psi\otimes\phi} = a_{\pi(x^{-1})^*(\psi\otimes\phi)\pi(x^{-1})} = a_{\pi(x)\psi\otimes\pi(x)\phi} = W(\pi(x)\psi, \pi(x)\phi)$$

which completes the proof. □

The weak expression for the Wigner distribution (C.4.6) shows that our quantization procedure does not preserve positivity for the same reason as in the Weyl-Heisenberg case: For a non-negative f, $\langle A_f\psi, \psi\rangle = \langle f, W(\psi)\rangle$ can be negative at a point if $W(\psi)$ is negative which is the case for most ψ in the Weyl-Heisenberg case [107, 108].

By letting f be another Wigner distribution in (C.4.6) we immediately get the following generalization of Theorem C.2.1.

Proposition C.4.15. Let $\psi_1, \phi_1, \psi_2, \phi_2 \in \mathcal{H}$, then

$$\big\langle W(\psi_1, \phi_1), W(\psi_2, \phi_2)\big\rangle_{L^2_r} = \langle \psi_1, \psi_2\rangle_{\mathcal{H}} \overline{\langle \phi_1, \phi_2\rangle_{\mathcal{H}}}.$$

From the above proposition it is easy to see that for an orthonormal sequence $(\phi_n)_n$, the sequence $\big(W(\phi_n, \phi_m)\big)_{n,m}$ is also orthonormal. In fact, the property of being complete is also preserved through this mapping which generalizes the results [104, Prop. 188] and [31, Lem. 7.4].

Corollary C.4.16. Let $(\phi_n)_n$ be an orthonormal basis for $\mathcal{H}$, then the sequence $(W(\phi_n, \phi_m))_{n,m}$ is an orthonormal basis for $L^2_r(G)$.

Proof. As remarked above, preserving orthonormality follows directly from Proposition C.4.15. For completeness, suppose $f \in L^2_r(G)$ is orthogonal to $W(\phi_n, \phi_m)$ for all n and m. Then by (C.4.6), we also have that

$$0 = \langle f, W(\phi_n, \phi_m)\rangle_{L^2_r} = \langle A_f \phi_m, \phi_n\rangle_{\mathcal{H}}$$

for all n, m. However, since $(\phi_n)_n$ is an orthonormal basis for $\mathcal{H}$, this means that A_f must be the zero operator which in turn implies that f is the zero function, finishing the proof. □

From Proposition C.4.15 we also get a proof that the Wigner distribution is uniquely determined by the window ψ up to a unimodular constant. This is well known in the Weyl-Heisenberg case, see e.g. [84, Prop. 1.98]. Our proof will follow that for the affine Wigner distribution in [31, Cor. 3.3].

Proposition C.4.17. Let $\psi, \phi \in \mathcal{H}$. Then $W(\psi) = W(\phi)$ if and only if $\psi = c \cdot \phi$ with $|c| = 1$.

Proof. If $\psi = c \cdot \phi$, the result is clear as $c\phi \otimes c\phi = \phi \otimes \phi$ if $|c| = 1$. For the other direction, note that

$$\big|\langle \psi, \phi\rangle_{\mathcal{H}}\big|^2 = \big\langle W(\psi), W(\phi)\big\rangle_{L^2_r} = \underbrace{\|W(\psi)\|^2_{L^2_r}}_{=\|\psi\|^4_{\mathcal{H}}} = \underbrace{\|W(\phi)\|^2_{L^2_r}}_{=\|\phi\|^4_{\mathcal{H}}} = \|\psi\|^2_{\mathcal{H}}\|\phi\|^2_{\mathcal{H}}.$$

Consequently, if we apply Cauchy-Schwarz to the left-hand side, we must have equality which implies that ψ and ϕ are collinear. □

We can also deduce the following fact about the sums of Wigner distributions.

Proposition C.4.18. Let $(\phi_k)_k$ be an orthonormal basis of $\mathcal{H}$, then $A_{\sum_k^n W(\phi_k)}$ converges weakly to $I_{\mathcal{H}}$ as $n \to \infty$.

Proof. This ends up being a straight-forward verification. Indeed, for any $\psi, \phi \in \mathcal{H}$,

$$\begin{aligned} \lim_{n\to\infty} \left\langle A_{\sum_k^n W(\phi_k)} \psi, \phi \right\rangle &= \lim_{n\to\infty} \sum_{k=1}^{n} \left\langle W(\phi_k), W(\phi, \psi) \right\rangle \\ &= \sum_{k=1}^{\infty} \langle \phi_k, \phi \rangle \overline{\langle \phi_k, \psi \rangle} = \langle \psi, \phi \rangle \end{aligned}$$

where we used the linearity of quantization, Proposition C.4.14 and Proposition C.4.15. □

Lastly we return to the discussion on positivity and note that the following characterization of positivity is immediate from Proposition C.4.14.

Proposition C.4.19. An operator $S \in \mathcal{S}^2$ is positive if and only if

$$\int_G a_S(x) W(\psi)(x) \, \mathrm{d}\mu_r(x) \geq 0 \qquad \text{for all } \psi \in \mathcal{H}.$$

Note that in the Weyl-Heisenberg case, Wigner distributions $W(\psi)$ are not positive everywhere in general [128] so positivity of an operator is not equivalent to $a_S \geq 0$.

C.4.4 Operator convolutions from quantum harmonic analysis

Our interest in the quantization map partly stems from its interaction with the operator convolutions from quantum harmonic analysis, discussed in Section C.2.6. In this section we will show that the convolution-quantization relations outlined in Section C.2.6 can be generalized beyond the Weyl-Heisenberg group. Corresponding results were earlier shown for affine quantization in [32, Sec. 3.2].

First we show that quantization respects function-operator convolutions in a nice manner.

Proposition C.4.20. Let $f \in L^1_r(G)$ and $g \in L^2_r(G)$. Then

$$A_{f * g} = f \star A_g.$$

Proof. We explicitly compute the dequantization $a_{f\star A_g}$ and show that it is equal to $f * g$ which suffices by injectivity. To move the dequantization inside the integral defining $f \star A_g$, we use that bounded operators commute with convergent Bochner integrals (see [85] or [153, Prop. 2.4]) which yields

$$\begin{aligned} a_{f\star A_g}(x) &= \mathcal{F}_{\mathrm{KO}}\left(\mathcal{F}_{\mathrm{W}}\left(\int_G f(y)\pi(y)^* A_g \pi(y)\,\mathrm{d}\mu_r(y)\right)\right)(x) \\ &= \int_G f(y)\mathcal{F}_{\mathrm{KO}}\big(\mathcal{F}_{\mathrm{W}}(\pi(y)^* A_g \pi(y))\big)(x)\,\mathrm{d}\mu_r(y) \\ &= \int_G f(y) R_{x^{-1}} g(y)\,\mathrm{d}\mu_r(y) \\ &= \int_G f(y) g(xy^{-1})\,\mathrm{d}\mu_r(y) = f * g \end{aligned}$$

where we used Proposition C.4.4 for the translation. □

For operator-operator convolutions we will need to keep track of an involution on one of the functions in the same way as for affine quantization [32, Prop. 3.7].

Proposition C.4.21. Let $f, g \in L^2_r(G)$, then

$$A_f \star A_g = f * \check{g}.$$

Proof. We compute

$$\begin{aligned} (A_f \star A_g)(x) &= \operatorname{tr}\big(A_f \pi(x)^* A_g \pi(x)\big) \\ &= \big\langle A_f, \pi(x)^* A_{\bar{g}} \pi(x)\big\rangle_{\mathcal{S}^2} \\ &= \big\langle f, R_{x^{-1}} \bar{g}\big\rangle_{L^2_r} \\ &= \int_G f(y)\check{g}(xy^{-1})\,\mathrm{d}\mu_r(y) = f * \check{g}(x) \end{aligned}$$

where we used that $A_f^* = A_{\bar{f}}$ from Proposition C.4.5. □

The rank-one realization of this result is of independent interest and implies a relation for the Wigner distribution which is well known in the Weyl-Heisenberg case.

Corollary C.4.22. Let $\psi_1, \psi_2, \phi_1, \phi_2 \in \mathcal{H}$, then

$$W(\psi_1, \psi_2) * \widecheck{W(\phi_2, \phi_1)}(x) = \mathcal{W}_{\phi_1}\psi_1(x)\overline{\mathcal{W}_{\phi_2}\psi_2(x)}.$$

In particular,

$$W(\psi) * \widecheck{W(\phi)}(x) = |\mathcal{W}_\phi \psi(x)|^2.$$

Proof. Since $A_{W(\psi_1,\psi_2)} = \psi_1 \otimes \psi_2$, we can use Proposition C.4.21 to write the convolution as

$$\begin{aligned}
W(\psi_1,\psi_2) * \widecheck{W(\phi_2,\phi_1)}(x) &= (\psi_1 \otimes \psi_2) \star (\phi_2 \otimes \phi_1)(x) \\
&= \operatorname{tr}\big((\psi_1 \otimes \psi_2)\pi(x)^*(\phi_2 \otimes \phi_1)\pi(x)\big) \\
&= \sum_n \langle \pi(x)^*(\phi_2 \otimes \phi_1)\pi(x)e_n, \psi_2\rangle\langle \psi_1, e_n\rangle \\
&= \sum_n \langle e_n, \pi(x)^*\phi_1\rangle\langle \phi_2, \pi(x)\psi_2\rangle\langle \psi_1, e_n\rangle \\
&= \langle \psi_1, \pi(x)^*\phi_1\rangle\overline{\langle \psi_2, \pi(x)^*\phi_2\rangle} = \mathcal{W}_{\phi_1}\psi_1(x)\overline{\mathcal{W}_{\phi_2}\psi_2(x)}
\end{aligned}$$

which is what we wished to show. □

By combining Proposition C.4.21 and Proposition C.4.20 we can also get an expression for $A_{T\star S}$. However, to generalize the classical result for the Weyl-Heisenberg group we need a notion of parity conjugating an operator. Since we don't have a parity operator in the general case, we cannot approach this issue head-on but can instead define $\check{S}$ using the quantization a_S of S so that it generalizes the classical property $A_{\check{a_S}} = \check{S}$ [153, Lem. 3.2 (ii)]. An immediate consequence of this definition is that $a_{\check{S}} = \check{a_S}$. We also get the following corollary from the earlier results.

Corollary C.4.23. Let $T, S \in \mathcal{S}^2$ be such that $a_T \in L^1_r(G)$, then

$$A_{T\star S} = a_T \star A_{\check{a_S}} = a_T \star \check{S}.$$

Proof. By first using that $T = A_{a_T}$, $S = A_{a_S}$ and then applying Proposition C.4.21 followed by Proposition C.4.20, we get

$$A_{T\star S} = A_{A_{a_T}\star A_{a_S}} = A_{a_T * \check{a_S}} = a_T \star A_{\check{a_S}} = a_T \star \check{S}$$

since this is how we defined $\check{S}$. □

C.4.5 Unimodular trace formula

In 2000, Du and Wong [63] were the first to show that for Weyl quantization on the Weyl-Heisenberg group, $\operatorname{tr}(A_f) = \int_{\mathbb{R}^{2d}} f(x)\,\mathrm{d}x$ which is called a *trace formula*, see [104, Prop. 286] and [153, Cor. 6.2.1] for other proofs. This result was generalized to the affine group in [32, Prop. 4.14] in two ways,

$$\int_G f(x)\,\mathrm{d}\mu_r(x) = \operatorname{tr}(A_f), \qquad \int_G f(x)\,\mathrm{d}\mu_l(x) = \operatorname{tr}(\mathcal{D}^{-1}A_f\mathcal{D}^{-1}). \qquad \text{(C.4.7)}$$

We will show the corresponding result under the assumption that the group is unimodular but remark that (C.4.7) is likely to hold in the general case as well.

Theorem C.4.24. *Assume that the underlying group is unimodular and let $f \in L^1(G)$ be such that $A_f \in \mathcal{S}^1$, then*

$$\frac{1}{\sqrt{\mathrm{Pf}_F}} \int_G f(x)\,\mathrm{d}\mu(x) = \mathrm{tr}(A_f). \tag{C.4.8}$$

Proof. We will begin by showing the first relation in the special case where f is continuous with compact support. Looking at the definition of the inverse Fourier-Kirillov transform (C.3.3), we see that the integral can then be identified with $\mathcal{F}_{\mathrm{KO}}^{-1}(f)(e)$. That it is okay to evaluate $\mathcal{F}_{\mathrm{KO}}^{-1}(f)$ at a point follows from the Riemann-Lebesgue lemma and that the mappings in Proposition C.3.11 are well behaved when the group is unimodular. Since $f \in L^2(G)$ we can also identify f with $a_{A_f} = \mathcal{F}_{\mathrm{KO}}(\mathcal{F}_{\mathrm{W}}(A_f))$. Now by applying Lemma C.3.15, we get

$$\begin{aligned}
\frac{1}{\sqrt{\mathrm{Pf}_F}} \int_G f(x)\,\mathrm{d}\mu(x) &= \mathcal{F}_{\mathrm{KO}}^{-1}(\mathcal{F}_{\mathrm{KO}}(\mathcal{F}_{\mathrm{W}}(A_f)))(e) \\
&= \mathcal{F}_{\mathrm{W}}(\mathcal{F}_{\mathrm{W}}^{-1}(\mathcal{F}_{\mathrm{W}}(A_f)))(e) \\
&= \mathcal{F}_{\mathrm{W}}(A_f)(e) \\
&= \mathrm{tr}(A_f)
\end{aligned}$$

where we used the continuity of $\mathcal{F}_{\mathrm{W}}(A_f)$ afforded by Proposition C.3.6 in the last step.

To drop the assumption of continuity and compact support of f, fix a continuous function g with compact support. From (C.2.15) in Section C.2.6 and Proposition C.4.21 above, we then get that

$$\begin{aligned}
\mathrm{tr}(A_f)\,\mathrm{tr}(A_g) &= \int_G A_f \star A_g(x)\,\mathrm{d}\mu(x) \\
&= \int_G f * \check{g}(x)\,\mathrm{d}\mu(x) \\
&= \int_G f(x)\,\mathrm{d}\mu(x) \int_G \check{g}(x)\,\mathrm{d}\mu(x) = \int_G f(x)\,\mathrm{d}\mu(x) \int_G g(x)\,\mathrm{d}\mu(x)
\end{aligned}$$

where we in the last step used the unimodularity of the group. Now dividing by $\mathrm{tr}(A_g)$ on both sides and using that $\mathrm{tr}(A_g) = \frac{1}{\sqrt{\mathrm{Pf}_F}} \int_G g(x)\,\mathrm{d}\mu(x)$, we get the desired equality. □

Remark C.4.25. One immediate consequence of this result is that for a Wigner distribution $W(\psi, \phi)$,

$$\frac{1}{\sqrt{\mathrm{Pf}_F}} \int_G W(\psi, \phi)(x)\,\mathrm{d}\mu(x) = \langle \psi, \phi \rangle.$$

With the above result, we can obtain a briefer (albeit less elementary) proof of [107, Thm. 4.4.6 (a)] and generalize it to all unimodular groups.

Corollary C.4.26. Assume that the underlying group is unimodular and suppose that $f \in L^2(G) \cap L^1(G)$ is such that $A_f \in \mathcal{S}^1$ and A_f is a positive operator, then

$$\|f\|_{L^2(G)} \leq \frac{1}{\sqrt{\mathrm{Pf}_F}} \int_G f(x)\, \mathrm{d}\mu(x).$$

Proof. Since $A_f \in \mathcal{S}^1$ is a positive operator, its singular value decomposition is of the form $A_f = \sum_n \lambda_n (\phi_n \otimes \phi_n)$ where $(\phi_n)_n$ is some orthonormal basis of $\mathcal{H}$ and all λ_n are non-negative. By Theorem C.4.24, we can write

$$\begin{aligned}\frac{1}{\sqrt{\mathrm{Pf}_F}} \int_G f(x)\, \mathrm{d}\mu(x) &= \mathrm{tr}(A_f)\\ &= \sum_n \lambda_n = \left\|(\lambda_n)_n\right\|_{\ell^1} \geq \left\|(\lambda_n)_n\right\|_{\ell^2} = \|A_f\|_{\mathcal{S}^2} = \|f\|_{L^2(G)}\end{aligned}$$

where we used the embedding $\ell^1 \subseteq \ell^2$ and that quantization is an isometry. □

This result implies another immediate corollary.

Corollary C.4.27. Suppose that $0 \neq f \in L^2(G) \cap L^1(G)$ is non-positive everywhere and such that $A_f \in \mathcal{S}^1$, then A_f cannot be a positive operator.

C.5 Applications

Classical Weyl quantization has historically proved to be a valuable tool to tackle a wide variety of problems. In this section, we set out to show how Weyl quantization on exponential groups can be used to treat analog problems. Specifically we will show three results which all have counterparts with the classical Weyl transform and the Weyl-Heisenberg group but for which the generalizations are novel and one proof which is simplified by the tools developed in this article.

C.5.1 Wavelet phase retrieval

The problem of *phase retrieval* in the context of time-frequency analysis may be summarized as inverting the mapping

$$|\mathcal{W}_\phi|^2 : \psi \mapsto |\langle \psi, \pi(\cdot)^* \phi \rangle|^2. \tag{C.5.1}$$

Since ψ can be recovered from $\mathcal{W}_\phi \psi$, this task is equivalent to recovering the phase from the (complex-valued) wavelet transform $\mathcal{W}_\phi \psi$, hence the name. Historically, this problem has been mainly considered for the Weyl-Heisenberg group where we are asked to recover the phase of the short-time Fourier transform $V_\phi\psi$. Outside of the Weyl-Heisenberg case, characterizing solvability of the problem has been notoriously difficult and has been an open question for many years. In the case of the affine group where $|\mathcal{W}_\phi\psi|^2$ is called the *scalogram*, there have been partial results involving sufficient conditions on the wavelet ϕ but there is no general condition as in the spectrogram case. This is an active area of research with many applications and recent results, see e.g. [11, 89, 159, 205].

In the Weyl-Heisenberg case, solvability is characterized by the Fourier transform of $W(\phi)$ being nonvanishing. This can be seen from the formula

$$|V_\phi\psi|^2 = W(\psi) * \widetilde{W(\phi)} \tag{C.5.2}$$

from Corollary C.4.22, by taking the Fourier transform on both sides and dividing away the Fourier transform of $\widetilde{W(\phi)}$. In the general case we do not have such a nice Fourier convolution theorem to rely on but we still have the relation

$$|\mathcal{W}_\phi\psi(x)|^2 = (\psi \otimes \psi) \star (\phi \otimes \phi)(x) = W(\psi) * \widetilde{W(\phi)}(x). \tag{C.5.3}$$

As we saw in Proposition C.4.17, the Wigner distribution $W(\psi)$ determines ψ up to a unimodular constant. Formally, we can recover ψ up to this phase constant by quantizing $W(\psi)$ to get $\psi \otimes \psi$ and applying the resulting operator to any element ξ which is not orthogonal to ψ as outlined in (C.4.5). This is written as

$$\psi = \frac{A_{W(\psi)}\xi}{\|A_{W(\psi)}\xi\|} \tag{C.5.4}$$

since $A_{W(\psi)} = \psi \otimes \psi$. Consequently, we wish to find conditions on ϕ so that we may manipulate (C.5.3) as to get an explicit expression for $W(\psi)$. This requires being able to "disentangle" the squared wavelet transform $|\mathcal{W}_\phi\psi|^2$, our tool for this will be the following lemma.

Lemma C.5.1. *Let $f, g \in L^1_l(G)$ be such that $f\Delta \in L^1_r(G)$ and π, π_r be the left and right integrated representations. Then*

$$\pi(f * g) = \pi(g)\pi_r(f\Delta).$$

Remark C.5.2. Recall that $f * g$ here and throughout the paper denotes *right* convolution.

Proof. Using the substitution $z = xy^{-1} \implies x = zy$, we compute

$$\begin{aligned}
\pi(f * g) &= \int_G (f * g)(x)\pi(x)\,\mathrm{d}\mu_l(x) = \int_G \left(\int_G f(y)g(xy^{-1})\,\mathrm{d}\mu_r(y) \right) \pi(x)\,\mathrm{d}\mu_l(x) \\
&= \int_G \int_G f(y)g(z)\pi(z)\pi(y)\,\mathrm{d}\mu_r(y)\,\mathrm{d}\mu_l(zy) \\
&= \left(\int_G g(z)\pi(z)\,\mathrm{d}\mu_l(z) \right) \left(\int_G f(y)\Delta(y)\pi(y)\,\mathrm{d}\mu_r(y) \right) = \pi(g)\pi_r(f\Delta).
\end{aligned}$$

□

We will also need the following auxiliary lemma which is easily verified.

Lemma C.5.3. *Let $f \in L^1_r(G)$ and $g \in L^1_l(G)$, then*

$$\widetilde{f * \check{g}}(x) = g * \check{f}(x).$$

Proof. With the substitution $z = yx$, we compute

$$\begin{aligned}
\widetilde{f * \check{g}}(x) = f * \check{g}(x^{-1}) &= \int_G f(y)\check{g}(x^{-1}y^{-1})\,\mathrm{d}\mu_r(y) \\
&= \int_G f(y)g(yx)\,\mathrm{d}\mu_r(y) \\
&= \int_G g(z)f(zx^{-1})\,\mathrm{d}\mu_r(zx^{-1}) \\
&= \int_G g(z)\check{f}(xz^{-1})\,\mathrm{d}\mu_r(z) = g * \check{f}(x)
\end{aligned}$$

which is what we wished to show. □

With these two lemmas, we will be able to wrangle (C.5.3) into a form where ψ can conceivably be recovered.

Theorem C.5.4. *If $\phi \in \mathcal{H}$ is such that $W(\phi) \in L^1_l(G)$, $W(\phi)\Delta \in L^1_r(G)$ and the right integrated representation $\pi_r(W(\phi)\Delta)$ has a right inverse, then the mapping $\psi \mapsto |\mathcal{W}_\phi\psi|^2$ for ψ such that $W(\psi) \in L^1_r(G)$ can be inverted up to a global phase as*

$$W(\psi) = \mathcal{F}_{\mathrm{W}}\Big(\pi\big(\widetilde{|\mathcal{W}_\phi\psi|^2}\big)\pi_r\big(W(\phi)\Delta\big)^{-1}\mathcal{D}\Big)$$

followed by

$$\psi = \frac{\mathcal{F}_{\mathrm{W}}^{-1}\big(\mathcal{F}_{\mathrm{KO}}^{-1}\big(W(\psi)\big)\big)\xi}{\|\mathcal{F}_{\mathrm{W}}^{-1}\big(\mathcal{F}_{\mathrm{KO}}^{-1}\big(W(\psi)\big)\big)\xi\|} \tag{C.5.5}$$

where $\xi \in \mathcal{H}$ is any vector which is not orthogonal to ψ.

Proof. We compute the left integrated representation of the squared modulus of the wavelet transform as

$$\begin{aligned}\pi\big(\widecheck{|\mathcal{W}_\phi\psi|^2}\big) &= \pi\big(W(\psi) * \widecheck{\widecheck{W(\phi)}}\big)\\ &= \pi\big(W(\phi) * \widecheck{W(\psi)}\big)\\ &= \pi\big(\widecheck{W(\psi)}\big)\pi_r\big(W(\phi)\Delta\big).\end{aligned}$$

Now if $\pi_r\big(W(\phi)\Delta\big)$ has a right inverse, we can apply it to $\pi\big(\widecheck{|\mathcal{W}_\phi\psi|^2}\big)$ to get that

$$\begin{aligned}\pi\big(\widecheck{|\mathcal{W}_\phi\psi|^2}\big)\pi_r\big(W(\phi)\Delta\big)^{-1} &= \pi\big(\widecheck{W(\psi)}\big)\\ \implies \mathcal{F}_{\mathrm{W}}^{-1}\big(W(\psi)\big) &= \pi\big(\widecheck{|\mathcal{W}_\phi\psi|^2}\big)\pi_r\big(W(\phi)\Delta\big)^{-1}\mathcal{D}\\ \implies W(\psi) &= \mathcal{F}_{\mathrm{W}}\Big(\pi\big(\widecheck{|\mathcal{W}_\phi\psi|^2}\big)\pi_r\big(W(\phi)\Delta\big)^{-1}\mathcal{D}\Big).\end{aligned}$$

From here the explicit expression (C.5.5) follows from applying (C.5.4). □

Remark C.5.5. The same argument can be used to show invertibility of the mapping $\psi \mapsto Q_S(\psi)$ where Q_S is the Cohen's class distribution with operator window S (see [115, Sec. 5.1] and [154, Sec. 7.5]). In that case, the condition becomes right invertibility of $\pi_r(a_S\Delta)$.

C.5.2 Wigner approximation problem

The problem of approximating arbitrary $L^2(\mathbb{R}^{2d})$ functions by Wigner distributions was first considered for the Weyl-Heisenberg group in [26] and later extended to the affine setting in [31, Sec. 8.1]. Since we have constructed a Wigner distribution in our setting, we can formulate the corresponding problem as follows: Given a group G with associated Hilbert space $\mathcal{H}$, how close is a function $f \in L_r^2(G)$ to the *Wigner space*

$$\mathfrak{W}(G) = \big\{W(\psi) : \psi \in \mathcal{H}\big\} \subset L_r^2(G)?$$

As self-adjoint rank-one operators are a proper subset of the space of Hilbert-Schmidt operators, the set $\mathfrak{W}(G)$ is also a proper subset of $L_r^2(G)$. Moreover, using the orthogonality relations for Wigner distributions, Proposition C.4.15, we can see that this space is closed.

In [31], the role of quantization in solving the problem is emphasized and their approach carries over with minor modifications to our setting since we have similar quantization properties. The main idea can be summarized as "*via quantization, being close to being a Wigner distribution corresponds to almost being a rank-one operator*".

Theorem C.5.6. *Fix a real-valued $f \in L^2_r(G)$ with quantization A_f and let $\lambda^+_{\max}(A_f)$ be the positive part of $\max_{\lambda \in \mathrm{Spec}(A_f)} \lambda$, then*

$$\inf_{g \in \mathfrak{W}(G)} \|f - g\|_{L^2_r} = \sqrt{\|f\|^2_{L^2_r} - \lambda^+_{\max}(A_f)^2}.$$

The infimum is always attained and the number of unique minimizers (up to multiplication by a unimodular constant) is equal to the multiplicity of $\lambda^+_{\max}(A_f)$.

In the interest of completeness, we write out a proof of the above theorem but remark that we follow the same strategy as in [31].

Proof. By quantization being a unitary isometry, we have that

$$\inf_{g \in \mathfrak{W}(G)} \|f - g\|_{L^2_r} = \inf_{\psi \in \mathcal{H}} \|A_f - \psi \otimes \psi\|_{\mathcal{S}^2}. \tag{C.5.6}$$

We wish to apply the spectral theorem to A_f and so first note that it is compact by virtue of being a Hilbert-Schmidt operator and self-adjoint by real-valuedness via Proposition C.4.5. Hence its eigendecomposition can be written as

$$A_f = \sum_k \lambda_k (\phi_k \otimes \phi_k)$$

with convergence in the Hilbert-Schmidt norm. We can now rewrite (C.5.6) as

$$\inf_{g \in \mathfrak{W}(G)} \|f - g\|_{L^2_r} = \inf_{\psi \in \mathcal{H}} \left\| \sum_k \lambda_k (\phi_k \otimes \phi_k) - \psi \otimes \psi \right\|_{\mathcal{S}^2}.$$

By the orthogonality of $(\phi_k)_k$, this quantity is minimized when $\psi = \sqrt{\lambda_j}\phi_j$ where λ_j and ϕ_j are the eigendata corresponding to the largest positive eigenvalue. The reformulation as $\sqrt{\|f\|^2_{L^2_r} - \lambda^+_{\max}(A_f)^2}$ also follows from the orthogonality of $(\phi_k)_k$. Lastly the statement about non-unique minimizers follows from the ambiguity associated with choosing λ_j and ϕ_j. □

C.5.3 Wavelet/Wigner spaces have trivial intersection

The intersections of wavelet spaces, meaning the images of $\mathcal{H}$ under the mapping $\mathcal{W}_\phi : \psi \mapsto \langle \psi, \pi(\cdot)\phi \rangle$, have been studied in [28, 98] where it was shown that the intersection $\mathcal{W}_{\phi_1}(\mathcal{H}) \cap \mathcal{W}_{\phi_2}(\mathcal{H})$ is trivial unless $\phi_1 = c\phi_2$ for some complex number c. Recently, Skrettingland and Luef provided a simplified proof for Gabor spaces in [156, Lem. 3.3] which depended on a Weyl-Heisenberg version of our Proposition C.3.3. Below we show that Proposition C.3.3 can simplify the proof of [98, Thm. 4.2] in our setting of exponential groups.

Proposition C.5.7. Let $\phi_1, \phi_2 \in \mathrm{Dom}(\mathcal{D}^{-1})$. If there exists $c \in \mathbb{C}$ such that $\phi_1 = c\phi_2$, then $\mathcal{W}_{\phi_1}(\mathcal{H}) = \mathcal{W}_{\phi_2}(\mathcal{H})$. Otherwise, $\mathcal{W}_{\phi_1}(\mathcal{H}) \cap \mathcal{W}_{\phi_2}(\mathcal{H}) = \{0\}$.

Proof. The forward implication follows from the relation $\mathcal{W}_{c\phi_2}(\xi) = \mathcal{W}_{\phi_2}(\bar{c}\xi)$. For the other direction, we have the following chain of implications

$$\begin{aligned}
& \mathcal{W}_{\phi_1}(\psi_1) = \mathcal{W}_{\phi_2}(\psi_2) & \\
\implies\ & \langle \psi_1, \pi(\cdot)^*\phi_1\rangle = \langle \psi_2, \pi(\cdot)^*\phi_2\rangle & \text{(Definition of } \mathcal{W}_\phi(\psi)) \\
\implies\ & \mathcal{F}_{\mathrm{W}}(\psi_1 \otimes \mathcal{D}^{-1}\phi_1) = \mathcal{F}_{\mathrm{W}}(\psi_2 \otimes \mathcal{D}^{-1}\phi_2) & \text{(Applying } \mathcal{F}_{\mathrm{W}} \text{ to both sides)} \\
\implies\ & \psi_1 \otimes \mathcal{D}^{-1}\phi_1 = \psi_2 \otimes \mathcal{D}^{-1}\phi_2 & \text{(Injectivity of } \mathcal{F}_{\mathrm{W}}) \\
\implies\ & \mathcal{D}^{-1}\phi_1 \otimes \psi_1 = \mathcal{D}^{-1}\phi_2 \otimes \psi_2 & \text{(Taking adjoints)} \\
\implies\ & \|\psi_1\|^2 \mathcal{D}^{-1}\phi_1 = \langle \psi_1, \psi_2\rangle \mathcal{D}^{-1}\phi_2 & \text{(Applying to } \psi_1) \\
\implies\ & \phi_1 = \frac{\langle \psi_1, \psi_2\rangle}{\|\psi_1\|^2}\phi_2 & \text{(Solving for } \phi_1)
\end{aligned}$$

which finishes the proof. □

Cross-Wigner distributions $W(\psi, \phi)$ are sometimes used as time-frequency distributions with ϕ being viewed as the window. For these purposes, define the cross-Wigner space $W(\mathcal{H}, \phi)$ as

$$W(\mathcal{H}, \phi) = \big\{W(\psi, \phi) : \psi \in \mathcal{H}\big\} \subset L^2_r(G).$$

In this case we have the same type of result as for wavelet spaces with an even shorter proof.

Proposition C.5.8. Let $\phi_1, \phi_2 \in \mathcal{H}$. If there exists $c \in \mathbb{C}$ such that $\phi_1 = c\phi_2$, then $W(\mathcal{H}, \phi_1) = W(\mathcal{H}, \phi_2)$. Otherwise $W(\mathcal{H}, \phi_1) \cap W(\mathcal{H}, \phi_2) = \{0\}$.

Proof. The forward implication again follows from $W(\xi, c\phi_2) = W(\bar{c}\xi, \phi_2)$ by the sesquilinearity of the cross-Wigner distribution from Proposition C.4.13.

Suppose there exist $\psi_1, \psi_2 \in \mathcal{H}$ such that $W(\psi_1, \phi_1) = W(\psi_2, \phi_2)$. We will show that this implies that $\phi_1 = c\phi_2$. Indeed, by the equality

$$\big\langle W(\psi_1, \phi_1), W(\psi_2, \phi_2)\big\rangle = \langle \psi_1, \psi_2\rangle \overline{\langle \phi_1, \phi_2\rangle}$$

from Proposition C.4.15, and by Cauchy-Schwarz, we must have proportionality in each of the inner products on the right hand side because we have it on the left hand side. □

C.5.4 L^p bounds on operator convolutions

Using the link between quantization and operator convolutions from quantum harmonic analysis established in Section C.4.4, we can obtain new results on integrability of operator convolutions. Before stating our result, we briefly survey what bounds have previously been available.

The original article on quantum harmonic analysis by Werner [211] included a version of Young's inequality stated purely for operators, see [211, Prop. 3.2 (5)], of the form

$$\|T \star S\|_{L^r(\mathbb{R}^{2d})} \leq \|T\|_{\mathcal{S}^p} \|S\|_{\mathcal{S}^q} \tag{C.5.7}$$

where $\frac{1}{p} + \frac{1}{q} = 1 + \frac{1}{r}$. It should be stressed that this was only shown for the Weyl-Heisenberg group.

In the generalization of quantum harmonic analysis to the affine group and locally compact groups [32, 115], the $p = q = r = 1$ base case of (C.5.7) needed to take into consideration the notion of admissibility, yielding $\|T \star S\|_{L^1_r(G)} \leq \|T\|_{\mathcal{S}^1} \|\mathcal{D}^{-1} S \mathcal{D}^{-1}\|_{\mathcal{S}^1}$. Interpolating this leads to the inequality

$$\|T \star S\|_{L^p_r(G)} \leq \|T\|_{\mathcal{S}^p} \|S\|_{\mathcal{S}^1}^{\frac{1}{q}} \left\|\mathcal{D}^{-1} S \mathcal{D}^{-1}\right\|_{\mathcal{S}^1}^{1/p}$$

for $\frac{1}{p} + \frac{1}{q} = 1$, see [115, Prop. 4.13 (ii)] or [32, Prop. 4.18 (2)] for the proof.

Interpolating the mapping $K_S : T \mapsto T \star S$ between the above and the standard Schatten norm inequality $\|T \star S\|_{L^\infty} \leq \|T\|_{\mathcal{S}^q} \|S\|_{\mathcal{S}^p}$ yields

$$\|T \star S\|_{L^{q_\theta}_r(G)} \leq \left(\|S\|_{\mathcal{S}^1}^{1/q} \left\|\mathcal{D}^{-1} S \mathcal{D}^{-1}\right\|_{\mathcal{S}^1}^{1/p} \right)^{1-\theta} \|S\|_{\mathcal{S}^p}^{\theta} \|T\|_{\mathcal{S}^{p_\theta}}$$

where $\frac{1}{p} + \frac{1}{q} = 1$, $\theta \in (0, 1)$, $\frac{1}{p_\theta} = \frac{\theta}{q} + \frac{1-\theta}{p}$ and $\frac{1}{q_\theta} = \frac{1-\theta}{p}$. However, we do not get the standard $1 + \frac{1}{r} = \frac{1}{p} + \frac{1}{q}$ structure from Young's inequality due to how we must treat the admissibility constant.

These approaches have all been operator-based and we now turn our attention to the construction of a quantization-based bound which essentially follows from Young's inequality for locally compact groups.

Proposition C.5.9. Let $T, S \in \mathcal{S}^2$ and $\frac{1}{p} + \frac{1}{q} = 1 + \frac{1}{r}$, $\frac{1}{q} + \frac{1}{q'} = 1$ with $1 \leq p, q, q', r \leq \infty$, then

$$\|T \star S\|_{L^r_r(G)} \leq \left\|a_T \Delta^{1/q'}\right\|_{L^p_r(G)} \left\|a_S\right\|_{L^q_l(G)}.$$

Proof. The key tool we will use is Young's inequality for locally compact groups [139, Lem. 2.1] which can be stated as

$$\|f *_G g\|_{L^r_r(G)} \leq \left\|f \Delta^{1/q'}\right\|_{L^p_r(G)} \|g\|_{L^q_r(G)} \tag{C.5.8}$$

with p, q, q', r as in the proposition. Proposition C.4.21 now immediately yields that

$$\|T \star S\|_{L^r_r(G)} = \|a_T * \check{a_S}\|_{L^r_r(G)}$$

and so our quantity is in the form of (C.5.8). We set $f = a_T,\ g = \check{a_S}$ and apply the inequality to obtain

$$\|T \star S\|_{L^r_r(G)} \leq \left\|a_T \Delta^{1/q'}\right\|_{L^p_r(G)} \|\check{a_S}\|_{L^q_r(G)} = \left\|a_T \Delta^{1/q'}\right\|_{L^p_r(G)} \|a_S\|_{L^q_l(G)}$$

which is what we wished to show. $\square$

Paper D

Five Ways to Recover the Symbol of a Non-Binary Localization Operator

Simon Halvdansson

Preprint Published in
Journal of Pseudo-Differential Operators and Applications,
2025, Volume 16

Paper D

Five Ways to Recover the Symbol of a Non-Binary Localization Operator

Abstract

Five constructive methods for recovering the symbol of a time-frequency localization operator with non-binary symbol are presented, two based on earlier work and three novel methods. For the two derivative methods which have previously been applied to binary symbols, we propose a changed symbol estimator and provide additional estimates that show how we can recover non-binary symbols. The three novel methods each have their own advantages and are all applicable to non-binary symbols. Two of them rely on prescribing the input of the localization operator and examining the output, allowing for targeting of the part of the symbol one wishes to recover while the last one relies on spectral information about the operator. All five methods are also implemented numerically and evaluated with the code available.

D.1 Introduction and main results

Arguably the main tool of time-frequency analysis is the *short-time Fourier transform*, defined for a signal $\psi \in L^2(\mathbb{R}^d)$ and window $g \in L^2(\mathbb{R}^d)$ as

$$V_g\psi(x,\omega) = \int_{\mathbb{R}^d} \psi(t)\overline{g(t-x)}e^{-2\pi i\omega t}\,dt$$

where the variables $x, \omega \in \mathbb{R}^d$ are referred to as the time and frequency, respectively. A standard result [107] states that this mapping can be inverted so that the signal

ψ can be recovered from $V_g\psi$ weakly as

$$\psi = \int_{\mathbb{R}^{2d}} V_g\psi(x,\omega)g(\cdot - x)e^{2\pi i\omega\cdot}\,dx\,d\omega$$

when g is normalized so that $\|g\|_{L^2} = 1$. Using a function $f : \mathbb{R}^{2d} \to \mathbb{R}$, usually referred to as the *symbol* or *mask*, we can weigh this reconstruction so that certain frequencies and time intervals have more or less priority than others. Formally, this happens via the *localization operator*

$$A_f^g : \psi \mapsto \int_{\mathbb{R}^{2d}} f(x,\omega)V_g\psi(x,\omega)g(\cdot - x)e^{2\pi i\omega t}\,dx\,d\omega. \tag{D.1.1}$$

Such operators have applications in signal analysis [141, 167, 178, 199], acoustics [38, 120, 206], pseudo-differential operators [109, 126], physics [27, 104] and operator theory [90, 109, 156] among others and their properties have been extensively studied [41, 45, 83, 220]. Abreu and Dörfler [7] first considered the inverse problem of recovering the symbol f from the localization operator A_f^g through various measurements related to A_f^g and this work has been continued in [7, 8, 155, 182]. All these investigations have been focused on the case where f is a *binary mask*, i.e., $f : \mathbb{R}^{2d} \to \{0,1\}$. The main contribution of this paper is showing corresponding results for more general classes of f as well as developing novel approaches to recovering f which have not been considered before.

There are several reasons to consider the case of non-binary symbols: If we view the inverse problem as a calibration, it is reasonable that imperfections in the system may cause the corresponding symbol to deviate from an intended binary design. Symbol discontinuities can also cause audible artifacts known as *musical noise* in the audio setting [35, 38] and it is therefore beneficial to design those systems with non-binary symbols in the first place. In some audio filtering contexts where binary masks are currently used such as in [120], a non-binary value is associated to each time-frequency coordinate and a mask is then constructed by thresholding. This approach, while straight-forward, is unlikely to be optimal which has motivated the use of non-binary masks in similar systems [38]. Audio processing systems may also include components such as preemphasis which cannot be represented using binary time-frequency masks. Moreover, localization operators can be identified with function-operator convolutions from quantum harmonic analysis [153] and Gabor-Toeplitz operators [156] and inverting the symbol to operator mapping is of independent theoretical interest in these settings. As localization operators with a fixed window function are dense in the class of trace-class operators [24, 153], a general operator can be completely described by its symbol which is unlikely to be binary-valued, hence non-binary symbol recovery provides a suitable setting for a general mapping from operators to functions. Indeed, the problem of approximating general Hilbert-Schmidt operators by Gabor multipliers has previously been investigated by Dörfler and Torrésani in [62].

Below we state all of our main results before having established all the relevant notation. In particular, we formulate some results with function-operator convolutions $f \star S$, Wigner distributions $W(\varphi)$, the Feichtinger algebra $M^1(\mathbb{R}^d)$ and Cohen's class distributions $Q_S(\psi)$. These are all detailed and properly defined in Section D.2 but hopefully the general idea of the theorems should be clear. If not, the reader can return to the formulations after finishing the preliminaries section.

White noise

Our first result shows how a smooth, non-negative, real-valued symbol can be approximated by looking at the spectrograms of the images of white noise under the localization operator. Due to specifics of the approximation procedure, we can only estimate the square of the symbol, f^2, meaning that sign information is lost. In particular, our estimator for f^2, the so called *average observed spectrogram*, is given by

$$\rho(z) = \frac{1}{K} \sum_{k=1}^{K} |V_\varphi(A_f^g \mathcal{N}_k)(z)|^2 \tag{D.1.2}$$

where $(\mathcal{N}_k)_{k=1}^K$ are K realizations of (complex) white noise and φ is our *reconstruction window* which does not necessarily have to coincide with g. This construction is from [182] in the binary case and follows previous work on white noise approaches in time-frequency analysis which have recently received attention [5, 22]. The notion and interpretation of white noise in this setting will be made precise in Section D.2.3. We will show how, as $K \to \infty$, this estimator converges with high probability to

$$\vartheta(z) = \sum_m \lambda_m^2 |V_\varphi h_m(z)|^2 \tag{D.1.3}$$

where $(\lambda_m)_m$ and $(h_m)_m$ are the eigenvalues and eigenfunctions of A_f^g, respectively. Using the framework of quantum harmonic analysis and asymptotics of products of localization operators, we will show how ϑ in turn is a good approximation of f^2.

Theorem D.1.1. *Let $f \in C_c^{d+2}(\mathbb{R}^{2d})$ be real-valued, ρ be given by* (D.1.2) *with* *white noise variance* σ^2, $g, \varphi \in \mathcal{S}(\mathbb{R}^d)$ *with* $\|g\|_{L^2} = \|\varphi\|_{L^2} = 1$ *and define*

$$B_1 = \left[\|G\|_{L^2} + \left(\sum_{j=1}^{2d} \|\partial_j^{d+1} G\|_{L^2}^2 \right)^{1/2} \right],$$

$$B_2 = \left(\int_{\mathbb{R}^{2d}} \left|(\nabla f^2)(z)\right| dz \right) \left(\int_{\mathbb{R}^{2d}} |z||V_\varphi g(z)|^2 \, dz \right)$$

where

$$G(y,z) = f(y) \left(\sum_{|\alpha|=1} \int_0^1 \partial^\alpha f(y + t(z-y)) \, dt (z-y) \right) V_g g(y-z).$$

Then given $t \in [t_0, 1]$, *there exists constants* $c, c' > 0$ *only dependent on* t_0, g, φ *such that*

$$\mathbb{P}\left(\int_{\mathbb{R}^{2d}} \left| \frac{\rho(z)}{\sigma^2} - f(z)^2 \right| dz > B + t \right) \leq c \|f\|_{L^2}^2 e^{-c'K} \tag{D.1.4}$$

where $B = AB_1 + B_2$ *and* A *is a constant independent of* f *and* g.

The above theorem should be read as *"the* L^1 *estimation error is bounded by* B *with high probability provided* K *is large enough"*. Note that the assumptions on f in the theorem are not fulfilled by binary symbols. An intermediate step in the proof of the theorem (Proposition D.3.1) shows how the average observed spectrogram still approximates f^2 under looser conditions on f but it does not supply a unified error estimate in the way that (D.1.4) does. This can be redeemed by convolving f with a suitably chosen mollifier and estimating the L^1 error this induces using Lemma D.2.7.

Weighted accumulated Cohen's class

Next we discuss two approaches which rely on spectral data about the localization operator. These are also stated for *mixed-state* localization operators, introduced and defined in Section D.2.2 below, as this stronger result follows directly from our methods. For the motivation of these operators beyond generalization, see [154, 155]. We also state the corresponding statement for the "pure" localization operators discussed above.

Theorem D.1.2. *Let* $f \in L^1(\mathbb{R}^{2d})$ *be real-valued and with bounded variation and* S, T *be positive trace-class operators with* $\operatorname{tr}(S) = \operatorname{tr}(T) = 1$. *Then if* $f \star S = \sum_m \lambda_m (h_m \otimes h_m)$,

$$\sum_{m=1}^{\infty} \lambda_m Q_T(h_m)(z) = f * (S \star \check{T})(z) \tag{D.1.5}$$

and the L^1 error can be bounded as

$$\left\| \sum_{m=1}^{N} \lambda_m Q_T(h_m) - f \right\|_{L^1} \leq \operatorname{Var}(f) \int_{\mathbb{R}^{2d}} |z| (S \star \check{T})(z)\, dz + \sum_{m=N+1}^{\infty} |\lambda_m|.$$

In particular, if $\varphi \in L^2(\mathbb{R}^d)$ with $\|\varphi\|_{L^2} = 1$ and $A_f^g = \sum_m \lambda_m (h_m \otimes h_m)$, then

$$\left\| \sum_{m=1}^{N} \lambda_m |V_\varphi(h_m)|^2 - f \right\|_{L^1} \leq \operatorname{Var}(f) \int_{\mathbb{R}^{2d}} |z| |V_\varphi g(z)|^2\, dz + \sum_{m=N+1}^{\infty} |\lambda_m|.$$

Again, the notation Q_S for Cohen's class is explained later in Section D.2.1 but reduces down to the spectrogram when S is a rank-one operator as highlighted above.

Similar sums have previously been investigated in [8] and [155] as accumulated spectrograms and Cohen's class distributions without the factor λ_m in front of the Cohen's class distribution or spectrogram. In Section D.4.1, we make the argument as to why the above formulation, which we refer to as the *weighted* accumulated Cohen's class or spectrogram, is preferable and leads to smaller errors.

Weighted accumulated Wigner distribution

For the weighted accumulated Cohen's class as well as the white noise approach, we had to use a reconstruction window φ. In the next theorem, we are able to sidestep this. Essentially the weighted accumulated Wigner distribution is the Weyl symbol of the localization operator and its properties are easy to deduce from this perspective.

Theorem D.1.3. *Let $f \in L^1(\mathbb{R}^{2d})$ be real-valued with bounded variation and $S = \sum_n s_n (g_n \otimes g_n)$ a positive trace-class operator. Then if $f \star S = \sum_m \lambda_m (h_m \otimes h_m)$,*

$$\sum_m \lambda_m W(h_m)(z) = f * \sum_n s_n W(g_n)(z)$$

where $W(g)$ is the Wigner distribution of g. In particular, if $S = g \otimes g$ so that $f \star S = A_f^g$,

$$\sum_m \lambda_m W(h_m)(z) = f * W(g)(z).$$

Moreover, if the window functions $(g_n)_n$ are in the Schwartz space, the convergence is in L^1 with the error bounds

$$\left\| \sum_m \lambda_m W(h_m) - f \right\|_{L^1} \leq \operatorname{Var}(f) \int_{\mathbb{R}^{2d}} |z| \left| \sum_n s_n W(g_n)(z) \right| dz$$

and the corresponding statement holds for the rank-one case $S = g \otimes g$.

Plane tiling

Next, we discuss an approach based on noting that adding up all the spectrograms of an orthonormal basis yields the function which is identically one in a manner that be likened to a tiling of phase space via a partition of unity. If we then apply our localization operator with symbol f to each basis element, this tiling should only make a contribution proportional to the size of f^2. This intuition turns out to be correct and is quantified in the following theorem.

Theorem D.1.4. *Let $f \in C_c^{d+2}(\mathbb{R}^{2d})$ be real-valued and $g, \varphi \in \mathcal{S}(\mathbb{R}^{2d})$ with $\|g\|_{L^2} = \|\varphi\|_{L^2} = 1$. Then for any orthonormal basis $\{e_n\}_n$ of $L^2(\mathbb{R}^d)$,*

$$\left\| \sum_n |V_\varphi(A_f^g e_n)|^2 - f^2 \right\|_{L^1} \leq B$$

where B is as in Theorem D.1.1.

The reason the above error estimate is so similar to that in Theorem D.1.1 is that it turns out that the plane tiling estimator is precisely the limit ϑ (D.1.3) which the average observed spectrogram converges to pointwise as $K \to \infty$ almost surely.

Gabor space projection

The last method is fundamentally different from those above in that it estimates f pointwise in a parallelizable way. Given a point $z \in \mathbb{R}^{2d}$, we estimate $f(z)$ as follows: Translate the window function g by z so that $V_g(\pi(z)g)$ is centered at z and takes the value 1 there, i.e., $V_g(\pi(z)g)(z) = 1$. Next apply the localization operator to $\pi(z)g$ so that the value of the STFT at z is scaled by approximately $f(z)$. The reconstruction window used can differ from the original window g as long as $V_\varphi g$ reaches its maximum close to 0. As we show in Section D.6, when $\varphi = g$ this procedure may be interpreted as projecting $f \cdot V_g g(\cdot - z)$ onto the Gabor space $V_g(L^2)$ which is the reason for the name.

Theorem D.1.5. *Let $f \in L^1(\mathbb{R}^{2d})$ and $g, \varphi \in L^2(\mathbb{R}^d)$ with $\|g\|_{L^2} = \|\varphi\|_{L^2} = 1$. Then*

$$V_\varphi(A_f^g(\pi(z)\varphi))(z) = f * |V_\varphi g|^2(z)$$

and consequently the estimator satisfies the error estimate

$$\left\|V_\varphi(A_f^g(\pi(\cdot)\varphi)) - f\right\|_{L^1} \leq \operatorname{Var}(f) \int_{\mathbb{R}^{2d}} |z||V_\varphi g(z)|^2 \, dz.$$

Due to the pointwise nature of this method, it can be used in conjunction with the white noise or plane tiling estimator which only estimates f^2 to supply sign information for f.

Note that the above estimator is identical to that for the weighted accumulated spectrogram (D.1.5). Consequently they can be combined as the weighted accumulated spectrogram is far more susceptible to numerical instabilities as discussed later in Section D.7.1. Moreover, the estimator is the convolution of the spectrogram and then kernel $|V_\varphi g|^2$ and so if we know g and can set $\varphi = g$, we can recover f precisely by a deconvolution procedure. This is of course a very unstable procedure but we implement it numerically in Section D.7.1 to show that it is feasible.

Outline

In Section D.2, we go through all of the required preliminaries to follow the proofs of the above theorems. Sections D.3, D.4, D.5 and D.6 are devoted to discussing the details, proofs and appropriate considerations for all of the recovery methods outlined above. Lastly in Section D.7, all of the five methods are implemented in MATLAB using the Large Time/Frequency Analysis Toolbox [175] with the code available on GitHub[1]. We also discuss implementation details and compare the performance of the different methods.

Notational conventions

The Schatten p-class of operators with singular values in ℓ^p will be denoted by $\mathcal{S}^p$ while the larger class of bounded operators on $L^2(\mathbb{R}^d)$ will be denoted by $\mathcal{L}(L^2)$. The adjoint of such an operator A will be denoted by A^* and we will write $\check{A} = PAP$ where P is the parity operator $P : f(t) \mapsto f(-t)$. For the open ball centered at z with radius r we will write $B_r(z)$ and for a function f of several variables, we will write $\partial_j^n f$ for the n:th derivative in the j:th variable. We will also use a multiindex $\alpha = (\alpha_1, \dots, \alpha_d)$ for the derivative $\partial^\alpha f = \partial_{\alpha_1} \cdots \partial_{\alpha_d} f$ and denote by C^n the set of functions f for which $\partial^\alpha f$ is continuous for all $\alpha \in \mathbb{N}^d$ with $|\alpha| \leq n$. The associated subspace C_c^n will specify those functions which have compact support. For the Schwartz functions on $\mathbb{R}^d$ we will write $\mathcal{S}(\mathbb{R}^d)$ while indicator functions of sets Ω will be denoted by χ_Ω. Inner products with no subscript will always refer to $L^2(\mathbb{R}^d)$ inner products so that $\langle \cdot, \cdot \rangle = \langle \cdot, \cdot \rangle_{L^2(\mathbb{R}^d)}$. For matrices of size $N \times M$ with entries in $\mathbb{C}$, we will write $\mathbb{C}^{N \times M}$.

[1] `https://github.com/SimonHalvdansson/Localization-Operator-Symbol-Recovery`

D.2 Preliminaries

D.2.1 Time-frequency analysis

We highlight some important facts from time-frequency analysis which we will have use for, in a compact form. For a more complete introduction the reader is referred to [107, 220].

Short-time Fourier transform

One of the main ideas underlying time-frequency analysis is that real world signals are correlated in time and frequency. This is utilized by the short-time Fourier transform (STFT) which allows us to analyze the frequency content of a signal ψ around a specific time by multiplying ψ by a window function g which is well-localized in time. It is defined as

$$V_g\psi(z) = \langle \psi, \pi(z)g \rangle = \int_{\mathbb{R}^d} \psi(t)\overline{g(t-x)}e^{-2\pi i \omega t}\, dt.$$

where the *time-frequency shift* $\pi(z)$ is a projective unitary square integrable representation acting as $\pi(z)\psi(t) = \pi(x,\omega)\psi(t) = e^{2\pi i \omega t}\psi(t-x)$ and the point $z = (x,\omega) \in \mathbb{R}^{2d}$ is said to belong to *phase space*. A standard result known as *Moyal's formula* states that we can compute inner products of two signals using their associated short-time Fourier transforms as

$$\langle V_{g_1}\psi_1, V_{g_2}\psi_2 \rangle_{L^2(\mathbb{R}^{2d})} = \langle \psi_1, \psi_2 \rangle \overline{\langle g_1, g_2 \rangle}.$$

In particular, the mapping $V_g : L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^{2d})$ is an isometry provided $\|g\|_{L^2} = 1$. As a consequence of the above relation, we have the reconstruction formula from the introduction which tells us how we can get a signal back from its STFT

$$\psi = \int_{\mathbb{R}^{2d}} V_g\psi(z)\pi(z)g\, dz. \tag{D.2.1}$$

Often in application, the square modulus $|V_g\psi|^2$ of the STFT, the *spectrogram*, is used as it is real-valued and non-negative and thus represents the *energy* distribution of the signal.

Localization operators

Provided we want to localize the support of a signal ψ to a region of phase space, an obvious idea is to limit the reconstruction in (D.2.1) to some subset $\Omega \subset \mathbb{R}^{2d}$. By the uncertainty principle, this is impossible but works up to a small error depending on the size of Ω. Generalizing this idea, we can add a weighing factor, or *symbol*, $f \in L^1(\mathbb{R}^{2d})$ to the reconstruction (D.2.1) which tells us how much we want to reconstruct different parts of the phase space representation of ψ. Formally, we write the application of the localization operator A_f^g to ψ as

$$A_f^g \psi(t) = \int_{\mathbb{R}^{2d}} f(z) V_g \psi(z) \pi(z) g(t)\, dz. \tag{D.2.2}$$

Localization operators in the time-frequency context were originally investigated by I. Daubechies [51, 52].

Gabor spaces

The range of the short-time Fourier transform with a specific window g is denoted by $V_g(L^2) \subset L^2(\mathbb{R}^{2d})$ and called the *Gabor space* associated to g. These spaces are reproducing kernel Hilbert spaces which fulfill

$$F = F * V_g g \qquad \text{for all } F \in V_g(L^2).$$

We already mentioned that $V_g^* V_g$ is the identity but reversed composition, $V_g V_g^* : L^2(\mathbb{R}^{2d}) \to V_g(L^2)$ is the orthogonal projection onto the Gabor space, often denoted by $P_{V_g(L^2)}$.

Modulation spaces

Feichtinger's algebra $M^1(\mathbb{R}^d)$, originally introduced in [68], is defined as the set of tempered distributions ψ such that $V_g \psi \in L^1(\mathbb{R}^{2d})$ where g is a Schwartz window function. It is a special case of the *modulation spaces* [68, 78] which are defined by integrability properties of short-time Fourier transforms and have the convenient property that they are independent of the window function g used.

Cohen's class of time-frequency distributions

There are several quadratic time-frequency distributions with properties similar to those of the spectrogram. Those that fulfill some basic desirable properties are commonly referred to as *Cohen's class distributions* [39] and include the spectrogram as a special case. They can all be written as

$$Q_\Phi(\psi) = \Phi * W(\psi)$$

where Φ is a tempered distribution and $W(\psi) = W(\psi, \psi)$ is the *Wigner distribution* of ψ, defined as

$$W(\psi, \phi)(x, \omega) = \int_{\mathbb{R}^d} \psi(t + x/2) \overline{\phi(t - x/2)} e^{-2\pi i \omega \cdot t}\, dt. \tag{D.2.3}$$

D.2.2 Quantum harmonic analysis

The theory of quantum harmonic analysis, first developed by Werner in [211], will play a central role in several of our main proofs. Its main components are definitions of convolutions between functions and operators and pairs of operators. For a function f and two operators T and S, these take the form

$$f \star S = \int_{\mathbb{R}^{2d}} f(z)\pi(z)S\pi(z)^*\,dz, \qquad T \star S(z) = \operatorname{tr}\left(T\pi(z)\check{S}\pi(z)^*\right) \tag{D.2.4}$$

where the first integral should be interpreted as a Bochner integral and $\check{S} = PSP$. Note in particular that with this formalism, $f \star S$ is an operator while $T \star S$ is a function. As we will see below, both of these definitions satisfy versions of Young's inequality which we will make use of. However, we first compute the prototypical rank-one function-operator and operator-operator convolutions since these serve as our main motivation for using the framework of quantum harmonic analysis. We remind the reader that a rank-one operator $\psi \otimes \phi$ acts as $f \mapsto \langle f, \phi\rangle\psi$.

Example D.2.1. The function-operator convolution $f \star (g \otimes g)$ for $f \in L^1(\mathbb{R}^{2d})$ and $g \in L^2(\mathbb{R}^d)$ is precisely the localization operator A_f^g. Indeed,

$$\begin{aligned} f \star (g \otimes g)\psi &= \int_{\mathbb{R}^{2d}} f(z)\pi(z)(g \otimes g)\pi(z)^*\psi\,dz \\ &= \int_{\mathbb{R}^{2d}} f(z)\langle \pi(z)^*\psi, g\rangle \pi(z)g\,dz \\ &= \int_{\mathbb{R}^{2d}} f(z)V_g\psi(z)\pi(z)g\,dz = A_f^g\psi. \end{aligned}$$

Example D.2.2. The simplest case of operator-operator convolutions reduces down to the spectrogram. For $\psi, \phi \in L^2(\mathbb{R}^d)$, we have that

$$\begin{aligned} (\psi \otimes \psi) \star (\phi \otimes \phi)^{\check{}}(z) &= \operatorname{tr}\left((\psi \otimes \psi)\pi(z)(\phi \otimes \phi)\pi(z)^*\right) \\ &= \sum_n \left\langle (\psi \otimes \psi)\pi(z)(\phi \otimes \phi)\pi(z)^* e_n, e_n \right\rangle \\ &= \sum_n \langle e_n, \pi(z)\phi\rangle\langle \pi(z)\phi, \psi\rangle\langle \psi, e_n\rangle \\ &= |\langle \psi, \pi(z)\phi\rangle|^2 = |V_\phi\psi(z)|^2 \end{aligned}$$

where $(e_n)_n$ was an arbitrary orthonormal basis used to compute the trace.

Many properties of function-operator and operator-operator convolutions are analogues of corresponding statements for classical function-function convolutions as we will see below.

Just as the integral is replaced by the trace in the definition of operator-operator convolutions (D.2.4), when measuring the size of operators, we will use the Schatten p-norms which are defined as

$$\|A\|_{\mathcal{S}^p} = \operatorname{tr}(|A|^p)^{1/p}$$

where $|A| = \sqrt{A^*A}$ is the absolute value of A. These norms induce the Schatten p-classes of operators, the most notable of which are the trace-class operators $\mathcal{S}^1$ and the Hilbert-Schmidt operators $\mathcal{S}^2$. As these operators are compact, they have a spectral decomposition of the form

$$A = \sum_n a_n(\psi_n \otimes \phi_n)$$

where $(\psi_n)_n$ and $(\phi_n)_n$ are orthonormal bases and $(a_n)_n$ is in ℓ^p if $A \in \mathcal{S}^p$. Note in particular that if A is self-adjoint, we have that $\psi_n = \phi_n$ for all n.

Next we collect some basic properties of these convolutions, the proofs of which can be found in [153].

Proposition D.2.3. Let $f, g \in L^1(\mathbb{R}^{2d})$, $S \in \mathcal{S}^p$, $T \in \mathcal{S}^q$ for $1 \leq p, q \leq \infty$ with $\frac{1}{p} + \frac{1}{q} = 1$ and $R \in \mathcal{S}^1$. Then

(i) $(f \star S)^* = \bar{f} \star S^*$,

(ii) $(f \star R) \star T = f * (R \star T)$,

(iii) $(f * g) \star S = f \star (g \star S)$,

(iv) $\|f \star S\|_{\mathcal{S}^p} \leq \|f\|_{L^1} \|S\|_{\mathcal{S}^p}$,

(v) $\|h \star R\|_{\mathcal{S}^p} \leq \|h\|_{L^p} \|R\|_{\mathcal{S}^1}$,

(vi) $\|S \star R\|_{L^p} \leq \|S\|_{\mathcal{S}^p} \|R\|_{\mathcal{S}^1}$.

In the next subsections, we dive deeper into some topics in quantum harmonic analysis which will be of use. For a more thorough introduction with more motivation and results, the reader is referred to [153].

Mixed-state localization operators

Localization operators reconstruct a function with respect to a single window or pair of windows in the non self-adjoint case. This construction has been generalized to multiple windows by considering the function-operator convolution $f \star S$ which can be seen as a weighted sum of localization operators [154]. Indeed, if $S = \sum_n s_n(g_n \otimes g_n) \in \mathcal{S}^1$, then

$$f \star S = f \star \sum_n s_n(g_n \otimes g_n) = \sum_n s_n A_f^{g_n}.$$

Later on in our results, we will need for our (mixed-state) localization operators to be self-adjoint. In view of Proposition D.2.3 (i), this requires the symbol f to be real-valued and the window operator S to be self-adjoint.

Fourier-Wigner transform

Another central tool of quantum harmonic analysis is the Fourier-Wigner transform, mapping operators to functions, defined for $S \in \mathcal{S}^1$ as

$$\mathcal{F}_W(S)(z) = e^{-\pi i x\omega} \operatorname{tr}(\pi(-z)S).$$

Our interest in the Fourier-Wigner transform is primarily based on its convolution properties which mirror those of the classical Fourier transform. To state the relevant result, we first need to define the *symplectic Fourier transform* which essentially is a rotated two-dimensional Fourier transform

$$\mathcal{F}_\sigma(f)(z) = \int_{\mathbb{R}^{2d}} f(z')e^{-2\pi i \sigma(z,z')}\, dz'$$

where $z = (x, \omega)$, $z' = (x', \omega')$ and $\sigma(z, z') = \omega x' - \omega' x$ is the standard symplectic form. We can now state the result which is analogous to the classical convolution theorem.

Proposition D.2.4. Let $f \in L^1(\mathbb{R}^{2d})$ and $S, T \in \mathcal{S}^1$. Then

$$\mathcal{F}_W(f \star S) = \mathcal{F}_\sigma(f) \cdot \mathcal{F}_W(S),$$
$$\mathcal{F}_\sigma(T \star S) = \mathcal{F}_W(T) \cdot \mathcal{F}_W(S).$$

Weyl quantization

A *quantization procedure* provides a mapping between functions and operators such as the mapping $f \mapsto f \star S$ which we invert in this paper. In time-frequency analysis and quantum harmonic analysis, we often make use of *Weyl quantization* which can be defined weakly as

$$\langle L_f \psi, \phi \rangle = \langle f, W(\phi, \psi) \rangle$$

where we refer to the mapping $f \mapsto L_f$ as the *Weyl transform*. For the inverse mapping, meaning the function associated to the operator S, we write a_S and call it the *Weyl symbol* of S.

Weyl quantization has a particularly nice formulation in quantum harmonic analysis where it can be written as

$$a_S = \mathcal{F}_\sigma(\mathcal{F}_W(S)).$$

In particular, it can be shown that $a_{\psi \otimes \phi} = W(\psi, \phi)$. It also holds that Weyl quantization is compatible with the convolutions of quantum harmonic analysis in the sense that

$$T \star S = a_T * a_S, \qquad a_{f \star S} = f * a_S \tag{D.2.5}$$

for $T, S \in \mathcal{S}^1$ and $f \in L^1(\mathbb{R}^{2d})$.

Cohen's class as operator-operator convolutions

The class of quadratic time-frequency distributions discussed in Section D.2.1 has a convenient formulation in quantum harmonic analysis using the Weyl quantization relations (D.2.5) above. By letting $\check{S}$ be the Weyl quantization of the tempered distribution Φ defining Q_Φ and using that $a_{\psi \otimes \psi} = W(\psi)$, we get that

$$Q_\Phi(\psi) = (\psi \otimes \psi) \star \check{S} =: Q_S(\psi). \tag{D.2.6}$$

This point of view makes it particularly easy to deduce properties of Cohen's class distributions such as bounding L^p norms or characterizing positivity.

D.2.3 Functional analytic and probabilistic aspects of white noise

The core of our approach to symbol recovery using white noise is computing spectrograms of random noise. This is inspired by recent theoretical work in [22, 101, 113] and others. For further details we refer the reader to the discussion in [182] as we follow their proof strategy. The two main result which we need are stated in [182] and we also state them here for the sake of completeness. The first is a version of the Hanson-Wright inequality [10, 184].

Theorem D.2.5 ([182, Theorem 3.1])**.** *Let X be an m-dimensional complex Gaussian random variable with $X \sim \mathcal{CN}(0, \Sigma)$ and $A \in \mathbb{C}^{m \times m}$ Hermitian. Then there exists an universal constant $C_{hw} > 0$ such that for every $t > 0$,*

$$\mathbb{P}\big(|\langle AX, X\rangle - \mathbb{E}\{\langle AX, X\rangle\}| > t\big) \le 2\exp\left(-C_{hw} \min\left\{\frac{t^2}{\|\Sigma\|_s^2 \|A\|_F^2}, \frac{t}{\|\Sigma\|_s \|A\|_s}\right\}\right)$$

where $\|\cdot\|_s$ and $\|\cdot\|_F$ are the spectral and Frobenius norms, respectively.

Secondly, the next lemma gives a constructive way to deal with the application of an operator to white noise.

Lemma D.2.6 ([182, Lemma 4.2]). *Let g be a Schwartz function with $\|g\|_{L^2} = 1$, $f \in L^1(\mathbb{R}^{2d})$ and $\mathcal{N}$ a realization of complex white noise with variance σ^2. Then there exists a sequence $(\alpha_m)_m$ where $\alpha_m \sim C\mathcal{N}(0, \sigma^2)$ of independent complex normal variables such that almost surely,*

$$A_f^g(\mathcal{N}) = \sum_m \lambda_m \alpha_m h_m \qquad \text{where } A_f^g = \sum_m \lambda_m (h_m \otimes h_m)$$

with almost sure absolute convergence in $L^2(\mathbb{R}^{2d})$.

We also remark that if our white noise is not complex but rather real valued, all results will still hold but with possibly larger constants. See [182, Section 2.2] for a discussion on this.

D.2.4 Approximate identities

The *variation* of a function $f \in L^1(\mathbb{R}^{2d})$ is defined as

$$\mathrm{Var}(f) = \sup \left\{ \int_{\mathbb{R}^{2d}} f(z) \,\mathrm{div}\, \phi(z) \, dz : \phi \in C_c^1(\mathbb{R}^{2d}, \mathbb{R}^{2d}),\ \|\phi\|_\infty \leq 1 \right\}$$

and in the special case where $f \in C^1(\mathbb{R}^{2d})$, it can be written as

$$\mathrm{Var}(f) = \int_{\mathbb{R}^{2d}} |\nabla f(z)| \, dz.$$

We say that functions f with $\mathrm{Var}(f) < \infty$ have *bounded variation*. In the case where f is the indicator function of some compact subset $\Omega \subset \mathbb{R}^{2d}$ with smooth boundary, the variation of f is equal to the Hausdorff measure of the boundary $\partial\Omega$ [66].

In what follows, we will want to measure how much a function is changed when it is convolved with some kernel. The next lemma quantifies this using the concept of variation introduced above.

Lemma D.2.7 ([8, Lemma 3.2]). *Let $\psi \in L^1(\mathbb{R}^{2d})$ have bounded variation and $\phi \in L^1(\mathbb{R}^{2d})$ with $\int_{\mathbb{R}^{2d}} \phi(z) \, dz = 1$, then*

$$\|\psi * \phi - \psi\|_{L^1} \leq \mathrm{Var}(\psi) \int_{\mathbb{R}^{2d}} |z| |\phi(z)| \, dz.$$

In the following, we will sometimes refer to ϕ as the *blurring kernel*.

D.3 Recovery via white noise

The idea behind symbol recovery via white noise is easy to summarize; on average, the spectrograms of white noise are constant so by looking at filtered white noise, the spectrograms reveal the characteristics of the filter. As spectrograms are inherently quadratic, this essentially limits us to real valued non-negative symbols as we can only estimate the squared modulus $|f|^2$.

D.3.1 Preliminary results

Before proceeding with a proof of Theorem D.1.1, we collect some preliminary results in the following proposition which is stated under looser conditions than Theorem D.1.1 and details all the intermediate steps from the average observed spectrogram $\rho(z) = \frac{1}{K}\sum_{k=1}^{K} |V_\varphi(A_f^g \mathcal{N}_k)(z)|^2$ to f^2.

Proposition D.3.1. Let f be a real-valued, bounded, integrable function with bounded derivative and bounded variation, ρ the average observed spectrogram (D.1.2) with white noise variance σ^2 and $g, \varphi \in \mathcal{S}(\mathbb{R}^d)$ with $\|g\|_{L^2} = \|\varphi\|_{L^2} = 1$. Then there exists a constant $C > 1$ such that

$$\mathbb{P}\left(\left|\frac{\rho(z)}{\sigma^2} - \vartheta(z)\right| > t\right) \leq 3\exp\left(-CK\min\left(\frac{t^2}{\vartheta(z)^2}, \frac{t}{\vartheta(z)}\right)\right), \tag{D.3.1}$$

$$\left\|\vartheta - f^2 * |V_\varphi g|^2\right\|_{L^\infty} \leq \|f\|_{L^\infty}\left(\sum_{|\alpha|=1}\|\partial^\alpha f\|_{L^\infty}\right)\|g\|_{M^1}^4, \tag{D.3.2}$$

$$\left\|f^2 * |V_\varphi g|^2 - f^2\right\|_{L^1} \leq \left(\int_{\mathbb{R}^{2d}}\left|(\nabla f^2)(z)\right| dz\right)\left(\int_{\mathbb{R}^{2d}}|z||V_\varphi g(z)|^2\, dz\right). \tag{D.3.3}$$

For the first part of the proposition, we will need a version of [182, Lemma 5.1] with unknown variance and non-binary symbols. This requires minimal modifications to the original proof in [182] and so we leave out the proof in the interest of brevity.

Lemma D.3.2. *Let $f, \rho, \sigma, g, \varphi$ and ϑ be as in Proposition D.3.1. Then there exists $C > 0$ such that for every $z \in \mathbb{R}^{2d}$,*

$$\mathbb{P}\left(\left|\frac{\rho(z)}{\sigma^2} - \vartheta(z)\right| > t\right) \leq 3\exp\left(-CK\min\left(\frac{t^2}{\vartheta(z)^2}, \frac{t}{\vartheta(z)}\right)\right).$$

We can now proceed with the proof of the proposition.

Proof of Proposition D.3.1. The first estimate of the proposition is precisely Lemma D.3.2 as stated above.

For (D.3.2), we first claim that

$$\vartheta(z) = \left(A_f^g\right)^2 \star (\varphi \otimes \varphi)\check{}(z). \tag{D.3.4}$$

Indeed, as $A_f^g = \sum_m \lambda_m (h_m \otimes h_m) \in \mathcal{S}^1$, it follows that $\left(A_f^g\right)^2 = \sum_m \lambda_m^2 (h_m \otimes h_m)$. Hence

$$\begin{aligned}\left(A_f^g\right)^2 \star (\varphi \otimes \varphi)\check{}(z) &= \sum_m \lambda_m^2 (h_m \otimes h_m) \star (\varphi \otimes \varphi)\check{}(z) \\ &= \sum_m \lambda_m^2 |V_\varphi h_m(z)|^2 = \vartheta(z)\end{aligned} \tag{D.3.5}$$

by Example D.2.2. To proceed from here, we need to relate the product $(A_f^g)^2$ to $A_{f^2}^g$, a problem which has been studied by Cordero, Rodino and Gröchenig among others. In our situation, their results in [44], [41, Theorem 4 (i)] and [41, Lemma 5] can be summarized as follows:

$$A_f^g A_f^g = A_{f^2}^g + V_g^* T V_g \tag{D.3.6}$$

where T is an integral operator with kernel $G : \mathbb{R}^{2d} \times \mathbb{R}^{2d} \to \mathbb{C}$ given by

$$G(y,z) = f(y) \sum_{|\alpha|=1} \int_0^1 (1-t)\partial^\alpha f(y + t(z-y))\, dt \frac{(z-y)^\alpha}{\alpha!} \langle \pi(z)g, \pi(y)g\rangle.$$

Moreover, the norm of $V_g^* T V_g$ can be bounded as

$$\|V_g^* T V_g\|_{\mathcal{L}(L^2)} \leq \|a\|_{L^\infty} \left(\sum_{|\alpha|=1} \|\partial^\alpha f\|_{L^\infty} \right) \|g\|_{M^1}^4. \tag{D.3.7}$$

Applying this result to $\left(A_f^g\right)^2$ yields

$$\left(A_f^g\right)^2 = A_{f^2}^g + V_g^* T V_g = f^2 \star (g \otimes g) + V_g^* T V_g.$$

Plugging this into (D.3.5) and applying Example D.2.2 yields

$$\begin{aligned}\vartheta(z) &= \left(f^2 \star (g \otimes g) + V_g^* T V_g\right) \star (\varphi \otimes \varphi)\check{}(z) \\ &= f^2 * |V_\varphi g|^2(z) + (V_g^* T V_g) \star (\varphi \otimes \varphi)\check{}(z).\end{aligned}$$

Rearranging the above and applying Proposition D.2.3 (vi) together with (D.3.7), we get the estimate

$$\begin{aligned}\left\|\vartheta - f^2 * |V_\varphi g|^2\right\|_{L^\infty} &= \|(V_g^* T V_g) \star (\varphi \otimes \varphi)\check{}\|_{L^\infty} \leq \|V_g^* T V_g\|_{\mathcal{L}(L^2)} \|\varphi\|_{L^2}^2 \\ &\leq \|f\|_{L^\infty} \left(\sum_{|\alpha|=1} \|\partial^\alpha f\|_{L^\infty} \right) \|g\|_{M^1}^4.\end{aligned}$$

Lastly for (D.3.3), applying Lemma D.2.7 with $\psi = f^2$ and $\phi = |V_\varphi g|^2$ yields the desired conclusion. □

D.3.2 Proof of Theorem D.1.1

For Theorem D.1.1, much of the machinery from the proof of Proposition D.3.1 can be reused but we will need an additional estimate on the localization operator product asymptotics and a lemma turning the estimate in Lemma D.3.2 into an L^1 error which is similar to [182, Lemma 5.5].

As a first step, we state a simplified version of [198, Theorem 2] adapted to a context in which we will soon need it.

Lemma D.3.3. *Let $T : L^2(\mathbb{R}^{2d}) \to L^2(\mathbb{R}^{2d})$ be an integral operator with kernel $G : \mathbb{R}^{2d} \times \mathbb{R}^{2d} \to \mathbb{C}$ that has compact support in the first variable. Then*

$$\|T\|_{\mathcal{S}^1} \leq A \left[\|G\|_{L^2} + \left(\sum_{j=1}^{2d} \left\| \partial_j^{d+1} G \right\|_{L^2}^2 \right)^{1/2} \right]$$

where the constant A is independent of G.

Armed with this lemma, we can bound the trace norm of the $V_\varphi^* T V_\varphi$ error operator from (D.3.6) above.

Lemma D.3.4. *Let $f \in C_c^{d+2}(\mathbb{R}^{2d})$ and $g \in \mathcal{S}(\mathbb{R}^d)$ with $\|g\|_{L^2} = 1$, then there exists a constant A independent of f and g such that*

$$\|V_g^* T V_g\|_{\mathcal{S}^1} \leq A \left[\|G\|_{L^2} + \left(\sum_{j=1}^{2d} \left\| \partial_j^{d+1} G \right\|_{L^2}^2 \right)^{1/2} \right] < \infty$$

where

$$G(y, z) = f(y) \left(\sum_{|\alpha|=1} \int_0^1 \partial^\alpha f(y + t(z - y))\, dt (z - y) \right) V_g g(y - z).$$

Proof. Since V_g is an isometry and $\|SR\|_{\mathcal{S}^1} \leq \|S\|_{\mathcal{S}^1} \|R\|_{\mathcal{L}(L^2)}$ for operators S and R, we conclude that it suffices to bound the trace norm of the integral operator T. The bound in the formulation follows directly upon applying Lemma D.3.3.

The finiteness of the error bound follows from the support of f being compact and that $g \in \mathcal{S}(\mathbb{R}^d)$ via [107, Theorem 11.2.5]. □

Lastly we formulate the promised L^1-estimate, based on Lemma D.3.2. Its formulation and proof is similar to that of [182, Lemma 5.5].

Lemma D.3.5. *Let* $f \in L^1(\mathbb{R}^{2d}) \cap L^\infty(\mathbb{R}^{2d})$ *and* $\gamma \in [\gamma_0, 1]$, $\gamma_0 > 0$. *Then there exists constants* c, c' *only dependent on* g, φ, γ_0 *such that*

$$\mathbb{P}\left(\int_{\mathbb{R}^{2d}} \left|\frac{\rho(z)}{\sigma^2} - \vartheta(z)\right| dz \geq \gamma\right) \leq c\|f\|_{L^2}^2 e^{-c'K}.$$

Proof. By Chebyshev's inequality with an arbitrary $p > 1$ applied to the random variable $\int_{\mathbb{R}^{2d}} \left|\frac{\rho(z)}{\sigma^2} - \vartheta(z)\right| dz$ followed by the Minkowski inequality, we have

$$\begin{aligned}
\mathbb{P}\left(\int_{\mathbb{R}^{2d}} \left|\frac{\rho(z)}{\sigma^2} - \vartheta(z)\right| dz \geq \gamma\right) &\leq \frac{1}{\gamma^p}\mathbb{E}\left\{\left(\int_{\mathbb{R}^{2d}} \left|\frac{\rho(z)}{\sigma^2} - \vartheta(z)\right| dz\right)^p\right\} \\
&\leq \frac{1}{\gamma^p}\mathbb{E}\left\{\int_{\mathbb{R}^{2d}} \left|\frac{\rho(z)}{\sigma^2} - \vartheta(z)\right|^p dz\right\} \\
&= \frac{1}{\gamma^p}\int_{\mathbb{R}^{2d}}\int_0^\infty pt^{p-1}\mathbb{P}\left\{\left|\frac{\rho(z)}{\sigma^2} - \vartheta(z)\right| \geq t\right\} dt\, dz.
\end{aligned} \tag{D.3.8}$$

Next we estimate the inner integral for each $z \in \mathbb{R}^{2d}$ using Lemma D.3.2 as

$$\begin{aligned}
&\int_0^\infty pt^{p-1}\mathbb{P}\left\{\left|\frac{\rho(z)}{\sigma^2} - \vartheta(z)\right| \geq t\right\} dt \\
&\quad\leq 3\int_0^\infty pt^{p-1}\exp\left(-CK\min\left(\frac{t^2}{\vartheta(z)^2}, \frac{t}{\vartheta(z)}\right)\right) dt \\
&\quad= 3\int_0^{\vartheta(z)} pt^{p-1}\exp\left(-\frac{CKt^2}{\vartheta(z)^2}\right) dt + 3\int_{\vartheta(z)}^\infty pt^{p-1}\exp\left(-\frac{CKt}{\vartheta(z)}\right) dt \\
&\quad= \frac{3p}{2}\vartheta(z)^p\left(\frac{1}{CK}\right)^{p/2}\int_0^{CK} s^{\frac{p}{2}-1}e^{-s}\,ds + 3p\left(\frac{\vartheta(z)}{CK}\right)^p\int_{CK}^\infty s^{p-1}e^{-s}\,ds \\
&\quad\leq \frac{3p}{2}\vartheta(z)(CK)^{-p/2}\Gamma\left(\frac{p}{2}\right) + 3p\vartheta(z)(CK)^{-p}\Gamma(p) \\
&\quad\leq 3p\vartheta(z)\frac{1}{C^p}\left(\frac{\Gamma(p/2+1)}{K^{p/2}} + \frac{\Gamma(p+1)}{K^p}\right)
\end{aligned}$$

where we used that $\vartheta(z) \leq 1$ and $C > 1$. Plugging this back into the original double integral allows us to conclude that for every $p > 1$,

$$\mathbb{P}\left(\int_{\mathbb{R}^{2d}} \left|\frac{\rho(z)}{\sigma^2} - \vartheta(z)\right| dz \geq \gamma\right) \leq 3p\|\vartheta\|_{L^1}\frac{1}{(\gamma C)^p}\left[\frac{\Gamma(p/2+1)}{K^{p/2}} + \frac{\Gamma(p+1)}{K^p}\right]. \tag{D.3.9}$$

The $\|\vartheta\|_{L^1}$ norm can be bounded using (D.3.4) and Proposition D.2.3 (vi) with $p = 1$ as

$$\begin{aligned}\|\vartheta\|_{L^1} &= \left\|(A_f^g)^2 \star (\varphi \otimes \varphi)\right\|_{L^1} \\ &\leq \left\|(A_f^g)^2\right\|_{\mathcal{S}^1} \|\varphi \otimes \varphi\|_{\mathcal{S}^1} = \|A_f^g\|_{\mathcal{S}^2}^2 \leq \|f\|_{L^2}^2 \|\varphi \otimes \varphi\|_{\mathcal{S}^1}^2 = \|f\|_{L^2}^2\end{aligned}$$

where we used Proposition D.2.3 (v) with $p = 2$ for the second to last step. Meanwhile to fold the last two factors of (D.3.9) into an exponential in K, we can repeat an argument from the proof of [182, Lemma 5.5] with minimal modifications to conclude the existence constants $c, c' > 0$ such that

$$\frac{1}{(\gamma C)^p}\left[\frac{\Gamma\,(p/2+1)}{K^{p/2}} + \frac{\Gamma(p+1)}{K^p}\right] \leq c e^{-c' K \gamma^2}.$$

Using that $\gamma \geq \gamma_0$, we can also absorb γ^2 into the constant by making c dependent on γ_0. Lastly the $3p$ factor can be absorbed into c which yields exactly the estimate in the lemma. □

We are now ready to complete the proof of Theorem D.1.1.

Proof of Theorem D.1.1. We first claim that

$$\left\|\vartheta - f^2 * |V_\varphi g|^2\right\|_{L^1} \leq A\left[\|G\|_{L^2} + \left(\sum_{j=1}^{2d} \|\partial_j^{d+1} G\|_{L^2}^2\right)^{1/2}\right].$$

Indeed, this follows from Lemma D.3.4 as

$$\begin{aligned}\left\|\vartheta - f^2 * |V_\varphi g|^2\right\|_{L^1} &= \left\|\left(A_f^g\right)^2 \star (\varphi \otimes \varphi)\check{} - A_{f^2}^g \star (\varphi \otimes \varphi)\check{}\right\|_{L^1} \\ &= \|(V_g^* T V_g) \star (\varphi \otimes \varphi)\check{}\|_{L^1} \\ &\leq \|V_g^* T V_g\|_{\mathcal{S}^1} \|\varphi \otimes \varphi\|_{\mathcal{S}^1} \\ &\leq A\left[\|G\|_{L^2} + \left(\sum_{j=1}^{2d} \left\|\partial_j^{d+1} G\right\|_{L^2}^2\right)^{1/2}\right]\end{aligned}$$

where we used Proposition D.2.3 (vi) for the second to last step.

We now expand the left hand side in the $\left\|\frac{\rho}{\sigma^2} - f^2\right\|_{L^1} > AB_1 + B_2 + t$ inequality using the above and Lemma D.2.7 with $\psi = f^2$ and $\phi = |V_\varphi g|^2$ to find

$$\begin{aligned}
&\mathbb{P}\left(\left\|\frac{\rho}{\sigma^2} - f^2\right\|_{L^1} > B_1 + B_2 + t\right) \\
&\quad \leq \mathbb{P}\left(\left\|\frac{\rho}{\sigma^2} - \vartheta\right\|_{L^1} + \left\|\vartheta - f^2 * |V_\varphi g|^2\right\|_{L^1} + \left\|f^2 * |V_\varphi g|^2 - f^2\right\|_{L^1} > B_1 + B_2 + t\right) \\
&\quad \leq \mathbb{P}\left(\left\|\frac{\rho}{\sigma^2} - \vartheta\right\|_{L^1} + B_1 + B_2 > B_1 + B_2 + t\right) \\
&\quad = \mathbb{P}\left(\left\|\frac{\rho}{\sigma^2} - \vartheta\right\|_{L^1} > t\right) \leq c\|f\|_{L^2}^2 e^{-c'K}
\end{aligned}$$

where we in the last step used Lemma D.3.5. □

Remark D.3.6. Both Proposition D.3.1 and Theorem D.1.1 have clear analogues in the Cohen's class case which we believe to hold true. Indeed, it is straight-forward to show that

$$\mathbb{E}\left(\frac{1}{K}\sum_{k=1}^{K} Q_S(f \star S(\mathcal{N}_k))\right) \xrightarrow[K\to\infty]{} \sum_{m=1}^{\infty} \lambda_m^2 Q_S(h_m)(z),$$

but controlling the error estimates requires generalizing Lemma D.3.2 to the non rank-one case which is considerably more difficult.

Remark D.3.7. The quantity $\int_{\mathbb{R}^{2d}} |z||V_\varphi g(z)|^2\,dz$ which appears in Proposition D.3.1 and Theorem D.1.1 should be seen as punishing the case $\varphi \neq g$, i.e., the reconstruction window differing from the window function g.

D.4 Recovery via spectral data

In this section we discuss and prove the two recovery results Theorem D.1.2 and Theorem D.1.3 which are dependent on the eigenvalues and eigenfunctions of the localization operator. As we will see in Section D.7, the instability of eigenvalues and eigenfunctions can cause issues for these approaches but they still perform well.

D.4.1 Weighted accumulated Cohen's class

The accumulated Cohen's class, introduced in [155], is a generalization of accumulated spectrograms from [8] where it was used for symbol recovery for binary localization operators $A^g_{\chi_\Omega}$. There, a central idea was that the eigenvalues can be separated into two groups with the first $\approx \lceil|\Omega|\rceil$ being close to 1, followed by a sharp "plunge region" after which the remaining eigenvalues are all close to 0. This fact was originally proved in [74]. Based on this fact, the quantity

$$\sum_{m=1}^{\lceil|\Omega|\rceil} |V_g(h_m)(z)|^2 \approx \chi_\Omega(z) \tag{D.4.1}$$

was defined as the *accumulated spectrogram*. Later on, [155] extended the concept to mixed-state localization operators $f \star S$ by replacing the spectrograms in (D.4.1) by Cohen's class distributions by approaching the proof from a quantum harmonic analysis perspective.

Much of the work in these papers is focused on showing that the accumulated Cohen's class (D.4.1) is close to the quantity

$$\sum_m \lambda_m Q_S(h_m)(z) \tag{D.4.2}$$

by going into specifics on the decay of the eigenvalues. However, since computing the accumulated spectrogram already requires knowing the eigenfunctions, we (almost always) have exact knowledge of the eigenvalues and can bypass this approximation step and include the eigenvalues in the estimator. In this way, the error of the approximation can be decreased with no loss in performance or increase in runtime. Moreover, we do not require a priori knowledge of $|\Omega|$ to decide the number of eigenpairs to include. A consequence of this approach is that the resulting estimator also works well for non-binary localization operators whose eigenvalues do not follow the same $0-1$ dichotomy. We refer to the quantity (D.4.2) as the *weighted* accumulated Cohen's class to highlight the addition of the eigenvalue weights.

Both [8] and [155] restricted their attention to the case where the window g or the operator window S was known a priori. We lift this restriction by introducing a reconstruction window φ or reconstruction operator window T which does not have to agree with the original window g or S in the same way as we did for the average observed spectrogram. As we will see in the proof below, the proper estimator then instead becomes $\sum_m \lambda_m Q_T(h_m)(z)$.

Proof of Theorem D.1.2. The key observation for the proof is that $\sum_m \lambda_m Q_T(h_m) = f * (S \star \check{T})$. To see this, expand $f \star S$ in its singular value decomposition $f \star S = \sum_m \lambda_m (h_m \otimes h_m)$ and note that

$$\begin{aligned} f * (S \star \check{T}) = (f \star S) \star \check{T} &= \left(\sum_m \lambda_m (h_m \otimes h_m) \right) \star \check{T} \\ &= \sum_m \lambda_m (h_m \otimes h_m) \star \check{T} = \sum_m \lambda_m Q_T(h_m) \end{aligned}$$

where we used (D.2.6) for the last step. We can now compute

$$\begin{aligned}\left\| \sum_{m=1}^{N} \lambda_m Q_T(h_m) - f \right\|_{L^1} &\leq \left\| \sum_{m=1}^{N} \lambda_m Q_T(h_m) - \sum_{m=1}^{\infty} \lambda_m Q_T(h_m) \right\|_{L^1} \\ &\qquad + \left\| f * (S \star \check{T}) - f \right\|_{L^1} \\ &\leq \sum_{m=N+1}^{\infty} |\lambda_m| \|Q_T(h_m)\|_{L^1} + \left\| f * (S \star \check{T}) - f \right\|_{L^1} \\ &\leq \sum_{m=N+1}^{\infty} |\lambda_m| + \mathrm{Var}(f) \int_{\mathbb{R}^{2d}} |z| (S \star \check{T})(z) \, dz\end{aligned}$$

where we used that $\|Q_T(h_m)\|_{L^1} \leq 1$ by Proposition D.2.3 (vi) with $p = 1$ and the estimate in Lemma D.2.7. □

Remark D.4.1. Ideally, we would want $S \star \check{T}$ to be a Dirac delta to make the above reconstruction exact in the sense that $\sum_m \lambda_m Q_T(h_m) = f$. The closest we can get to this is in the lattice setting where such a construction is possible which is discussed in [192, Section 6.1]. The error incurred from $S \star \check{T} \neq \delta_0$ is partially captured in the $\int_{\mathbb{R}^{2d}} |z| (S \star \check{T})(z) \, dz$ factor which simplifies to $\int_{\mathbb{R}^{2d}} |z| |V_\varphi g(z)|^2 \, dz$ in the rank-one setting which we recognize from Section D.3. The specifics of this error were discussed and illustrated in [8, Figure 3].

The reader familiar with [155] will note that we essentially followed the exact same path for the proof as in that paper without restricting ourselves to indicator functions $f = \chi_\Omega$ and allowing $T \neq S$.

The recovery procedure detailed above is clearly linear and hence it is easy to see that it is continuous on $\mathcal{S}^1$. We mean this in the sense that if I is the map $f \star S \mapsto f * (S \star \check{T})$ and $A \in \mathcal{S}^1$ is a perturbation, then

$$\left\| I(f \star S + \varepsilon A) - I(f \star S) \right\|_{L^1} = \varepsilon \|A \star \check{T}\|_{L^1} \leq \varepsilon \|A\|_{\mathcal{S}^1} \tag{D.4.3}$$

by linearity and Proposition D.2.3 (vi).

As the estimator converges to a convolution in the $N \to \infty$ case, we can attempt to perform a deconvolution procedure to recover f exactly if we know the blurring kernel and its Fourier transform is zero-free. This is investigated numerically in Section D.7.1.

D.4.2 Weighted accumulated Wigner distribution

The approach in Theorem D.1.3 is perhaps the simplest of those detailed in this paper once framed as just computing the Weyl symbol of the localization operator and comparing with f (see (D.2.5)). There is also no requirement for a reconstruction window in this situation as our construction only depends on the spectral data of the localization operator. Note that we again adopt the *weighted* terminology to highlight the eigenvalue dependence as in Section D.4.1.

Proof of Theorem D.1.3. We prove the full case where S is a positive trace-class operator and note that the special rank-one case follows from it.

As discussed in Section D.2.2, the Weyl symbol of the function-operator convolution $f \star S$ is given by $f * a_S$ where a_S is the Weyl symbol of S. By the linearity of the Weyl symbol mapping $S \mapsto a_S$, we can compute this by using the spectral decomposition of $S = \sum_n s_n(g_n \otimes g_n)$ and the fact that $a_{g\otimes g} = W(g)$:

$$\sum_m \lambda_m W(h_m) = a_{f\star S} = f * a_S = f * \sum_n s_n W(g_n).$$

In order for the sum in the left-hand side to converge in L^1, we need for the Wigner distribution of each eigenfunction h_m to be integrable. This is equivalent to $h_m \in M^1(\mathbb{R}^d)$ which follows from $g_n \in \mathcal{S}(\mathbb{R}^d)$ by [23, Theorem 4.1].

The L^1-error estimate now follows by applying Lemma D.2.7 with $\psi = f$ and blurring kernel $\phi = \sum_n s_n W(\varphi_n)$. □

Just as in Theorem D.1.2, we can possibly deconvolve $\sum_m \lambda_m W(h_m) = f * \sum_n s_n W(g_n)$ to recover f exactly provided the Fourier transform $\mathcal{F}\left(\sum_n s_n W(g_n)\right)$ is zero-free. It turns out that yields precisely the same expression for the deconvolution as if we would naively deconvolve $A_f^g = f \star (g \otimes g)$ using the Fourier-Wigner transform from Section D.2.2.

Note that the sum $\sum_m \lambda_m W(h_m)$ is easily seen to converge pointwise by the bound $|W(h_m)(z)| \leq 2^d \|h_m\|_{L^2}^2$ while we need the extra condition on the window for L^1-convergence. This is why we could not formulate Theorem D.1.3 with partial sums as we did for Theorem D.1.2.

The above argument can be taken another step to show that the reconstruction procedure is not stable as was the case for accumulated spectrograms as shown in (D.4.3). To see that the inverse mapping $I : \mathcal{S}^1 \to L^1(\mathbb{R}^{2d})$, $f \star S \mapsto \sum_m \lambda_m W(h_m)$ is not continuous, fix $\psi \in L^2(\mathbb{R}^d) \setminus M^1(\mathbb{R}^d)$ and consider the perturbation operator $A = \psi \otimes \psi \in \mathcal{S}^1$ for which we have

$$\|I(f \star S + \varepsilon A) - I(f \star S)\|_{L^1} = \varepsilon \|I(\psi \otimes \psi)\|_{L^1} = \varepsilon \|W(\psi)\|_{L^1} = \infty.$$

In Section D.7.1 we provide an example showing the performance of the estimator $\sum_m \lambda_m W(h_m)$ and discuss aspects of the numerical implementation.

D.5 Recovery via plane tiling

Using quantum harmonic analysis, it is easy to show that the spectrograms of an orthonormal basis add up to the function which is identically 1. Indeed, using the relation $1 \star (\varphi \otimes \varphi) = \|\varphi\|_{L^2}^2 I_{L^2}$ from [211, Proposition 3.2 (3)], we get for a normalized $\varphi \in L^2(\mathbb{R}^d)$ that

$$\begin{aligned}
\sum_n |V_\varphi(e_n)|^2 &= \sum_n (e_n \otimes e_n) \star (\varphi \otimes \varphi)\check{} \\
&= \left(\sum_n (e_n \otimes e_n)\right) \star (\varphi \otimes \varphi)\check{} \\
&= I \star (\varphi \otimes \varphi) \\
&= 1 * (\varphi \otimes \varphi) \star (\varphi \otimes \varphi)\check{} = 1 * |V_\varphi \varphi|^2 = \|V_\varphi \varphi\|_{L^2}^2 = 1.
\end{aligned}$$

Intuitively, we should expect that those basis elements whose spectrograms are primarily supported away from the support of f should lose most of their mass when we apply A_f^g to them and the rest should remain intact or be scaled by something proportional to f. This is the motivation for the plane tiling approach which we prove below. The proof is rather straight-forward and we are able to inherit the main error estimate from Theorem D.1.1 as the sum approaches the same quantity ϑ from (D.1.3) in the $K \to \infty$ situation in that theorem.

Proof of Theorem D.1.4. We first rework the estimator $\sum_n |V_\varphi(A_f^g e_n)(z)|^2$ into a more manageable form using the self-adjointness of A_f^g and Example D.2.2 as

$$\begin{aligned}
\sum_n |V_\varphi(A_f^g e_n)(z)|^2 &= \sum_n \left(A_f^g e_n \otimes A_f^g e_n\right) \star (\varphi \otimes \varphi)\check{}(z) \\
&= A_f^g \left(\sum_n e_n \otimes e_n\right) A_f^g \star (\varphi \otimes \varphi)\check{}(z) \\
&= \left(A_f^g I A_f^g\right) \star (\varphi \otimes \varphi)\check{}(z) \\
&= \left(A_f^g\right)^2 \star (\varphi \otimes \varphi)\check{}(z) = \vartheta(z)
\end{aligned}$$

where ϑ is the same as in Section D.3. The same analysis on the size of $\|\vartheta - f^2\|_{L^1}$ from the proof of Theorem D.1.1 again applies and yields the desired conclusion. □

Remark D.5.1. The above result can be extended to mixed-state localization operators as was done in Section D.4 through some technical considerations. More specifically, it is possible to control the error $\| \sum_n Q_T((f \star S)e_n) - f^2\|_{L^1}$ if S

and T are positive rank-one operators whose spectral decomposition consists of Schwartz functions. This is done by bounding the trace norm of the error operator in the expansion $(f \star S)^2 = f^2 \star S + V_g^* T V_g$ in a similar way to how it was done in the proof of Theorem D.1.1.

If one has a choice between this method and the average observed spectrogram from Section D.3, the better method depends on the size of the support of the symbol. A good choice for the orthonormal basis are time-frequency shifted Hermite functions as these will tile out a growing circle [51]. By choosing the time-frequency shift appropriately, we can thus approximate $f(z_0)$ well by $\sum_{n=1}^N |V_\varphi(A_f^g(\pi(z_0)h_n))(z)|^2$ for small N. This is illustrated in Section D.7.1 below.

D.6 Recovery via Gabor projection

The main difference between Gabor projection and the other proposed methods is that we are estimating each pixel of f in an independent way. Notably we do this without any L^∞ continuity guarantees. The idea of the procedure laid out in the introduction is purely intuitive and we can strengthen this intuition by investigating the special case $\varphi = g$ in detail. In particular, this is when the procedure is a (pointwise) projection to the Gabor space $V_g(L^2)$ in the sense that

$$V_g(A_f^g(\pi(z)g))(z) = V_g V_g^*(f \cdot V_g g(\cdot - z))(z) = P_{V_g(L^2)}[f \cdot V_g g(\cdot - z)](z). \tag{D.6.1}$$

If we had some L^∞ continuity conditions for the orthogonal projection onto the Gabor space we could therefore get more guarantees on the performance of this method. Note also that since $V_g g$ is continuous, we could use $V_g(A_f^g(\pi(z)g))$ to estimate the value of f for a neighborhood of z which would lead to a shorter runtime but worse recovery performance.

Proof of Theorem D.1.5. The proof essentially boils down to expanding the short-time Fourier transform and synthesis V_g^* in (D.6.1), changing the order of integration, and identifying the blurring kernel. Indeed,

$$\begin{aligned} V_\varphi\big(A_f^g(\pi(z)g)\big)(z) &= \int_{\mathbb{R}^d} V_g^*\big(f \cdot V_g(\pi(z)\varphi)\big)(t)\overline{\pi(z)\varphi(t)}\,dt \\ &= \int_{\mathbb{R}^d}\int_{\mathbb{R}^{2d}} f(w)V_g(\pi(z)\varphi)(w)\pi(w)g(t)\overline{\pi(z)\varphi(t)}\,dw\,dt. \end{aligned}$$

From here we may use Fubini as can be seen by considering

$$\begin{aligned}
&\int_{\mathbb{R}^d}\int_{\mathbb{R}^{2d}} \left| f(w) V_g(\pi(z)\varphi)(w)\pi(w)g(t)\overline{\pi(z)\varphi(t)} \right| dw\, dt \\
&\qquad = \int_{\mathbb{R}^{2d}} |f(w)||V_g\varphi(w-z)| \underbrace{\int_{\mathbb{R}^d} \left|\pi(w)g(t)\overline{\pi(z)\varphi(t)}\right| dt}_{\leq \|g\|_{L^2}\|\varphi\|_{L^2}}\, dw \\
&\qquad \leq \|g\|_{L^2}\|\varphi\|_{L^2}\|f\|_{L^1}\|V_g\varphi\|_{L^\infty} < \infty.
\end{aligned}$$

Hence we may continue with a similar computation as

$$\begin{aligned}
V_\varphi(A_f^g(\pi(z)g))(z) &= \int_{\mathbb{R}^{2d}} f(w)e^{-2\pi i x\cdot\eta} V_g\varphi(w-z)\left(\int_{\mathbb{R}^d} \pi(w)g(t)\overline{\pi(z)\varphi(t)}\, dt\right) dw \\
&= \int_{\mathbb{R}^{2d}} f(w)e^{-2\pi i x\cdot\eta} V_g\varphi(w-z)\overline{V_g(\pi(z)\varphi)(w)}\, dw \\
&= \int_{\mathbb{R}^{2d}} f(w)|V_\varphi g(z-w)|^2\, dw = f * |V_\varphi g|^2(z)
\end{aligned}$$

where we used that $V_g(\pi(z)\varphi)(w) = e^{-2\pi i x\cdot\eta}V_g\varphi(w-z)$ for $z = (x,\omega),\ w = (y,\eta)$ and cancelled out the two phase factors. □

Note in particular that the above construction handles noise in the input $\pi(z)\varphi$ gracefully in the sense that if we add a perturbation, the effect on the absolute value of the estimator is proportional to the L^2 energy of the noise as can be seen by the linearity of the estimator and the boundedness of A_f^g and V_φ.

D.7 Numerical implementation

In computer applications there is no continuum and the integral in the the localization operator definition (D.1.1) is replaced by a sum, most often over some lattice, yielding what is referred to as a *Gabor multiplier* [75]. While showing that the results of this paper carry over to this setting is a non-trivial undertaking which we do not attempt, we settle for investigating the numerical behavior and draw only empirical conclusions. Do note however that many results on localization operators do carry over to the Gabor multiplier setting [43, 75, 81] and that in particular localization operators can be approximated in the $\mathcal{S}^1$ norm by Gabor multipliers [81].

Before looking at details, we present a brief visual comparison between the different methods for a collection of symbols and lattice parameters in Figure D.1. All of the figures presented in this section were generated using the Large

Time/Frequency Analysis Toolbox (LTFAT) [175] and the associated code can be found in the GitHub repository[2].

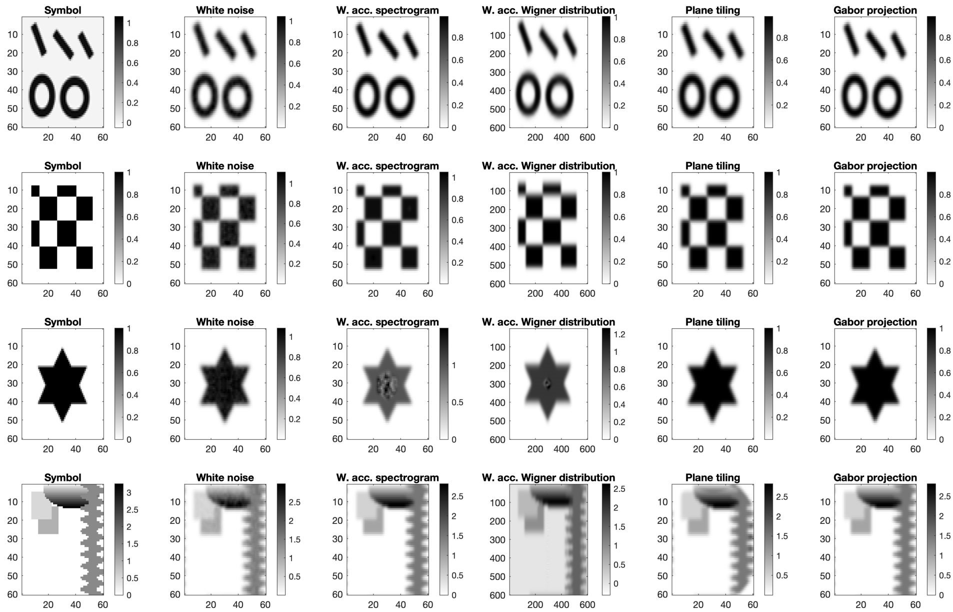

Figure D.1: Overview of all methods for a collection of four symbols.

D.7.1 Implementation details and examples

In this section we go through all methods and discuss implementation details.

White noise

Looking at the formulation of Proposition D.3.1 and Theorem D.1.1 in detail, one sees that the process for how the average observed spectrogram approximates f has many intermediate steps. Essentially, $\rho_K \to \vartheta \approx f^2 * |V_g g|^2 \approx f^2$. To highlight the differences between these steps and to show how the estimator handles non-binary symbols, we present plots of all the quantities next to each other in Figure D.2.

[2] `https://github.com/SimonHalvdansson/Localization-Operator-Symbol-Recovery`

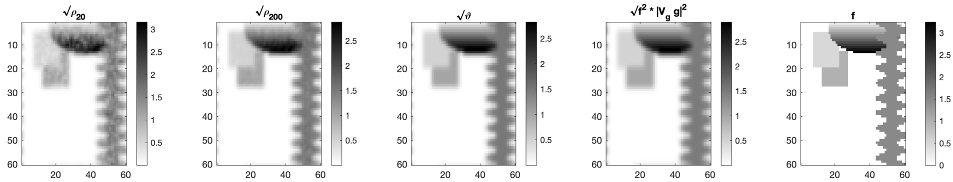

Figure D.2: All the intermediate steps in the white noise approximation detailed in Section D.3 for $\varphi = g$. The subscript on ρ indicates the number of samples K used.

Note in particular that the difference between ϑ and $f^2 * |V_g^g|$ which corresponds to the error estimate in Lemma D.3.4 is negligible in the above example. However, the averaging of 200 observations in ρ_{200} leads to a visible difference from ϑ which is not present in the limit of an infinite number of observations. The conclusion here is that the convergence, which is shown to have L^1 error proportional to $\frac{1}{\sqrt{K}}$ in Theorem D.1.1 where K is the number of observations, really is slow to converge in practice.

Since our implicit estimator for $|f|$ in Proposition D.3.1 and Theorem D.1.1 is $\sqrt{\frac{\rho}{\sigma^2}}$, we need some knowledge of σ^2 in order to be able to scale our estimator correctly. If we have the ability to inspect the noise realizations $(\mathcal{N}_k)_{k=1}^K$, this is straight-forward, either via the variance of each $\mathcal{N}_k$ or as the average value of $|V_\varphi(\mathcal{N}_k)|^2$ as can be seen from the calculations in the proof of Lemma D.3.2. Note however that the variance σ^2 affects our estimator linearly so even without an estimate for σ^2 we can correctly estimate $|f|$ up to a constant factor.

In [182], there is no need to estimate the variance explicitly as the mask is assumed to be binary. Implicitly, they use the maximum value of the estimator $\|\rho\|_\infty$ as an estimate of the variance as they use this to normalize ρ.

Weighted accumulated spectrogram

In the following example, we set both the reconstruction and original window to be the standard Gaussian for convenience. Consequently, the blurring kernel or equivalently the impulse response of the system $f \mapsto \sum_m \lambda_m |V_g h_m|^2$ is given by a two-dimensional Gaussian as this is the STFT of a Gaussian with respect to itself. In the finite setting, it is easy to also find the impulse response of the system by letting the input symbol be a Dirac delta. This approach is employed in Figure D.3 where we have deconvolved the estimator to recover the original symbol precisely as

$$\mathcal{F}^{-1}\left(\frac{\mathcal{F}\left(\sum_m \lambda_m |V_g h_m|^2\right)}{\mathcal{F}\left(\sum_m \lambda_m^0 |V_g h_m^0|^2\right)}\right) \qquad \text{where } A_{\delta_0}^g = \sum_m \lambda_m^0 (h_m^0 \otimes h_m^0).$$

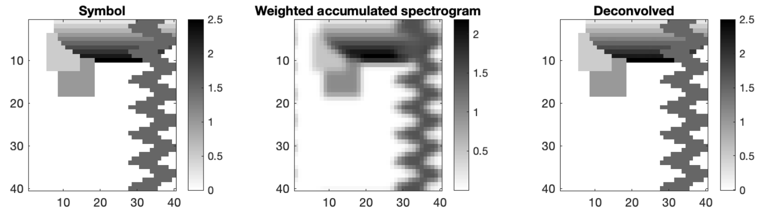

Figure D.3: A symbol, the associated accumulated spectrogram and the deconvolved estimate of the symbol.

This deconvolution procedure is clearly very unstable and if we increase the resolution of the frame we eventually encounter noticeable errors as can be seen below.

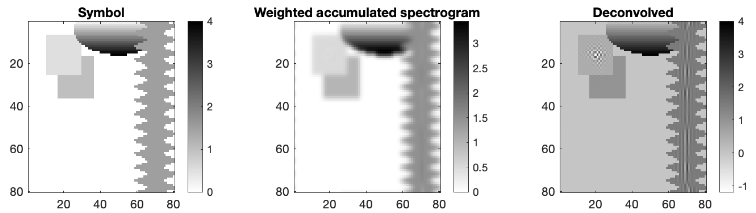

Figure D.4: Example of a deconvolution with visible errors.

The performance of the estimator is also dependent on the orthogonality of the eigenfunctions. In the following example, we have a large amount of eigenfunctions with very similar eigenvalues and the eigenfunctions from LTFAT are not perfectly orthogonal, resulting in visible errors, see Figure D.5.

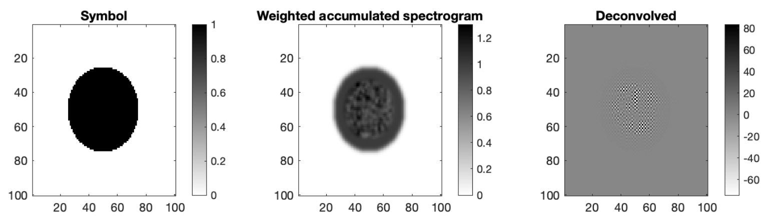

Figure D.5: Example of a failed deconvolution due to non-orthogonal eigenfunctions.

Weighted accumulated Wigner distribution

The Wigner-based approach in Theorem D.1.3 is notably more direct than the spectrogram approach in Theorem D.1.2 in that there is no reconstruction window.

Numerically however, there is some ambiguity as to how the integral defining $W(\phi)$ (D.2.3) should be discretized. The implementations in LTFAT and MATLAB [216] differ but provided the symbol f is only supported on the positive frequency half of phase space, we get similar results. In the implementation, we employ a procedure which compresses the matrix defining the symbol so that it is zeroed out for negative frequency components to avoid these issues.

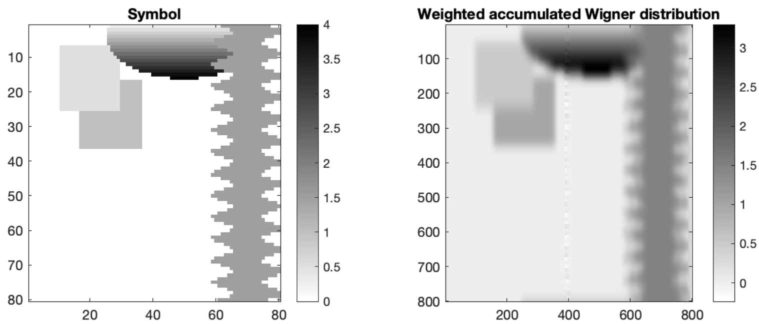

Figure D.6: A symbol and the corresponding weighted accumulated Wigner distribution estimator.

Due to the aforementioned discretization problem and the positive frequency restriction, finding the system impulse response to perform a deconvolution is not as straight-forward as in the weighted accumulated spectrogram case and we do not attempt it.

Note that in the example above and those in Figure D.1, the weighted accumulated Wigner distribution is considerably more blurry in the frequency direction than in the time direction. For both cases, the window function is a standard Gaussian.

Plane tiling

As discussed briefly near the end of Section D.5, by time-frequency shifting a collection of Hermite functions we can get an orthonormal basis which has an appropriate center in the time-frequency plane. This is illustrated in the following figure where we use different number of basis elements to highlight the incremental nature of the approximation.

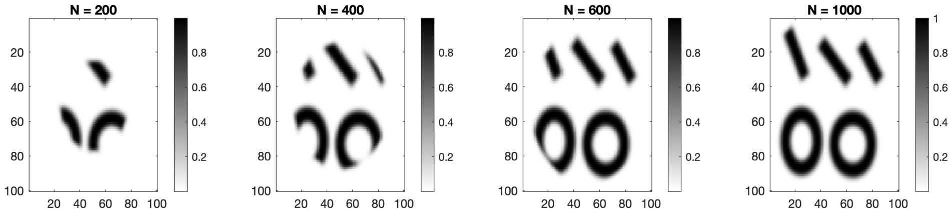

Figure D.7: The plane tiling estimator for the same orthonormal basis with different number of terms N.

As each spectrogram has total L^1 energy 1, the number of required basis elements is dependent on the size of the support of the symbol and the resolution of the frame.

To obtain an orthonormal basis which tiles the plane differently from the Hermites, one can use the eigenfunctions of a localization operator with a prescribed symbol.

Gabor projection

As discussed in the introduction and as is clear from the formulation of Theorem D.1.5, Gabor projection yields an identical estimator to the weighted accumulated spectrogram. From the numerical side, the main differences are that we can choose which regions of the symbol we want to recover. Moreover, since we are not relying on the eigendata of the operator, we are not susceptible to the type of problems that lead to the issue highlighted in Figure D.5 above. This allows us to perform deconvolution for a wider set of symbols and lattice parameters as can be seen by comparing Figure D.5 with the one below.

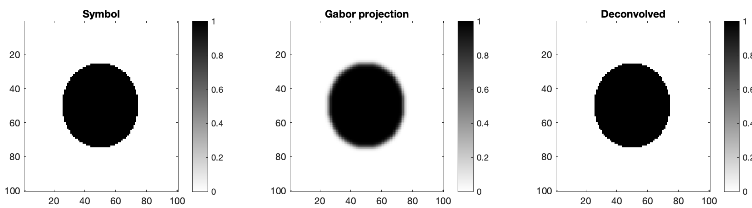

Figure D.8: Example of Gabor projection at high frame resolution with a circle as a symbol and a deconvolved version.

D.7.2 Performance comparison

Lastly we numerically investigate the L^1 error for the different methods. For an estimator ρ, we compute the error as

$$E = \frac{\|\rho - f\|_{L^1}}{\|f\|_{L^1}}$$

where the $\|f\|_{L^1}$ denominator is for normalization. As was discussed in [8] and as is the case for all estimators discussed in this paper, the estimators mostly differ from the symbol by some sort of blurring kernel which is constant in size in the sense that if f is dilated, the blurring kernel has a smaller effect. In the discrete setting, dilating the symbol is equivalent to increasing the resolution of the time-frequency frame and so for small lattice parameters we should expect smaller errors. For this reason, all of the errors reported below are for the same time-frequency frame and the absolute sizes of the errors should not be given too much weight. Instead, one should look at how the different methods compare.

Table D.1: Reconstruction L^1 error for different methods and symbols. Errors are reported in percent of L^1 energy of the symbol. Abbreviations: **WN** = White Noise, **WAS** = Weighted Accumulated Spectrogram, **WAWD** = Weighted Accumulated Wigner Distribution, **PT** = Plane Tiling, **GP** = Gabor Projection

Symbol	**WN**	**WAS**	**WAWD**	**PT**	**GP**
Circle	14.8	13.4	15.7	12.1	10.0
Sum of Gaussians	7.2	4.5	20.1	6.2	4.5
Star	13.7	13.1	16.0	12.0	9.7
Lines & circles	30.0	23.5	35.8	29.6	23.5
Blurred lines & circles	22.9	17.4	32.8	22.8	17.4
Tiles	29.3	23.7	29.1	28.1	23.6
NTNU	16.9	13.1	19.2	15.3	13.1

By comparing the errors for Gabor projection with the accumulated spectrogram we see the benefits of Gabor projection discussed in Section D.7.1 above. The reason the errors for plane tiling are generally lower than those for white noise is that we only use $K = 200$ samples.

All of the symbols used in the comparison are collected in the figure below.

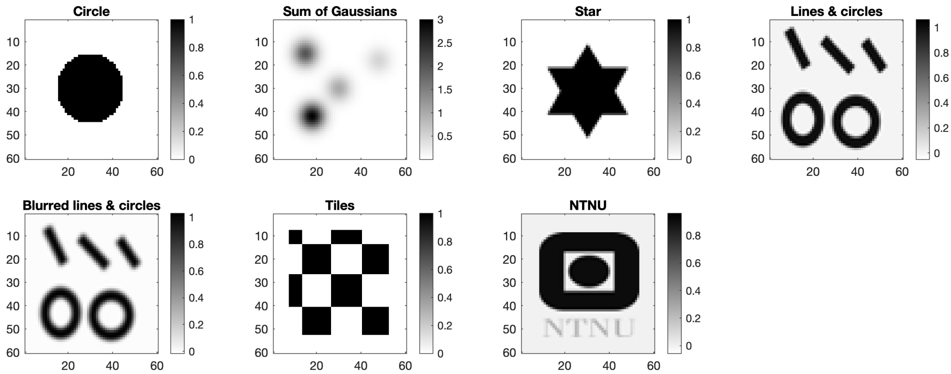

Figure D.9: The symbols used for the comparison in Table D.1.

Paper E

On Accumulated Spectrograms for Gabor Frames

Simon Halvdansson

Published in
Journal of Mathematical Analysis and Applications,
2025, Volume 543

Paper E

On Accumulated Spectrograms for Gabor Frames

Abstract

Analogs of classical results on accumulated spectrograms, the sum of spectrograms of eigenfunctions of localization operators, are established for Gabor multipliers. We show that the lattice ℓ^1 distance between the accumulated spectrogram and the indicator function of the Gabor multiplier mask is bounded by the number of lattice points near the boundary of the mask and that this bound is sharp in general. The methods developed for the proofs are also used to show that the Weyl-Heisenberg ensemble restricted to a lattice is hyperuniform when the Gabor frame is tight.

E.1 Introduction and main results

In time-frequency analysis, localization operators restrict a signal f to a subset Ω of the time-frequency plane [51] by means of a restricted resolution of the identity via the short-time Fourier transform (STFT) $V_g f$ as

$$A_\Omega^g f = \int_\Omega V_g f(z)\pi(z)g\,dz, \qquad V_g f(z) = \langle f, \pi(z)g\rangle$$

where $g \in L^2(\mathbb{R}^d)$ is a *window function*, $z = (x, \omega) \in \mathbb{R}^{2d}$ is a point in *time-frequency space* and $\pi(z)f(t) = \pi(x, \omega)f(t) = e^{2\pi i\omega\cdot t} f(t-x)$ is a *time-frequency shift* [107]. The spectral behavior of such operators has been studied extensively [7, 74, 76, 124, 162, 164, 180], showing that there are approximately $|\Omega|$ eigenvalues close to 1, followed by a *plunge region* of size comparable to the length of the perimeter of Ω, after which the remaining eigenvalues are close to 0. In [8], Abreu, Gröchenig and Romero showed that Ω can be estimated from only the spectrograms of the first $\lceil|\Omega|\rceil$ eigenfunctions using the *accumulated spectrogram*

$$\rho_\Omega(z) = \sum_{k=1}^{\lceil|\Omega|\rceil} |V_g h_k^\Omega(z)|^2 \tag{E.1.1}$$

where $(h_k^\Omega)_{k=1}^\infty$ are the eigenfunctions of the localization operator. That result was eventually improved by a sharp estimate in [4] to

$$\|\rho_\Omega - \chi_\Omega\|_{L^1(\mathbb{R}^{2d})} \leq C_g|\partial\Omega| \tag{E.1.2}$$

where χ_Ω is the indicator function of Ω and C_g is a constant depending only on g.

While these results have been numerically verified and used in the discrete setting [8, 9, 57, 60, 118], there have been no proofs that corresponding results hold for Gabor multipliers, the discrete version of localization operators. It is the goal of this article to fill this gap by establishing versions of the main results of [4, 8] which are valid in the setting of Gabor frames. Before stating the results, we establish some notation and conventions.

A *Gabor frame* is a collection $\{\pi(\lambda)g\}_{\lambda\in\Lambda}$, induced by the pair (g, Λ), where $g \in L^2(\mathbb{R}^d)$ is the window function and $\Lambda \subset \mathbb{R}^{2d}$ is a lattice, satisfying the inequalities

$$A\|f\|_{L^2}^2 \leq \sum_{\lambda\in\Lambda} |\langle f, \pi(\lambda)g\rangle|^2 \leq B\|f\|_{L^2}^2 \qquad \text{for all } f \in L^2(\mathbb{R}^d)$$

for a pair of *frame bounds* $A, B > 0$ [107]. For a given Gabor frame there always exists a *dual window* h such that the reconstruction formula

$$f = \sum_{\lambda\in\Lambda} V_g f(\lambda)\pi(\lambda)h$$

holds for $f \in L^2(\mathbb{R}^d)$ in the L^2 sense. When $A = B$ the frame is said to be *tight* and the dual window is a scalar multiple of g. The analog of localization operators in this setting, *Gabor multipliers* [75], are then constructed by restricting the above formula to a subset $\Omega \subset \mathbb{R}^{2d}$ as

$$G_{\Omega,\Lambda}^g f = \sum_{\lambda\in\Lambda} \chi_\Omega(\lambda) V_g f(\lambda)\pi(\lambda)g \qquad \text{for } f \in L^2(\mathbb{R}^d). \tag{E.1.3}$$

While the tight $A = B$ situation is preferable, we will throughout the article allow for non-tight frames but still take the Gabor multipliers to be defined as in the formula above with $h = g$. This makes the operator self-adjoint and together with a compact mask Ω this guarantees that $G_{\Omega,\Lambda}^g$ is both a compact and self-adjoint operator whose eigendecomposition can be written as

$$G_{\Omega,\Lambda}^g = \sum_{k=1}^\infty \lambda_k^\Omega (h_k^\Omega \otimes h_k^\Omega)$$

where $(h_k^\Omega)_{k=1}^\infty$ is an orthonormal basis of $L^2(\mathbb{R}^d)$, $(h_k^\Omega \otimes h_k^\Omega)(f) = \langle f, h_k^\Omega \rangle h_k^\Omega$ is a rank-one projection operator [190] and $(\lambda_k^\Omega)_{k=1}^\infty$ are the eigenvalues. The accumulated spectrogram on Λ is then defined, analogously to (E.1.1), as

$$\rho_\Omega(\lambda) = \frac{1}{\|g\|_{L^2}^2} \sum_{k=1}^{A_\Omega} |V_g h_k^\Omega(\lambda)|^2, \qquad A_\Omega = \left\lceil \frac{\#(\Omega \cap \Lambda)\|g\|_{L^2}^2}{B} \right\rceil$$

since it is at most the first A_Ω eigenvalues which are close to the upper frame bound B in this setting [81]. For technical reasons, all our main results will require the window function g to belong to the space

$$M_\Lambda^*(\mathbb{R}^d) = \left\{ g \in L^2(\mathbb{R}^d) : \|g\|_{M_\Lambda^*} = \left(\sum_{\lambda \in \Lambda} |\lambda||V_g g(\lambda)|^2 \right)^{1/2} < \infty \right\}.$$

By [77, Proposition 2.1], $M_\Lambda^*(\mathbb{R})$ can be embedded in $L^2(\mathbb{R}) \cap M_{v_{1/2}}^{4/3}(\mathbb{R})$ where $M_{v_s}^p(\mathbb{R})$ is the weighted modulation space with weight function $v_s(z) = (1 + |z|)^s$ [107].

The boundary measure $|\partial\Omega|$ used in (E.1.2) cannot be used in the discrete setting in general apart from when $d = 1$ due to some pathological counterexamples. Details of this and how we can circumvent the problem by assuming that Ω has *maximally Ahlfors regular boundary* are discussed in Section E.2.2. To state our results generally, we need the lattice-dependent boundary measure

$$\partial_\Lambda^r \Omega = \Lambda \cap (\partial\Omega + B(0, r)).$$

With it, we are ready to state our main results, the first of which is analogous to the sharp growth bound (E.1.2).

Theorem E.1.1. *Let $g \in M_\Lambda^*(\mathbb{R}^d)$ and Λ be such that (g, Λ) induces a frame with frame constants $A, B > 0$, $r > 0$ and $\Omega \subset \mathbb{R}^{2d}$ be compact. Then there exists a constant C depending only on r and d such that*

$$\|\rho_\Omega - \chi_\Omega\|_{\ell^1(\Lambda)} \leq C \left(\|g\|_{M_\Lambda^*}^2 + 1 \right) \#\partial_\Lambda^{r_\Lambda} \Omega + 2\frac{B - A}{B} \#(\Omega \cap \Lambda) + \frac{B}{\|g\|_{L^2}^2}$$

where $r_\Lambda = r + l_M$ and l_M is the diameter of the fundamental domain of Λ.

Note in particular that when the frame is tight, the error grows as $\#\partial_\Lambda^{r_\Lambda}\Omega$. This result can be used to approximate Ω directly from ρ_Ω as a level set.

Corollary E.1.2. Let $g \in M_\Lambda^*(\mathbb{R}^d)$ and Λ be such that (g, Λ) induces a frame with frame constants $A, B > 0$, $r > 0$, $\Omega \subset \mathbb{R}^{2d}$ be compact and

$$\tilde{\Omega} = \left\{ \lambda \in \Lambda : \rho_\Omega(\lambda) > 1/2 \right\}.$$

Then there exists an constant C dependent only on r and d such that

$$\#\big((\Omega\Delta\tilde{\Omega})\cap\Lambda\big) \leq C\left(\|g\|_{M^*_\Lambda}^2+1\right)\#\partial_\Lambda^{r_\Lambda}\Omega + 4\frac{B-A}{B}\#(\Omega\cap\Lambda) + \frac{2B}{\|g\|_{L^2}^2}$$

where Δ denotes the symmetric difference of two sets, $r_\Lambda = r + l_M$ and l_M is the diameter of the fundamental domain of Λ.

In general, it is impossible to establish a tighter bound on the $\ell^1(\Lambda)$ norm than Theorem E.1.1 which we prove in Theorem E.1.3 below where a special case is investigated. This result is analogous to [4, Theorem 1.6] but in the lattice setting we need to assume some additional conditions on the window g.

Theorem E.1.3. *Let $g \in M^*_\Lambda(\mathbb{R}^d)$ and Λ be such (g,Λ) induces a tight frame with frame constant* 1 *and*

$$|V_g g(z)| \leq C(1+|z|)^{-s}, \qquad V_g g(\lambda) \neq 0 \quad \text{for } \lambda \in \Lambda\cap B(0, r+3l_M)$$

for some $C > 0$ and $s > 2d+1$ where l_M is the diameter of the fundamental domain of Λ. Then there exists constants C_1, C_2 only dependent on g, the lattice Λ and the radius r such that

$$C_1\#\partial_\Lambda^{r_\Lambda}B(0,R) \leq \|\rho_{B(0,R)} - \chi_{B(0,R)}\|_{\ell^1(\Lambda)} \leq C_2\#\partial_\Lambda^{r_\Lambda}B(0,R)$$

for all $R > 0$, where $r_\Lambda = r + l_M$ and l_M is the diameter of the fundamental domain of Λ.

The proofs of these results follow similar paths to the original results in [4, 8] and the main novelty of the present work lies in relating the eigenvalues of Gabor multipliers to our boundary measure.

Our final main result is not directly related to accumulated spectrograms, but its proof uses parts of the same method used to prove the three theorems above.

The Weyl-Heisenberg ensemble, originally introduced in [9] and further studied in [2, 6, 132], is a determinantal point process induced by a window function which generalizes the Ginibre ensemble. While not previously mentioned in the literature to the best of our knowledge, there is a clear discrete counterpart for Gabor frames where the point process is restricted to Λ. We are able to show that when the Gabor frame is tight, the point process is hyperuniform which an analog to one of the main results for the continuous case in [9].

Theorem E.1.4. *Let $g \in M^*_\Lambda(\mathbb{R}^d)$ and Λ be such that (g,Λ) induces a tight frame, then the determinantal point process $\mathcal{X}_g$ on Λ with correlation kernel $K_g(\lambda,\lambda') = \langle\pi(\lambda')g, \pi(\lambda)g\rangle$ is hyperuniform.*

In Section E.4, we give a proper definition of the point process, define hyperuniformity and prove the theorem.

Notational conventions

The ball centered at $z \in \mathbb{R}^{2d}$ with radius r will be denoted by $B(z, r)$. When measuring the size of a set, we will write # for cardinalities of discrete sets and $|\cdot|$ for Lebesgue measures of sets with interiors as well as the arc length of paths or the $(d-1)$-dimensional Hausdorff measure of a boundary $\partial\Omega$. For complex numbers z, $|z|$ will denote the absolute value as customary. The symmetric difference between two sets A, B will be denoted by $A\Delta B := (A \setminus B) \cup (B \setminus A)$. The values of constants will be allowed to change between inequalities so that factors can be absorbed. For a lattice Λ, the set of summable sequences will be denoted by $\ell^1(\Lambda)$ with $\|c\|_{\ell^1(\Lambda)} = \sum_{\lambda\in\Lambda} |c(\lambda)|$ and discrete convolutions between elements of $\ell^1(\Lambda)$ will be denoted by $*_\Lambda$.

E.2 Background and tools

In this section we collect some common tools and results which will be used throughout the article. For more background on the motivations and interpretations of accumulated spectrograms, see the original article [8], and for more properties of Gabor multipliers and a more thorough introduction, see [75, 107].

E.2.1 Bounding regularization error

In forthcoming proofs, we will repeatedly need to bound the difference $\chi_\Omega - \chi_\Omega *_\Lambda \phi$. The main tool for the continuous version of this is [8, Lemma 3.2]. With the goal of establishing a version of that result for the lattice setting, we prove the following lemma showing that the characteristic function χ_Ω can be well approximated by a Schwartz function f.

Lemma E.2.1. *Given a compact set $\Omega \subset \mathbb{R}^{2d}$ and $r > 0$, there exists a Schwartz function f and a constant C only dependent on the dimension d such that*

(i) $f(z) = 1$ *for* $z \in \Omega$,

(ii) $\mathrm{supp}(f) \subset \Omega + B(0, r)$,

(iii) $\|\nabla f\|_{L^\infty} \leq C/r$,

(iv) $\|\chi_\Omega - f\|_{\ell^1(\Lambda)} \leq \#\partial^r_\Lambda\Omega$.

Proof. Using the smooth bump function

$$\phi(x) = \chi_{[-1,1]}(x)e^{\frac{-1}{1-x^2}},$$

we can define the Schwartz function $\phi_r(z) = \frac{c}{r^{2d}}\phi(|z|/r)$ supported in $B(0,r)$ and by choosing c appropriately we can guarantee that it integrates to 1. Indeed, with ω_d the surface area of the unit sphere in $\mathbb{R}^d$, the integral

$$\begin{aligned}\frac{c}{r^{2d}}\int_{B(0,r)}\phi(|z|/r)\,dz &= \frac{c}{r^{2d}}\omega_{2d-1}\int_0^r x^{2d-1}\phi(x/r)\,dx\\ &= \frac{c}{r^{2d}}\omega_{2d-1}\int_0^1 y^{2d-1}r^{2d-1}\phi(y)r\,dy\\ &= c\omega_{2d-1}\int_0^1 y^{2d-1}\phi(y)\,dy\end{aligned}$$

is independent of r. We are now ready to define f as

$$f(z) = \chi_{\Omega+B(0,r/2)} * \phi_{r/2}(z).$$

Since $\phi_{r/2}$ is Schwartz, so is f and properties (i) and (ii) follow from standard properties of convolutions using that $\mathrm{supp}(\phi_{r/2}) \subset B(0,r/2)$. Next (iv) follow from $\mathrm{supp}(\chi_\Omega - f) \subset \partial\Omega + B(0,r)$ and that $\|f\|_{L^\infty} \le \|\chi_\Omega\|_{L^\infty}\|\phi_{r/2}\|_{L^1} = 1$ by Young's inequality. Finally for the bound on $|\nabla f|$, we can estimate

$$\begin{aligned}\|\nabla f\|_{L^\infty} &= \|\chi_{\Omega+B(0,r/2)} * (\nabla\phi_{r/2})\|_{L^\infty}\\ &\le \|\chi_{\Omega+B(0,r/2)}\|_{L^\infty}\|\nabla\phi_{r/2}\|_{L^1} = \int_{B(0,r/2)}|\nabla\phi_{r/2}(z)|\,dz\\ &= \frac{c}{(r/2)^{2d}}\int_{B(0,r/2)}\frac{1}{r/2}\left|\phi'\left(\frac{|z|}{r/2}\right)\right|\,dz\\ &= \frac{c}{(r/2)^{2d+1}}\omega_{2d-1}\int_0^{r/2}x^{2d-1}\left|\phi'\left(\frac{x}{r/2}\right)\right|\,dx\\ &= \frac{c}{(r/2)^{2d+1}}\omega_{2d-1}\frac{r^{2d}}{2^{2d}}\int_0^1 y^{2d-1}|\phi'(y)|\,dy = \frac{C}{r}\end{aligned}$$

where C is a constant that collects terms independent of r.

□

Using this lemma, we can establish the promised lattice version of [8, Lemma 3.2].

Proposition E.2.2. Let $\phi \in \ell^1(\Lambda)$ be non-negative and satisfy

$$1-\delta \le \sum_{\lambda\in\Lambda}\phi(\lambda) \le 1, \qquad \sum_{\lambda\in\Lambda}|\lambda||\phi(\lambda)| < \infty$$

for some $0 \le \delta \le 1$. Then there exists a constant C dependent on $r > 0$ such that

$$\|\chi_\Omega - \chi_\Omega *_\Lambda \phi\|_{\ell^1(\Lambda)} \le C \left(\sum_{\lambda\in\Lambda} |\lambda||\phi(\lambda)| + 1 \right) \#\partial_\Lambda^{r_\Lambda}\Omega + \delta \#(\Omega \cap \Lambda)$$

for any compact set $\Omega \subset \mathbb{R}^{2d}$, where $r_\Lambda = r + l_M$ and l_M is the diameter of the fundamental domain of Λ.

Proof. Applying Lemma E.2.1, we can replace χ_Ω by f using a triangle inequality argument as

$$\begin{aligned}\|\chi_\Omega - \chi_\Omega *_\Lambda \phi\|_{\ell^1(\Lambda)} \le \underbrace{\|\chi_\Omega - f\|_{\ell^1(\Lambda)}}_{\le \#\partial_\Lambda^r \Omega} + \|f - f *_\Lambda \phi\|_{\ell^1(\Lambda)} \\ + \underbrace{\|f *_\Lambda \phi - \chi_\Omega *_\Lambda \phi\|_{\ell^1(\Lambda)}}_{\le \#\partial_\Lambda^r \Omega}\end{aligned}$$

where we used Young's inequality for the estimate on the last term, Lemma E.2.1 and that $\|\phi\|_{\ell^1(\Lambda)} \le 1$. Now it remains to show that the middle term can be bounded by a similar quantity. To do so, we will adapt [8, Lemma 3.2] to the lattice setting. Note that

$$\begin{aligned}\|f - f *_\Lambda \phi\|_{\ell^1(\Lambda)} &= \sum_{\lambda\in\Lambda} \left| \sum_{\lambda'\in\Lambda} f(\lambda')\phi(\lambda - \lambda') - f(\lambda) \right| \\ &= \sum_{\lambda\in\Lambda} \left| \sum_{\lambda'\in\Lambda} f(\lambda')\phi(\lambda - \lambda') - f(\lambda) \left(\sum_{\lambda'\in\Lambda} \phi(\lambda - \lambda') + 1 - \sum_{\lambda'\in\Lambda} \phi(\lambda - \lambda') \right) \right| \\ &\le \sum_{\lambda\in\Lambda} \left| \sum_{\lambda'\in\Lambda} \left[f(\lambda) - f(\lambda') \right] \phi(\lambda - \lambda') \right| + \sum_{\lambda\in\Lambda} \left| f(\lambda) \left(1 - \sum_{\lambda'\in\Lambda} \phi(\lambda - \lambda') \right) \right| \\ &\le \sum_{\lambda\in\Lambda} \left| \sum_{\lambda'\in\Lambda} \left[f(\lambda) - f(\lambda') \right] \phi(\lambda - \lambda') \right| + \delta \|f\|_{\ell^1(\Lambda)}\end{aligned}$$

since $1 = \sum_{\lambda'\in\Lambda} \phi(\lambda') + 1 - \sum_{\lambda'\in\Lambda} \phi(\lambda')$. By elementary calculus, it holds that

$$f(\lambda) - f(\lambda') = \int_0^1 \langle \nabla f(t(\lambda - \lambda') + \lambda'), \lambda - \lambda' \rangle \, dt,$$

and so we can write

$$\begin{aligned}\sum_{\lambda\in\Lambda}\left|\sum_{\lambda'\in\Lambda}\big[f(\lambda)-f(\lambda')\big]\phi(\lambda-\lambda')\right|&\\ &=\sum_{\lambda\in\Lambda}\left|\sum_{\lambda'\in\Lambda}\int_0^1\langle\nabla f(t(\lambda-\lambda')+\lambda'),\lambda-\lambda'\rangle\phi(\lambda-\lambda')\,dt\right|\\ &\le\int_0^1\sum_{\lambda\in\Lambda}\sum_{\lambda'\in\Lambda}|\nabla f(t(\lambda-\lambda')+\lambda')||\lambda-\lambda'||\phi(\lambda-\lambda')|\,dt\\ &=\int_0^1\sum_{\lambda\in\Lambda}\sum_{\lambda'\in\Lambda}|\nabla f(t\lambda+\lambda')||\lambda||\phi(\lambda)|\,dt\\ &=\sum_{\lambda\in\Lambda}|\lambda||\phi(\lambda)|\int_0^1\sum_{\lambda'\in\Lambda}|\nabla f(t\lambda+\lambda')|\,dt\end{aligned}$$

by Tonelli and Cauchy-Schwarz followed by a change of variables. We now claim that the final integral and sum can be uniformly bounded over all t and λ. Indeed, ∇f is supported in $\partial\Omega + B(0,r)$ and $|\nabla f|$ is uniformly bounded by C/r by Lemma E.2.1 so

$$\int_0^1\sum_{\lambda'\in\Lambda}|\nabla f(t\lambda+\lambda')|\,dt\le C\#\big((\Lambda+\lambda t)\cap(\partial\Omega+B(0,r))\big)$$

once we absorb r in C. Now λt can be written as $\lambda t=\lambda_0+z_0$ where $\lambda_0\in\Lambda$ and z_0 is in the fundamental domain of Λ. This can be used to bound the quantity using the inclusion

$$\begin{aligned}(\Lambda+\lambda t)\cap(\partial\Omega+B(0,r))&=(\Lambda+z_0)\cap(\partial\Omega+B(0,r))\\ &=\Lambda\cap(\partial\Omega+B(z_0,r))\subset\Lambda\cap(\partial\Omega+B(0,r_\Lambda)).\end{aligned}$$

Plugging this back into the $\|f-f*_\Lambda\phi\|_{\ell^1(\Lambda)}$ estimate yields

$$\|f-f*_\Lambda\phi\|_{\ell^1(\Lambda)}\le C\left(\sum_{\lambda\in\Lambda}|\lambda||\phi(\lambda)|\right)\#\partial_\Lambda^{r_\Lambda}\Omega+\delta\|f\|_{\ell^1(\Lambda)}.$$

Adding back the two $\#\partial_\Lambda^r\Omega$ terms from earlier and using that $\#\partial_\Lambda^r\Omega\le\#\partial_\Lambda^{r_\Lambda}\Omega$, we get

$$\|\chi_\Omega-\chi_\Omega*_\Lambda\phi\|_{\ell^1(\Lambda)}\le\left(C\sum_{\lambda\in\Lambda}|\lambda||\phi(\lambda)|+2\right)\#\partial_\Lambda^{r_\Lambda}\Omega+\delta\|f\|_{\ell^1(\Lambda)}.$$

The factor $\|f\|_{\ell^1(\Lambda)}$ can be expanded as

$$\|f\|_{\ell^1(\Lambda)} \leq \|f - \chi_\Omega\|_{\ell^1(\Lambda)} + \|\chi_\Omega\|_{\ell^1(\Lambda)} \leq \#\partial^r_\Lambda\Omega + \#(\Omega \cap \Lambda)$$

and once we add this additional $\#\partial^r_\Lambda\Omega \leq \#\partial^{r_\Lambda}_\Lambda\Omega$ term and use that $\delta \leq 1$ we can absorb into C to get the desired formulation. □

E.2.2 Relating $\#\partial^r_\Lambda\Omega$ to $|\partial\Omega|$

As mentioned in the introduction, our results are formulated with the more lattice-oriented boundary measure ∂^r_Λ given by

$$\partial^r_\Lambda\Omega = \Lambda \cap (\partial\Omega + B(0,r))$$

The following proposition clarifies the connection to the standard Lebesgue measure $|\partial\Omega|$ in the general case and shows why we cannot use it for $d > 1$ without additional assumptions.

Proposition E.2.3. For each radius r, lattice Λ and integer $k > 0$, there exist constants $C, D > 0$ such that

$$\#\partial^r_\Lambda\Omega \leq C|\partial\Omega| + D$$

for all compact $\Omega \subset \mathbb{R}^{2d}$ whose boundary consists of at most k closed curves if and only if $d = 1$.

Note that the constant C needs to be dependent on the number of closed curves which make up $\partial\Omega$ as one could otherwise construct a counterexample as $\Omega = \big(\Lambda + B(0,\varepsilon)\big) \cap B(0,R)$ in which case the left hand side would grow as R^2 and the right hand side as $R^2\varepsilon$.

Proof. Let l_m be the largest number so that all lattice points are separated by at least l_m. We first prove the inequality for $d = 1$ and then present a counterexample for $d > 1$.

Since Ω is bounded, the set $\partial\Omega + B(0,r)$ can be covered by a finite collection Q of squares with side length $l_m/\sqrt{2}$. Then each element of $\Lambda \cap (\partial\Omega + B(0,r))$ is in no more than one of these squares since the points of Λ are spaced by at least l_m and two points in a square with side length $l_m/\sqrt{2}$ are at most l_m apart from each other. Formally,

$$\#\big(\Lambda \cap (\partial\Omega + B(0,r))\big) \leq \#Q.$$

Meanwhile, the squares can be encapsulated in a bigger dilation around $\partial\Omega$:

$$\bigcup_{q\in Q} q \subset \partial\Omega + B(0, r + l_m).$$

As the total area of the squares is given by $\#Q\frac{l_m^2}{2}$, we can estimate

$$\#\big(\Lambda \cap (\partial\Omega + B(0,r))\big) \leq \#Q \leq \frac{2}{l_m^2}\big|\partial\Omega + B\big(0, r + l_m\big)\big|. \tag{E.2.1}$$

Now it only remains to relate this quantity to $|\partial\Omega|$. By assumption, $\partial\Omega$ can be decomposed into a collection of closed curves $\gamma_1, \dots, \gamma_k$. For each such closed curve, we claim that

$$|\gamma + B(0, R)| \leq C|\gamma| + D$$

where we have written R for $r + l_m$. Indeed, if $|\gamma| = \infty$ we are done so without loss of generality, we can place a finite collection of points $z_1, \dots, z_n$ along γ, spaced by R. It then holds that

$$\gamma + B(0, R) \subset \bigcup_{i=1}^{n} B(z_i, 2R)$$

since any point in $\gamma + B(0, R)$ is within R distance to a point in γ and any point in γ is within R distance to a point z_i.

The number of balls, n, can be related to $|\gamma|$ as

$$|\gamma| \leq R \cdot n \leq |\gamma| + R$$

since the distance between the centers along γ is R. Putting all of this together, we can estimate

$$\begin{aligned}
|\gamma + B(0, R)| &\leq \left|\bigcup_{i=1}^{n} B(z_i, 2R)\right| \\
&\leq n \cdot 4\pi R^2 \\
&\leq \left(\frac{|\gamma|}{R} + 1\right) 4\pi R^2
\end{aligned}$$

which proves the claim. Applying this result to (E.2.1), we obtain

$$\begin{aligned}
\#\big(\Lambda \cap (\partial\Omega + B(0,r))\big) &\leq \frac{2}{l_m^2}\sum_{i=1}^{k} |\gamma_i + B(0, r + l_m)| \\
&\leq \frac{2}{l_m^2}\sum_{i=1}^{k}\left(\frac{|\gamma_i|}{r + l_m} + 1\right) 4\pi\big(r + l_m\big)^2 \\
&= \frac{8\pi}{l_m^2}\left(\frac{|\partial\Omega|}{r + l_m} + k\right)\big(r + l_m\big)^2
\end{aligned}$$

from which we see that both constants only depend on Λ, r and k.

We will now show that equivalence does not hold for $d > 1$ using an example where Ω is particularly elongated. Specifically, let Ω be a hyperrectangle in $2d$ dimensions with all side lengths ε except for one of length L, i.e.,

$$\Omega = \left\{(x_1, \dots, x_{2d}) : 0 \le |x_1|, |x_2|, \dots, |x_{2d-1}| \le \varepsilon/2,\ 0 \le |x_{2d}| \le L/2\right\}.$$

Without loss of generality by rotating Ω if necessary, we can assume that there is an infinite collection of lattice points spaced by at most l_M, the maximum distance between two lattice points, along the axis where Ω has a side length L. Consequently, we can choose L large enough so that $\#\partial^r_\Lambda\Omega > D + 1$. Meanwhile, the surface area $|\partial\Omega|$ can be bounded by $cL\varepsilon^{2d-2}$ so by choosing ε small enough, we can make the $C|\partial\Omega|$ term arbitrarily small. In particular, if $C|\partial\Omega| < 1$ then $C|\partial\Omega| + D < D + 1$ which contradicts the $\#\partial^r_\Lambda\Omega$ bound. □

If we assume additional regularity of $\partial\Omega$, we can relate $\#\partial^r_\Lambda\Omega$ to $|\partial\Omega|$ in all dimensions. The following proposition was contributed by an anonymous referee.

A set Ω is said to have *maximally Ahlfors regular boundary* with constant $\kappa_{\partial\Omega}$ if

$$\mathcal{H}^{d-1}(\partial\Omega \cap B(z, r)) \ge \kappa_{\partial\Omega} r^{d-1} \qquad \text{for every } z \in \partial\Omega,\ 0 < r < |\partial\Omega|^{1/(d-1)}$$

where $\mathcal{H}^{d-1}$ is the $(d-1)$-dimensional Hausdorff measure. This condition is not too strict and has been used for similar purposes in [162, 163].

Proposition E.2.4. Let Ω be a compact set with maximally Ahlfors regular boundary with constant $\kappa_{\partial\Omega}$. For each radius $r > 0$ and lattice Λ, there exist constants C, D such that

$$\#\partial^r_\Lambda\Omega \le C\frac{|\partial\Omega|}{\kappa_{\partial\Omega}}\left(1 + \frac{D}{|\partial\Omega|}\right).$$

Proof. Let F be a fundamental domain of Λ with $\operatorname{diam}(F) = l_M$. Since Ω is compact it follows that the collection

$$\Lambda^* = \left\{\lambda \in \Lambda : (\lambda + F) \cap (\partial\Omega + B(0, r)) \ne \emptyset\right\}$$

is finite and the union of all translates $\lambda + F, \lambda \in \Lambda^*$ covers $\partial\Omega + B(0, r)$. This implies that

$$\left(\partial\Omega + B(0, r)\right) \subset \bigcup_{\lambda \in \Lambda^*} (\lambda + F) \subset \left(\partial\Omega + B(0, r + l_M)\right).$$

Since each set $\lambda + F$ contains exactly one point in Λ, it follows that $\#\partial_\Lambda^r \Omega = \#\big(\Lambda \cap (\partial\Omega + B(0,r))\big) \leq \#\Lambda^*$ and it remains to bound $\#\Lambda^*$. An application of [163, Lemma 2.1] yields

$$\begin{aligned}\#\Lambda^* = \frac{1}{|F|}\left|\bigcup_{\lambda\in\Lambda^*}(\lambda + F)\right| &\leq \frac{1}{|F|}\big|\partial\Omega + B(0, r + l_m)\big| \\ &\leq \frac{C_d}{|F|}\frac{|\partial\Omega|}{\kappa_{\partial\Omega}}(r + l_M)\left(1 + \frac{(r + l_M)^{d-1}}{|\partial\Omega|}\right)\end{aligned}$$

which concludes the proof. □

This result implies alternative versions of all the main results formulated with $|\partial\Omega|$ instead of $\#\partial_\Lambda^{r_\Lambda}\Omega$.

Remark E.2.5. The hyperrectangle from Proposition E.2.3 actually has maximally Ahlfors regular boundary but the constant goes to zero as $\varepsilon \to 0$ or $L \to \infty$ which is why it worked as a counterexample.

E.2.3 Spectral properties of Gabor multipliers

Our key to proving all of the main theorems will be to relate them to spectral properties of Gabor multipliers. For this reason, we collect some results on the eigenvalues in this section, the first of which is the simple bound

$$\begin{aligned}0 \leq \lambda_k^\Omega = \big\langle G_{\Omega,\Lambda}^g h_k^\Omega, h_k^\Omega\big\rangle &= \sum_{\lambda\in\Lambda}\chi_\Omega(\lambda)V_g h_k^\Omega(\lambda)\langle\pi(\lambda)g, h_k^\Omega\rangle \\ &\leq \sum_{\lambda\in\Lambda}|\langle h_k^\Omega, \pi(\lambda)g\rangle|^2 \leq B\|h_k^\Omega\|_{L^2}^2 = B.\end{aligned} \tag{E.2.2}$$

In the upcoming proofs we will repeatedly have use for the fact that for any orthonormal basis $(e_n)_{n=1}^\infty$,

$$\sum_{n=1}^\infty |V_g e_n(z)|^2 = \|g\|_{L^2}^2 \qquad \text{for all } z \in \mathbb{R}^{2d}. \tag{E.2.3}$$

To see that it holds, simply write out $|V_g e_n(z)|^2 = \langle e_n, \pi(z)g\rangle\langle\pi(z)g, e_n\rangle$ and sum. Note that this places an upper bound of 1 on the accumulated spectrogram ρ_Ω.

The next results on the trace and Hilbert-Schmidt norm of Gabor multipliers are standard but we provide a proof in the interest of completion.

Lemma E.2.6. *The eigenvalues $\{\lambda_k^\Omega\}_{k=1}^\infty$ of $G_{\Omega,\Lambda}^g$ satisfy*

(i) $\sum_{k=1}^\infty \lambda_k^\Omega = \#(\Omega\cap\Lambda)\|g\|_{L^2}^2$,

(ii) $\sum_{k=1}^{\infty} (\lambda_k^{\Omega})^2 = \sum_{\lambda \in \Omega\cap\Lambda} \sum_{\lambda' \in \Omega\cap\Lambda} |V_g g(\lambda - \lambda')|^2.$

Proof. For the trace, we can compute

$$\begin{aligned}\sum_{k=1}^{\infty} \left\langle G_{\Omega,\Lambda}^{g} h_k^{\Omega}, h_k^{\Omega} \right\rangle &= \sum_{k=1}^{\infty} \sum_{\lambda \in \Omega\cap\Lambda} V_g h_k^{\Omega}(\lambda) \langle \pi(\lambda) g, h_k^{\Omega} \rangle \\ &= \sum_{\lambda \in \Omega\cap\Lambda} \sum_{k=1}^{\infty} |V_g h_k^{\Omega}(\lambda)|^2 = \sum_{\lambda \in \Omega\cap\Lambda} \|g\|_{L^2}^2 = \#(\Omega \cap \Lambda) \|g\|_{L^2}^2\end{aligned}$$

where we used (E.2.3). Meanwhile for the sum of the squares of the eigenvalues, we can write

$$\begin{aligned}\sum_{k=1}^{\infty} \left\langle G_{\Omega,\Lambda}^{g} h_k^{\Omega}, G_{\Omega,\Lambda}^{g} h_k^{\Omega} \right\rangle &= \sum_{k=1}^{\infty} \sum_{\lambda \in \Omega\cap\Lambda} V_g h_k^{\Omega}(\lambda) \langle \pi(\lambda) g, G_{\Omega,\Lambda}^{g} h_k^{\Omega} \rangle \\ &= \sum_{k=1}^{\infty} \sum_{\lambda \in \Omega\cap\Lambda} V_g h_k^{\Omega}(\lambda) \overline{\sum_{\lambda' \in \Omega\cap\Lambda} V_g h_k^{\Omega}(\lambda') \langle \pi(\lambda') g, \pi(\lambda) g \rangle} \\ &= \sum_{\lambda \in \Omega\cap\Lambda} \sum_{\lambda' \in \Omega\cap\Lambda} \sum_{k=1}^{\infty} \langle h_k^{\Omega}, \pi(\lambda) g \rangle \langle \pi(\lambda') g, h_k^{\Omega} \rangle \langle \pi(\lambda) g, \pi(\lambda') g \rangle \\ &= \sum_{\lambda \in \Omega\cap\Lambda} \sum_{\lambda' \in \Omega\cap\Lambda} |V_g g(\lambda - \lambda')|^2\end{aligned}$$

since $(h_k^{\Omega})_{k=1}^{\infty}$ is an orthonormal basis. □

The next lemma is a version of [81, Lemma 4.14] which works for non-tight frames.

Lemma E.2.7. *Let $g \in L^2(\mathbb{R}^d)$ and Λ be such that (g, Λ) induces a frame with frame constants A, B and $\Omega \subset \mathbb{R}^{2d}$ be compact. Then for each $\delta \in (0, B)$,*

$$\begin{aligned}&\left| B\#\left\{k : \lambda_k^{\Omega} > B(1-\delta)\right\} - \#(\Omega \cap \Lambda)\|g\|_{L^2}^2 \right| \\ &\qquad \le \max\left\{\frac{1}{\delta}, \frac{1}{1-\delta}\right\} \left| \#(\Omega \cap \Lambda)\|g\|_{L^2}^2 - \frac{1}{B} \sum_{\lambda \in \Omega\cap\Lambda} \sum_{\lambda' \in \Omega\cap\Lambda} |V_g g(\lambda - \lambda')|^2 \right|.\end{aligned}$$

Proof. By the eigenvalues bound (E.2.2), the operator H defined as

$$H\big(G_{\Omega,\Lambda}^{g}\big) = \sum_{k=1}^{\infty} H(\lambda_k^{\Omega})(h_k^{\Omega} \otimes h_k^{\Omega}), \qquad H(t) = \begin{cases} -t & \text{if } 0 \le t \le B(1-\delta), \\ B - t & \text{if } B(1-\delta) < t \le B \end{cases}$$

is well-defined. By applying H to $G^g_{\Omega,\Lambda}$ and taking the trace we get that

$$\operatorname{tr}\left(H\big(G^g_{\Omega,\Lambda}\big)\right) = \sum_{k=1}^{\infty} H(\lambda_k^\Omega) = B\#\big\{k : \lambda_k^\Omega > B(1-\delta)\big\} - \#(\Omega\cap\Lambda)\|g\|_{L^2}^2.$$

As a function, H can be bounded by

$$|H(t)| \leq \max\left\{\frac{1}{\delta}, \frac{1}{1-\delta}\right\}\left(t - \frac{t^2}{B}\right)$$

and hence

$$\begin{aligned}\left|B\#\big\{k : \lambda_k^\Omega > B(1-\delta)\big\} - \#(\Omega\cap\Lambda)\|g\|_{L^2}^2\right| &= \left|\operatorname{tr}\left(H\big(G^g_{\Omega,\Lambda}\big)\right)\right| \leq \operatorname{tr}(|H|(G^g_{\Omega,\Lambda})) \\ &\leq \max\left\{\frac{1}{\delta}, \frac{1}{1-\delta}\right\}\left(\operatorname{tr}\left(G^g_{\Omega,\Lambda}\right) - \frac{1}{B}\operatorname{tr}\left(\left(G^g_{\Omega,\Lambda}\right)^2\right)\right) \\ &= \max\left\{\frac{1}{\delta}, \frac{1}{1-\delta}\right\}\left|\#(\Omega\cap\Lambda)\|g\|_{L^2}^2 - \frac{1}{B}\sum_{\lambda\in\Omega\cap\Lambda}\sum_{\lambda'\in\Omega\cap\Lambda}|V_g g(\lambda-\lambda')|^2\right|.\end{aligned}$$

□

The next property is essentially a standard result [8, Lemma 4.1] specialized to the setting of lattice convolutions $*_\Lambda$.

Lemma E.2.8. *Let $g \in L^2(\mathbb{R}^d)$ and $\Omega \subset \mathbb{R}^{2d}$ be compact. Then*

$$\sum_{k=1}^{\infty} \lambda_k^\Omega |V_g h_k^\Omega(\lambda)|^2 = \chi_\Omega *_\Lambda |V_g g|^2(\lambda).$$

Proof. We will compute the trace $\operatorname{tr}\left(G^g_{\Omega,\Lambda}\pi(\lambda)(g\otimes g)\pi(\lambda)^*\right)$ using both the singular value decomposition and the definition (E.1.3) and equate the results. For the singular value decomposition, we have

$$\begin{aligned}\operatorname{tr}\left(\sum_{k=1}^{\infty}\lambda_k^\Omega (h_k^\Omega\otimes h_k^\Omega)\pi(\lambda)(g\otimes g)\pi(\lambda)^*\right) &= \sum_{k=1}^{\infty}\lambda_k^\Omega \operatorname{tr}\left((h_k^\Omega\otimes h_k^\Omega)\pi(\lambda)(g\otimes g)\pi(\lambda)^*\right) \\ &= \sum_{k=1}^{\infty}\lambda_k^\Omega \sum_{n=1}^{\infty}\left\langle (h_k^\Omega\otimes h_k^\Omega)\pi(\lambda)(g\otimes g)\pi(\lambda)^* e_n, e_n\right\rangle \\ &= \sum_{k=1}^{\infty}\lambda_k^\Omega \sum_{n=1}^{\infty}\left\langle e_n, \pi(\lambda)g\right\rangle\left\langle \pi(\lambda)g, h_k^\Omega\right\rangle\left\langle h_k^\Omega, e_n\right\rangle \\ &= \sum_{k=1}^{\infty}\lambda_k^\Omega |V_g h_k^\Omega(\lambda)|^2.\end{aligned}$$

Meanwhile the trace can also be computed as

$$
\begin{aligned}
\operatorname{tr}\left(G^g_{\Omega,\Lambda}\pi(\lambda)(g\otimes g)\pi(\lambda)^*\right) &= \sum_{n=1}^{\infty}\left\langle G^g_{\Omega,\Lambda}\pi(\lambda)(g\otimes g)\pi(\lambda)^* e_n, e_n\right\rangle \\
&= \sum_{n=1}^{\infty}\left\langle \pi(\lambda)^* e_n, g\right\rangle\left\langle G^g_{\Omega,\Lambda}\pi(\lambda)g, e_n\right\rangle \\
&= \sum_{n=1}^{\infty}\left\langle e_n, \pi(\lambda)g\right\rangle\left\langle \sum_{\lambda'\in\Lambda}\chi_\Omega(\lambda')\langle\pi(\lambda)g,\pi(\lambda')g\rangle\right\rangle\left\langle \pi(\lambda')g, e_n\right\rangle \\
&= \sum_{\lambda'\in\Lambda}\chi_\Omega(\lambda')|\langle\pi(\lambda')g,\pi(\lambda)g\rangle|^2 \\
&= \chi_\Omega *_\Lambda |V_g g|^2(\lambda)
\end{aligned}
$$

which finishes the proof. □

We this result, we are ready to apply Proposition E.2.2 on spectral quantities.

Lemma E.2.9. *Let $g\in M^*_\Lambda(\mathbb{R}^d)$ and Λ be such that (g,Λ) induces a frame with frame constants $A,B>0$, and $\Omega\subset\mathbb{R}^{2d}$ be compact. Then*

$$
\begin{aligned}
&\left|\frac{1}{B}\sum_{\lambda\in\Lambda}\sum_{\lambda'\in\Lambda}\chi_\Omega(\lambda)\chi_\Omega(\lambda')|V_g g(\lambda-\lambda')|^2 - \#(\Omega\cap\Lambda)\|g\|^2_{L^2}\right| \\
&\qquad\le C\|g\|^2_{L^2}\left(\|g\|^2_{M^*_\Lambda}+1\right)\#\partial^{r_\Lambda}_\Lambda\Omega + \frac{B-A}{B}\#(\Omega\cap\Lambda)\|g\|^2_{L^2}
\end{aligned}
$$

for a constant C depending only on r and d.

Proof. We estimate

$$
\begin{aligned}
&\left|\frac{1}{B}\sum_{\lambda\in\Lambda}\sum_{\lambda'\in\Lambda}\chi_\Omega(\lambda)\chi_\Omega(\lambda')|V_g g(\lambda-\lambda')|^2 - \#(\Omega\cap\Lambda)\|g\|^2_{L^2}\right| \\
&\qquad= \left|\frac{1}{B}\sum_{\lambda\in\Lambda}\chi_\Omega(\lambda)\big(\chi_\Omega *_\Lambda |V_g g|^2(\lambda)\big) - \|g\|^2_{L^2}\sum_{\lambda\in\Lambda}\chi_\Omega(\lambda)\right| \\
&\qquad= \|g\|^2_{L^2}\left|\sum_{\lambda\in\Lambda}\chi_\Omega(\lambda)\left(\frac{1}{B\|g\|^2_{L^2}}\chi_\Omega *_\Lambda |V_g g|^2(\lambda) - \chi_\Omega(\lambda)\right)\right| \\
&\qquad\le \|g\|^2_{L^2}\sum_{\lambda\in\Lambda}\left|\frac{1}{B\|g\|^2_{L^2}}\chi_\Omega *_\Lambda |V_g g|^2(\lambda) - \chi_\Omega(\lambda)\right| \\
&\qquad= \|g\|^2_{L^2}\left\|\chi_\Omega *_\Lambda \frac{|V_g g|^2}{B\|g\|^2_{L^2}} - \chi_\Omega\right\|_{\ell^1(\Lambda)}.
\end{aligned}
$$

Now since $A\|g\|_{L^2}^2 \leq \sum_{\lambda\in\Lambda} |V_g g(\lambda)^2| \leq B\|g\|_{L^2}^2$ by the frame inequality, we can apply Proposition E.2.2 with $\phi = \frac{|V_g g|^2}{B\|g\|_{L^2}^2}$ and $\frac{A}{B} = 1 - \delta$ to the final norm to obtain the desired bound. $\square$

E.2.4 Lattice Gabor space

The image space of the standard short-time Fourier transform, the so called *Gabor space* $V_g(L^2) \subset L^2(\mathbb{R}^{2d})$, is a reproducing kernel Hilbert space (RKHS) with reproducing kernel $K_g(z,w) = \langle \pi(w)g, \pi(z)g\rangle$ [76]. The Toeplitz operators on this space, Gabor-Toeplitz operators, are unitarily equivalent to localization operators via conjugation with the STFT. Similarly, it can be shown that the image of $L^2(\mathbb{R}^d)$ under our STFT, which maps to $\ell^2(\Lambda)$, also is a reproducing kernel Hilbert space [76, Section 5] with reproducing kernel

$$K_g(\lambda, \lambda') = \langle \pi(\lambda')g, \pi(\lambda)g\rangle. \tag{E.2.4}$$

E.3 Accumulated spectrograms

In this section we prove all the theorems on accumulated spectrograms. As we will see, the proofs generally follow those from [8] and [4].

E.3.1 ℓ^1 estimate

We first set out to prove our most important result, Theorem E.1.1, using the approach from [4].

The following lemma allows us to move from $\|\rho_\Omega - \chi_\Omega\|_{\ell^1(\Lambda)}$ to a purely spectral quantity since $\mathrm{tr}(G^g_{\Omega,\Lambda}) = \#(\Omega\cap\Lambda)\|g\|_{L^2}^2$ by Lemma E.2.6.

Lemma E.3.1. *Let $g \in L^2(\mathbb{R}^d)$ and Λ be such that (g,Λ) induces a frame with frame constants $A, B > 0$, and $\Omega \subset \mathbb{R}^{2d}$ be compact. Then*

$$\|\rho_\Omega - \chi_\Omega\|_{\ell^1(\Lambda)} \leq \frac{2}{\|g\|_{L^2}^2}\left(\#(\Omega\cap\Lambda)\|g\|_{L^2}^2 - \sum_{k=1}^{A_\Omega} \lambda_k^\Omega\right) + \frac{B}{\|g\|_{L^2}^2}.$$

Proof. We first rewrite the difference $\|\rho_\Omega - \chi_\Omega\|_{\ell^1(\Lambda)}$ as

$$\|\rho_\Omega - \chi_\Omega\|_{\ell^1(\Lambda)} = \frac{1}{\|g\|_{L^2}^2}\left\|\sum_{k=1}^{A_\Omega} |V_g h_k^\Omega|^2 - \|g\|_{L^2}^2\chi_\Omega\right\|_{\ell^1(\Lambda)}. \tag{E.3.1}$$

Now note that the eigenvalues of $G^g_{\Omega,\Lambda}$ can be written as

$$\lambda_k^\Omega = \left\langle G^g_{\Omega,\Lambda} h_k^\Omega, h_k^\Omega \right\rangle = \sum_{\lambda\in\Omega\cap\Lambda} V_g h_k^\Omega(\lambda)\langle \pi(\lambda)g, h_k^\Omega\rangle = \sum_{\lambda\in\Omega\cap\Lambda} |V_g h_k^\Omega(\lambda)|^2. \quad \text{(E.3.2)}$$

The $\ell^1(\Lambda)$ difference in (E.3.1) can be split into two parts, the interior and exterior of Ω. For the interior, we have that $\chi_\Omega(\lambda) = 1$ and $\sum_{k=1}^{A_\Omega} |V_g h_k^\Omega(\lambda)|^2 \leq \|g\|_{L^2}^2$ by (E.2.3), so

$$\begin{aligned}\sum_{\lambda\in\Omega\cap\Lambda} \left| \sum_{k=1}^{A_\Omega} |V_g h_k^\Omega(\lambda)|^2 - \|g\|_{L^2}^2 \chi_\Omega(\lambda) \right| &= \#(\Omega\cap\Lambda)\|g\|_{L^2}^2 - \sum_{k=1}^{A_\Omega} \sum_{\lambda\in\Omega\cap\Lambda} |V_g h_k^\Omega(\lambda)|^2 \\ &= \#(\Omega\cap\Lambda)\|g\|_{L^2}^2 - \sum_{k=1}^{A_\Omega} \lambda_k^\Omega\end{aligned}$$

where we used (E.3.2) for the second step. Meanwhile for the exterior where $\chi_\Omega(\lambda) = 0$,

$$\begin{aligned}\sum_{\lambda\in\Omega^c\cap\Lambda} \left| \sum_{k=1}^{A_\Omega} |V_g h_k^\Omega(\lambda)|^2 - \|g\|_{L^2}^2 \chi_\Omega(\lambda) \right| &= \sum_{k=1}^{A_\Omega} \sum_{\lambda\in\Lambda} |V_g h_k^\Omega(\lambda)|^2 - \sum_{k=1}^{A_\Omega} \sum_{\lambda\in\Omega\cap\Lambda} |V_g h_k^\Omega(\lambda)|^2 \\ &\leq BA_\Omega - \sum_{k=1}^{A_\Omega} \lambda_k^\Omega \qquad \text{(E.3.3)} \\ &\leq B + \#(\Omega\cap\Lambda)\|g\|_{L^2}^2 - \sum_{k=1}^{A_\Omega} \lambda_k^\Omega\end{aligned}$$

where we used that the upper frame bound is B. Combining these two results, we get the expression in the statement of the lemma. □

Note that the only inequalities in the above proof stem from the frame being non-tight and $\frac{\#(\Omega\cap\Lambda)\|g\|_{L^2}^2}{B}$ not being an integer.

Theorem E.1.1. Let $g \in M^*_\Lambda(\mathbb{R}^d)$ and Λ be such that (g, Λ) induces a frame with frame constants $A, B > 0$, $r > 0$ and $\Omega \subset \mathbb{R}^{2d}$ be compact. Then there exists a constant C depending only on r and d such that

$$\|\rho_\Omega - \chi_\Omega\|_{\ell^1(\Lambda)} \leq C\left(\|g\|_{M^*_\Lambda}^2 + 1\right) \#\partial_\Lambda^{r_\Lambda}\Omega + 2\frac{B-A}{B}\#(\Omega\cap\Lambda) + \frac{B}{\|g\|_{L^2}^2}$$

where $r_\Lambda = r + l_M$ and l_M is the diameter of the fundamental domain of Λ.

Proof. By Lemma E.2.6 followed by Lemma E.2.9, it holds that

$$
\begin{aligned}
0 &\leq \operatorname{tr}\left(G^g_{\Omega,\Lambda}\right) - \frac{1}{B}\operatorname{tr}\left((G^g_{\Omega,\Lambda})^2\right) \\
&\leq C\|g\|^2_{L^2}\left(\|g\|^2_{M^*_\Lambda} + 1\right)\#\partial^{r_\Lambda}_\Lambda \Omega + \frac{B-A}{B}\#(\Omega\cap\Lambda)\|g\|^2_{L^2}.
\end{aligned} \tag{E.3.4}
$$

We can also write the trace difference (E.3.4), with $A_\Omega = \left\lceil \frac{\#(\Omega\cap\Lambda)\|g\|^2_{L^2}}{B} \right\rceil$, as

$$
\begin{aligned}
\operatorname{tr}\left(G^g_{\Omega,\Lambda}\right) - \frac{1}{B}\operatorname{tr}\left((G^g_{\Omega,\Lambda})^2\right) &= \frac{1}{B}\sum_{k=1}^{\infty}\lambda^\Omega_k(B-\lambda^\Omega_k) \\
&= \frac{1}{B}\sum_{k=1}^{A_\Omega}\lambda^\Omega_k(B-\lambda^\Omega_k) + \frac{1}{B}\sum_{k=A_\Omega+1}^{\infty}\lambda^\Omega_k(B-\lambda^\Omega_k) \\
&\geq \frac{\lambda^\Omega_{A_\Omega}}{B}\sum_{k=1}^{A_\Omega}(B-\lambda^\Omega_k) + \left(B-\lambda^\Omega_{A_\Omega}\right)\frac{1}{B}\sum_{k=A_\Omega+1}^{\infty}\lambda^\Omega_k \\
&= \lambda^\Omega_{A_\Omega}A_\Omega - \frac{\lambda^\Omega_{A_\Omega}}{B}\sum_{k=1}^{A_\Omega}\lambda^\Omega_k + \left(B-\lambda^\Omega_{A_\Omega}\right)\frac{1}{B}\left(\#(\Omega\cap\Lambda)\|g\|^2_{L^2} - \sum_{k=1}^{A_\Omega}\lambda^\Omega_k\right) \\
&= \lambda^\Omega_{A_\Omega}A_\Omega + \#(\Omega\cap\Lambda)\|g\|^2_{L^2}\left(1 - \frac{\lambda^\Omega_{A_\Omega}}{B}\right) - \sum_{k=1}^{A_\Omega}\lambda^\Omega_k \\
&= \#(\Omega\cap\Lambda)\|g\|^2_{L^2} - \sum_{k=1}^{A_\Omega}\lambda^\Omega_k + \lambda^\Omega_{A_\Omega}\left(A_\Omega - \frac{\#(\Omega\cap\Lambda)\|g\|^2_{L^2}}{B}\right) \\
&\geq \#(\Omega\cap\Lambda)\|g\|^2_{L^2} - \sum_{k=1}^{A_\Omega}\lambda^\Omega_k.
\end{aligned}
$$

Combining the above with (E.3.4) yields

$$
\#(\Omega\cap\Lambda)\|g\|^2_{L^2} - \sum_{k=1}^{A_\Omega}\lambda^\Omega_k \leq C\|g\|^2_{L^2}\left(\|g\|^2_{M^*_\Lambda}+1\right)\#\partial^{r_\Lambda}_\Lambda\Omega + \frac{B-A}{B}\#(\Omega\cap\Lambda)\|g\|^2_{L^2}.
$$

We can use this estimate together with Lemma E.3.1 to conclude that

$$
\begin{aligned}
\left\|\rho_\Omega - \chi_\Omega\right\|_{\ell^1(\Lambda)} &\leq \frac{2}{\|g\|^2_{L^2}}\left(\#(\Omega\cap\Lambda)\|g\|^2_{L^2} - \sum_{k=1}^{A_\Omega}\lambda^\Omega_k\right) + \frac{B}{\|g\|^2_{L^2}} \\
&\leq 2C\left(\|g\|^2_{M^*_\Lambda}+1\right)\#\partial^{r_\Lambda}_\Lambda\Omega + 2\frac{B-A}{B}\#(\Omega\cap\Lambda) + \frac{B}{\|g\|^2_{L^2}}
\end{aligned}
$$

which, after absorbing the factor 2, is what we wished to show. □

Remark E.3.2. The above proof can be extended to apply to multi-window Gabor multipliers [62, 154] or, more generally, mixed-state Gabor multipliers [81, 154, 192] in the same way as was done in [155] as Lemma E.2.6 and Lemma E.2.9 both are valid in this setting. In the notation of [81], this means that if S is a trace-class operator such that (S, Λ) is a mixed-state Gabor frame and $G^S_{\Omega,\Lambda} = \sum_{k=1}^{\infty} \lambda_k^{\Omega}(h_k^{\Omega} \otimes h_k^{\Omega})$, we can set $\rho^S_{\Omega} = \frac{1}{\operatorname{tr}(S)} \sum_{k=1}^{A_{\Omega}} Q_S(h_k^{\Omega})$ with $A_{\Omega} = \lceil \frac{\#(\Omega\cap\Lambda)\operatorname{tr}(S)}{B} \rceil$ and it will hold that

$$\|\rho^S_{\Omega} - \chi_{\Omega}\|_{\ell^1(\Lambda)} \leq C \left(\sum_{\lambda\in\Lambda} |\lambda| \tilde{S}(\lambda) + 1 \right) \#\partial^{r_\Lambda}_{\Lambda}\Omega + 2\frac{B-A}{B}\#(\Omega \cap \Lambda) + \frac{B}{\operatorname{tr}(S)}$$

for the same constant C. This type of generalization is analogous to that for the accumulated Cohen's class in [155].

Remark E.3.3. Parallel to the development of accumulated spectrograms, there have been corresponding results for the eigenvalues and eigenfunctions of *wavelet* localization operators, see [7, 99, 115, 164, 220]. It is likely that results similar to Theorem E.1.1 hold in the frame setting for accumulated scalograms but we make no attempts to prove this here.

Theorem E.1.1 can be used to approximate Ω as the set where $\rho_{\Omega} > 1/2$. The task of approximately inverting the mapping $\Omega \mapsto G^g_{\Omega,\Lambda}$ has been studied elsewhere in the continuous setting [7, 8, 118, 182], but not in the discrete case.

Corollary E.1.2. Let $g \in M^*_{\Lambda}(\mathbb{R}^d)$ and Λ be such that (g, Λ) induces a frame with frame constants $A, B > 0$, $r > 0$, $\Omega \subset \mathbb{R}^{2d}$ be compact and

$$\tilde{\Omega} = \left\{\lambda \in \Lambda : \rho_{\Omega}(\lambda) > 1/2\right\}.$$

Then there exists a constant C dependent only on r and d such that

$$\#\big((\Omega\Delta\tilde{\Omega}) \cap \Lambda\big) \leq C \left(\|g\|^2_{M^*_{\Lambda}} + 1\right) \#\partial^{r_\Lambda}_{\Lambda}\Omega + 4\frac{B-A}{B}\#(\Omega \cap \Lambda) + \frac{2B}{\|g\|^2_{L^2}}$$

where Δ denotes the symmetric difference of two sets, $r_\Lambda = r + l_M$ and l_M is the diameter of the fundamental domain of Λ.

Proof. Define $E = \{\lambda \in \Lambda : |\rho_{\Omega}(\lambda) - \chi_{\Omega}(\lambda)| \geq 1/2\}$, then we can bound the cardinality of E using Chebyshev's inequality as

$$\begin{aligned}\#E &= \#\left\{\lambda \in \Lambda : |\rho_{\Omega}(\lambda) - \chi_{\Omega}(\lambda)| \geq 1/2\right\} \leq 2\|\rho_{\Omega} - \chi_{\Omega}\|_{\ell^1(\Lambda)} \\ &\leq 2C \left(\|g\|^2_{M^*_{\Lambda}} + 1\right) \#\partial^{r_\Lambda}_{\Lambda}\Omega + 4\frac{B-A}{B}\#(\Omega \cap \Lambda) + \frac{2B}{\|g\|^2_{L^2}}\end{aligned}$$

using Theorem E.1.1. We claim that $(\Omega\Delta\tilde{\Omega}) \cap \Lambda \subset E$. Indeed, if $\lambda \in \Lambda$ is in Ω but not $\tilde{\Omega}$, then $\chi_\Omega(\lambda) = 1$ and $\rho_\Omega(\lambda) \leq 1/2$ so $|\rho_\Omega(\lambda) - \chi_\Omega(\lambda)| \geq 1/2$. Meanwhile if λ is in $\tilde{\Omega}$ but not Ω, then $\chi_\Omega(\lambda) = 0$ and $\rho_\Omega(\lambda) > 1/2$ so $|\rho_\Omega(\lambda) - \chi_\Omega(\lambda)| > 1/2$. This proves the inclusion from which the result follows. □

E.3.2 Sharpness of estimate

On top of improving the L^1 estimate of $\rho_\Omega - \chi_\Omega$, [4] also showed that it is impossible to establish stronger bounds on the L^1 distance by bounding the growth from below in the special case of dilated balls. In this section, we do the same by following a similar strategy as that used to prove [4, Theorem 1.6] specialized to the lattice case. From now on we will assume that the Gabor frame is tight with frame constant 1 as we can only guarantee that $\|\rho_{B(0,R)} - \chi_{B(0,R)}\|_{\ell^1(\Lambda)}$ grows slower than the area of Ω when this is the case. Note that the frame constant being 1 is not a restriction as one can multiply g by a constant if this is not the case.

Theorem E.1.3. Let $g \in M^*_\Lambda(\mathbb{R}^d)$ and Λ be such (g, Λ) induces a tight frame with frame constant 1 and

$$|V_g g(z)| \leq C(1 + |z|)^{-s}, \qquad V_g g(\lambda) \neq 0 \quad \text{for } \lambda \in \Lambda \cap B(0, r + 3l_M)$$

for some $C > 0$ and $s > 2d+1$ where l_M is the diameter of the fundamental domain of Λ. Then there exists constants C_1, C_2 only dependent on g, the lattice Λ and the radius r such that

$$C_1 \# \partial^{r_\Lambda}_\Lambda B(0, R) \leq \|\rho_{B(0,R)} - \chi_{B(0,R)}\|_{\ell^1(\Lambda)} \leq C_2 \# \partial^{r_\Lambda}_\Lambda B(0, R)$$

for all $R > 0$, where $r_\Lambda = r + l_M$ and l_M is the diameter of the fundamental domain of Λ.

We will prove the theorem by showing that the size of the plunge region grows at least as fast as the boundary of the balls. The theorem on which the following lemma is based uses a different measure of the size of the boundary so we have to use a geometric argument to show that it is equivalent to the $\partial^r_\Lambda \Omega$ measure that we use.

Lemma E.3.4. *Let $g \in L^2(\mathbb{R}^d)$ be such that*

$$|V_g g(z)| \leq C(1 + |z|)^{-s}$$

for some $C > 0$ and $s > 2d + 1$. Assume further that there is a positive number r such that

$$V_g g(\lambda) \neq 0 \qquad \text{for } \lambda \in B(0, r + 2l_M) \cap \Lambda.$$

Then there exists a $\delta > 0$ *and a constant* c *such that*

$$c\#\partial^r_\Lambda B(0,R) \leq \#\{k : \delta < \lambda_k^{B(0,R)} < 1-\delta\}$$

for all $R > l_M$.

The proof of the lemma essentially boils down to translating [75, Theorem 5.5.3] to our case where we measure the size of the boundary by $\#\partial^r_\Lambda\Omega$. For the readers convenience, we repeat a version of it here.

Theorem E.3.5 ([75, Theorem 5.5.3])**.** *Let* $g \in L^2(\mathbb{R}^d)$ *be such that*

$$|V_g g(z)| \leq C(1+|z|)^{-s}$$

for some $C > 0$ *and* $s > 2d+1$*. Assume further that there is a positive number* r *such that the fundamental domain of* Λ *is contained in* $B(0,r)$ *and*

$$V_g g(\lambda) \neq 0 \qquad \text{for } \lambda \in B(0,r)\cap\Lambda.$$

If $\mathcal{O}$ *be a family of finite subsets* $\Omega_\Lambda \subset \Lambda$ *satisfying*

$$\#S^k_{\Omega_\Lambda} \leq C\#\partial^r\Omega_\Lambda, \tag{E.3.5}$$

for some $C > 0$*, where*

$$\partial^r\Omega_\Lambda = \{\lambda \in \Omega_\Lambda : B(\lambda,r)\cap\Omega^c_\Lambda \neq \emptyset\} \cup \{\lambda \in \Omega^c_\Lambda : B(\lambda,r)\cap\Omega_\Lambda \neq \emptyset\}$$

and

$$S^k_{\Omega_\Lambda} = \{\lambda \in \Omega_\Lambda : k \leq d(\lambda,\Omega^c_\Lambda) < k+1\},$$

then for any $\delta > 0$ *sufficiently close to* 0*, there is a positive constant* c *such that for all* $\Omega_\Lambda \in \mathcal{O}$*,*

$$c\#\partial^r\Omega_\Lambda \leq \#\{k : \delta < \lambda_k^\Omega < 1-\delta\}.$$

Proof of Lemma E.3.4. Ultimately our goal is to apply the above theorem to the collection of balls $(B(0,R))_{R>0}$ although due to the difference in boundary measures we cannot apply it directly. We relate them by showing that

$$\#\partial^r_\Lambda B(0,R) \leq \#\partial^{r+2l_M}(B(0,R)\cap\Lambda).$$

Indeed, if $\lambda \in \partial^r_\Lambda B(0,R)$ we know that λ is within distance r to a point $\omega \in \partial\Omega$. If $\lambda \in B(0,R)$, let $\omega_c = \frac{\omega}{|\omega|}(R+l_M)$ and if $\lambda \in B(0,R)^c$, let $\omega_c = \frac{\omega}{|\omega|}(R-l_M)$. Then since $R > l_M$ by assumption, ω_c is a point on the other side of $\partial B(0,R)$. Now

$B(\omega_c, l_M)$ is fully contained in the opposite side of $\partial B(0, R)$ again by $R > l_M$ and since $B(\omega_c, l_M)$ has radius l_M, it must contain a lattice point λ_e. By the triangle inequality,

$$d(\lambda, \lambda_e) \leq d(\lambda, \omega) + d(\omega, \omega_c) + d(\omega_c, \lambda_e) \leq r + 2l_M$$

and so we conclude that $\lambda \in \partial^{r+2l_M} B(0, R)$.

From [75, Remark 5.5.4 (ii)] we know that (E.3.5) is satisfied for dilated balls so we can apply the theorem with the radius $r + 2l_M$ since the theorem assumptions have been included in the assumptions of the lemma. □

Remark E.3.6. In [75] it is stated that the proofs of the results of [75, Section 5.5], including [75, Theorem 5.5.3], "will appear elsewhere" but this has not yet appeared.

With the above lemma in place, we are ready to proceed with the proof of the theorem.

Proof of Theorem E.1.3. Using Theorem E.1.1 with $A = B = 1$ we get the upper bound with an additional term $\frac{1}{\|g\|_{L^2}^2}$. Since the theorem is trivially true if $R = 0$, we can assume that $\#\partial_\Lambda^r B(0, R) \geq 1$ and this means we can absorb the additional term in the C_2 constant and thus it only remains to prove the lower bound.

From the proof of Lemma E.3.1 we have that

$$\begin{aligned}
\|\rho_\Omega - \chi_\Omega\|_{\ell^1(\Lambda)} &= \frac{1}{\|g\|_{L^2}^2} \left(A_\Omega - \sum_{k=1}^{A_\Omega} \lambda_k^\Omega + \#(\Omega \cap \Lambda)\|g\|_{L^2}^2 - \sum_{k=1}^{A_\Omega} \lambda_k^\Omega \right) \\
&= \frac{1}{\|g\|_{L^2}^2} \left(A_\Omega - \sum_{k=1}^{A_\Omega} \lambda_k^\Omega + \sum_{k=1}^{\infty} \lambda_k^\Omega - \sum_{k=1}^{A_\Omega} \lambda_k^\Omega \right)
\end{aligned}$$

where we used that the frame is tight with frame constant 1 which yields equality in (E.3.3) for the first equality and Lemma E.2.6 for the last equality. This can be bounded using purely eigenvalues as

$$\begin{aligned}
A_\Omega - \sum_{k=1}^{A_\Omega} \lambda_k^\Omega + \sum_{k=1}^{\infty} \lambda_k^\Omega - \sum_{k=1}^{A_\Omega} \lambda_k^\Omega &= \sum_{k=1}^{A_\Omega} (1 - \lambda_k^\Omega) + \sum_{k=A_\Omega+1}^{\infty} \lambda_k^\Omega \\
&\geq \sum_{k=1}^{\infty} \lambda_k^\Omega (1 - \lambda_k^\Omega). \qquad \text{(E.3.6)}
\end{aligned}$$

Meanwhile, by applying Lemma E.3.4 with radius r_Λ we know that there exists a $\delta > 0$ and $c > 0$ such that

$$c\#\partial_\Lambda^{r_\Lambda} B(0, R) \leq \#\{k : \delta < \lambda_k^{B(0,R)} < 1 - \delta\}$$

for all $R > l_M$. Letting $P \subset \mathbb{N}$ denote the indices k such that $\delta < \lambda_k^{B(0,R)} < 1 - \delta$ for this δ, we have that

$$\begin{aligned}\sum_{k=1}^{\infty} \lambda_k^{B(0,R)}\big(1 - \lambda_k^{B(0,R)}\big) &\geq \sum_{k \in P} \lambda_k^{B(0,R)}\big(1 - \lambda_k^{B(0,R)}\big) \\ &\geq \delta^2 |P| \geq \delta^2 c \# \partial_\Lambda^{r_\Lambda} B(0,R).\end{aligned}$$

Plugging this into (E.3.6), we can relate it to $\|\rho_\Omega - \chi_\Omega\|_{\ell^1(\Lambda)}$ with $\Omega = B(0,R)$ to get

$$\|\rho_{B(0,R)} - \chi_{B(0,R)}\|_{\ell^1(\Lambda)} \geq \frac{1}{\|g\|_{L^2}^2} \delta^2 |P| \geq \delta^2 c \# \partial_\Lambda^{r_\Lambda} B(0,R)$$

which finishes the proof in the $R > l_M$ case.

The $R \leq l_M$ case can be treated manually by first noting that $\#\partial_\Lambda^{r_\Lambda} B(0,R)$ can be bounded from above by $\#\big(\Lambda \cap B(0, R + r_\lambda)\big)$. Meanwhile $\|\rho_{B(0,R)} - \chi_{B(0,R)}\|_{\ell^1(\Lambda)}$ can be bounded from below as follows. From the beginning of this proof, we know that we can crudely bound

$$\|\rho_\Omega - \chi_\Omega\|_{\ell^1(\Lambda)} = \sum_{k=1}^{A_\Omega} (1 - \lambda_k^\Omega) + \sum_{k=A_\Omega+1}^{\infty} \lambda_k^\Omega \geq 1 - \lambda_1^{B(0,R)} \geq 1 - \lambda_1^{B(0,l_M)} > 0$$

where we in the last step used the general eigenvalue bound from (E.2.2). Then by choosing $C_1 = \frac{1 - \lambda_1^{B(0,l_M)}}{\#(\Lambda \cap B(0, R + r_\lambda))}$, we get that

$$\begin{aligned}C_1 \# \partial_\Lambda^{r_\Lambda} B(0,R) &\leq C_1 \# \big(\Lambda \cap B(0, R + r_\lambda)\big) \\ &= 1 - \lambda_1^{B(0,l_M)} \leq \|\rho_{B(0,R)} - \chi_{B(0,R)}\|_{\ell^1(\Lambda)}\end{aligned}$$

finishing the proof. □

Note that if we choose g to be the standard Gaussian, the conditions of Lemma E.3.4 are fulfilled so the above theorem is not vacuous.

E.4 Hyperuniformity of Weyl-Heisenberg ensembles on lattices

As we saw in Section E.2.4, the image space $V_g(L^2) \subset \ell^2(\Lambda)$ is a reproducing kernel Hilbert space. The corresponding RKHS for the continuous STFT was used to induce a determinantal point process on $\mathbb{R}^{2d}$ in [2, 6, 9], the Weyl-Heisenberg ensemble, and by the same procedure we can induce a determinantal point process $\mathcal{X}_g$ on Λ with k-point intensities

$$\rho_k(\lambda_1, \ldots, \lambda_k) = \det\big([K_g(\lambda_i, \lambda_j)]_{1\leq i,j\leq k}\big)$$

where K_g is the reproducing kernel $K_g(\lambda, \lambda') = \langle \pi(\lambda')g, \pi(\lambda)g\rangle$ from (E.2.4) [76, 127].

A point process $\mathcal{X}$ is said to be *hyperuniform* if the variance of the number of points in a large ball grows slower than the volume [201, 202]. Letting $\mathcal{X}(\Omega)$ denote the points in a set Ω, this means that

$$\mathbb{V}\big[\mathcal{X}(B(0,R))\big] = o(R^{2d}). \tag{E.4.1}$$

In particular, when the growth is on the order of R^{2d-1}, the point process is said to be *class I* hyperuniform.

A standard formula (see e.g. [110, Proposition 1.E.1]) specialized to the case of a lattice Λ states that for a determinantal point process with correlation kernel K,

$$\mathbb{V}\left[\sum_{x\in\mathcal{X}} f(x)\right] = \sum_{\lambda\in\Lambda} f(\lambda)^2 K(\lambda,\lambda) - \sum_{\lambda\in\Lambda}\sum_{\lambda'\in\Lambda} f(\lambda)f(\lambda')K(\lambda,\lambda')K(\lambda',\lambda). \tag{E.4.2}$$

We will use this formula to show that our determinantal point process $\mathcal{X}_g$ on Λ, induced by a tight Gabor frame, is class I hyperuniform.

Theorem E.1.4. Let $g \in M^*_\Lambda(\mathbb{R}^d)$ and Λ be such that (g, Λ) induces a tight frame, then the determinantal point process $\mathcal{X}_g$ on Λ with correlation kernel $K_g(\lambda, \lambda') = \langle \pi(\lambda')g, \pi(\lambda)g\rangle$ is hyperuniform.

Proof. We can assume that the frame constant is 1 without loss of generality by multiplying g by a constant if necessary. With $K_g(\lambda, \lambda') = \langle \pi(\lambda')g, \pi(\lambda)g\rangle$, we have that $K_g(\lambda,\lambda) = \|g\|_{L^2}^2$ and $K_g(\lambda, \lambda') = \overline{K_g(\lambda', \lambda)}$. Hence, with $\phi(\lambda - \lambda') = |K(\lambda, \lambda')|^2$, we can write (E.4.2) as

$$\mathbb{V}\big[\mathcal{X}_g(\Omega)\big] = \sum_{\lambda\in\Lambda} \|g\|_{L^2}^2 \chi_\Omega(\lambda) - \sum_{\lambda\in\Lambda}\sum_{\lambda'\in\Lambda} \chi_\Omega(\lambda)\chi_\Omega(\lambda')\phi(\lambda-\lambda').$$

From here Lemma E.2.9 applies and bounds the variance by $C\|g\|_{L^2}^2(\|g\|_{M^*_\Lambda}^2 + 1)\#\partial_\Lambda^{r_\Lambda}\Omega$. We claim that this implies the desired growth behavior (E.4.1) when specialized to the case $\Omega = B(0,R)$. Indeed, we can write

$$\#\partial_\Lambda^{r_\Lambda} B(0,R) = \#\big(\Lambda \cap (B(0, R+r_\Lambda) \setminus B(0, R-r_\Lambda))\big)$$

whenever $R > r_\Lambda$ and this cardinality can be bounded in the same way as was done in Proposition E.2.3. Let A denote the annulus $B(0, R + r_\Lambda) \setminus B(0, R - r_\Lambda)$ and cover A by a collection Q of hypercubes of side length $l_m/\sqrt{2d}$ where l_m is the smallest distance between two points in Λ. Then each hypercube contains at most one lattice point so $\#(\Lambda \cap A) \leq \#Q$. Meanwhile, the hypercubes can be contained in a larger annulus so by comparing volumes we find

$$\begin{aligned}\bigcup_{q\in Q} q \subset A + B(0, l_m) \implies& \#Q \left(\frac{l_m}{\sqrt{2d}}\right)^{2d} \leq \frac{\pi^d}{d!}\left((R + r_\Lambda + l_m)^{2d} - (R - r_\Lambda - l_m)^{2d}\right)\\ \implies& \#(\Lambda \cap A) \leq \#Q = O(R^{2d-1})\end{aligned}$$

which is what we wished to show. □

Acknowledgments

The author appreciates the reviewers' detailed and thoughtful comments, which led to significant improvements of the manuscript.

Paper F

On a Time-Frequency Blurring Operator with Applications in Data Augmentation

Simon Halvdansson

Published in
Journal of Fourier Analysis and Applications,
2025, Volume 31

Paper F

On a Time-Frequency Blurring Operator with Applications in Data Augmentation

Abstract

Inspired by the success of recent data augmentation methods for signals which act on time-frequency representations, we introduce an operator which convolves the short-time Fourier transform of a signal with a specified kernel. Analytical properties including boundedness, compactness and positivity are investigated from the perspective of time-frequency analysis. A convolutional neural network and a vision transformer are trained to classify audio signals using spectrograms with different augmentation setups, including the above mentioned time-frequency blurring operator, with results indicating that the operator can significantly improve test performance, especially in the data-starved regime.

F.1 Introduction and motivation

In time-frequency analysis, functions are analyzed in the *phase space* of time and frequency. The phase space representation of a function can be modified to e.g. change the pitch or mask out an unwanted hiss and a signal can then be synthesized back. In this paper, we investigate the action of *blurring* or *spreading* a function in phase space. More precisely, we convolve the short-time Fourier transform (STFT)

$$V_\varphi \psi(x, \omega) = \langle \psi, \pi(x, \omega)\varphi \rangle = \int_{\mathbb{R}^d} \psi(t)\overline{\varphi(t-x)}e^{-2\pi i \omega \cdot t}\, dt$$

of a signal ψ with a kernel μ and then synthesize a new function from the result

$$B_\mu^\varphi \psi(t) = V_\varphi^*(\mu * V_\varphi \psi)(t) = \int_{\mathbb{R}^{2d}} \mu * V_\varphi \psi(x, \omega)\varphi(t - x)e^{2\pi i \omega \cdot t} \, dx \, d\omega. \quad \text{(F.1.1)}$$

We refer to B_μ^φ as a *time-frequency blurring operator* or *STFT convolver* with window φ and kernel μ.

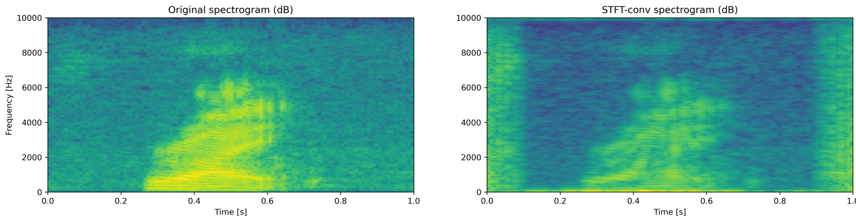

Figure F.1: Action of a time-frequency blurring operator with Gaussian kernel on an audio recording, illustrated using spectrograms.

A similar operator was investigated as part of earlier work by Dörfler and Torrésani [61, 62] where a twisted convolution was instead used to represent general Hilbert-Schmidt operators in the time-frequency domain. As we will see, the change to a regular convolution greatly affects the properties of the operator.

Our main motivation for studying the operator (F.1.1) is its use as a data augmentation method for signals in machine learning. This area has seen much interest in recent years with applications in several areas of signal processing such as automatic speech recognition [171, 208], EEG analysis [183], keyword spotting [179] and sound classification [1], and several augmentation libraries released [130, 133, 157, 165, 168, 169]. With the rapid rise in popularity of vision transformer (ViT) based methods for signal analysis [56, 103, 140], which require more training data compared to convolutional methods [140], there is an increased need for robust augmentation methods. Standard techniques in this field include but are not limited to time-frequency masking, adding noise, preemphasis, simulating room impulse response and changing volume, pitch and speed.

Augmentation methods which act directly on the spectrogram, the squared modulus of the STFT, have recently received a great deal of attention [171, 203, 207, 210] and have been used to achieve state of the art results in a variety of tasks [37, 177, 204, 222]. These methods can be easy to implement in the training pipeline as neural networks often use the spectrogram as an input feature. An obvious drawback is that the resulting augmented spectrograms do not necessarily correspond to spectrograms of actual signals which possess certain smoothness properties. Consequently, the training data may be out of distribution with respect to actual real world data. Should one wish to construct the closest corresponding waveform of the augmented spectrogram, a costly and inexact reconstruction method such as the Griffin–Lim algorithm [106] must be used.

The proposed blurring operator instead acts on the STFT of the signal which allows for computationally efficient waveform synthesis since phase information is not lost as is the case for spectrogram augmentations. Upon computing the STFT, we obtain an in-distribution spectrogram. Many of the standard methods for signal data augmentation can be realized as STFT multipliers with different symbols but have traditionally been implemented in the waveform domain instead, c.f. [130, 133]. In [19], the quantitative differences between these approaches were investigated. It is our belief that phase space based augmentation methods are a valuable tool and from this perspective, the introduction of a STFT convolver is a natural next step after STFT multipliers whose associated methods have seen widespread use and success.

The reason we expect our operator to be suitable for signal data augmentation is that it preserves high level phase space features while only changing the local structure. By this we mean that features such as a clear base note with regularly spaced harmonics or a chirp are preserved while the high resolution patterns in phase space are altered significantly. While these patterns certainly define an important part of the characteristic of the signal, the higher level structure stays intact.

The applied reader may focus their attention on Sections F.3 and F.5 for a deeper discussion on implementation considerations for which much of the preliminaries in Section F.2 are not required.

Notational conventions

For functions on $\mathbb{R}^{2d}$, we will use $L^{p,q}(\mathbb{R}^{2d})$ to denote the mixed-norm Lebesgue spaces with p-norm in the first d variables and q-norm in the remaining d variables. The Fourier transform of a function f will be denoted by $\hat{f}$ and use the standard normalization $\hat{f}(\omega) = \int_{\mathbb{R}^d} f(t)e^{-2\pi i\omega\cdot t}\,dt$. By $\mathcal{F}L^p(\mathbb{R}^d)$ we will mean the Fourier-Lebesgue space consisting of functions whose Fourier transform is in $L^p(\mathbb{R}^d)$. A check over a function, $\check{\psi}$, will mean that the argument is negated i.e., $\check{\psi}(t) = \psi(-t)$. We will take our ambient space to be $L^2(\mathbb{R}^d)$ and so norms $\|\cdot\|$ and inner products $\langle\cdot,\cdot\rangle$ without subscripts are to be understood to be taken in this space. The Schwartz space on $\mathbb{R}^d$ will be denoted by $\mathcal{S}(\mathbb{R}^d)$.

F.2 Time-frequency preliminaries

In this section we go over some of the key objects from time-frequency analysis which we will make use of below. For a more comprehensive overview, the reader is referred to [45, 107, 170, 218, 220].

F.2.1 Short-time Fourier transform

For a signal $\psi \in L^2(\mathbb{R}^d)$ and a window $\varphi \in L^2(\mathbb{R}^d)$, the short-time Fourier transform of ψ with respect to φ is defined as

$$V_\varphi \psi(x, \omega) = \langle \psi, \pi(x, \omega)\varphi \rangle = \int_{\mathbb{R}^d} \psi(t)\overline{\varphi(t-x)}e^{-2\pi i \omega \cdot t}\, dt \tag{F.2.1}$$

where $\pi(x, \omega)$ is a time-frequency shift by time x and frequency ω, defined as $\pi(x, \omega) = M_\omega T_x$ where $M_\omega f(t) = e^{2\pi i \omega \cdot t}$ is the modulation operator and $T_x f(t) = f(t-x)$ is the translation operator. We often write $z = (x, \omega) \in \mathbb{R}^{2d}$ for the coordinates of x, ω in time-frequency space $\mathbb{R}^{2d}$. As is indicated in the middle step of the definition (F.2.1), this transform projects the signal onto time-frequency shifted versions of the window and the interpretation is that $V_\varphi \psi(x, \omega)$ measures the importance of the time t and frequency ω to ψ.

One of the key properties of the STFT is *Moyal's identity* which states that

$$\left\langle V_{\varphi_1}\psi_1, V_{\varphi_2}\psi_2 \right\rangle_{L^2(\mathbb{R}^{2d})} = \langle \psi_1, \psi_2 \rangle \overline{\langle \varphi_1, \varphi_2 \rangle}. \tag{F.2.2}$$

As a consequence, for a normalized window φ, $V_\varphi : L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^{2d})$ is an isometry and the STFT mapping can be inverted in the sense that

$$\psi = \frac{1}{\langle \varphi_1, \varphi_2 \rangle} V_{\varphi_2}^*(V_{\varphi_1}\psi) = \frac{1}{\langle \varphi_1, \varphi_2 \rangle} \int_{\mathbb{R}^{2d}} V_{\varphi_1}\psi(z)\pi(z)\varphi_2\, dz \tag{F.2.3}$$

weakly.

The STFT is in general complex valued and in many cases the phase information is superfluous. For this reason, the squared modulus $|V_\varphi \psi|^2$, which is called the spectrogram, is often used in applications. The process of recovering the STFT from the spectrogram, meaning to invert the mapping $V_\varphi \psi \mapsto |V_\varphi \psi|^2$, is called *phase retrieval* and is an area of rich research.

F.2.2 Gabor spaces

The image of the short-time Fourier transform, $V_\varphi(L^2(\mathbb{R}^d)) \subset L^2(\mathbb{R}^{2d})$, is called the *Gabor space* associated to the window φ. We saw earlier in (F.2.3) how the adjoint of the STFT mapping V_φ can be used to synthesize a signal from a function on $\mathbb{R}^{2d}$. If that original function is an STFT, the synthesized signal will be precisely the original signal used to compute the STFT. However, if we take a function in $L^2(\mathbb{R}^{2d})$, synthesize a signal from it with V_φ^* and then compute the STFT, this is precisely orthogonally projecting the original function onto the Gabor space. We can write both of these facts as

$$V_\varphi^* V_\varphi = I_{L^2}, \qquad V_\varphi V_\varphi^* = P_{V_\varphi(L^2)} \tag{F.2.4}$$

where $P_{V_\varphi(L^2)}$ is the orthogonal projection onto the Gabor space $V_\varphi(L^2(\mathbb{R}^d))$.

F.2.3 Localization operators

By introducing a multiplication operator in between V_φ and V_φ^* in $V_\varphi^* V_\varphi = I_{L^2}$, we can effectively choose which regions of the time-frequency plane we want to prioritize when reconstructing a signal. The resulting operator, called a localization operator or STFT multiplier, was first investigated by I. Daubechies in [51] and has since seen wide applications in signal analysis, pseudo-differential operators and partial differential equations. We write it formally as

$$A_m^\varphi \psi = V_\varphi^*(m \cdot V_\varphi \psi) = \int_{\mathbb{R}^{2d}} m(z) V_\varphi \psi(z) \pi(z) \varphi \, dz \tag{F.2.5}$$

and refer to the function $m : \mathbb{R}^{2d} \to \mathbb{C}$ as the *mask* or *symbol* of the operator. The mask is often taken to be an indicator function of some subset of the time-frequency plane. In this way, we can restrict a signal to this subset when reconstructing it.

At times, we separate the analysis window in V_φ and the synthesis window in V_φ^* and use the relation $V_{\varphi_2}^* V_{\varphi_1} = \langle \varphi_1, \varphi_2 \rangle I_{L^2}$ from (F.2.3), leading to the more general operator

$$A_m^{\varphi_1,\varphi_2} \psi = V_{\varphi_2}^*(m \cdot V_{\varphi_1} \psi) = \int_{\mathbb{R}^{2d}} m(z) V_{\varphi_1} \psi(z) \pi(z) \varphi_2 \, dz. \tag{F.2.6}$$

F.2.4 Modulation spaces

Often when dealing with mapping properties of localization operators and other problems in time-frequency analysis, modulation spaces turn out to be a suitable setting. They are a class of function spaces first introduced by Feichtinger in [78] which are characterized by the integrability properties of short-time Fourier transforms. For a non-zero Schwartz window $\varphi \in \mathcal{S}(\mathbb{R}^d)$, we can define the modulation space $M^{p,q}(\mathbb{R}^d)$ for $1 \leq p, q \leq \infty$ as the set of all tempered distributions ψ such that

$$\|\psi\|_{M^{p,q}(\mathbb{R}^d)} = \left(\int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} |V_\varphi \psi(x, \omega)|^p \, dx \right)^{q/p} d\omega \right)^{1/q} < \infty.$$

When p or q are equal to ∞, we make suitable adjustments and for $p = q$ we use the shorthand $M^{p,p}(\mathbb{R}^d) = M^p(\mathbb{R}^d)$. The norms induced by different Schwartz windows are all equivalent and so we do not indicate the window when denoting a modulation space. From (F.2.2), it is clear that $M^2(\mathbb{R}^d) = L^2(\mathbb{R}^d)$. The case of $p = 1$ is of special importance and the space $M^1(\mathbb{R}^d)$ is often called *Feichtinger's algebra*.

We can interpolate between different modulation spaces in the same way as we do for Lebesgue spaces as described in the following simplified lemma which follows from [112, Theorem 1.1].

Lemma F.2.1. *Let* $1 \leq p_1, p_2 \leq \infty$ *and* p_θ *be defined by the relation*

$$\frac{1}{p_\theta} = \frac{1-\theta}{p_1} + \frac{\theta}{p_2}.$$

Then we can interpolate between $M^{p_1}(\mathbb{R}^d)$ *and* $M^{p_2}(\mathbb{R}^d)$ *as*

$$\Big[M^{p_1}(\mathbb{R}^d),\, M^{p_2}(\mathbb{R}^d)\Big]_\theta = M^{p_\theta}(\mathbb{R}^d).$$

Later we will also have use for the following mapping property which is a special case of [107, Proposition 11.3.7].

Lemma F.2.2. *If* $\varphi \in M^1(\mathbb{R}^d)$*, then* V_φ^* *maps* $L^{p,q}(\mathbb{R}^{2d})$ *to* $M^{p,q}(\mathbb{R}^d)$ *for* $1 \leq p, q \leq \infty$ *with*

$$\|V_\varphi^* F\|_{M^{p,q}(\mathbb{R}^d)} \lesssim \|\varphi\|_{M^1(\mathbb{R}^d)} \|F\|_{L^{p,q}(\mathbb{R}^{2d})}$$

and $\psi \mapsto \|V_\varphi \psi\|_{L^{p,q}(\mathbb{R}^{2d})}$ *is an equivalent norm on* $M^{p,q}(\mathbb{R}^d)$.

F.2.5 Quadratic time-frequency distributions

The spectrogram $|V_\varphi \psi|^2$ is just one example of a time-frequency distribution which has a quadratic dependence on ψ. Another widely used object is the Wigner distribution

$$W(\psi)(x,\omega) = \int_{\mathbb{R}^d} \psi\left(x + \frac{t}{2}\right) \overline{\psi\left(x - \frac{t}{2}\right)} e^{-2\pi i \omega \cdot t}\, dt$$

defined for $\psi \in L^2(\mathbb{R}^d)$. More generally, one can characterize a class of well-behaved quadratic time-frequency representations as those of the form

$$Q_\Phi(\psi) = W(\psi) * \Phi$$

for some tempered distribution Φ. These time-frequency distributions are said to belong to *Cohen's class of quadratic time-frequency distributions*.

F.3 Related objects and adaptations

For applications, the formulation (F.1.1) may not always be the optimal incarnation of the idea of blurring a signal in time and frequency. In this section, we motivate this statement by detailing alternatives which may be better suited to certain applications.

F.3.1 Spectrogram blurring

As discussed in the introduction, many contemporary signal augmentation methods act directly on the spectrogram. In particular, the widely popular SpecAugment augmentation method [171] multiplies the spectrogram by a time-frequency mask. Here there is a clear analog in localization operators and we should expect that differences between SpecAugment and applying a localization operator with the same mask should only be visible near the edges of the mask. For time-frequency blurring operators however, performing the blurring action on the STFT instead of the spectrogram leads to a noticeable difference as the phase data of the STFT greatly affects the convolution. For example, the STFT of white noise is largely unstructured while sinusoids and impulses give rise to predictable phase gradients, see e.g [174, Figure 1] and [102, Figure 2], and hence a blurring operation which takes phase into account should affect white noise differently than such a signal.

Performing the convolution on the spectrogram is easier to implement and the intuition behind "spreading" the time-frequency contents of the signal stays intact. To the best of our knowledge, this approach has previously only been mentioned in the GitHub repository [203] but never studied in any detail.

Analyzing spectrogram blurring as a mapping on $L^2(\mathbb{R}^d)$ is in general intractable as it involves a phase retrieval step to obtain a waveform. By instead looking at spectrogram blurring as a mapping from $L^2(\mathbb{R}^d)$ to $L^1(\mathbb{R}^{2d})$, we see that blurring a spectrogram exactly corresponds to switching quadratic time-frequency distribution from the spectrogram to a Cohen's class distribution [107, 125, 170]. Indeed,

$$\mu * |V_\varphi \psi|^2 = \mu * \big(W(\psi) * W(\check{\varphi})\big) = W(\psi) * \big(\mu * W(\check{\varphi})\big)$$

and so a blurred spectrogram is a well-behaved quadratic time-frequency distribution.

In Section F.5.1, we present a brief numerical comparison between spectrogram blurring and the time-frequency blurring operator where we refer to the spectrogram blurring procedure as SpecBlur. There we see that blurred spectrograms differ significantly from the spectrograms of real signals.

F.3.2 Position-dependent kernel

An obvious generalization of B_μ^φ is to allow the kernel μ to depend on $z \in \mathbb{R}^{2d}$. We can write this as

$$B_{\boldsymbol{\mu}}^\varphi \psi = \int_{\mathbb{R}^{2d}} \mu_z * V_\varphi \psi(z) \pi(z) \varphi \, dz$$

where we use a bold $\boldsymbol{\mu}$ to indicate that it is a function on the double phase space $\mathbb{R}^{4d}$ with $\boldsymbol{\mu}(z, w) = \mu_z(w)$. This is a very general form of operator and e.g. localization operators can be realized as a special case by

$$\mu_z = m(z)\delta_0 \implies B_{\boldsymbol{\mu}}^\varphi = A_m^\varphi \tag{F.3.1}$$

since $B_{\delta_0}^\varphi = V_\varphi^* V_\varphi = I_{L^2}$. In view of this, these generalized time-frequency blurring operators may provide a nice model for approximating operators or interpolating between them, see e.g. [166] for work in a similar direction for localization operators. Indeed, such an object has already been used in the discrete setting in [158, 187] for noise reduction under the name "Deep Filtering". In [158, 187] and related works, the STFT and engineered additional features are used as the input to a neural network which outputs the position dependent kernels μ_z with the goal of minimizing some distance function

$$d\big(B_{\boldsymbol{\mu}}^\varphi \psi_{\text{noisy}}, \psi_{\text{clean}}\big).$$

Earlier work on noise reduction, e.g., [217], has focused on learning an optimal time-frequency mask which, in view of (F.3.1), is just a special case of a time-frequency blurring operator. On a high level, the deep filtering papers have exchanged a function on $\mathbb{R}^{2d}$ (m) for one on $\mathbb{R}^{4d}$ $(\boldsymbol{\mu})$ or, equivalently, imposed structure on the final layer of their neural network.

We will not investigate the properties of these more heavily parameterized operators further but note that the present paper is likely to be a good first step in this direction.

F.3.3 Window generalizations

We briefly discuss two generalizations of time-frequency blurring operators which both have clear counterparts for localization operators. The first is separating the analysis and synthesis windows as was done for localization operators in (F.2.6). This leads to the following operator

$$B_\mu^{\varphi_1, \varphi_2} \psi = V_{\varphi_2}^* (\mu * V_{\varphi_1} \psi) = \int_{\mathbb{R}^{2d}} \mu * V_{\varphi_1} \psi(z) \, \pi(z) \varphi_2 \, dz.$$

Many of the results which will be derived in Section F.4 can be adapted to this generalized operator. The required changes to the proofs are minimal and routine and, in the interest of brevity, we leave them to the interested reader.

The concept of multi-window localization operators, defined as linear combinations of localization operators, $\sum_{n=1}^{N} A_m^{\varphi_1^n,\varphi_2^n}$, was further generalized to permit *operator windows* S of trace-class in [154] as a part of the framework of quantum harmonic analysis [211]. Without going into too much detail, we note that these are defined via the Bochner integral

$$A_m^S = \int_{\mathbb{R}^{2d}} m(z)\pi(z)S\pi(z)^* \, dz$$

which reduces down to multi-window localization operator when the operator S is of finite rank. This formulation of localization operators has proven useful [115, 154–156]. In our case, the generalization to operator windows S of trace-class can be written as

$$B_\mu^S \psi = \int_{\mathbb{R}^{2d}} \mu * \big(\pi(z)S\pi(\cdot)^*\psi\big)(z) \, dz.$$

Note specifically that if S has the singular value decomposition $S = \sum_n s_n(\varphi_n^1 \otimes \varphi_n^2)$, it can be written as

$$B_\mu^S = \sum_n s_n B_\mu^{\varphi_n^1,\varphi_n^2},$$

i.e., a linear combination of time-frequency blurring operators.

F.4 Analytical properties

The numerical realizations of the blurring operator discussed above and in Section F.5 are inherently discrete and, as such, harder to investigate beyond looking at examples. In this section, we look at boundedness, positivity and compactness of the operator to further inform our intuitions of how the operator behaves. Beyond boundedness, these results are, at face value, of limited value for practitioners since every discrete function can be considered as a sampling of a Schwartz function. Still, intuition about the properties of the operator is likely to be informative for choosing which kernel to use and understanding in which contexts the operator is likely to be useful. Specifics around the correspondence of the continuous and discrete settings of time-frequency analysis have been developed in [82, 131]

F.4.1 Mapping properties

Up until now we have left the spaces which φ and μ belong to ambiguous. We will see that the most natural assumptions are that the kernel μ is a bounded measure and that the window φ is square integrable or in Feichtinger's algebra, similar to the convention for localization operators. Changing these assumptions affects which spaces the operator maps to and from. While making no claims of completeness, we collect some of the more easily available results on these mapping properties in this subsection.

Our first result provides the most natural boundedness condition on $L^2(\mathbb{R}^d)$.

Proposition F.4.1. Let $\mu \in M(\mathbb{R}^{2d})$ and $\varphi \in L^2(\mathbb{R}^d)$, then B_μ^φ is a bounded operator on $L^2(\mathbb{R}^d)$ with

$$\|B_\mu^\varphi\|_{B(L^2(\mathbb{R}^d))} \le \|\mu\|_{M(\mathbb{R}^{2d})}\|\varphi\|^2.$$

Proof. For $\psi \in L^2(\mathbb{R}^d)$ we compute

$$\begin{aligned}
\|B_\mu^\varphi\psi\| &= \frac{1}{\|\varphi\|}\left\|V_\varphi(V_\varphi^*(\mu * V_\varphi\psi))\right\|_{L^2(\mathbb{R}^{2d})}\\
&\le \|\varphi\|\|\mu * V_\varphi\psi\|_{L^2(\mathbb{R}^{2d})}\\
&\le \|\mu\|_{M(\mathbb{R}^{2d})}\|\varphi\|\|V_\varphi\psi\|_{L^2(\mathbb{R}^{2d})}\\
&= \|\mu\|_{M(\mathbb{R}^{2d})}\|\varphi\|^2\|\psi\|
\end{aligned}$$

where we used that $V_\varphi V_\varphi^* = P_{V_\varphi(L^2)}$ is an orthogonal projection, Young's inequality and Moyal's identity (F.2.2) twice. □

Next we look at mapping between modulation spaces for general exponents.

Proposition F.4.2. Let $\mu \in L^{p_1,p_2}(\mathbb{R}^{2d})$, $\psi \in M^{q_1,q_2}(\mathbb{R}^d)$ and $\varphi \in M^1(\mathbb{R}^d)$ with $\frac{1}{p_i} + \frac{1}{q_i} = 1 + \frac{1}{r_i}$ and $1 \le p_i, q_i, r_i \le \infty$ for $i = 1, 2$. Then $B_\mu^\varphi\psi \in M^{r_1,r_2}(\mathbb{R}^d)$ with

$$\|B_\mu^\varphi\psi\|_{M^{r_1,r_2}(\mathbb{R}^d)} \lesssim \|\mu\|_{L^{p_1,p_2}(\mathbb{R}^{2d})}\|\psi\|_{M^{q_1,q_2}(\mathbb{R}^d)}$$

where the implicit constant depends on φ and the p, q, r constants.

Proof. For $\psi \in M^{q_1,q_2}(\mathbb{R}^d)$, we have that

$$\begin{aligned}
\|B_\mu^\varphi\psi\|_{M^{r_1,r_2}(\mathbb{R}^d)} &= \|V_\varphi^*(\mu * V_\varphi\psi)\|_{M^{r_1,r_2}(\mathbb{R}^d)}\\
&\lesssim \|\mu * V_\varphi\psi\|_{L^{r_1,r_2}(\mathbb{R}^{2d})} && \big(\text{Lemma F.2.2}\big)\\
&\le \|\mu\|_{L^{p_1,p_2}(\mathbb{R}^{2d})}\|V_\varphi\psi\|_{L^{q_1,q_2}(\mathbb{R}^{2d})} && \big(\text{Young's inequality}\big)\\
&\lesssim \|\mu\|_{L^{p_1,p_2}(\mathbb{R}^{2d})}\|\psi\|_{M^{q_1,q_2}(\mathbb{R}^d)}
\end{aligned}$$

where we used the equivalence of norms for modulation spaces in the last step. □

A proof of the general version of Young's inequality which we used above can be found in e.g. [107, Proposition 11.1.3].

If we want the output $B_\mu^\varphi\psi$ to be in an L^p space instead of a modulation space, we can get this out of the above result with some minor work.

Proposition F.4.3. Fix $1 \leq p \leq 2$ and let $\varphi \in M^1(\mathbb{R}^d)$, $\mu \in L^1(\mathbb{R}^{2d})$ and $\psi \in M^p(\mathbb{R}^d)$, then

$$\|B_\mu^\varphi\psi\|_{L^p(\mathbb{R}^d)} \lesssim \|\mu\|_{L^1(\mathbb{R}^{2d})}\|\psi\|_{M^p(\mathbb{R}^d)}$$

where the implicit constant is dependent on φ and p.

Proof. We wish to interpolate between the $p = 1$ and $p = 2$ cases of this. Since $M^2(\mathbb{R}^d) = L^2(\mathbb{R}^d)$ as noted earlier, the $p = 2$ version is just Proposition F.4.1. Meanwhile for $p = 1$ it suffices to note that $L^1(\mathbb{R}^d)$ can be embedded in $M^1(\mathbb{R}^d)$ by [107, Proposition 12.1.4]. Now an application of Lemma F.2.1 yields the desired result. □

If we want $B_\mu^\varphi\psi$ to belong to $L^\infty(\mathbb{R}^d)$, we can place additional assumptions on the window function.

Proposition F.4.4. Let $\varphi \in L^2(\mathbb{R}^d) \cap L^\infty(\mathbb{R}^d)$, $\mu \in L^1(\mathbb{R}^{2d})$ and $\psi \in M^1(\mathbb{R}^d)$, then

$$\|B_\mu^\varphi\psi\|_{L^\infty(\mathbb{R}^d)} \lesssim \|\mu\|_{L^1(\mathbb{R}^{2d})}\|\varphi\|_{L^\infty(\mathbb{R}^d)}\|\psi\|_{M^1(\mathbb{R}^d)}.$$

Proof. The simplest way to show this is to consider the dual formulation of the $L^\infty(\mathbb{R}^d)$ norm. Indeed,

$$\begin{aligned}
\|B_\mu^\varphi\psi\|_{L^\infty(\mathbb{R}^d)} &= \sup_{\|\phi\|_{L^1}=1} \left|\left\langle \int_{\mathbb{R}^{2d}} \mu * V_\varphi\psi(z)\pi(z)\varphi\, dz, \phi\right\rangle\right| \\
&= \sup_{\|\phi\|_{L^1}=1} \left|\int_{\mathbb{R}^{2d}} \mu * V_\varphi\psi(z)\langle\pi(z)\varphi, \phi\rangle\, dz\right| \\
&\leq \sup_{\|\phi\|_{L^1}=1} \left\|\mu * V_\varphi\psi(\cdot)\langle\pi(\cdot)\varphi, \phi\rangle\right\|_{L^1(\mathbb{R}^{2d})} \\
&\leq \sup_{\|\phi\|_{L^1}=1} \left\|\mu * V_\varphi\psi\right\|_{L^1(\mathbb{R}^{2d})} \|\langle\pi(\cdot)\varphi, \phi\rangle\|_{L^\infty(\mathbb{R}^{2d})} \\
&\lesssim \|\mu\|_{L^1(\mathbb{R}^{2d})}\|\psi\|_{M^1(\mathbb{R}^d)}\|\varphi\|_{L^\infty(\mathbb{R}^d)}.
\end{aligned}$$

□

In particular, by combining Proposition F.4.3 and Proposition F.4.4 we see that when $\mu \in L^1(\mathbb{R}^{2d}), \varphi \in M^1(\mathbb{R}^d) \subset L^2(\mathbb{R}^d) \cap L^\infty(\mathbb{R}^d)$ (this follows from [107, Proposition 12.1.4]) and $\psi \in M^1(\mathbb{R}^d)$, $B_\mu^\varphi \psi$ is in $L^1(\mathbb{R}^d) \cap L^\infty(\mathbb{R}^d)$.

Lastly we show that if the window and kernel are in the Schwartz space, B_μ^φ maps the Schwartz space to itself. To do so, we will need two preliminary results.

Lemma F.4.5 ([107, Proposition 11.2.4])**.** *Fix $\varphi \in \mathcal{S}(\mathbb{R}^d)$ and assume $F : \mathbb{R}^{2d} \to \mathbb{C}$ has rapid decay on $\mathbb{R}^{2d}$, then the integral*

$$\psi(t) = \int_{\mathbb{R}^{2d}} F(z)\pi(z)\varphi(t)\, dz$$

defines a function $\psi \in \mathcal{S}(\mathbb{R}^d)$.

In particular, the above lemma implies that if $F \in \mathcal{S}(\mathbb{R}^{2d})$, then $V_\varphi^* F \in \mathcal{S}(\mathbb{R}^d)$.

Lemma F.4.6 ([107, Theorem 11.2.5])**.** *Fix $\varphi \in \mathcal{S}(\mathbb{R}^d)$, then ψ is in the Schwartz space if and only if $V_\varphi \psi \in \mathcal{S}(\mathbb{R}^{2d})$.*

We can now proceed with the main proposition.

Proposition F.4.7. Let $\mu \in \mathcal{S}(\mathbb{R}^{2d})$ and $\varphi \in \mathcal{S}(\mathbb{R}^d)$, then

$$B_\mu^\varphi : \mathcal{S}(\mathbb{R}^d) \to \mathcal{S}(\mathbb{R}^d).$$

Proof. Let $\psi \in \mathcal{S}(\mathbb{R}^d)$, then by Lemma F.4.6 $V_\varphi \psi \in \mathcal{S}(\mathbb{R}^d)$. Now $B_\mu^\varphi \psi = V_\varphi^*(\mu * V_\varphi \psi)$ and so by Lemma F.4.5, the result follows upon noting that the convolution $\mu * V_\varphi \psi$ between two Schwartz functions is another Schwartz function. $\square$

F.4.2 Weak action and positivity

In the case of localization operators, computing the weak action has proven valuable [42, 153]. Inspired by this, we look at the weak action of B_μ^φ in this section. By moving the adjoint of V_φ to the other side of the inner product, we immediately get

$$\left\langle B_\mu^\varphi \psi, \phi \right\rangle = \left\langle \mu * V_\varphi \psi, V_\varphi \phi \right\rangle_{L^2(\mathbb{R}^{2d})}.$$

Of course, we cannot move over the convolution but in an effort to obtain something more workable, we take the Fourier transform of both sides and rearrange, yielding

$$\left\langle B_\mu^\varphi \psi, \phi \right\rangle = \left\langle \hat{\mu} \cdot \widehat{V_\varphi \psi}, \widehat{V_\varphi \phi} \right\rangle_{L^2(\mathbb{R}^{2d})} = \left\langle \hat{\mu}, \overline{\widehat{V_\varphi \psi}} \cdot \widehat{V_\varphi \phi} \right\rangle_{L^2(\mathbb{R}^{2d})}. \tag{F.4.1}$$

From this formulation we can obtain a general bound on $\langle B_\mu^\varphi \psi, \phi \rangle$. Before the proof, we recall the Hausdorff-Young inequality which states that if $\frac{1}{p} + \frac{1}{q} = 1$ with $p \in [1, 2]$ and $F \in L^p$, then

$$\|\hat{F}\|_{L^q} \leq \|F\|_{L^p}.$$

Proposition F.4.8. Let $1 \leq p \leq \infty$ and $1 \leq q, r \leq 2 \leq q', r' \leq \infty$ be such that $\frac{1}{p} + \frac{1}{q'} + \frac{1}{r'} = 1$, $\frac{1}{q} + \frac{1}{q'} = 1$ and $\frac{1}{r} + \frac{1}{r'} = 1$. If $\mu \in \mathscr{F}L^p(\mathbb{R}^{2d})$, $\psi \in M^q(\mathbb{R}^d)$, $\phi \in M^r(\mathbb{R}^d)$ and $\varphi \in M^1(\mathbb{R}^d)$, then

$$\left|\left\langle B_\mu^\varphi \psi, \phi \right\rangle\right| \lesssim \|\hat{\mu}\|_{L^p(\mathbb{R}^{2d})} \|\psi\|_{M^q(\mathbb{R}^d)} \|\phi\|_{M^r(\mathbb{R}^d)}.$$

Proof. From (F.4.1), we have that

$$\left|\left\langle B_\mu^\varphi \psi, \phi \right\rangle\right| \leq \left\|\hat{\mu} \cdot \widehat{V_\varphi \psi} \cdot \widehat{V_\varphi \phi}\right\|_{L^1(\mathbb{R}^{2d})}$$

on which we can apply generalized Hölder. This yields

$$\left|\left\langle B_\mu^\varphi \psi, \phi \right\rangle\right| \leq \left\|\hat{\mu}\right\|_{L^p(\mathbb{R}^{2d})} \left\|\widehat{V_\varphi \psi}\right\|_{L^{q'}(\mathbb{R}^{2d})} \left\|\widehat{V_\varphi \phi}\right\|_{L^{r'}(\mathbb{R}^{2d})}.$$

Upon applying the Hausdorff-Young inequality followed by Lemma F.2.2 to the two last quantities, the desired result follows. □

From (F.4.1), we can also get a condition for the positivity of B_μ^φ.

Proposition F.4.9. Let $\varphi \in L^2(\mathbb{R}^d)$ and $\mu \in M(\mathbb{R}^{2d})$ with $\hat{\mu} \geq 0$. Then B_μ^φ is a positive operator on $L^2(\mathbb{R}^d)$.

Proof. If $\hat{\mu} \geq 0$, non-negativity of $\langle B_\mu^\varphi \psi, \psi \rangle$ clearly follows from (F.4.1) as

$$\left\langle B_\mu^\varphi \psi, \psi \right\rangle = \left\langle \hat{\mu}, \left|\widehat{V_\varphi \psi}\right|^2 \right\rangle_{L^2(\mathbb{R}^{2d})}$$

and finiteness follows from that B_μ^φ is bounded on $L^2(\mathbb{R}^d)$ by Proposition F.4.2. □

The next proposition essentially states that there exists nontrivial kernels such that the convolution operator is the zero operator on the Gabor space $V_\varphi(L^2(\mathbb{R}^d))$, which is not the case for $L^2(\mathbb{R}^{2d})$.

Proposition F.4.10. There exists non-zero $\varphi \in L^2(\mathbb{R}^d)$ and $\mu \in L^1(\mathbb{R}^{2d})$ such that B_μ^φ is the zero operator.

Proof. We will tacitly choose μ, φ such that the integrand in the integral defining the inner product (F.4.1) is always zero so that $\langle B_\mu^\varphi \psi, \phi \rangle = 0$ for all $\psi, \phi \in L^2(\mathbb{R}^d)$. First, we compute the Fourier transform of a general STFT $V_\varphi \psi$. The identity

$$V_\varphi \psi(x, \omega) = M_{-\omega}(\psi * M_\omega \varphi^*)(x),$$

where $\varphi^*(x) = \overline{\varphi(-x)}$, from [107, Lemma 3.1.1] simplifies this considerably. Let $\mathcal{F}_1$ denote the Fourier transform in the first d variables and $\mathcal{F}_2$ the Fourier transform in the last d variables, we then have that

$$\begin{aligned}\mathcal{F}_1(V_\varphi\psi)(x',\omega) &= \mathcal{F}_1\big(M_{-\omega}(\psi * M_\omega\varphi^*)\big)(x')\\ &= T_{-\omega}\big(\hat{\psi}\cdot\widehat{M_\omega\varphi^*}\big)(x')\\ &= \hat{\psi}(x'+\omega)\widehat{M_\omega\varphi^*}(x'+\omega)\\ &= \hat{\psi}(x'+\omega)\widehat{\varphi^*}(x').\end{aligned}$$

The full Fourier transform can now be computed as

$$\begin{aligned}\widehat{V_\varphi\psi}(x',\omega') &= \mathcal{F}_2\big(\hat{\psi}(x'+\cdot)\widehat{\varphi^*}(x')\big)(\omega')\\ &= \check{\psi}(\omega')e^{2\pi i x'\cdot\omega'}\widehat{\varphi^*}(x').\end{aligned}$$

From this we see that the signal ψ has no effect on the support in the first d variables of $\widehat{V_\varphi\psi}$. If we now choose φ such that $\widehat{\varphi^*}$ is zero on some set $E \subset \mathbb{R}^d$ and μ such that the first d variables of $\hat{\mu}$ is supported on the same set E, it will hold that

$$\begin{aligned}\big\langle B_\mu^\varphi\psi,\phi\big\rangle &= \Big\langle \hat{\mu}, \overline{\widehat{V_\varphi\psi}}\cdot\widehat{V_\varphi\phi}\Big\rangle_{L^2(\mathbb{R}^{2d})}\\ &= \int_{\mathbb{R}^{2d}} \hat{\mu}(x',\omega')\widehat{V_\varphi\psi}(x',\omega')\overline{\widehat{V_\varphi\phi}(x',\omega')}\,dx'\,d\omega'\\ &= \int_{\mathbb{R}^{2d}} \hat{\mu}(x',\omega')\check{\psi}(\omega')\overline{\check{\phi}(\omega')}|\widehat{\varphi^*}(x')|^2\,dx'\,d\omega' = 0\end{aligned}$$

for all $\psi,\phi\in L^2(\mathbb{R}^d)$, implying that B_μ^φ is the zero operator. □

F.4.3 Non-compactness

As the blurring operator is based on a convolution, it should come as no surprise that it is not compact. However, a proof requires some careful considerations due to the involvement of the synthesis V_φ^*. Before proceeding with a proof, we establish two preliminary lemmas.

Lemma F.4.11. *Let $F \in L^1(\mathbb{R}^{2d})$ and $\varphi \in L^1(\mathbb{R}^d)$, then*

$$\widehat{V_\varphi^* F}(\xi) = \int_{\mathbb{R}^{2d}} F(x,\omega)\hat{\varphi}(\xi-\omega)e^{-2\pi i x\cdot(\xi-\omega)}\,dx\,d\omega.$$

Proof. We compute

$$\begin{aligned}
\widehat{V_\varphi^* F}(\xi) &= \int_{\mathbb{R}^d} V_\varphi^* F(t) e^{-2\pi i t\cdot\xi}\,dt \\
&= \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^{2d}} F(x,\omega)\varphi(t-x)e^{2\pi i t\cdot\omega}\,dx\,d\omega \right) e^{-2\pi i t\cdot\xi}\,dt \\
&= \int_{\mathbb{R}^{2d}} F(x,\omega) \left(\int_{\mathbb{R}^d} \varphi(t-x)e^{-2\pi i t\cdot(\xi-\omega)}\,dt \right) dx\,d\omega \\
&= \int_{\mathbb{R}^{2d}} F(x,\omega) \left(\int_{\mathbb{R}^d} \varphi(s)e^{-2\pi i s\cdot(\xi-\omega)}\,ds \right) e^{-2\pi i x\cdot(\xi-\omega)}\,dx\,d\omega \\
&= \int_{\mathbb{R}^{2d}} F(x,\omega)\hat{\varphi}(\xi-\omega)e^{-2\pi i x\cdot(\xi-\omega)}\,dx\,d\omega
\end{aligned}$$

where the exchange of order of integration is justified by Fubini. □

Lemma F.4.12. *Let $F \in L^2(\mathbb{R}^{2d})$, then*

$$\left\| P_{V_\varphi(L^2)} F \right\|_{L^2(\mathbb{R}^{2d})} = \left\| P_{V_\varphi(L^2)} F(\cdot - (0,\xi)) \right\|_{L^2(\mathbb{R}^{2d})}$$

for all $\xi \in \mathbb{R}^d$.

Proof. Write $F = F_1 + F_2$ where $F_1 \in V_\varphi(L^2(\mathbb{R}^d))$ and $F_2 \in V_\varphi(L^2(\mathbb{R}^d))^\perp$ for the orthogonal decomposition of F. The first function F_1 can be identified with $V_\varphi\eta$ for some $\eta \in L^2(\mathbb{R}^d)$. In general, we have that

$$V_\varphi(M_\xi\eta)(x,\omega) = \langle M_\xi\eta, M_\omega T_x\varphi\rangle = \langle \eta, M_{\omega-\xi}T_x\varphi\rangle = V_\varphi\eta(x,\omega-\xi). \quad \text{(F.4.2)}$$

As a consequence, $F_1(\cdot - (0,\xi)) \in V_\varphi(L^2(\mathbb{R}^d))$ and if we can show that $F_2(\cdot - (0,\xi)) \in V_\varphi(L^2(\mathbb{R}^d))^\perp$, we are done. To see this, note that for any $\chi \in L^2(\mathbb{R}^d)$,

$$\begin{aligned}
\left\langle F_2(\cdot - (0,\xi)), V_\varphi\chi \right\rangle_{L^2(\mathbb{R}^{2d})} &= \left\langle F_2, V_\varphi\chi(\cdot + (0,\xi)) \right\rangle_{L^2(\mathbb{R}^{2d})} \\
&= \left\langle F_2, V_\varphi(M_{-\xi}\chi) \right\rangle_{L^2(\mathbb{R}^{2d})} = 0
\end{aligned}$$

where the final equality is due to $F_2 \in V_\varphi(L^2(\mathbb{R}^d))^\perp$. Combining these facts, we have that

$$\begin{aligned}
\left\| P_{V_\varphi(L^2)} F(\cdot - (0,\xi)) \right\|_{L^2(\mathbb{R}^{2d})} &= \left\| P_{V_\varphi(L^2)} V_\varphi(M_\xi\eta) \right\|_{L^2(\mathbb{R}^{2d})} \\
&\quad + \left\| P_{V_\varphi(L^2)} F_2(\cdot - (0,\xi)) \right\|_{L^2(\mathbb{R}^{2d})} \\
&= \|\eta\| = \left\| P_{V_\varphi(L^2)} F \right\|_{L^2(\mathbb{R}^{2d})}
\end{aligned}$$

which is what we wished to show. □

We are now ready to prove that the operator is non-compact through an example of a bounded sequence of functions whose image under B_μ^φ has no convergent subsequence.

Theorem F.4.13. *Let $\mu \in M(\mathbb{R}^{2d})$ and $\varphi \in M^1(\mathbb{R}^d)$ be such that B_μ^φ is not the zero operator, then B_μ^φ is non-compact.*

Proof. Since B_μ^φ is not the zero operator, there exists a function $\psi_0 \in L^2(\mathbb{R}^d)$ such that $B_\mu^\varphi \psi_0 \neq 0$. From Proposition F.4.1 we know that B_μ^φ is a bounded operator on $L^2(\mathbb{R}^d)$ and so we can approximate ψ_0 by a function $\psi_c \in M^1(\mathbb{R}^d)$ with compact support in the Fourier domain such that $\|B_\mu^\varphi \psi_c\| \neq 0$. By rescaling, we can assume that $\|\psi_c\| = 1$ without loss of generality. For technical reasons, we will need to assume that both μ and $\hat{\varphi}$ also have compact support so that we can choose a positive number R such that the supports of $\hat{\psi_c}$ and $\hat{\varphi}$ are contained in the $\mathbb{R}^d$ ball of radius R, B_R, and the support of μ is contained in the ball of the same radius in $\mathbb{R}^{2d}$.

Consider the bounded $L^2(\mathbb{R}^d)$ sequence $(\psi_n)_n$ defined by $\psi_n = M_{8Rn}\psi_c$. We will show that the sequence $(B_\mu^\varphi \psi_n)_n$ is uniformly separated by means of disjoint supports in the frequency domain, contradicting compactness, and lastly show that we can remove the assumption of compact supports of μ and $\hat{\varphi}$.

For the disjointness of the supports, we first claim that if the function $\hat{\psi}$ has support in $E \subset \mathbb{R}^d$ and the Fourier transform of the window, $\hat{\varphi}$, has support in B_R, the support of the last d coordinates of $V_\varphi \psi$ is contained in $E + B_R$. Indeed,

$$V_\varphi \psi(x, \omega) = \left\langle \hat{\psi}, T_\omega M_{-x} \hat{\varphi} \right\rangle = \int_E \hat{\psi}(\xi) \overline{\hat{\varphi}(\xi - \omega)} e^{2\pi i (\xi - \omega) \cdot x} \, d\xi$$

where the relation $V_\varphi \psi(x, \omega) = \langle \hat{\psi}, T_\omega M_{-x} \hat{\varphi} \rangle$ can be found in [107, Lemma 3.1.1]. Note now that the statement $V_\varphi(x, \omega) = 0$ if $\omega \notin E + B_R$ holds if $\xi \in E, \omega \notin E + B_R \implies \xi - \omega \notin B_R$ which in turn is easily verified.

Next since the support of μ is contained in B_R, this time a $2d$-dimensional ball, we can conclude that convolution $\mu * V_\varphi \psi$ has support in the last d variables contained in $E + B_{2R}$.

Finally for the synthesized signals $B_\mu^\varphi \psi_n$, we claim that if $F \in L^1(\mathbb{R}^{2d})$ has support in the last d variables contained in $E + B_{2R}$, it follows that the support of $\widehat{V_\varphi^* F}$ is contained in $E + B_{3R}$. To see this, we use Lemma F.4.11 and expand as

$$\begin{aligned} \widehat{V_\varphi^* F}(\xi) &= \int_{\mathbb{R}^{2d}} F(x, \omega) \hat{\varphi}(\xi - \omega) e^{-2\pi i x \cdot (\xi - \omega)} \, dx \, d\omega \\ &= \int_{\mathbb{R}^d} \left(\int_{E + B_{2R}} F(x, \omega) \hat{\varphi}(\xi - \omega) e^{-2\pi i x \cdot (\xi - \omega)} \, d\omega \right) dx. \end{aligned}$$

We claim that this quantity is zero whenever $\xi \notin E + B_{3R}$. Indeed, $\omega \in E + B_{2R}$, $\xi \notin E + B_{3R} \implies \xi - \omega \notin B_R$ as is easily verified.

By these three steps we have shown that the support of the Fourier transform of $B_\mu^\varphi \psi_n = V_\varphi^*(\mu * V_\varphi \psi_n)$ is contained in $\operatorname{supp} \psi_n + B_{3R}$ and by the construction of ψ_n, these supports are disjoint for different n. Consequently, $\|B_\mu^\varphi \psi_n - B_\mu^\varphi \psi_m\| = \|B_\mu^\varphi \psi_n\| + \|B_\mu^\varphi \psi_m\|$ whenever $n \neq m$. We claim that this norm is unchanged when ψ is modulated. Indeed, as we saw in (F.4.2), a modulation of the signal corresponds to a translation in phase space and convolutions respect translations. Hence the uniform boundedness follows from Lemma F.4.12 since

$$\|B_\mu^\varphi \psi\| = \|P_{V_\varphi(L^2)}(\mu * V_\varphi \psi)\|_{L^2(\mathbb{R}^{2d})}$$

by Moyal's formula (F.2.2) and (F.2.4). This means that

$$\|B_\mu^\varphi \psi_n - B_\mu^\varphi \psi_m\| = 2\|B_\mu^\varphi \psi_c\| > 0$$

for all $n \neq m$ and there can be no convergent subsequence of $(B_\mu^\varphi \psi_n)_n$.

Now to lift the assumption of compactness of the supports of μ and $\hat{\varphi}$, we write $\mu = \mu_0 + \mu_1$ and $\varphi = \varphi_0 + \varphi_1$ where μ_0 and $\hat{\varphi}_0$ have compact support and

$$\begin{aligned} \|\mu_1\|_{M(\mathbb{R}^{2d})} &< \min\left\{ \frac{\|B_{\mu_0}^{\varphi_0}\psi_c\|}{3\|\varphi_0\|^2}, \sqrt{\frac{\|B_{\mu_0}^{\varphi_0}\psi_c\|}{3}} \right\}, \\ \|\varphi_1\|^2 &< \min\left\{ \frac{\|B_{\mu_0}^{\varphi_0}\psi_c\|}{3\|\mu_0\|_{M(\mathbb{R}^{2d})}}, \sqrt{\frac{\|B_{\mu_0}^{\varphi_0}\psi_c\|}{3}} \right\}. \end{aligned}$$

We can then decompose the operator as

$$B_\mu^\varphi = B_{\mu_0}^{\varphi_0} + \underbrace{B_{\mu_0}^{\varphi_1} + B_{\mu_1}^{\varphi_0} + B_{\mu_1}^{\varphi_1}}_{=C}.$$

Now by applying the above proof to $B_{\mu_0}^{\varphi_0}$, we get a sequence $(\psi_n)_n$, all with norm 1, such that

$$\|B_{\mu_0}^{\varphi_0}\psi_n - B_{\mu_0}^{\varphi_0}\psi_m\| = 2\|B_{\mu_0}^{\varphi_0}\psi_c\|.$$

From here, we can apply the triangle inequality to find that

$$\begin{aligned} \|B_\mu^\varphi(\psi_n - \psi_m)\| &\geq \|B_{\mu_0}^{\varphi_0}(\psi_n - \psi_m)\| - \|C(\psi_n - \psi_m)\| \\ &\geq 2\big(\|B_{\mu_0}^{\varphi_0}\psi_c\| - \|C\|_{B(L^2(\mathbb{R}^d))}\big) \end{aligned} \tag{F.4.3}$$

and so it will follow that $(B_\mu^\varphi \psi_n)_n$ is uniformly bounded if we can bound the norm of C from above by $\|B_{\mu_0}^{\varphi_0}\psi_c\|$. Indeed, by the bound in Proposition F.4.1 and the conditions imposed on the norms of μ_1 and φ_1 above,

$$\begin{aligned}\|C\|_{B(L^2(\mathbb{R}^d))} &\leq \|B_{\mu_0}^{\varphi_1}\|_{B(L^2(\mathbb{R}^d))} + \|B_{\mu_1}^{\varphi_0}\|_{B(L^2(\mathbb{R}^d))} + \|B_{\mu_1}^{\varphi_1}\|_{B(L^2(\mathbb{R}^d))} \\ &< 3\frac{\|B_{\mu_0}^{\varphi_0}\|}{3} = \|B_{\mu_0}^{\varphi_0}\|.\end{aligned}$$

Consequently, the right hand side of (F.4.3) is positive and $(B_\mu^\varphi \psi_n)_n$ is uniformly separated, implying non-compactness. □

F.5 Implementations and applications

F.5.1 Examples

We briefly discuss some specifics of implementing the time-frequency blurring operator as well as spectrogram blurring. The code used to produce the figures in this section can be found on GitHub[1] where details around the kernels, windows and specific settings also are available.

STFT blurring

The purest implementation of (F.1.1) consists of simply a STFT, a convolution operation and an inverse STFT. In code, we can implement this as a function which acts on the waveform of the signal and this is what we use as the augmentation operation in Section F.5.2. For visualizing the result, we can use either the regular spectrogram or a log-mel spectrogram. Both versions can be seen in Figure F.2 and another example without mel rescaling is illustrated in Figure F.1.

In general, the discrete time-frequency blurring operator significantly reduces the ℓ^2 energy of the input waveform since the phase of the STFT varies rapidly. In the figures we have performed $0-1$ normalization on the spectrograms to compensate for this phenomenon.

Kernel dependence

So far, we have only looked at Gaussian kernels. In Figure F.3 we look at the action of 5 different kernels to the same audio signal. The square and circle are similar to what we have seen before and mainly blur the spectrograms. Still, the outputs are guaranteed to be absolute values of elements of the Gabor space $V_\varphi(L^2)$ which is not the case for spectrogram blurring. For the translated Gaussian kernel, we

[1] `https://www.github.com/SimonHalvdansson/Time-Frequency-Blurring-Operator`

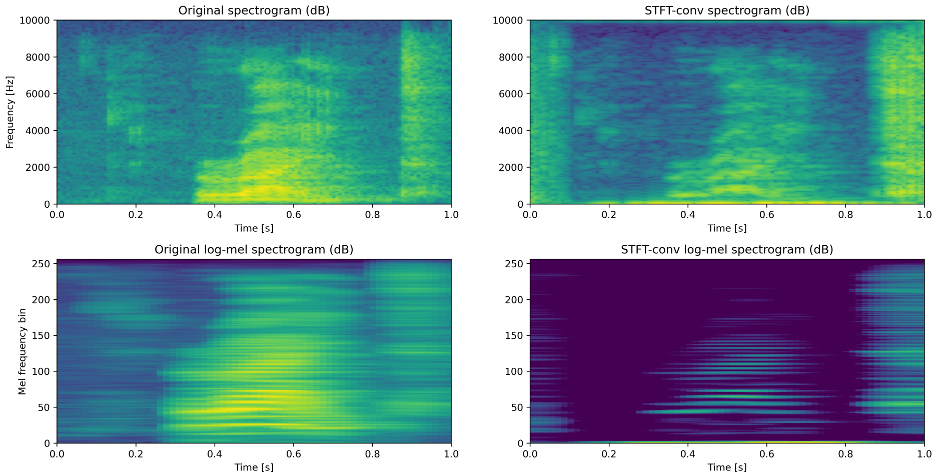

Figure F.2: Spectrograms and log-mel spectrograms of an audio recording and the same recording with a time-frequency blurring operator with Gaussian kernel applied to it.

have used zero padding to clarify the effect which is mostly a translation of the spectrogram. Still, since $B^{\varphi}_{T_z\mu}(\psi) \neq B^{\varphi}_{\mu}(\pi(z)\psi)$ due to $\pi(z)\pi(w) \neq \pi(z+w)$, this is not precisely the effect. The two straight line kernels mainly have the effect of stretching the spectrograms. On a very simple level, one might expect that horizontal blurring should correspond to convolving the original waveform ψ and that vertical blurring should correspond to a convolution on $\hat{\psi}$. This is not quite what we observe partially due to the $B^{\varphi}_{T_z\mu}(\psi) \neq B^{\varphi}_{\mu}(\pi(z)\psi)$ phenomena mentioned earlier.

Spectrogram blurring

Spectrogram blurring as discussed in Section F.3.1 is implemented by first computing the spectrogram, rescaling it to logarithmic decibel scale, and then applying a convolution. In Figure F.4, we illustrate this on the spectrogram of an audio clip in both normal and mel scale. It is the mel-rescaled version which we use for data augmentation in Section F.5.2.

In Figure F.5 we lastly look at the difference between the blurring operator and spectrogram blurring on white noise. The zeros of spectrograms of white noise have been studied in e.g. [21]. A noticeable difference in the figure is that the middle spectrogram has considerably more zeros than the blurred spectrogram and the zeros behave similarly to those of pure white noise. Meanwhile the blurred spectrogram clearly does not correspond to any actual signal.

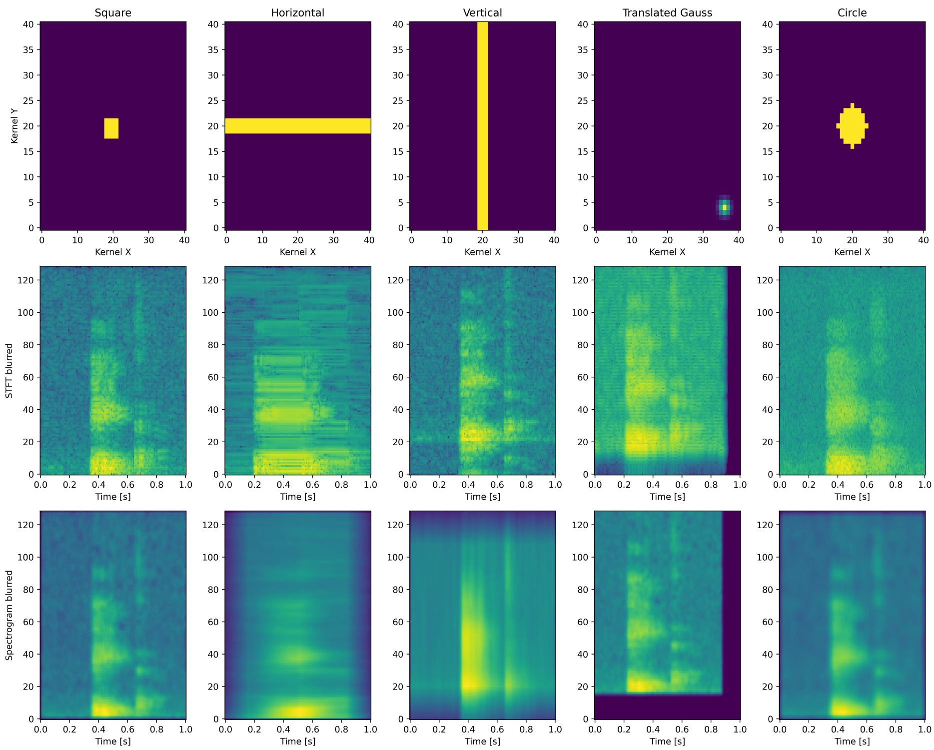

Figure F.3: Five different kernels (top row), the results of time-frequency blurring operators applied with those kernels (middle row) and spectrogram blurring with the same kernels (bottom row).

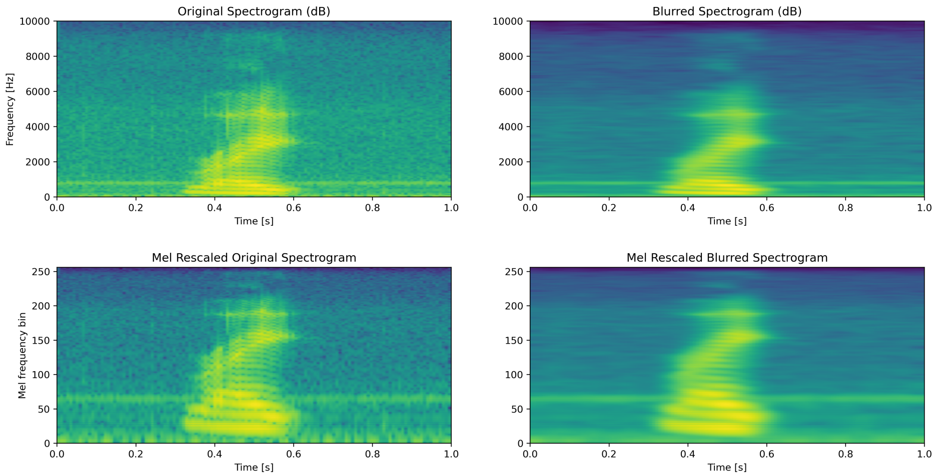

Figure F.4: Spectrograms and log-mel spectrograms of an audio recording and the same spectrograms blurred with a Gaussian kernel.

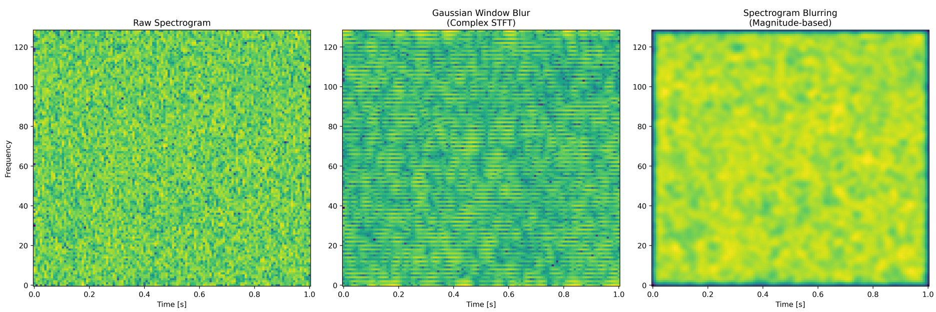

Figure F.5: Spectrogram of white noise, the same white noise with a blurring operator applied to it, and a blurred version of the original spectrogram with the same kernel.

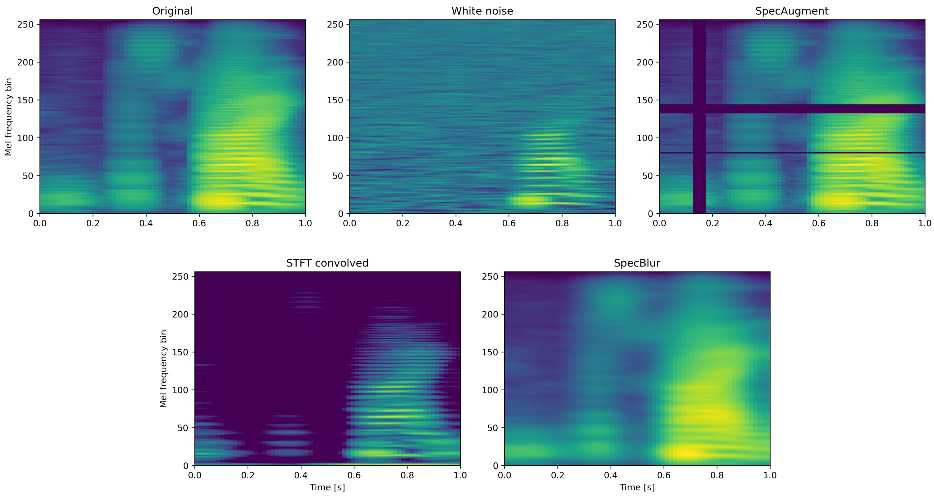

Figure F.6: Log-mel spectrograms of an audio recording from the SpeechCommands V2 dataset [209] with different augmentation techniques applied to it.

Figure F.3 also illustrates the difference to the time-frequency blurring operator where the outputs of the blurring operator are significantly more similar to actual spectrograms than the blurred spectrograms in the middle row.

F.5.2 Evaluation as a data augmentation method

To investigate the performance of the blurring operator as well as spectrogram blurring for data augmentation, a convolutional neural network (CNN) using the ResNet-34 architecture [121] and a vision transformer (ViT) using the TinyViT-11M architecture [221] was trained on the SpeechCommands V2 dataset [209] for 35-class classification using different augmentation setups. These two models were chosen as they represent the two main contemporary paradigms for image recognition systems [58, 103]. The full code used to produce all the plots and tables of this section is available on GitHub[2].

Setup

The SpeechCommands V2 dataset [209] contains audio recordings of utterances of 35 different words which we aim to classify using a neural network. As our focus is on data augmentation, we limit the number of training examples available to the model to learn from to 100, 300, 600 or 1000 recordings from each class. We use

[2] https://www.github.com/SimonHalvdansson/Time-Frequency-Blurring-Operator

an 80/20 train/validation split and use the validation set to decide when to stop training. We always use 200 examples from each class as the test data set.

For preprocessing, all recordings are padded to be 1 second long and then the log-mel spectrogram of resolution 63×256 (time $\times$ frequency) is computed and used as input to the network. As the purpose is to evaluate the effect of using STFT convolving and spectrogram blurring for data augmentation, we also use different augmentation methods to compare against. Specifically, we consider the augmentation techniques of adding white noise to the waveform and performing basic time and frequency masking as described in the SpecAugment paper [171]. Finally for each of the different counts of input examples, we evaluate the model with no augmentation, all augmentations, once for each augmentation, once with only white noise and SpecAugment and once with both blurring augmentations, making for a total of 8 reported accuracies per input size.

There is a natural variation to the test accuracies as we pick the dataset which we train on randomly and randomly initialize the weights. We therefore train the network several times with different randomly selected datasets each time. In order to compare augmentation methods, we estimate the mean of the accuracies for each configuration. Here the *standard error* quantifies our uncertainty in the mean and is given by $\sigma/\sqrt{n}$ where σ is the natural standard deviation of the distribution we sample from and n is the number of samples. We estimate σ as the standard deviation of the accuracies we record over the n training runs. For each configuration the training and test procedures are repeated until the estimated standard errors are so small that we can compare the accuracies of different augmentation methods.

For both the time-frequency blurring operator and spectrogram blurring, we use a two-dimensional Gaussian kernel with manually fine-tuned spread in time and frequency, see the code for exact details. Spectrogram blurring is performed on the log-mel spectrogram as in Section F.5.1 while the time-frequency blurring operator is applied with a non-rescaled STFT as in Section F.5.1. We stress that the exact details of the implementation should be considered secondary as we are mainly interested in comparing the novel augmentation to the baseline and classical augmentation techniques.

CNN results

We summarize the results from 1780 runs with the ResNet-34 CNN [121] in Table F.1 which can be reproduced using the code in the GitHub repository.

Table F.1: Average CNN test accuracies with standard errors (%) for 100, 300, 600 and 1000 input examples per class for different augmentation setups.

Augmentation	**Acc-100**	**Acc-300**	**Acc-600**	**Acc-1000**
None	47.05±0.26	75.13±0.38	83.13±0.25	87.27±0.19
White noise	62.67±0.48	78.10±0.27	84.74±0.33	88.77±0.21
SpecAugment	67.70±0.41	80.77±0.19	85.86±0.18	88.99±0.16
STFT-blur	63.31±0.50	78.69±0.29	84.90±0.27	88.88±0.18
SpecBlur	64.50±0.45	80.38±0.20	85.83±0.25	88.91±0.29
White noise + SpecAug	66.57±0.34	81.62±0.19	87.38±0.19	89.90±0.14
STFT-blur + SpecBlur	68.24±0.30	81.67±0.19	86.45±0.20	89.36±0.15
All	70.48±0.23	83.17±0.15	87.98±0.11	90.51±0.14

We see the greatest improvements compared to the baseline for the lower example configurations which is to be expected. Notably, there is a significant improvement for both STFT-blur and SpecBlur compared to the baseline for all training sizes ($p < 0.001$) as well as a significant improvement going from white noise + SpecAugment to all augmentations ($p < 0.05$). Combining STFT-blur with SpecBlur offers modest improvements compared to just using one of them. However, this could be because the augmentation is applied to each example in the training set leading to some overfitting in the direction of blurred spectrograms.

ViT results

The experiments were repeated with a TinyViT-11M vision transformer [221] and the results of 1310 runs are in Table F.2. The code to produce the table can be found in the main GitHub repository.

Table F.2: Average ViT test accuracies with standard errors (%) for 100, 300, 600 and 1000 input examples per class for different augmentation setups.

Augmentation	**Acc-100**	**Acc-300**	**Acc-600**	**Acc-1000**
None	25.85±0.29	71.15±0.46	84.32±0.23	89.17±0.20
White noise	41.64±0.32	80.84±0.22	87.94±0.15	90.72±0.09
SpecAugment	46.97±0.33	81.26±0.22	87.55±0.08	90.61±0.14
STFT-blur	50.46±0.28	81.00±0.24	87.56±0.18	90.40±0.15
SpecBlur	52.67±0.30	84.08±0.14	89.00±0.12	91.29±0.13
White noise + SpecAug	56.61±0.33	84.46±0.15	89.61±0.15	91.80±0.15
STFT-blur + SpecBlur	67.54±0.29	85.65±0.16	89.22±0.17	91.72±0.12
All	73.38±0.19	86.89±0.14	90.60±0.13	92.70±0.08

As vision transformers generally are more sensitive to the amount of training data, we see larger accuracy gains from using augmentation than in the CNN case. The improvement over baseline for STFT-blur and SpecBlur is statistically significant for all training sizes ($p < 0.001$). STFT-blur and SpecBlur also provide statistically significant improvements when comparing white noise together with SpecAugment against all augmentation methods combined ($p < 0.001$). While STFT-blur + SpecBlur consistently outperformed SpecBlur alone, when only using one of the two augmentation methods, SpecBlur performed better ($p < 0.001$).

Conclusions

We have seen that both STFT-blur and SpecBlur are promising augmentation techniques that can be used to improve the performance of both convolutional neural networks and vision transformers on spectrograms, at least at smaller scales. The computationally more efficient SpecBlur generally performs better than STFT-blur alone but combining several augmentation methods results in superior performance. In particular, using some form of blurring on top of white noise and SpecAugment results in significantly improved performance in the situations tested.

Paper G

Empirical Plunge Profiles of Time-Frequency Localization Operators

Simon Halvdansson

Preprint

Paper G

Empirical Plunge Profiles of Time-Frequency Localization Operators

Abstract

For time-frequency localization operators with symbol $R\Omega$, we work out the exact large R eigenvalue behavior for rotationally invariant Ω and conjecture that the same relation holds for all scaled symbols $R\Omega$ as long as the window is the standard Gaussian. Specifically, we conjecture that the k-th eigenvalue of the localization operator with symbol $R\Omega$ converges to $\frac{1}{2}\operatorname{erfc}\left(\sqrt{2\pi}\frac{k-|R\Omega|}{|\partial R\Omega|}\right)$ as $R \to \infty$. To support the conjecture, we compute the eigenvalues of discrete frame multipliers with various symbols using LTFAT and find that they agree with the behavior of the conjecture to a large degree.

G.1 Introduction and background

When restricting a signal $f \in L^2(\mathbb{R}^d)$ to a subset of the time-frequency plane, there are two main approaches. The simpler is to consider a spatial cutoff, followed by a Fourier multiplier, followed by the same spatial cutoff once more. For sets $E, F \subset \mathbb{R}^{2d}$ and $\mathcal{F}$ the Fourier transform, we can write such operators as

$$Sf = \chi_F \mathcal{F}^{-1} \chi_E \mathcal{F} \chi_F f \tag{G.1.1}$$

where χ_Ω is the indicator function of the set Ω. While such an operator does not yield a function compactly supported in both time and frequency, as that is prohibited by the uncertainty principle, it approximately does so provided E, F are large enough. This line of work goes back to Landau, Pollak and Slepian, starting in the 1960s [146, 147, 195–197], who showed many of the classical properties of these operators which we will refer to as *Fourier concentration operators* following [163].

Another more general way to restrict a signal to a subset $\Omega \subset \mathbb{R}^{2d}$ of the time-frequency plane is to apply the multiplication operator on a time-frequency representation of f. Specifically, using the *short-time Fourier transform* (STFT) [107], defined with a window function $g \in L^2(\mathbb{R}^d)$ as

$$V_g f(x,\omega) = \int_{\mathbb{R}^d} f(t)\overline{g(t-x)}e^{-2\pi i\omega t}\,dt = \langle f, \pi(x,\omega)g\rangle,$$

where $\pi(x,\omega)f(t) = M_\omega T_x f(t) = e^{2\pi i\omega\cdot t}f(t-x)$ is a *time-frequency shift*, we can define the *localization operator* A_Ω^g as

$$A_\Omega^g f = \int_\Omega V_g f(x,\omega)\pi(x,\omega)g\,dx\,d\omega.$$

If $\Omega = \mathbb{R}^{2d}$, this is just the identity operator on $L^2(\mathbb{R}^d)$, provided g is normalized, but for a compact Ω we get a compact self-adjoint operator. Localization operators were first considered in the time-frequency context by Daubechies [51].

In both cases, the operators can be interpreted as projection operators onto either the subspace $F \times E$ or Ω of the time-frequency plane. The number of orthogonal functions which fit in these subspaces is approximately equal to the area of the subset of the time-frequency plane and consequently the corresponding eigenvalues are close to 1. This is followed by what is commonly referred to as the *plunge region* where the eigenvalues rapidly decay to 0. All eigenvalues in the plunge region $\delta < \lambda_k < 1-\delta$ correspond to eigenfunctions which are partially supported inside and outside the subset Ω of the time-frequency plane. As these eigenfunctions are orthogonal, they each occupy a part of $\partial\Omega$ and we therefore expect the number of eigenvalues in the plunge region to depend on the size of $\partial\Omega$.

G.1.1 Earlier results

For Fourier concentration operators, these intuitions have been quantified with quite some success. The number of eigenvalues close to one was shown to be approximately equal to $|E|\cdot|F|$ for E, F intervals by Landau in [145]. In particular, for any $\delta > 0$, the quantity

$$\frac{\#\{k : \lambda_k > 1-\delta\}}{|E|\cdot|F|} \tag{G.1.2}$$

approaches 1 as we dilate E and F. The size of the plunge region, meaning the number of eigenvalues between δ and $1-\delta$, was also bounded by $\log(|E|\cdot|F|)$ up to a constant.

In the more general setting of compact E and F, Marceca, Romero and Speckbacher [163] showed under mild conditions that the size of the plunge region is bounded by

$$\frac{|\partial E|}{\kappa_{\partial E}}\frac{|\partial F|}{\kappa_{\partial F}}\log\left(\frac{|\partial E||\partial F|}{\kappa_{\partial E}\delta}\right)^{2d(1+\alpha)+1}$$

up to a constant factor, where $\kappa_{\partial E}$ is the maximal Ahlfors regular boundary constant of E, α is some number in $(0, 1/2)$ and $|\partial E|$ is the $(d-1)$-dimensional Hausdorff measure of the boundary ∂E.

There are also more detailed asymptotics on the eigenvalue behavior near the plunge region, see [142] and references therein for an overview of these results.

Less is known in the case of time-frequency localization operators and this is what we aim to start to address in this paper. The number of eigenvalues close to 1 was first bounded by Ramanathan and Topiwala in [180] by showing that

$$\frac{\#\{k : \lambda_k^\Omega > 1-\delta\}}{|\Omega|} \tag{G.1.3}$$

also converges to 1 as Ω is dilated. As a byproduct of the proof, one also finds the upper limit

$$\left|\#\{k : \lambda_k^\Omega > 1-\delta\} - |\Omega|\right| \leq C|\partial\Omega|$$

but no equivalence. Similar results are also available for Gabor multipliers, the discrete variant of localization operators, see e.g. [74, 75].

G.1.2 Our contribution

We will show that in the case of a rotationally invariant symbol, meaning a disk, annuli or union of annuli, the eigenvalues of localization operators can be computed explicitly and asymptotically exhibit an erfc (complementary error function) decay after $|\Omega|$ eigenvalues, over a range proportional to $|\partial\Omega|$. We conjecture that this behavior is universal for all symbols Ω as long as the window is the standard Gaussian and support the conjecture by verifying it numerically for a diverse collection of Ω with small error.

G.2 Eigenvalue behavior

G.2.1 Eigenvalues on disks, annuli, and rotationally invariant sets

In the original article on time-frequency localization operators [51], a general formula for computing the eigenvalues of localization operators with Gaussian window

$g_0(t) = 2^{1/4}e^{-\pi t^2}$ and a rotationally invariant symbol was given. Specialized to the case $\Omega = B(0, R)$, $d = 1$ and with our normalization conventions, we get

$$\lambda_k^{B(0,R)} = 1 - e^{-\pi R^2} \sum_{j=0}^{k} \frac{(\pi R^2)^j}{j!}. \tag{G.2.1}$$

from [51, Eq. (19c)]. Through a connection with the Poisson distribution, the large R asymptotics of this can be computed neatly.

Theorem G.2.1. *Let λ_k^R be the k-th eigenvalue of the localization operator with symbol $B(0, R)$. It then holds that*

$$\left|\lambda_k^R - \frac{1}{2}\operatorname{erfc}\left(\frac{k - \pi R^2}{\sqrt{2\pi}R}\right)\right| = O\left(\frac{1}{R}\right) \tag{G.2.2}$$

where erfc *is the complementary error function.*

Proof. We recognize (G.2.1) as the Poisson cumulative distribution function (CDF) with parameter πR^2. Specifically, if $X \sim \mathrm{Po}(\pi R^2)$, then

$$\lambda_k^R = 1 - \mathbb{P}(X \leq k).$$

For large R, the Poisson distribution $\mathrm{Po}(\pi R^2)$ can be approximated by a normal distribution with mean and variance πR^2 due to the central limit theorem. Therefore, we can approximate the CDF of X as

$$\mathbb{P}(X \leq k) \approx \Phi\left(\frac{k - \pi R^2}{\sqrt{\pi R^2}}\right),$$

where Φ is the standard normal CDF.

Substituting this approximation into our expression for λ_k^R, we obtain

$$\lambda_k^R \approx 1 - \Phi\left(\frac{k - \pi R^2}{\sqrt{\pi R^2}}\right).$$

Now recall that the complementary error function is related to the standard normal Φ by

$$\Phi(z) = \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{z}{\sqrt{2}}\right)\right] \implies 1 - \Phi(z) = \frac{1}{2}\operatorname{erfc}\left(\frac{z}{\sqrt{2}}\right).$$

Applying this to our expression for λ_k^R, we find that

$$\lambda_k^R \approx \frac{1}{2}\operatorname{erfc}\left(\frac{k - \pi R^2}{\sqrt{2\pi}R}\right).$$

To quantify the error in this approximation, we employ the Berry-Esseen theorem, which provides a bound on the difference between the Poisson CDF and its normal approximation. For the Poisson distribution, the Berry-Esseen bound states that

$$\left|\mathbb{P}(X \leq k) - \Phi\left(\frac{k - \pi R^2}{\sqrt{\pi R^2}}\right)\right| = O\left(\frac{1}{\sqrt{\pi R^2}}\right) = O\left(\frac{1}{R}\right).$$

Consequently, the difference between λ_k^R and its normal approximation satisfies

$$\left|\lambda_k^R - \frac{1}{2}\operatorname{erfc}\left(\frac{k - \pi R^2}{\sqrt{2\pi} R}\right)\right| = O\left(\frac{1}{R}\right)$$

which is what we wished to show. □

Remark G.2.2. In [51] the large R eigenvalue behavior is only investigated for fixed k which just tells us how quickly $\lambda_k^R \to 1$ as $R \to \infty$, not the full erfc behavior. Still, the eigenvalue formula (G.2.1) and the rest of the results of [51] are so well known that Theorem G.2.1 should perhaps be considered folklore in the field. Still, we have found no reference for it in the literature and so we include a proof in the interest of completion while making no claim of originality.

Remark G.2.3. The corresponding eigenvalue formula for $d > 1$ is also available in [51] but is dependent on a multiindex k. To make computations simpler, we have chosen to restrict ourselves to the $d = 1$ case.

While the above theorem was specialized to the case of a disk centered at 0, the same eigenvalue behavior can be observed irrespective of the center of the disk. To see this, recall that for disks centered at 0 it is the Hermite functions $(h_k)_k$ which are the eigenfunctions. Now using that

$$\langle A^{g_0}_{B(0,R)} h_k, h_k\rangle = \lambda_k,$$

we can see that $\pi(z_0)h_k$ is an eigenfunction with the same eigenvalue for the localization operator $A^{g_0}_{B(z_0,R)}$. Indeed, with the change of variables $w = z - z_0$,

$$\begin{aligned}\left\langle A^{g_0}_{B(z_0,R)}(\pi(z_0)h_k), \pi(z_0)h_k\right\rangle &= \int_{B(z_0,R)} V_{g_0}(\pi(z_0)h_k)(z)\langle \pi(z)g_0, \pi(z_0)h_k\rangle\, dz \\ &= \int_{B(0,R)} \langle \pi(z_0)h_k, \pi(z_0+w)g_0\rangle\langle \pi(z_0+w)g_0, \pi(z_0)h_k\rangle\, dz \\ &= \int_{B(0,R)} \langle h_k, \pi(w)g_0\rangle\langle \pi(w)g_0, h_k\rangle\, dz = \lambda_k\end{aligned}$$

where we in the second to last step canceled out two phase factors. This means that we have the same erfc eigenvalue decay no matter where the disk is centered.

In the case where Ω is an annulus, which we will take to be centered at 0 in the interest of brevity, we also have an erfc eigenvalue decay but we will have to work a little harder to show it. For a deeper discussion on localization operators with annuli as symbols, see [7].

Proposition G.2.4. Let λ_k^R be the k-th eigenvalue of the localization operator with symbol $B(0,R) \setminus B(0,rR)$ for $r < 1$. It then holds that

$$\left|\lambda_k^R - \frac{1}{2}\operatorname{erfc}\left(\frac{k - \pi R^2(1-r^2)}{\sqrt{2\pi}R(1+r)}\right)\right| = O\left(\frac{1}{R}\right). \tag{G.2.3}$$

Proof. In this situation, the eigenfunctions of $A^{g_0}_{B(0,R)\setminus B(0,rR)}$ are still the Hermite functions and the unordered eigenvalues can be written as

$$\lambda_k^{B(0,R)\setminus B(0,rR)} = \lambda_k^{B(0,R)} - \lambda_k^{B(0,rR)} \tag{G.2.4}$$

where $\lambda_k^{B(0,R)}$ are the eigenvalues from Theorem G.2.1. As $R \to \infty$, this quantity will converge to

$$f(k) = \frac{1}{2}\left[\operatorname{erfc}\left(\frac{k - \pi R^2}{\sqrt{2\pi}R}\right) - \operatorname{erfc}\left(\frac{k - \pi (rR)^2}{\sqrt{2\pi}rR}\right)\right] \tag{G.2.5}$$

with error bounded by $O(1/R)$. However, (G.2.4) is not ordered decreasingly. Writing $\mu_k = \lambda_k^{B(0,R)} - \lambda_k^{B(0,rR)}$ and letting μ_k^* denote the k-th element of the ordered collection $(\mu_k)_k$, then μ_k^* is the k-th eigenvalue λ_k^R. To relate this to f we will let f^* denote the decreasing rearrangement $f^* : \mathbb{R}^+ \to \mathbb{R}$ of f and show that

$$\mu_k = f(k) \implies |\mu_k^* - f^*(k)| \leq \|f'\|_\infty.$$

Let $\bar{f}$ be defined so that $\bar{f}(k) = \mu_k$ for integers k and $\bar{f}$ is constant on all intervals $[k, k+1)$. Then

$$\left|\bar{f}(x) - f(x)\right| = \left|f(\lfloor x \rfloor) - f(x)\right| \leq \|f'\|_\infty \left|x - \lfloor x \rfloor\right| \leq \|f'\|_\infty,$$

i.e., $\|\bar{f} - f\|_\infty < \|f'\|_\infty$. Meanwhile $\mu_k^* = \bar{f}^*(k)$ so

$$\left|\mu_k^* - f^*(k)\right| = \left|\bar{f}^*(k) - f^*(k)\right| \leq \left\|\bar{f}^* - f^*\right\|_\infty \leq \left\|\bar{f} - f\right\|_\infty$$

by $\|g^* - h^*\|_\infty \leq \|g - h\|_\infty$ for general functions g, h, see [151, Chapter 3] for a proof. It is easy to see that $\|f'\|_\infty = O(1/R)$ so we can conclude that

$$|\lambda_k - f^*(k)| = O\left(\frac{1}{R}\right). \tag{G.2.6}$$

To construct the rearrangement f^*, we will let two cursors traverse the two erfc functions at the same height but different locations along the x-axis. Specifically, let

$$x_1(t) = t\sqrt{2\pi}R + \pi R^2, \qquad x_2(t) = -t\sqrt{2\pi}rR + \pi(rR)^2.$$

These two cursors meet at $t = t_0 = \frac{-\pi R^2(1-r)}{\sqrt{2\pi}R}$ however this is not necessarily the highest point of f. We would like to be able to traverse the two sides of f at the same time at the same heights but this is not quite possible with our x_1, x_2. Consider instead the alternative function $\tilde{f}$ which is defined as

$$\tilde{f}(x_2(t)) := f(x_1(t))$$

for $t \in [t_0, \infty)$ and $\tilde{f} = f$ on $x_1([t_0, \infty))$. This construction is well-defined because x_1 and x_2 perfectly divide up all of $\mathbb{R}$. We can now bound the error $|f - \tilde{f}|$ as

$$\begin{aligned}\left\|f - \tilde{f}\right\|_\infty &= \sup_{t \geq t_0} |f(x_1(t)) - f(x_2(t))| \\ &= \frac{1}{2} \sup_{t \geq t_0} \left|2 - \operatorname{erfc}\left(\frac{t\sqrt{2\pi}R + \pi R^2(1 - r^2)}{\sqrt{2\pi}rR}\right) - \operatorname{erfc}\left(\frac{-t\sqrt{2\pi}rR - \pi R^2(1 - r^2)}{\sqrt{2\pi}R}\right)\right|\end{aligned}$$

where we used that $\operatorname{erfc}(t) + \operatorname{erfc}(-t) = 2$. Since erfc is a decreasing function, we can bound this quantity using first that

$$\begin{aligned}\operatorname{erfc}\left(\frac{t\sqrt{2\pi}R + \pi R^2(1 - r^2)}{\sqrt{2\pi}rR}\right) &\leq \operatorname{erfc}\left(\frac{t_0\sqrt{2\pi}R + \pi R^2(1 - r^2)}{\sqrt{2\pi}rR}\right) \\ &= \operatorname{erfc}\left(\frac{\pi R^2(r - r^2)}{\sqrt{2\pi}rR}\right) = O\left(\frac{1}{R}\right) \qquad \text{(G.2.7)}\end{aligned}$$

where we used that $\operatorname{erfc}(z) = \frac{e^{-z^2}}{\sqrt{\pi}z} + O\left(\frac{1}{z^2}\right)$ [55, §7.12(i)]. Now $2 - \operatorname{erfc}(x)$ is an increasing function so it can be bounded from above by plugging in $t = t_0$ since the argument is decreasing in t. This means that

$$\begin{aligned}2 - \operatorname{erfc}\left(\frac{-t\sqrt{2\pi}rR - \pi R^2(1 - r^2)}{\sqrt{2\pi}R}\right) &\leq 2 - \operatorname{erfc}\left(\frac{-t_0\sqrt{2\pi}rR - \pi R^2(1 - r^2)}{\sqrt{2\pi}R}\right) \\ &= 2 - \operatorname{erfc}\left(\frac{-\pi R^2(1 - r)}{\sqrt{2\pi}R}\right) \\ &= \operatorname{erfc}\left(\frac{\pi R^2(1 - r)}{\sqrt{2\pi}R}\right) = O\left(\frac{1}{R}\right).\end{aligned}$$

Having established $\|f - \tilde{f}\|_\infty$ is small, it also follows that the two rearrangements f^* and $\tilde{f}^*$ satisfy

$$\left\|f^* - \tilde{f}^*\right\|_\infty \leq \left\|f - \tilde{f}\right\|_\infty$$

from [151, Chapter 3] again. It still remains to write out the rearrangement $\tilde{f}^*(x)$. To do so, we introduce another cursor $x(t)$ which should map t_0 to 0. We also make the ansatz that $x(t)$ is linear in t, i.e., of the form $x(t) = at + b$. To preserve the total mass, for each dt we should add the mass from $\tilde{f}(x_1(t))$ and $\tilde{f}(x_2(t))$ combined to $\tilde{f}^*(x(t))$. Using that $\tilde{f}(x_2(t)) = \tilde{f}(x_1(t)) = f(x_1(t))$, this means that

$$\begin{aligned}
\tilde{f}^*(x(t))\frac{dx}{dt}dt &= \tilde{f}(x_1(t))\left|\frac{dx_1}{dt}\right|dt + \tilde{f}(x_2(t))\left|\frac{dx_2}{dt}\right|dt \\
&= f(x_1(t))\left(\left|\frac{dx_1}{dt}\right| + \left|\frac{dx_2}{dt}\right|\right)dt \\
&= \frac{1}{2}\left[\operatorname{erfc}(t) - \operatorname{erfc}\left(\frac{t\sqrt{2\pi}R + \pi R^2(1-r^2)}{\sqrt{2\pi}rR}\right)\right]\sqrt{2\pi}R(1+r)dt
\end{aligned}$$

for all $t \in [t_0, \infty)$. Note that we put absolute value bars on the derivatives $\frac{dx_1}{dt}$ and $\frac{dx_2}{dt}$ since even if we are traversing the x-axis in the negative axis, the mass should still be added to $\tilde{f}^*$.

Plugging in $t = t_0$ in the above, using that $\tilde{f}(0) = \tilde{f}(x(t_0)) = f(x_1(t_0))$ and that $\frac{dx}{dt} = a$, we find that

$$\underbrace{\tilde{f}^*(x(t_0))}_{=f(x_1(t_0))}\, a = f(x_1(t_0))\sqrt{2\pi}R(1+r) \implies a = \sqrt{2\pi}R(1+r).$$

Combining this with $x(t_0) = 0$ yields

$$\begin{aligned}
x(t_0) = \sqrt{2\pi}R(1+r)t_0 + b = 0 &\implies b = -\pi R^2(1-r^2) \\
&\implies x(t) = t\sqrt{2\pi}R(1+r) - \pi R^2(1-r^2).
\end{aligned}$$

Plugging this back, we see that

$$\begin{aligned}
\tilde{f}^*(t\sqrt{2\pi}R(1+r) - \pi R^2(1-r^2)) &= \frac{1}{2}\operatorname{erfc}(t) + O\left(\frac{1}{R}\right) \\
\implies \tilde{f}^*(x) &= \frac{1}{2}\operatorname{erfc}\left(\frac{x - \pi R^2(1-r^2)}{\sqrt{2\pi}R(1+r)}\right) + O\left(\frac{1}{R}\right)
\end{aligned}$$

where we used that $\operatorname{erfc}\left(\frac{t\sqrt{2\pi}R+\pi R^2(1-r^2)}{\sqrt{2\pi}rR}\right) = O(1/R)$ from (G.2.7). Combining this with (G.2.6), we conclude that

$$\left|\lambda_k - \frac{1}{2}\operatorname{erfc}\left(\frac{k-\pi R^2(1-r^2)}{\sqrt{2\pi}R(1+r)}\right)\right| = O\left(\frac{1}{R}\right)$$

which is what we wished to show. □

The rearrangement from Proposition G.2.4 is implemented numerically in Figure G.1.

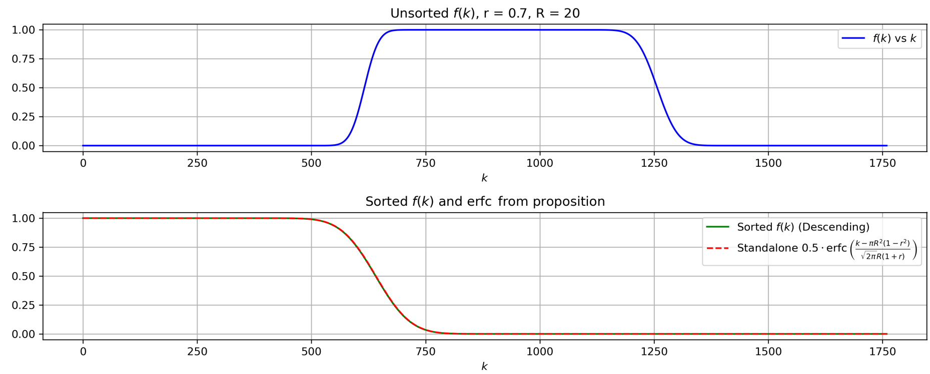

Figure G.1: Numerical verification of Proposition G.2.4 with $R = 20$, $r = 0.7$ comparing a manual sorting of samples of f with the proposed $\tilde{f}^*$.

The same translation argument we mentioned for disks applies to annuli since the proof is only based on the collection of eigenvalues $(\lambda_k^R)_k$.

Our final generalization of this result is a lifting to finite unions of annuli. To formulate this in a manner susceptible to generalization, note that all rotationally invariant Ω can be written as

$$\Omega = \{z \in \mathbb{R}^2 : |z| \in A\}$$

for some $A \subset \mathbb{R}^+$. To rule out the cases where A has an isolated point which changed $\partial\Omega$ but not the localization operator, we can assume that Ω is *regular closed*, i.e., $\Omega = \overline{\operatorname{int}(\Omega)}$. Lastly for technical reasons in the proof below we will need to assume that the number of annuli is finite so that the distance between two annuli is bounded away from 0 which we formulate as Ω having a finite number of connected components.

Proposition G.2.5. Let $\Omega \subset \mathbb{R}^2$ be a compact, regular closed and rotationally invariant set with a finite number of connected components, and let $\lambda_k^{R\Omega}$ the k-th eigenvalue of the localization operator with symbol $R\Omega$. It then holds that

$$\left|\lambda_k^{R\Omega} - \frac{1}{2}\operatorname{erfc}\left(\sqrt{2\pi}\frac{k - |R\Omega|}{|\partial R\Omega|}\right)\right| = O\left(\frac{1}{R}\right). \tag{G.2.8}$$

Proof. Write $\Omega = \cup_{n=1}^N \Omega_n$ for which $|\Omega| = \sum_{n=1}^N |\Omega_n|$ and $|\partial\Omega| = \sum_{n=1}^N |\partial\Omega_n|$. If we let r_{i}^n and r_{o}^n denote the inner and outer radii of Ω_n, we can write the unordered eigenvalues of $A_\Omega^{g_0}$ as

$$\mu_k^\Omega = \sum_{n=1}^N \lambda_k^{r_{\mathrm{o}}^n} - \lambda_k^{r_{\mathrm{i}}^n}$$

where λ_k^R is k-th eigenvalue of $A_{B(0,R)}^{g_0}$. If we let f be the function

$$f(x) = \sum_{n=1}^N \frac{1}{2}\left(\operatorname{erfc}\left(\frac{x - \pi(r_{\mathrm{o}}^n)^2}{\sqrt{2\pi}r_{\mathrm{o}}^n}\right) - \operatorname{erfc}\left(\frac{x - \pi(r_{\mathrm{i}}^n)^2}{\sqrt{2\pi}r_{\mathrm{i}}^n}\right)\right) = \sum_{n=1}^N f_n(x),$$

the unordered μ_k^Ω are (asymptotically) samples of f at the integers. We will need for these f_n to be orthogonal and to that end set out to construct functions $\tilde{f}_n$ which are compactly supported on disjoint subsets of $\mathbb{R}^+$. For each $n \geq 1$, define the expanded radii and the set

$$\tilde{r}_{\mathrm{i}}^n = r_{\mathrm{i}}^n - \frac{r_{\mathrm{i}}^n - r_{\mathrm{o}}^{n-1}}{2}, \qquad \tilde{r}_{\mathrm{o}}^n = r_{\mathrm{o}}^n + \frac{r_{\mathrm{i}}^{n+1} - r_{\mathrm{o}}^n}{2}, \qquad E_n = \left[\tilde{r}_{\mathrm{i}}^n, \tilde{r}_{\mathrm{o}}^n\right]$$

where we set $r_{\mathrm{o}}^0 = 0$ to treat the edge case. Then if we define the compactly supported functions $\tilde{f}_n(x) = \chi_{E_n}(x) f_n(x)$, it holds that

$$\left\|f_n - \tilde{f}_n\right\|_\infty = \max\left\{f_n(\tilde{r}_{\mathrm{i}}^n), f_n(\tilde{r}_{\mathrm{o}}^n)\right\}.$$

To see that this quantity is $O(\frac{1}{R})$ when we scale Ω, note that since $\tilde{r}_{\mathrm{i}}^n < r_{\mathrm{i}}^n$ and $\tilde{r}_{\mathrm{o}}^n > r_{\mathrm{o}}^n$, when we plug these values into f_n it will be bounded by $\operatorname{erfc}(\varepsilon R)$ for some $\varepsilon > 0$ which is $O(\frac{1}{R})$ by the same expansion from [55, §7.12(i)] we used in Proposition G.2.4.

Now using Proposition G.2.4, the rearrangement of each f_n can be written as

$$\begin{aligned} f_n^*(x) &= \frac{1}{2}\operatorname{erfc}\left(\frac{x - \pi\left((r_{\mathrm{o}}^n)^2 - (r_{\mathrm{i}}^n)^2\right)}{\sqrt{2\pi}(r_{\mathrm{o}}^n + r_{\mathrm{i}}^n)}\right) + O\left(\frac{1}{|\partial\Omega_n|}\right) \\ &= \frac{1}{2}\operatorname{erfc}\left(\sqrt{2\pi}\frac{x - |\Omega_n|}{|\partial\Omega_n|}\right) + O\left(\frac{1}{|\partial\Omega_n|}\right). \end{aligned}$$

Since $\|f_n - \tilde{f}_n\|_\infty = O(\frac{1}{R})$, we can conclude that $\|f_n^* - \tilde{f}_n^*\|_\infty = O(\frac{1}{R})$. Define

$$g_n(x) = \frac{1}{2}\operatorname{erfc}\left(\sqrt{2\pi}\frac{x - |\Omega_n|}{|\partial\Omega_n|}\right)$$

so that $\|g_n - f_n^*\|_\infty = O(\frac{1}{R})$ and, in turn, $\|g_n - \tilde{f}_n^*\|_\infty = O(\frac{1}{R})$. If we also define

$$\tilde{f}(x) = \sum_{n=1}^{N} \tilde{f}_n(x) \implies \|\tilde{f}^* - f^*\|_\infty = \left\|\sum_{n=1}^{N}(\tilde{f}_n - f_n)\right\|_\infty = O\left(\frac{1}{R}\right),$$

we see that it suffices to show that $\|f^* - \frac{1}{2}\operatorname{erfc}\big(\sqrt{2\pi}\frac{\cdot - |\Omega|}{|\partial\Omega|}\big)\|_\infty = O(\frac{1}{R})$. Since the collection $(\tilde{f}_n)_n$ all have disjoint supports, we can write

$$|\{x \geq 0 : \tilde{f}^*(x) > \gamma\}| = \sum_{n=1}^{N} |\{x \geq 0 : \tilde{f}_n(x) > \gamma\}| = \sum_{n=1}^{N} |\{x \geq 0 : \tilde{f}_n^*(x) > \gamma\}| \tag{G.2.9}$$

where we in the last step used that the measures of level sets are unaffected by rearrangements. For the g_n functions, we can explicitly compute

$$|\{x \geq 0 : g_n(x) > \gamma\}| = g_n^{-1}(\gamma) = |\Omega_n| + \operatorname{erfc}^{-1}(2\gamma)\frac{|\partial\Omega_n|}{\sqrt{2\pi}}.$$

Now write ε for the largest difference between $|\tilde{f}_n^* - g_n|$ over all n, it then holds that

$$\begin{aligned}
&|\{x \geq 0 : g_n(x) > \gamma + \varepsilon\}| \leq |\{x \geq 0 : \tilde{f}_n^*(x) > \gamma\}| \leq |\{x \geq 0 : g_n(x) > \gamma - \varepsilon\}| \\
\implies\ & |\Omega_n| + \frac{|\partial\Omega_n|}{\sqrt{2\pi}}\operatorname{erfc}^{-1}(2(\gamma+\varepsilon)) \leq |\{x \geq 0 : \tilde{f}_n(x) > \gamma\}| \leq |\Omega_n| + \frac{|\partial\Omega_n|}{\sqrt{2\pi}}\operatorname{erfc}^{-1}(2(\gamma-\varepsilon)) \\
\implies\ & |\Omega| + \frac{|\partial\Omega|}{\sqrt{2\pi}}\operatorname{erfc}^{-1}(2(\gamma+\varepsilon)) \leq |\{x \geq 0 : \tilde{f}^*(x) > \gamma\}| \leq |\Omega| + \frac{|\partial\Omega|}{\sqrt{2\pi}}\operatorname{erfc}^{-1}(2(\gamma-\varepsilon))
\end{aligned}$$

where we in the last step summed over n and plugged in (G.2.9). Equivalently, since $\tilde{f}^*$ is decreasing (since no $\tilde{f}_n$ has derivative zero in an interval), we can write

$$\tilde{f}^*\left(|\Omega| + \frac{|\partial\Omega|}{\sqrt{2\pi}}\operatorname{erfc}^{-1}(2(\gamma+\varepsilon))\right) \leq \tilde{f}^*(x) = \gamma \leq \tilde{f}^*\left(|\Omega| + \frac{|\partial\Omega|}{\sqrt{2\pi}}\operatorname{erfc}^{-1}(2(\gamma-\varepsilon))\right).$$

Moreover, $\tilde{f}^*$ is invertible and so we can conclude that for any $\gamma = \tilde{f}^*(x)$,

$$
\begin{aligned}
|\Omega| + \frac{|\partial\Omega|}{\sqrt{2\pi}} \operatorname{erfc}^{-1}(2(\gamma+\varepsilon)) \leq (\tilde{f}^*)^{-1}(\gamma) \leq |\Omega| + \frac{|\partial\Omega|}{\sqrt{2\pi}} \operatorname{erfc}^{-1}(2(\gamma-\varepsilon)) \\
\implies \operatorname{erfc}^{-1}(2(\gamma+\varepsilon)) \leq \frac{\sqrt{2\pi}}{|\partial\Omega|}\left[(\tilde{f}^*)^{-1}(\gamma) - |\Omega|\right] \leq \operatorname{erfc}^{-1}(2(\gamma-\varepsilon)) \\
\implies \gamma - \varepsilon \leq \frac{1}{2}\operatorname{erfc}\left(\frac{\sqrt{2\pi}}{|\partial\Omega|}\left[(\tilde{f}^*)^{-1}(\gamma) - |\Omega|\right]\right) \leq \gamma + \varepsilon \\
\implies \tilde{f}^*(x) - \varepsilon \leq \frac{1}{2}\operatorname{erfc}\left(\sqrt{2\pi}\frac{x - |\Omega|}{|\partial\Omega|}\right) \leq \tilde{f}^*(x) + \varepsilon.
\end{aligned}
$$

With this we can finish the proof by noting that $\left|\lambda_k^{R\Omega} - f^*(k)\right| = O(\frac{1}{R})$ by the same argument as we used in Proposition G.2.4 and that $\varepsilon = O(\frac{1}{R})$ since N is finite. □

G.2.2 Universality

These proofs have ultimately led us to the erfc asymptotics by a central limit theorem argument which notoriously is universal in the sense that we get the same limit for a large class of probability distributions. In physics, the notion of universality [54] near a boundary point is a well studied phenomenon and in particular erfc universality is a very important result in random matrix theory [122]. This setting is of particular interest due to its strong connection to localization operators, see [3, 9].

An early piece of evidence in the direction of the boundary universality conjecture in random matrix theory [122] was the calculation of the eigenvalue asymptotics for the special case of the Gaussian unitary ensemble (GUE), where each entry in the random matrix is a Gaussian random variable. In this setup, Forrester and Honner [86] showed that the density of the eigenvalues near a boundary point will converge to an erfc kernel in the limiting case. The Gaussian unitary ensemble precisely corresponds to the case of a localization operator with Gaussian window functions and the disk as its symbol through an intricate procedure involving the Bargmann transform [9]. This correspondence inspires confidence that the link between random matrices and eigenvalues of localization operators may persist in the eigenvalues asymptotics for more general classes of symbols.

It is conceivable that for other Ω than those we have discussed, there could exist a random variable X_Ω such that $\lambda_k^\Omega = \mathbb{P}(X_\Omega \leq k)$ that has the property that this probability is related to the central limit theorem in the large R limit. To make a precise conjecture on these asymptotics, we first need to make the rescaling implicit in Theorem G.2.1, Proposition G.2.4 and Proposition G.2.5 more explicit.

Following the setup in e.g. [8], we define the dilation $R\Omega$ of Ω as

$$R\Omega = \{z \in \mathbb{R}^{2d} : z/R \in \Omega\}.$$

In both (G.2.2) and (G.2.3), the argument of the erfc function can be written as

$$\sqrt{2\pi}\frac{k - |R\Omega|}{|\partial R\Omega|}$$

and we conjecture that this behavior could be universal.

Conjecture G.2.6. Let $\Omega \subset \mathbb{R}^2$ be compact and regular closed, and let λ_k^Ω be the k-th eigenvalue of the localization operator with symbol Ω and window the standard Gaussian. Then

$$\left|\lambda_k^{R\Omega} - \frac{1}{2}\operatorname{erfc}\left(\sqrt{2\pi}\frac{k - |R\Omega|}{|\partial R\Omega|}\right)\right| = O\left(\frac{1}{R}\right)$$

where erfc is the complementary error function.

In particular, this conjecture implies that the plunge region has width comparable to $|\partial\Omega|$ which is a weaker result which we will investigate in Section G.3.

It is possible that we might have to require stronger conditions on Ω for the conjecture to hold. In particular the condition of Ω having maximally Ahlfors regular boundary has proven important in recent work by Marceca and Romero [162]. However in the absence of evidence to the contrary, we present the conjecture in full generality.

A reason to believe in plunge profile universality is the min-max formulation of the eigenvalues of A_Ω^g discussed in e.g. [8], namely

$$\lambda_k^\Omega = \max\left\{\int_\Omega |V_g f(z)|^2 \, dz : \|f\|_{L^2} = 1, f \perp h_1^\Omega, \dots, h_{k-1}^\Omega\right\}.$$

This implies that the plunge eigenvalues belong to the eigenfunctions which are supported around the boundary of Ω. In particular, the values λ_k^Ω depend on how the short-time Fourier transforms of orthogonal eigenfunctions h_k^Ω repel each other around the boundary. In the large R limit, the boundary $\partial R\Omega$ is approximately straight both for $\Omega = B(0,1)$ and general Ω, save for pathological examples. It is therefore not unreasonable that the eigenfunctions, which do not scale with R, would have the same behavior for any Ω when R is large as these local objects do not sense the global structure of $R\Omega$.

If the window function g induces some form of anisotropy in the time-frequency plane, the number of spectrograms $|V_g h_k^\Omega|^2$ which occupy a given stretch of $\partial\Omega$ could be dependent on the angle for this approximate line segment. This is not an issue for the interior, where there are no boundary effects to consider and all spectrograms take up the same area, 1. In [51, Section V.B], it is shown that when the window is a dilated Gaussian and the symbol is a corresponding ellipse, the eigenvalues are the same as for a symmetric disk and the standard Gaussian. Hence, we know that the conjecture is false if we remove the condition of the window being the standard Gaussian. This issue is investigated numerically in Section G.3.3 below.

Still, we have so far only presented heuristic arguments in favor of Conjecture G.2.6. Our main evidence comes in the form of computing $\lambda_k^\Omega - \frac{1}{2}\operatorname{erfc}\left(\sqrt{2\pi}\frac{k-|\Omega|}{|\partial\Omega|}\right)$ for a large collection of Ω for frame multipliers. The strong correspondence between results for localization operators and Gabor multipliers has been investigated for a long time and holds up well, from proving the same eigenvalue plunge behavior [74, 75] to showing trace-class convergence for dense lattices [81] and accumulated spectrogram behavior [119].

G.3 Numerical verification

In this section we attempt to verify Conjecture G.2.6 numerically using the Large Time-Frequency Analysis Toolbox (LTFAT) [175]. Obviously, we are not able to realize localization operators as they are continuous objects and even Gabor multipliers are based on samples of $L^2(\mathbb{R}^d)$ functions. However, the finite Gabor multipliers, or *frame multipliers*, we can realize in LTFAT which are based on vector representations of signals are likely to approximate Gabor multipliers well. Moreover, those Gabor multipliers in turn will have similar eigenvalue behavior as the corresponding localization operators as they are close in trace-class and Hilbert-Schmidt norms for dense lattices [75, 81].

G.3.1 Setup

In LTFAT, the `framemuleigs` function takes in a symbol, analysis frame and synthesis frame and returns the eigenvalues and optionally the eigenvectors of the corresponding frame multiplier. The frames are in turn determined by the time-hop distance `a`, the number of frequency channels `M` and the window function `g`. To avoid under- and oversampling, we always set the signal length `L` to `L` = `a` × `M` and unless stated otherwise, we use a standard Gaussian window function `g` = `pgauss(L)`.

The full code is available on GitHub[1] and should ideally be self-explanatory. In LTFAT, symbols are defined on a `M` × `M` grid and we have used a pipeline where

[1] `https://github.com/SimonHalvdansson/Time-Frequency-Plunge-Profiles`

images can be converted to binary masks to simplify experimentation with symbols which are difficult to define in code. The area $|\Omega|$ is computed by summing the symbol while for the perimeter length $|\partial\Omega|$ we use the built-in MATLAB function `regionprops`. To account for the different coordinates in discrete phase space, the symbol area is multiplied by $\mathsf{a}\,/\,\mathsf{M}$ and the perimeter by $\sqrt{\mathsf{a}/\mathsf{M}}$.

To support a weaker version of the conjecture numerically, we will write out the length of the perimeter as reported by `regionprops`, the number of eigenvalues in the plunge region with $\delta = 0.1$ as well as their quotient which should be approximately constant across different symbols and frames.

G.3.2 Symbol and frame dependence

We will first consider the case which we know best, that where the symbol is a disk. The error we observe here will serve as a benchmark for all upcoming experiments as they should only come from the following factors:

- Finite size symbol (not asymptotic limit),
- Localization operator to Gabor multiplier error (lattice effects),
- Gabor multiplier to frame multiplier error (discrete functions),
- Lattice boundary length (measuring $|\partial\Omega|$ using `regionprops`).

For this reason, we expect that the errors we observe for the disk should be a lower bound for the errors we observe, a form of noise floor.

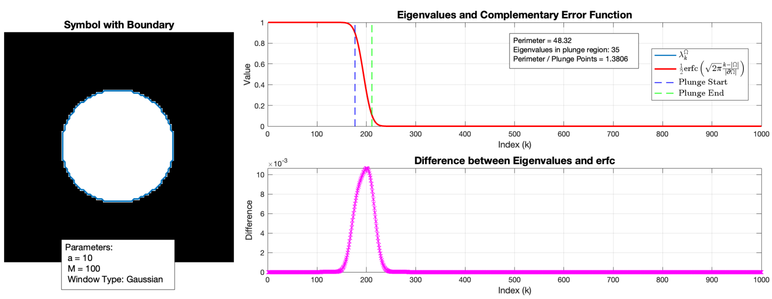

Figure G.2: Experiment with $\mathsf{a} = 10$, $\mathsf{M} = 100$, Gaussian window and a disk as the symbol. Maximum error is close to 1.0%.

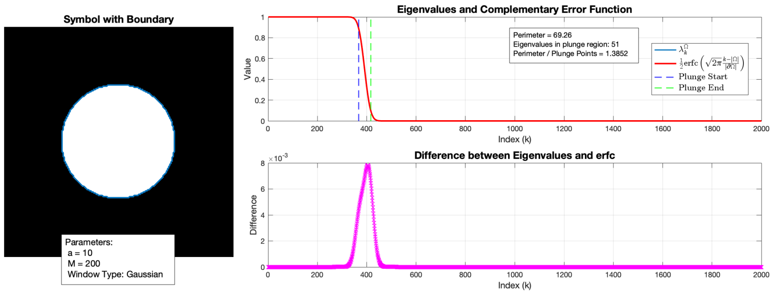

Figure G.3: Experiment with a = 10, M = 200, Gaussian window and a disk as the symbol. Maximum error is close to 0.8%.

The higher value for M corresponds to a denser lattice which explains the smaller peak discrepancy in Figure G.3 compared to Figure G.2.

Next we look at a collection of different symbols and frame parameters. For a star shape we observe considerably higher errors for a sparse lattice but for the a = 10, M = 100 lattice the error is comparable to that for the disk, see Figures G.4 and G.5.

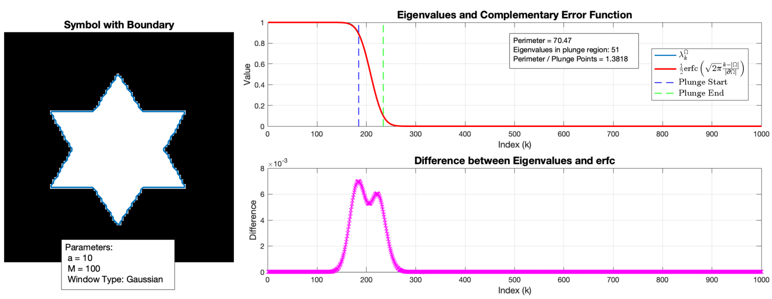

Figure G.4: Experiment with a = 10, M = 100, Gaussian window and a star shape as the symbol. Maximum error is close to 0.7%.

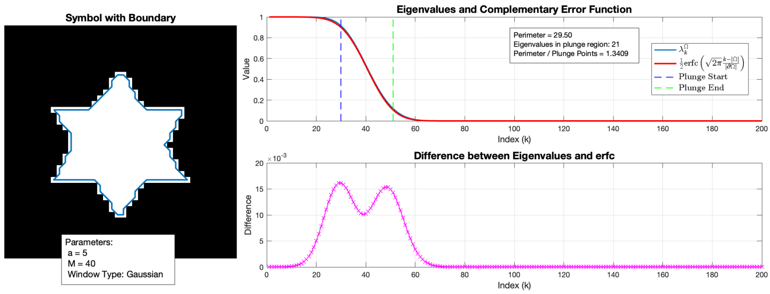

Figure G.5: Experiment with $\mathsf{a} = 5, \mathsf{M} = 40$, Gaussian window and a star shape as the symbol. Maximum error is close to 1.6%.

The symbols in Figures G.6 and G.7 are poorly conditioned as they are thin which means that eigenfunctions belonging to the plunge region are likely to be influenced by the symbol boundary on the opposite side. In this case, we see considerably higher errors.

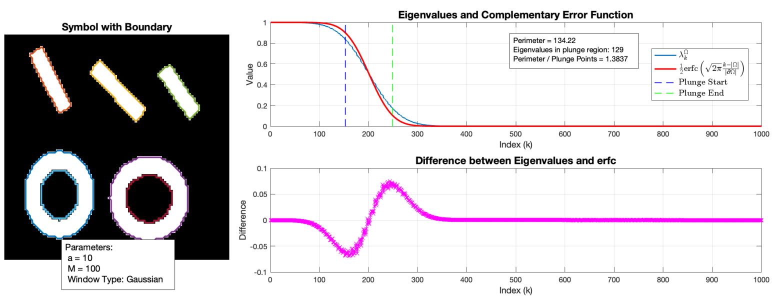

Figure G.6: Experiment with $\mathsf{a} = 10, \mathsf{M} = 100$, Gaussian window and lines and circles as the symbol. Maximum error is close to 7.5%.

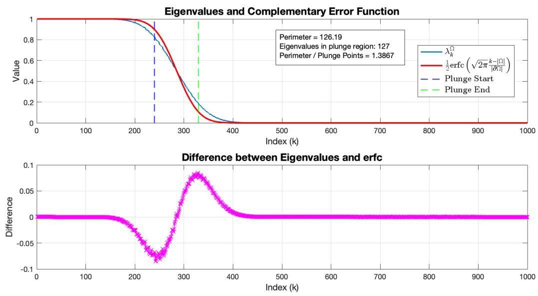

Figure G.7: Experiment with $\mathsf{a} = 10, \mathsf{M} = 100$, Gaussian window and tiles as the symbol. Maximum error is close to 8.0%.

For a more well behaved but still intricate symbol, see Figure G.8.

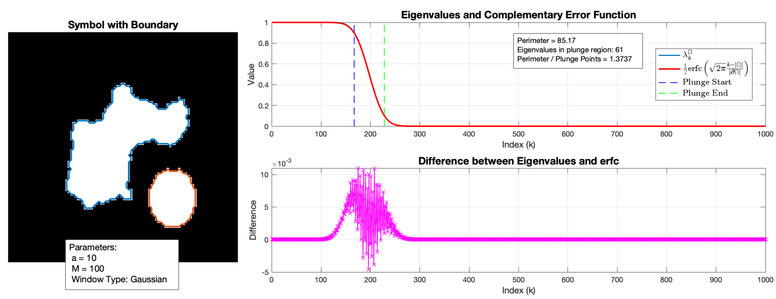

Figure G.8: Experiment with $\mathsf{a} = 10, \mathsf{M} = 100$, Gaussian window and blobs as the symbol. Maximum error is close to 1.1%.

The case of elliptical symbols was discussed in [51, Section V.B] where it was shown that if the Gaussian window was dilated appropriately, the eigenvalue behavior is the same as for the disk with the same area. However, the perimeter of an ellipse differs significantly from that of the disk with the same area, which is why we required the window to be the standard Gaussian. In Figure G.9, we verify that the conjecture still appears to hold in this case.

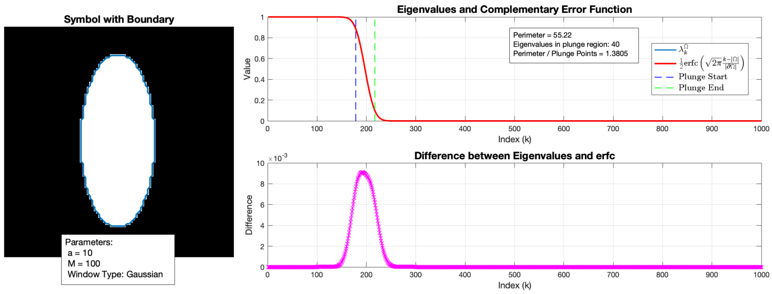

Figure G.9: Experiment with a = 10, M = 100, Gaussian window and an ellipse as the symbol. Maximum error is close to 0.9%.

Lastly we look at a square symbol (Figure G.10) where the results are similar to those for the disk or star.

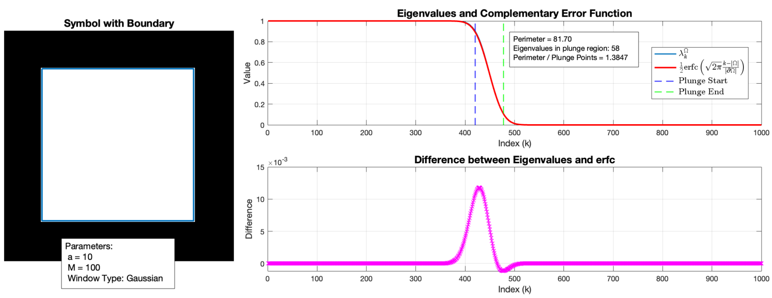

Figure G.10: Experiment with a = 10, M = 100, Gaussian window and a square as the symbol. Maximum error is close to 1.2%.

The results from all the above figures are summarized in Table G.1.

Table G.1: Summarized results for a collection of symbols and parameters. Here $\#P$ denotes the cardinality of the plunge region with parameter $\delta = 0.1$.

Symbol	**a**	**M**	L^∞ **error**	$\lvert\partial\Omega\rvert/\#P$
Disk	10	100	1.0%	1.3806
Disk	10	200	0.8%	1.3852
Star	10	100	1.0%	1.3818
Star	5	40	1.6%	1.3409
Lines and circles	10	100	7.5%	1.3837
Tiles	10	100	8.0%	1.3867
Blobs	10	100	1.1%	1.3737
Ellipse	10	100	0.9%	1.3805
Square	10	100	1.2%	1.3847

The tiles and lines and circles examples have considerably higher errors and were chosen to have a high $|\partial\Omega|/|\Omega|$ ratio. For the symbols which are interior-dominated, as all symbols are asymptotically as we scale R, the erfc curve is remarkably close to the true eigenvalue behavior.

G.3.3 Window dependence

All of the examples we have seen so far have been with a Gaussian window. In this section, we show that the fitted curve has a markedly larger discrepancy when the window is a box function, which has worse time-frequency concentration than the standard Gaussian, and offer an explanation for why.

In Figure G.11, we have repeated the experiment from the previous section with a different window and get a noticeably larger discrepancy.

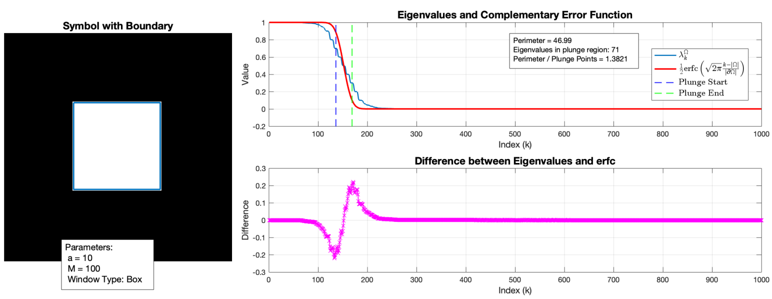

Figure G.11: Eigenvalue decay for frame multiplier with box window function.

Other symbols also have similarly uneven decay which is also wider than that for Gaussian windows.

As mentioned near the end of Section G.2, we have reason to believe that the spectrograms corresponding to different eigenfunctions are more separated when the window function has uneven concentration in time versus frequency. To investigate this, we consider a sparse time-frequency lattice and a disk symbol. By taking the three spectrograms corresponding to the eigenvalues in the middle of the plunge region, those closest to $\lambda = 1/2$, and mapping their brightness to different color channels this separation can be visualized, see Figure G.12.

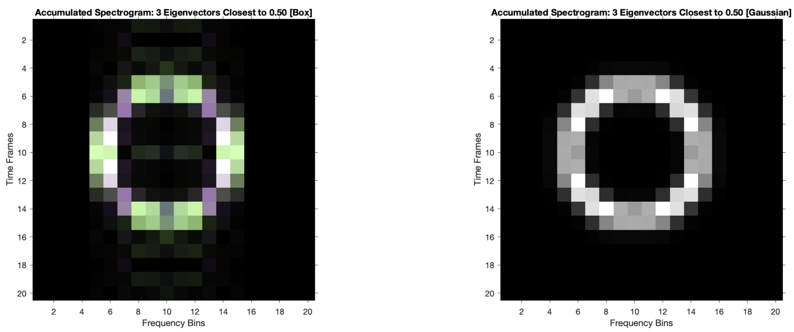

Figure G.12: Accumulated spectrograms of three eigenfunctions (colored red, green, blue) closest to eigenvalue $1/2$ for a box window (left) and a Gaussian window (right).

In Figure G.12 we see that the accumulated spectrogram with Gaussian window is almost monochrome, meaning that the spectrograms always intersect while the spectrograms with a box window are well separated.

Bibliography

[1] O. O. Abayomi-Alli, R. Damaševičius, A. Qazi, M. Adedoyin-Olowe, and S. Misra, "Data augmentation and deep learning methods in sound classification: A systematic review," *Electronics*, vol. 11, no. 22, p. 3795, 2022. DOI: 10.3390/electronics11223795.

[2] L. D. Abreu, K. Gröchenig, and J. L. Romero, "Harmonic analysis in phase space and finite Weyl–Heisenberg ensembles," *J. Stat. Phys.*, vol. 174, no. 5, pp. 1104–1136, 2019. DOI: 10.1007/s10955-019-02226-2.

[3] L. D. Abreu, J. Pereira, J. L. Romero, and S. Torquato, "The Weyl-Heisenberg ensemble: Statistical mechanics meets time-frequency analysis," *2017 International Conference on Sampling Theory and Applications (SampTA)*, IEEE, 2017, pp. 199–202. DOI: 10.1109/sampta.2017.8024413.

[4] L. D. Abreu, J. M. Pereira, and J. L. Romero, "Sharp rates of convergence for accumulated spectrograms," *Inverse Problems*, vol. 33, no. 11, p. 115 008, 2017. DOI: 10.1088/1361-6420/aa8d79.

[5] L. D. Abreu, "Local maxima of white noise spectrograms and Gaussian entire functions," *J. Fourier Anal. Appl.*, vol. 28, no. 6, 2022. DOI: 10.1007/s00041-022-09979-7.

[6] L. D. Abreu, "Entanglement entropy and hyperuniformity of Ginibre and Weyl– Heisenberg ensembles," *Lett. Math. Phys.*, vol. 113, no. 3, 2023. DOI: 10.1007/s11005-023-01674-y.

[7] L. D. Abreu and M. Dörfler, "An inverse problem for localization operators," *Inverse Problems*, vol. 28, no. 11, p. 115 001, 2012. DOI: 10.1088/0266-5611/28/11/115001.

[8] L. D. Abreu, K. Gröchenig, and J. L. Romero, "On accumulated spectrograms," *Trans. Amer. Math. Soc.*, vol. 368, no. 5, pp. 3629–3649, 2015. DOI: 10.1090/tran/6517.

[9] L. D. Abreu, J. M. Pereira, J. L. Romero, and S. Torquato, “The Weyl–Heisenberg ensemble: Hyperuniformity and higher Landau levels,” *J. Stat. Mech. Theory Exp.*, vol. 2017, no. 4, p. 043 103, 2017. DOI: 10.1088/1742-5468/aa68a7.

[10] R. Adamczak, “A note on the Hanson-Wright inequality for random vectors with dependencies,” *Electron. Commun. Probab.*, vol. 20, 2015. DOI: 10.1214/ecp.v20-3829.

[11] R. Alaifari, F. Bartolucci, and M. Wellershoff, “Unique wavelet sign retrieval from samples without bandlimiting,” *Proc. Amer. Math. Soc. Ser. B*, vol. 11, no. 30, pp. 330–344, 2024. DOI: 10.1090/bproc/201.

[12] S. T. Ali, N. M. Atakishiyev, S. M. Chumakov, and K. B. Wolf, “The Wigner function for general Lie groups and the wavelet transform,” *Ann. Henri Poincaré*, vol. 1, pp. 685–714, 2000. DOI: 10.1007/PL00001012.

[13] S. T. Ali, H. Führ, and A. E. Krasowska, “Plancherel inversion as unified approach to wavelet transforms and Wigner functions,” *Ann. Henri Poincare*, vol. 4, no. 6, pp. 1015–1050, 2003. DOI: 10.1007/s00023-003-0154-4.

[14] S. T. Ali, A. E. Krasowska, and R. Murenzi, “Wigner functions from the two-dimensional wavelet group,” *J. Opt. Soc. Am. A*, vol. 17, no. 12, p. 2277, 2000. DOI: 10.1364/josaa.17.002277.

[15] S. T. Ali, J.-P. Antoine, and J.-P. Gazeau, *Coherent States, Wavelets, and Their Generalizations*. Springer New York, 2014. DOI: 10.1007/978-1-4614-8535-3.

[16] J.-P. Antoine and R. Murenzi, “Two-dimensional directional wavelets and the scale-angle representation,” *Signal Process.*, vol. 52, no. 3, pp. 259–281, 1996. DOI: 10.1016/0165-1684(96)00065-5.

[17] J.-P. Antoine, R. Murenzi, P. Vandergheynst, and S. T. Ali, *Two-Dimensional Wavelets and their Relatives*. Cambridge University Press, 2004. DOI: 10.1017/cbo9780511543395.

[18] D. Arnal and B. Currey, *Representations of Solvable Lie Groups* (New Mathematical Monographs). Cambridge University Press, 2020, vol. 39. DOI: 10.1017/9781108552288.

[19] P. Balazs, F. Bastianoni, E. Cordero, H. G. Feichtinger, and N. Schweighofer, “Comparisons between Fourier and STFT multipliers: The smoothing effect of the short-time Fourier transform,” *J. Math. Anal. Appl.*, vol. 529, no. 1, p. 127 579, 2024. DOI: 10.1016/j.jmaa.2023.127579.

[20] P. Balazs, K. Gröchenig, and M. Speckbacher, "Kernel theorems in coorbit theory," *Trans. Amer. Math. Soc. Ser. B*, vol. 6, no. 11, pp. 346–364, 2019. DOI: 10.1090/btran/42.

[21] R. Bardenet, J. Flamant, and P. Chainais, "On the zeros of the spectrogram of white noise," *Appl. Comput. Harmon. Anal.*, vol. 48, no. 2, pp. 682–705, 2020. DOI: 10.1016/j.acha.2018.09.002.

[22] R. Bardenet and A. Hardy, "Time-frequency transforms of white noises and Gaussian analytic functions," *Appl. Comput. Harmon. Anal.*, vol. 50, pp. 73–104, 2021. DOI: 10.1016/j.acha.2019.07.003.

[23] F. Bastianoni, E. Cordero, and F. Nicola, "Decay and smoothness for eigenfunctions of localization operators," *J. Math. Anal. Appl.*, vol. 492, no. 2, p. 124 480, 2020. DOI: 10.1016/j.jmaa.2020.124480.

[24] D. Bayer and K. Gröchenig, "Time-frequency localization operators and a Berezin transform," *Integral Equ. Oper. Theory.*, vol. 82, no. 1, pp. 95–117, 2015. DOI: 10.1007/s00020-014-2208-z.

[25] M. E. B. Bekka, "Amenable unitary representations of locally compact groups," *Invent. Math.*, vol. 100, no. 1, pp. 383–401, 1990. DOI: 10.1007/bf01231192.

[26] J. S. Ben-Benjamin, L. Cohen, N. C. Dias, P. Loughlin, and J. N. Prata, "What is the Wigner function closest to a given square integrable function?" *SIAM J. Math. Anal.*, vol. 50, no. 5, pp. 5161–5197, 2018. DOI: 10.1137/18M116633X.

[27] F. A. Berezin, "Wick and anti-Wick operator symbols," *Mathematics of the USSR-Sbornik*, vol. 15, no. 4, pp. 577–606, 1971. DOI: 10.1070/sm1971v015n04abeh001564.

[28] E. Berge, "Interpolation in wavelet spaces and the HRT-conjecture," *J. Pseudo-Differ. Oper. Appl.*, vol. 12, no. 1, pp. 1–26, 2021. DOI: 10.1007/s11868-021-00386-y.

[29] E. Berge, "A primer on coorbit theory," *J. Fourier Anal. Appl.*, vol. 28, no. 1, 2021. DOI: 10.1007/s00041-021-09892-5.

[30] E. Berge, S. M. Berge, and R. Fulsche, "A quantum harmonic analysis approach to segal algebras," *Integral Equ. Oper. Theory.*, vol. 96, no. 3, 2024. DOI: 10.1007/s00020-024-02771-w.

[31] E. Berge, S. M. Berge, and F. Luef, "The affine Wigner distribution," *Appl. Comput. Harmon. Anal.*, vol. 56, pp. 150–175, 2022. DOI: 10.1016/j.acha.2021.08.006.

[32] E. Berge, S. M. Berge, F. Luef, and E. Skrettingland, "Affine quantum harmonic analysis," *J. Funct. Anal.*, vol. 282, no. 4, p. 109 327, 2022. DOI: 10.1016/j.jfa.2021.109327.

[33] S. M. Berge and S. Halvdansson, *Weyl quantization of exponential Lie groups for qquare integrable representations*, 2025. arXiv: 2502.18034 [math.FA].

[34] P. Bieliavsky, V. Gayral, S. Neshveyev, and L. Tuset, "Quantization of subgroups of the affine group," *J. Funct. Anal.*, vol. 280, no. 4, p. 108 844, 2021. DOI: 10.1016/j.jfa.2020.108844.

[35] I. Brons, R. Houben, and W. A. Dreschler, "Perceptual effects of noise reduction by time-frequency masking of noisy speech," *J. Acoust. Soc. Am.*, vol. 132, no. 4, pp. 2690–2699, 2012. DOI: 10.1121/1.4747006.

[36] L. Chen and J. Zhao, "Weyl transform and generalized spectrogram associated with quaternion heisenberg group," *Bull. Sci. Math.*, vol. 136, no. 2, pp. 127–143, 2012. DOI: 10.1016/j.bulsci.2011.12.002.

[37] S. Chen *et al.*, *BEATs: Audio Pre-Training with Acoustic Tokenizers*, 2022. arXiv: 2212.09058 [eess.AS].

[38] T.-S. Chi, C.-W. Huang, and W.-S. Chou, "A frequency bin-wise nonlinear masking algorithm in convolutive mixtures for speech segregation," *J. Acoust. Soc. Am.*, vol. 131, no. 5, EL361–EL367, 2012. DOI: 10.1121/1.3697530.

[39] L. Cohen, "Time-frequency distributions-a review," *Proc. IEEE.*, vol. 77, no. 7, pp. 941–981, 1989. DOI: 10.1109/5.30749.

[40] J. Conway, *A course in operator theory*. Providence, R.I: American Mathematical Society, 2000, ISBN: 978-0-8218-2065-0.

[41] E. Cordero and K. Grochenig, "Symbolic Calculus and Fredholm Property for Localization Operators," *J. Fourier Anal. Appl.*, vol. 12, no. 4, pp. 371–392, 2006. DOI: 10.1007/s00041-005-5077-7.

[42] E. Cordero and K. Gröchenig, "Time–frequency analysis of localization operators," *J. Funct. Anal.*, vol. 205, no. 1, pp. 107–131, 2003. DOI: 10.1016/s0022-1236(03)00166-6.

[43] E. Cordero and K. Gröchenig, "On the product of localization operators," *Modern Trends in Pseudo-Differential Operators*, Birkhäuser Basel, pp. 279–295. DOI: 10.1007/978-3-7643-8116-5_16.

[44] E. Cordero and L. Rodino, "Wick calculus: A time-frequency approach," *Osaka Journal of Mathematics*, vol. 42, no. 1, pp. 43–63, 2005. DOI: ojm/1153494314.

[45] E. Cordero, L. Rodino, and K. Gröchenig, "Localization operators and time-frequency analysis," *Harmonic, Wavelet and P-Adic Analysis*, World Scientific, 2007, pp. 83–110. DOI: 10.1142/9789812770707_0005.

[46] S. Dahlke, G. Kutyniok, P. Maass, C. Sagiv, H.-G. Stark, and G. Teschke, "The Uncertainty Principle Associated with the Continuous Shearlet Transform.," *Int. J. Wavelets Multiresolut. Inf. Process.*, vol. 06, no. 02, pp. 157–181, 2008. DOI: 10.1142/s021969130800229x.

[47] S. Dahlke, G. Kutyniok, G. Steidl, and G. Teschke, "Shearlet coorbit spaces and associated Banach frames," *Appl. Comput. Harmon. Anal.*, vol. 27, no. 2, pp. 195–214, 2009. DOI: 10.1016/j.acha.2009.02.004.

[48] S. Dahlke, G. Steidl, and G. Teschke, "The continuous shearlet transform in arbitrary space dimensions," *J. Fourier Anal. Appl.*, vol. 16, no. 3, pp. 340–364, 2009. DOI: 10.1007/s00041-009-9107-8.

[49] A. Dasgupta and S. K. Nayak, "Pseudo-differential operators, Wigner transform and Weyl transform on the Similitude group, $\mathbb{SIM}(2)$," *Bull. Sci. Math.*, vol. 174, p. 103 087, 2022. DOI: 10.1016/j.bulsci.2021.103087.

[50] A. Dasgupta and S. K. Nayak, "Pseudo-differential operators, Wigner transform and Weyl transform on the affine Poincaré group," *Bull. Sci. Math.*, vol. 184, p. 103 255, 2023. DOI: 10.1016/j.bulsci.2023.103255.

[51] I. Daubechies, "Time-frequency localization operators: A geometric phase space approach," *IEEE Trans. Inform. Theory*, vol. 34, no. 4, pp. 605–612, 1988. DOI: 10.1109/18.9761.

[52] I. Daubechies, "The wavelet transform, time-frequency localization and signal analysis," *IEEE Transactions on Information Theory*, vol. 36, no. 5, pp. 961–1005, 1990. DOI: 10.1109/18.57199.

[53] I. Daubechies, *Ten Lectures on Wavelets*. Society for Industrial and Applied Mathematics, 1992. DOI: 10.1137/1.9781611970104.

[54] P. Deift, "Universality for mathematical and physical systems," 25th International Congress of Mathematicians, ICM 2006, 2006, pp. 125–152.

[55] F. W. J. Olver *et al.*, *NIST Digital Library of Mathematical Functions*, Release 1.2.3 of 2024-12-15. [Online]. Available: https://dlmf.nist.gov/.

[56] L. Dong, S. Xu, and B. Xu, "Speech-transformer: A no-recurrence sequence-to-sequence model for speech recognition," *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2018. DOI: 10.1109/icassp.2018.8462506.

[57] M. Dörfler, "Gabor analysis for a class of signals called music," PhD thesis, University of Vienna, 2002.

[58] M. Dörfler, T. Grill, R. Bammer, and A. Flexer, "Basic filters for convolutional neural networks applied to music: Training or design?" *Neural Comput. Appl.*, vol. 32, no. 4, pp. 941–954, 2018. DOI: 10.1007/s00521-018-3704-x.

[59] M. Dörfler, F. Luef, H. McNulty, and E. Skrettingland, "Time-frequency analysis and coorbit spaces of operators," *J. Math. Anal. Appl.*, vol. 534, no. 2, p. 128 058, 2024. DOI: 10.1016/j.jmaa.2023.128058.

[60] M. Dörfler, F. Luef, and E. Skrettingland, "Local structure and effective dimensionality of time series data sets," *Appl. Comput. Harmon. Anal.*, vol. 73, p. 101 692, 2024. DOI: 10.1016/j.acha.2024.101692.

[61] M. Dörfler and B. Torrésani, "Spreading function representation of operators and gabor multiplier approximation," *SAMPTA'07, Sampling Theory and Applications*, https://hal.archives-ouvertes.fr/hal-00146274, Thessaloniki, Greece, 2007.

[62] M. Dörfler and B. Torrésani, "Representation of operators in the time-frequency domain and generalized Gabor multipliers," *J. Fourier Anal. Appl.*, vol. 16, no. 2, pp. 261–293, 2009. DOI: 10.1007/s00041-009-9085-x.

[63] J. Du and M. W. Wong, "A trace formula for Weyl transforms," *Approx. Theory Appl.*, vol. 16, no. 1, pp. 41–45, 2000. DOI: 10.1007/bf02845227.

[64] M. Duflo and C. C. Moore, "On the regular representation of a nonunimodular locally compact group," *J. Funct. Anal.*, vol. 21, no. 2, pp. 209–243, 1976. DOI: 10.1016/0022-1236(76)90079-3.

[65] Encyclopedia of Mathematics, *Pfaffian*, https://encyclopediaofmath.org/index.php?title=Pfaffian, Accessed: 2025-02-18.

[66] L. C. Evans and R. F. Gariepy, *Measure theory and fine properties of functions* (Studies in Advanced Mathematics). Boca Raton, FL: CRC Press, 1991. DOI: 10.1201/b18333.

[67] H. G. Feichtinger, "A characterization of Wiener's algebra on locally compact groups," *Arch. Math. (Basel)*, vol. 29, pp. 136–140, 1977. DOI: 10.1007/BF01220386.

[68] H. G. Feichtinger, "On a new Segal algebra," *Monatsh. Math.*, vol. 92, no. 4, pp. 269–289, 1981. DOI: 10.1007/bf01320058.

[69] H. G. Feichtinger, "Banach convolution algebras of Wiener type," *Proc. Conf. on Functions, Series, Operators, Budapest 1980*, ser. Colloq. Math. Soc. Janos Bolyai, Eds. B. Sz.-Nagy and J. Szabados., vol. 35, Amsterdam: North-Holland, 1983, pp. 509–524.

[70] H. G. Feichtinger, "Spline-type spaces in Gabor analysis," *Wavelet Analysis: Twenty Years Developments Proceedings of the International Conference of Computational Harmonic Analysis, Hong Kong, China, June 4–8, 2001*, ser. Ser. Anal. D. X. Zhou, Ed., vol. 1, World Sci.Pub., 2002, pp. 100–122. DOI: 10.1142/9789812776679_0006.

[71] H. G. Feichtinger, "Gabor multipliers with varying lattices," *Proc. SPIE Conf.*, M. A. Unser, A. Aldroubi, and A. F. Laine, Eds., NuHAG, 2003. DOI: 10.1117/12.507648.

[72] H. G. Feichtinger, "A novel mathematical approach to the theory of translation invariant linear systems," *Recent Applications of Harmonic Analysis to Function Spaces, Differential Equations, and Data Science: Novel Methods in Harmonic Analysis, Volume 2*. Cham: Springer International Publishing, 2017, pp. 483–516. DOI: 10.1007/978-3-319-55556-0_3.

[73] H. G. Feichtinger, "Homogeneous Banach Spaces as Banach Convolution Modules over $M(G)$," *Mathematics*, vol. 10, no. 3, p. 364, 2022. DOI: 10.3390/math10030364.

[74] H. G. Feichtinger and K. Nowak, "A Szegő-type theorem for Gabor-Toeplitz localization operators," *Michigan Math. J.*, vol. 49, no. 1, 2001. DOI: 10.1307/mmj/1008719032.

[75] H. G. Feichtinger and K. Nowak, "A first survey of Gabor multipliers," *Advances in Gabor Analysis*, H. G. Feichtinger and T. Strohmer, Eds. Boston, MA: Birkhäuser Boston, 2003, pp. 99–128. DOI: 10.1007/978-1-4612-0133-5_5.

[76] H. G. Feichtinger, K. Nowak, and M. Pap, "Spectral properties of Toeplitz operators acting on Gabor type reproducing kernel Hilbert spaces," *Mathematics Without Boundaries*. Springer New York, 2014, pp. 163–180. DOI: 10.1007/978-1-4939-1106-6_6.

[77] H. G. Feichtinger, K. Nowak, and M. Pap, "Asymptotic boundary forms for tight Gabor frames and lattice localization domains," *J. Appl. Math. Phys.*, vol. 03, no. 10, pp. 1316–1342, 2015. DOI: 10.4236/jamp.2015.310160.

[78] H. G. Feichtinger and K. Gröchenig, "Banach spaces related to integrable group representations and their atomic decompositions, I," *J. Funct. Anal.*, vol. 86, no. 2, pp. 307–340, 1989. DOI: 10.1016/0022-1236(89)90055-4.

[79] H. G. Feichtinger and K. H. Gröchenig, "Banach spaces related to integrable group representations and their atomic decompositions. part II," *Monatsh. Math.*, vol. 108, no. 2-3, pp. 129–148, 1989. DOI: 10.1007/bf01308667.

[80] H. G. Feichtinger and K. Gröchenig, "A unified approach to atomic decompositions via integrable group representations," *Function Spaces and Applications*, Springer Berlin Heidelberg, 1988, pp. 52–73. DOI: 10.1007/bfb0078863.

[81] H. G. Feichtinger, S. Halvdansson, and F. Luef, "Measure-operator convolutions and applications to mixed-state Gabor multipliers," *Sampl. Theory Signal Process. Data Anal.*, vol. 22, no. 2, 2024. DOI: 10.1007/s43670-024-00090-0.

[82] H. G. Feichtinger and N. Kaiblinger, "Quasi-interpolation in the Fourier algebra," *J. Approx. Theory*, vol. 144, no. 1, pp. 103–118, 2007. DOI: 10.1016/j.jat.2006.05.001.

[83] C. Fernández and A. Galbis, "Compactness of time-frequency localization operators on $L^2(\mathbb{R}^d)$," *J. Funct. Anal.*, vol. 233, no. 2, pp. 335–350, 2006. DOI: 10.1016/j.jfa.2005.08.008.

[84] G. B. Folland, *Harmonic Analysis in Phase Space*. Princeton University Press, 1989. DOI: 10.1515/9781400882427.

[85] G. B. Folland, *A Course in Abstract Harmonic Analysis*, Second. Chapman and Hall/CRC, 2015. DOI: 10.1201/b19172.

[86] P. J. Forrester and G. Honner, "Exact statistical properties of the zeros of complex random polynomials," *J. Phys. A*, vol. 32, no. 16, pp. 2961–2981, 1999. DOI: 10.1088/0305-4470/32/16/006.

[87] H. Führ, *Abstract Harmonic Analysis of Continuous Wavelet Transforms*. Springer-Verlag Berlin Heidelberg, 2005. DOI: 10.1007/b104912.

[88] H. Führ and K. Gröchenig, "Sampling theorems on locally compact groups from oscillation estimates," *Math. Z.*, vol. 255, no. 1, pp. 177–194, 2006. DOI: 10.1007/s00209-006-0019-x.

[89] H. Führ and V. Oussa, "Phase retrieval for nilpotent groups," *J. Fourier Anal. Appl.*, vol. 29, no. 4, 2023. DOI: 10.1007/s00041-023-10031-5.

[90] R. Fulsche, "Correspondence theory on p-Fock spaces with applications to Toeplitz algebras," *J. Funct. Anal.*, vol. 279, no. 7, p. 108 661, 2020. DOI: 10.1016/j.jfa.2020.108661.

[91] R. Fulsche and N. Galke, "Quantum harmonic analysis on locally compact abelian groups," *J. Fourier Anal. Appl.*, vol. 31, no. 1, 2025. DOI: 10.1007/s00041-024-10140-9.

[92] R. Fulsche and R. Hagger, "Quantum harmonic analysis for polyanalytic Fock spaces," *J. Fourier Anal. Appl.*, vol. 30, no. 6, 2024. DOI: 10.1007/s00041-024-10124-9.

[93] D. Gabor, "Theory of communication," *J. Inst. Electr. Eng.*, vol. 94, no. 73, pp. 58–58, 1947. DOI: 10.1049/ji-1.1947.0015.

[94] V. Gayral, J. M. Gracia-Bondía, and J. C. Várilly, "Fourier analysis on the affine group, quantization and noncompact Connes geometries," *J. Noncommut. Geom.*, vol. 2, pp. 215–261, 2008. DOI: 10.4171/JNCG/20.

[95] J. P. Gazeau and C. Habonimana, "Signal analysis and quantum formalism: Quantizations with no Planck constant," *Landscapes of Time-Frequency Analysis*, Springer International Publishing, 2020, pp. 135–155. DOI: 10.1007/978-3-030-56005-8_8.

[96] J. P. Gazeau and R. Murenzi, "Covariant affine integral quantization(s)," *J. Math. Phys.*, vol. 57, no. 5, p. 052 102, 2016. DOI: 10.1063/1.4948397.

[97] J.-P. Gazeau, T. Koide, and R. Murenzi, "2-D covariant affine integral quantization(s)," *Adv. Oper. Theory*, vol. 5, no. 3, pp. 901–935, 2020. DOI: 10.1007/s43036-020-00039-9.

[98] M. Ghandehari and K. F. Taylor, "Images of the continuous wavelet transform," *Operator Methods in Wavelets, Tilings, and Frames*, American Mathematical Society, 2014, pp. 55–65. DOI: 10.1090/conm/626/12508.

[99] S. Ghobber, "Some results on wavelet scalograms," *Int. J. Wavelets Multiresolut. Inf. Process.*, vol. 15, no. 03, p. 1 750 019, 2017. DOI: 10.1142/s0219691317500199.

[100] S. Ghosh and R. K. Srivastava, "Unbounded Weyl transform on the Euclidean motion group and Heisenberg motion group," *J. Operator Theory*, vol. 90, no. 2, pp. 605–623, 2023.

[101] S. Ghosh, M. Lin, and D. Sun, "Signal analysis via the stochastic geometry of spectrogram level sets," *IEEE Trans. Signal Process.*, vol. 70, pp. 1104–1117, 2022. DOI: 10.1109/tsp.2022.3153596.

[102] B. D. Giorgi, M. Levy, and R. Sharp, "Mel spectrogram inversion with stable pitch," *International Society for Music Information Retrieval Conference*, 2022. arXiv: 2208.12782 [cs.SD].

[103] Y. Gong, Y.-A. Chung, and J. Glass, "AST: Audio Spectrogram Transformer," *Interspeech 2021*, ISCA, 2021. DOI: 10.21437/interspeech.2021-698.

[104] M. A. de Gosson, *Symplectic Methods in Harmonic Analysis and in Mathematical Physics*. Springer Basel, 2011. DOI: 10.1007/978-3-7643-9992-4.

[105] M. Goto, "On a class of type I solvable Lie groups I," *J. Math. Soc. Japan*, vol. 30, no. 4, pp. 771–778, 1978. DOI: 10.2969/jmsj/03040771.

[106] D. Griffin and J. Lim, "Signal estimation from modified short-time Fourier transform," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 32, no. 2, pp. 236–243, 1984. DOI: 10.1109/tassp.1984.1164317.

[107] K. Gröchenig, *Foundations of Time-Frequency Analysis*. Birkhäuser Boston, 2001. DOI: 10.1007/978-1-4612-0003-1.

[108] K. Gröchenig, P. Jaming, and E. Malinnikova, "Zeros of the Wigner distribution and the short-time Fourier transform," *Rev. Mat. Complut.*, vol. 33, no. 3, pp. 723–744, 2019. DOI: 10.1007/s13163-019-00335-w.

[109] K. Gröchenig and J. Toft, "Isomorphism properties of Toeplitz operators and pseudo-differential operators between modulation spaces," *J. Anal. Math.*, vol. 114, no. 1, pp. 255–283, 2011. DOI: 10.1007/s11854-011-0017-8.

[110] M. Guillaume and J. Gautier, "On sampling determinantal point processes," PhD thesis, Ecole Centrale de Lille, 2020. [Online]. Available: https://theses.hal.science/tel-03132893.

[111] K. Guo, G. Kutyniok, and D. Labate, "Sparse multidimensional representations using anisotropic dilation and shear operators," *Wavelets and Splines*, (Athens, GA), Nashboro Press, 2006, pp. 189–201.

[112] W. Guo, D. Fan, H. Wu, and G. Zhao, "Sharpness of complex interpolation on α-modulation spaces," *J. Fourier Anal. Appl.*, vol. 22, no. 2, pp. 427–461, 2015. DOI: 10.1007/s00041-015-9424-z.

[113] A. Haimi, G. Koliander, and J. L. Romero, "Zeros of Gaussian Weyl–Heisenberg functions and hyperuniformity of charge," *J. Stat. Phys.*, vol. 187, no. 3, 2022. DOI: 10.1007/s10955-022-02917-3.

[114] B. C. Hall, *Quantum Theory for Mathematicians*. Springer-Verlag New York, 2013. DOI: 10.1007/978-1-4614-7116-5.

[115] S. Halvdansson, "Quantum harmonic analysis on locally compact groups," *J. Funct. Anal.*, vol. 285, no. 8, p. 110 096, 2023. DOI: 10.1016/j.jfa.2023.110096.

[116] S. Halvdansson, *On a time-frequency blurring operator with applications in data augmentation*, 2024. arXiv: 2405.12899 [math.FA].

[117] S. Halvdansson, *Empirical plunge profiles of time-frequency localization operators*, 2025. arXiv: 2502.11805 [math.FA].

[118] S. Halvdansson, "Five ways to recover the symbol of a non-binary localization operator," *J. Pseudo-Differ. Oper. Appl.*, vol. 16, no. 1, 2025. DOI: 10.1007/s11868-025-00677-8.

[119] S. Halvdansson, "On accumulated spectrograms for Gabor frames," *J. Math. Anal. Appl.*, vol. 543, no. 2, p. 129 044, 2025. DOI: 10.1016/j.jmaa.2024.129044.

[120] O. Hazrati, J. Lee, and P. C. Loizou, "Blind binary masking for reverberation suppression in cochlear implants," *J. Acoust. Soc. Am.*, vol. 133, no. 3, pp. 1607–1614, 2013. DOI: 10.1121/1.4789891.

[121] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, 2016. DOI: 10.1109/cvpr.2016.90.

[122] H. Hedenmalm and A. Wennman, "Planar orthogonal polynomials and boundary universality in the random normal matrix model," *Acta Math.*, vol. 227, no. 2, pp. 309–406, 2021. DOI: 10.4310/acta.2021.v227.n2.a3.

[123] C. Heil, "An introduction to weighted Wiener amalgams," *Wavelets and their Applications (Chennai, January 2002)*, M. Krishna, R. Radha, and S. Thangavelu, Eds., New Delhi: Allied Publishers, 2003, pp. 183–216.

[124] C. E. Heil, J. Ramanathan, and P. N. Topiwala, "Asymptotic singular value decay of time-frequency localization operators," *Wavelet Applications in Signal and Image Processing II*, A. F. Laine and M. A. Unser, Eds., SPIE, 1994. DOI: 10.1117/12.188774.

[125] F. Hlawatsch and G. Boudreaux-Bartels, "Linear and quadratic time-frequency signal representations," *IEEE Signal Process. Mag.*, vol. 9, no. 2, pp. 21–67, 1992. DOI: 10.1109/79.127284.

[126] L. Hörmander, "The Weyl calculus of pseudo-differential operators," *Comm. Pure Appl. Math.*, vol. 32, no. 3, pp. 359–443, 1979. DOI: 10.1002/cpa.3160320304.

[127] J. Hough, M. Krishnapur, Y. Peres, and B. Virág, *Zeros of Gaussian analytic functions and determinantal point processes*. Providence, RI: American Mathematical Society, 2009. DOI: 10.1090/ULECT/051.

[128] R. L. Hudson, "When is the Wigner quasi-probability density non-negative?" *Rep. Math. Phys.*, vol. 6, no. 2, pp. 249–252, 1974. DOI: 10.1016/0034-4877(74)90007-X.

[129] T. Hytönen, J. van Neerven, M. Veraar, and L. Weis, *Analysis in Banach Spaces*. Springer International Publishing, 2016. DOI: 10.1007/978-3-319-48520-1.

[130] I. Jordal, *Audiomentations*, https://github.com/iver56/audiomentations, 2023.

[131] N. Kaiblinger, "Approximation of the Fourier transform and the dual Gabor window," *J. Fourier Anal. Appl.*, vol. 11, no. 1, pp. 25–42, 2005. DOI: 10.1007/s00041-004-3070-1.

[132] M. Katori and T. Shirai, "Partial isometries, duality, and determinantal point processes," *Random Matrices Theory Appl.*, vol. 11, no. 03, 2021. DOI: 10.1142/s2010326322500253.

[133] E. Kharitonov *et al.*, "Data augmenting contrastive learning of speech representations in the time domain," 2020. arXiv: 2007.00991 [eess.AS].

[134] A. A. Kirillov, *Lectures on the Orbit Method*. American Mathematical Society, 2004. DOI: 10.1090/gsm/064.

[135] J. Kiukas, "Covariant observables on a nonunimodular group," *J. Math. Anal. Appl.*, vol. 324, no. 1, pp. 491–503, 2006. DOI: 10.1016/j.jmaa.2005.12.016.

[136] J. Kiukas, P. Lahti, J. Schultz, and R. F. Werner, "Characterization of informational completeness for covariant phase space observables," *J. Math. Phys.*, vol. 53, no. 10, p. 102 103, 2012. DOI: 10.1063/1.4754278.

[137] J. Kiukas, P. Lahti, and K. Ylinen, "Normal covariant quantization maps," *J. Math. Anal. Appl.*, vol. 319, no. 2, pp. 783–801, 2006. DOI: 10.1016/j.jmaa.2005.06.069.

[138] J. R. Klauder and B.-S. K. Skagerstam, "Extension of Berezin-Lieb inequalities," *Excursions in Harmonic Analysis, Volume 2*, Birkhäuser Boston, 2012, pp. 251–266. DOI: 10.1007/978-0-8176-8379-5_13.

[139] A. Klein and B. Russo, "Sharp inequalities for Weyl operators and Heisenberg groups," *Math. Ann.*, vol. 235, no. 2, pp. 175–194, 1978. DOI: 10.1007/bf01405012.

[140] A. Kolesnikov *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," *International Conference on Learning Representations*, 2021. arXiv: 2010.11929 [cs.CV].

[141] A. M. Krémé, V. Emiya, C. Chaux, and B. Torrésani, "Time-Frequency Fading Algorithms Based on Gabor Multipliers," *IEEE J. Sel. Topics Signal Process.*, vol. 15, no. 1, pp. 65–77, 2021. DOI: 10.1109/JSTSP.2020.3045938.

[142] A. Kulikov, "Exponential lower bound for the eigenvalues of the time-frequency localization operator before the plunge region," *Appl. Comput. Harmon. Anal.*, vol. 71, p. 101 639, 2024. DOI: 10.1016/j.acha.2024.101639.

[143] V. Kumar and S. S. Mondal, "Trace class and Hilbert-Schmidt pseudo differential operators on step two nilpotent Lie groups," *Bull. Sci. Math.*, vol. 171, p. 103 015, 2021. DOI: 10.1016/j.bulsci.2021.103015.

[144] G. Kutyniok and D. Labate, "Introduction to Shearlets," *Shearlets*, Birkhäuser Boston, 2012, pp. 1–38. DOI: 10.1007/978-0-8176-8316-0_1.

[145] H. J. Landau, "On Szegö's eigenvalue distribution theorem and non-Hermitian kernels," *J. Anal. Math.*, vol. 28, no. 1, pp. 335–357, 1975. DOI: 10.1007/bf02786820.

[146] H. J. Landau and H. O. Pollak, "Prolate Spheroidal Wave Functions, Fourier Analysis and Uncertainty - II," *Bell System Tech. J.*, vol. 40, no. 1, pp. 65–84, 1961. DOI: 10.1002/j.1538-7305.1961.tb03977.x.

[147] H. J. Landau and H. O. Pollak, "Prolate Spheroidal Wave Functions, Fourier Analysis and Uncertainty-III: The Dimension of the Space of Essentially Time- and Band-Limited Signals," *Bell System Tech. J.*, vol. 41, no. 4, pp. 1295–1336, 1962. DOI: 10.1002/j.1538-7305.1962.tb03279.x.

[148] P. D. Lax, *Functional Analysis*. Wiley-Interscience, 2002, p. 608, ISBN: 978-0471556046.

[149] H. Leptin, "Ideal theory in group algebras of locally compact groups," *Invent. Math.*, vol. 31, no. 3, pp. 259–278, 1976. DOI: 10.1007/bf01403147.

[150] J. Li and M. W. Wong, "Localization operators for ridgelet transforms," *Mathematical Modelling of Natural Phenomena*, vol. 9, no. 5, D. Damanik, M. Ruzhansky, V. Vougalter, and M. Wong, Eds., pp. 194–203, 2014. DOI: 10.1051/mmnp/20149513.

[151] E. H. Lieb and M. Loss, *Analysis* (Graduate studies in mathematics). Providence, RI: American Mathematical Society, 2001. DOI: 10.1090/gsm/014.

[152] F. Luef and E. Skrettingland, "Convolutions for Berezin quantization and Berezin-Lieb inequalities," *J. Math. Phys.*, vol. 59, no. 2, p. 023 502, 2018. DOI: 10.1063/1.5023241.

[153] F. Luef and E. Skrettingland, "Convolutions for localization operators," *J. Math. Pures Appl.*, vol. 118, pp. 288–316, 2018. DOI: 10.1016/j.matpur.2017.12.004.

[154] F. Luef and E. Skrettingland, "Mixed-state localization operators: Cohen's class and trace class operators," *J. Fourier Anal. Appl.*, vol. 25, no. 4, pp. 2064–2108, 2019. DOI: 10.1007/s00041-019-09663-3.

[155] F. Luef and E. Skrettingland, "On accumulated Cohen's class distributions and mixed-state localization operators," *Constr. Approx.*, vol. 52, no. 1, pp. 31–64, 2019. DOI: 10.1007/s00365-019-09465-2.

[156] F. Luef and E. Skrettingland, "A Wiener Tauberian theorem for operators and functions," *J. Funct. Anal.*, vol. 280, no. 6, p. 108 883, 2021. DOI: 10.1016/j.jfa.2020.108883.

[157] E. Ma, *nlpaug*, https://github.com/makcedward/nlpaug, 2022.

[158] W. Mack and E. A. P. Habets, "Deep filtering: Signal extraction and reconstruction using complex time-frequency filters," *IEEE Signal Process. Lett.*, vol. 27, pp. 61–65, 2020. DOI: 10.1109/lsp.2019.2955818.

[159] S. Mallat and I. Waldspurger, "Phase retrieval for the Cauchy wavelet transform," *J. Fourier Anal. Appl.*, vol. 21, no. 6, pp. 1251–1309, 2015. DOI: 10.1007/s00041-015-9403-4.

[160] M. Mantoiu and M. Ruzhansky, "Pseudo-differential operators, Wigner transform and Weyl systems on type I locally compact groups," *Doc. Math.*, vol. 22, pp. 1539–1592, 2017. DOI: 10.4171/dm/604.

[161] M. Măntoiu and M. Sandoval, "Pseudo-differential operators associated to general type I locally compact groups," *Analysis and Partial Differential Equations: Perspectives from Developing Countries*, Springer International Publishing, 2019, pp. 172–190. DOI: 10.1007/978-3-030-05657-5_11.

[162] F. Marceca and J. L. Romero, "Spectral deviation of concentration operators for the short-time fourier transform," *Studia Math.*, vol. 270, no. 2, pp. 145–173, 2023. DOI: 10.4064/sm220214-17-10.

[163] F. Marceca, J. L. Romero, and M. Speckbacher, "Eigenvalue estimates for Fourier concentration operators on two domains," *Arch. Rat. Mechan. Anal.*, vol. 248, no. 3, 2024. DOI: 10.1007/s00205-024-01979-9.

[164] F. D. Mari, H. G. Feichtinger, and K. Nowak, "Uniform eigenvalue estimates for time-frequency localization operators," *J. Lond. Math. Soc.*, vol. 65, no. 03, pp. 720–732, 2002. DOI: 10.1112/s0024610702003101.

[165] B. McFee, E. Humphrey, and J. Bello, "A software framework for musical data augmentation," *Proceedings of the 16th International Society for Music Information Retrieval Conference, ISMIR 2015*, M. Muller and F. Wiering, Eds., ser. Proceedings of the 16th International Society for Music Information Retrieval Conference, ISMIR 2015, International Society for Music Information Retrieval, 2015, pp. 248–254. DOI: 10.5281/zenodo.1418365.

[166] A. Olivero, B. Torrésani, P. Depalle, and R. Kronland-Martinet, "Sound morphing strategies based on alterations of time-frequency representations by Gabor multipliers," *Proceedings of the AES International Conference*, 2012. [Online]. Available: https://hal.science/hal-00682959.

[167] A. Olivero, B. Torrésani, and R. Kronland-Martinet, "A new method for Gabor multipliers estimation: Application to sound morphing," *2010 18th European Signal Processing Conference*, 2010, pp. 507–511. [Online]. Available: https://hal.science/hal-00493711/en/.

[168] E. Onuonga, *Sigment*, https://github.com/eonu/sigment, 2020.

[169] Z. Papakipos and J. Bitton, *AugLy: Data Augmentations for Robustness*, 2022. arXiv: 2201.06494 [cs.AI].

[170] A. Papandreou-Suppappola, Ed., *Applications in Time-Frequency Signal Processing*. CRC Press, 2018. DOI: 10.1201/9781315220017.

[171] D. S. Park *et al.*, "SpecAugment: A simple data augmentation method for automatic speech recognition," *Interspeech 2019*, ISCA, 2019. DOI: 10.21437/interspeech.2019-2680.

[172] L. Peng and J. Zhao, "Weyl transforms associated with the Heisenberg group," *Bull. Sci. Math.*, vol. 132, no. 1, pp. 78–86, 2008. DOI: 10.1016/j.bulsci.2007.07.002.

[173] J. C. T. Pool, "Mathematical aspects of the Weyl correspondence," *J. Math. Phys.*, vol. 7, no. 1, pp. 66–76, 1966. DOI: 10.1063/1.1704817.

[174] Z. Prusa and N. Holighaus, "Phase vocoder done right," *2017 25th European Signal Processing Conference (EUSIPCO)*, IEEE, 2017. DOI: 10.23919/eusipco.2017.8081353.

[175] Z. Průša, P. L. Søndergaard, N. Holighaus, C. Wiesmeyr, and P. Balazs, "The Large Time-Frequency Analysis Toolbox 2.0," *Sound, Music, and Motion*, ser. LNCS, Springer International Publishing, 2014, pp. 419–442. DOI: 10.1007/978-3-319-12976-1_25.

[176] L. T. Rachdi and K. Trimèche, "Weyl transforms associated with the spherical mean operator," *Anal. Appl. (Singap.)*, vol. 01, no. 02, pp. 141–164, 2003. DOI: 10.1142/s0219530503000156.

[177] A. Radford, J. W. Kim, T. Xu, G. Brockman, C. McLeavey, and I. Sutskever, "Robust speech recognition via large-scale weak supervision," 2022. arXiv: 2212.04356 [eess.AS].

[178] S. Rajbamshi, G. Tauböck, P. Balazs, and L. D. Abreu, "Random Gabor Multipliers for Compressive Sensing: A Simulation Study," *2019 27th European Signal Processing Conference (EUSIPCO)*, 2019, pp. 1–5. DOI: 10.23919/EUSIPCO.2019.8903092.

[179] A. Raju, S. Panchapagesan, X. Liu, A. Mandal, and N. Strom, *Data augmentation for robust keyword spotting under playback interference*, 2018. arXiv: 1808.00563 [cs.CL].

[180] J. Ramanathan and P. Topiwala, "Time–frequency localization and the spectrogram," *Appl. Comput. Harmon. Anal.*, vol. 1, no. 2, pp. 209–215, 1994. DOI: 10.1006/acha.1994.1008.

[181] J. L. Romero, "Characterization of coorbit spaces with phase-space covers," *J. Funct. Anal.*, vol. 262, no. 1, pp. 59–93, 2012. DOI: 10.1016/j.jfa.2011.09.005.

[182] J. L. Romero and M. Speckbacher, "Estimation of binary time-frequency masks from ambient noise," *SIAM J. Math. Anal.*, vol. 56, no. 3, pp. 3559–3587, 2024. DOI: 10.1137/22m149805x.

[183] C. Rommel, J. Paillard, T. Moreau, and A. Gramfort, "Data augmentation for learning predictive models on EEG: A systematic comparison," *J. Neural Eng.*, vol. 19, no. 6, p. 066020, 2022. DOI: 10.1088/1741-2552/aca220.

[184] M. Rudelson and R. Vershynin, "Hanson-Wright inequality and sub-gaussian concentration," *Electron. Commun. Probab.*, vol. 18, 2013. DOI: 10.1214/ecp.v18-2865.

[185] W. Rudin, *Functional Analysis, 2nd Edition*. New York: McGraw-Hill, 1991, ISBN: 978-0070542365.

[186] C. Sagiv, N. A. Sochen, and Y. Y. Zeevi, "The uncertainty principle: Group theoretic approach, possible minimizers and scale-space properties," *J. Math. Imaging Vision*, vol. 26, no. 1-2, pp. 149–166, 2006. DOI: 10.1007/s10851-006-8301-4.

[187] H. Schroter, A. N. Escalante-B, T. Rosenkranz, and A. Maier, "Deepfilternet: A low complexity speech enhancement framework for full-band audio based on deep filtering," *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2022. DOI: 10.1109/icassp43922.2022.9747055.

[188] H. S. Shapiro, *Topics in Approximation Theory* (Lecture Notes in Mathematics). Springer-Verlag, 1971, vol. 187, pp. viii+275. DOI: 10.1007/bfb0058976.

[189] B. Simon, "The Weyl transform and L^p functions on phase space," *Proc. Amer. Math. Soc.*, vol. 116, no. 4, pp. 1045–1047, 1992. DOI: 10.1090/s0002-9939-1992-1100663-7.

[190] B. Simon, *Trace ideals and their applications*. Providence, R.I: American Mathematical Society, 2005. DOI: 10.1090/surv/120.

[191] E. Skrettingland, "Convolutions for localization operators," M.S. thesis, NTNU, 2017. [Online]. Available: https://folk.ntnu.no/franzl/supervision.html.

[192] E. Skrettingland, "Quantum harmonic analysis on lattices and Gabor multipliers," *J. Fourier Anal. Appl.*, vol. 26, no. 3, 2020. DOI: 10.1007/s00041-020-09759-1.

[193] E. Skrettingland, "On Gabor g-frames and Fourier series of operators," *Studia Math.*, vol. 259, no. 1, pp. 25–78, 2021. DOI: 10.4064/sm191115-24-9.

[194] E. Skrettingland, "Equivalent norms for modulation spaces from positive Cohen's class distributions," *J. Fourier Anal. Appl.*, vol. 28, no. 2, 2022. DOI: 10.1007/s00041-022-09930-w.

[195] D. Slepian, "Prolate Spheroidal Wave Functions, Fourier Analysis, and Uncertainty-V: The Discrete Case," *Bell System Tech. J.*, vol. 57, no. 5, pp. 1371–1430, 1978. DOI: 10.1002/j.1538-7305.1978.tb02104.x.

[196] D. Slepian and H. O. Pollak, "Prolate Spheroidal Wave Functions, Fourier Analysis and Uncertainty - I," *Bell System Tech. J.*, vol. 40, no. 1, pp. 43–63, 1961. DOI: 10.1002/j.1538-7305.1961.tb03976.x.

[197] D. Slepian, "Prolate Spheroidal Wave Functions, Fourier Analysis and Uncertainty - IV: Extensions to Many Dimensions; Generalized Prolate Spheroidal Functions," *Bell System Tech. J.*, vol. 43, no. 6, pp. 3009–3057, 1964. DOI: 10.1002/j.1538-7305.1964.tb01037.x.

[198] W. F. Stinespring, "A sufficient condition for an integral operator to have a trace.," *J. Reine Angew. Math.*, vol. 1958, no. 200, pp. 200–207, 1958. DOI: 10.1515/crll.1958.200.200.

[199] T. Strohmer, "Pseudodifferential operators and Banach algebras in mobile communications," *Appl. Comput. Harmon. Anal.*, vol. 20, no. 2, pp. 237–249, 2006. DOI: 10.1016/j.acha.2005.06.003.

[200] O. Takenouchi, "Sur les facteurs représentations d'un groupe de Lie resoluble de type E," *Math. J. Okayama Univ.*, vol. 7, pp. 151–161, 1957.

[201] S. Torquato, "Hyperuniform states of matter," *Phys. Rep*, vol. 745, pp. 1–95, 2018. DOI: 10.1016/j.physrep.2018.03.001.

[202] S. Torquato and F. H. Stillinger, "Local density fluctuations, hyperuniformity, and order metrics," *Phys. Rev. E*, vol. 68, no. 4, 2003. DOI: 10.1103/physreve.68.041113.

[203] A. Variar, *SABER - Semi-Supervised Audio Baseline for Easy Reproduction*, https://github.com/SABER-labs/SABER, 2021.

[204] R. Vygon and N. Mikhaylovskiy, "Learning efficient representations for keyword spotting with triplet loss," *Speech and Computer*, Springer International Publishing, 2021, pp. 773–785. DOI: 10.1007/978-3-030-87802-3_69.

[205] I. Waldspurger, "Phase retrieval for wavelet transforms," *IEEE Trans. Inf. Theory*, vol. 63, no. 5, pp. 2993–3009, 2017. DOI: 10.1109/TIT.2017.2672727.

[206] D. Wang, "On ideal binary mask as the computational goal of auditory scene analysis," *Speech Separation by Humans and Machines*, Kluwer Academic Publishers, pp. 181–197. DOI: 10.1007/0-387-22794-6_12.

[207] H. Wang, Y. Zou, and W. Wang, "SpecAugment++: A hidden space data augmentation method for acoustic scene classification," *Interspeech 2021*, ISCA, 2021. DOI: 10.21437/interspeech.2021-140.

[208] J. Wang, S. Kim, and Y. Lee, "Speech Augmentation Using Wavenet in Speech Recognition," *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019, pp. 6770–6774. DOI: 10.1109/ICASSP.2019.8683388.

[209] P. Warden, "Speech Commands: A Dataset for Limited-Vocabulary Speech Recognition," 2018. arXiv: 1804.03209 [cs.CL].

[210] S. Wei, S. Zou, F. Liao, and W. Lang, "A comparison on data augmentation methods based on deep learning for audio classification," *J. Phys.: Conf. Ser.*, vol. 1453, no. 1, p. 012 085, 2020. DOI: 10.1088/1742-6596/1453/1/012085.

[211] R. Werner, "Quantum harmonic analysis on phase space," *J. Math. Phys.*, vol. 25, no. 5, pp. 1404–1411, 1984. DOI: 10.1063/1.526310.

[212] H. Weyl, "Quantenmechanik und gruppentheorie," *Zeitschrift für Physik*, vol. 46, no. 1–2, pp. 1–46, 1927. DOI: 10.1007/BF02055756.

[213] H. Weyl, *The theory of groups and quantum mechanics* (Dover Books on Mathematics). Mineola, NY: Dover Publications, 1950.

[214] N. Wiener, "Tauberian Theorems," *The Annals of Mathematics*, vol. 33, no. 1, p. 1, 1932. DOI: 10.2307/1968102.

[215] E. P. Wigner, "On the quantum correction for thermodynamic equilibrium," *Phys. Rev.*, vol. 40, no. 5, pp. 749–759, 1932. DOI: 10.1103/PhysRev.40.749.

[216] *Wigner-Ville distribution and smoothed pseudo Wigner-Ville distribution*, https://se.mathworks.com/help/signal/ref/wvd.html, Accessed: 2023-09-12.

[217] D. S. Williamson, Y. Wang, and D. Wang, "Complex ratio masking for monaural speech separation," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 24, no. 3, pp. 483–492, 2016. DOI: 10.1109/taslp.2015.2512042.

[218] M. W. Wong, *Weyl Transforms* (Universitext). Springer-Verlag, 1998. DOI: 10.1007/b98973.

[219] M. W. Wong, "L^p boundedness of localization operators associated to left regular representations," *Proc. Amer. Math. Soc.*, vol. 130, no. 10, pp. 2911–2919, 2002. DOI: 10.1090/S0002-9939-02-06685-6.

[220] M. W. Wong, *Wavelet Transforms and Localization Operators*. Birkhäuser Basel, 2002. DOI: 10.1007/978-3-0348-8217-0.

[221] K. Wu *et al.*, "TinyViT: Fast pretraining distillation for small vision transformers," *Computer Vision – ECCV 2022*. Springer Nature Switzerland, 2022, pp. 68–85. DOI: 10.1007/978-3-031-19803-8_5.

[222] Y. Zhang *et al.*, *Pushing the limits of semi-supervised learning for automatic speech recognition*, 2020. arXiv: 2010.10504 [eess.AS].

[223] M. Zibulski and Y. Y. Zeevi, "Analysis of multiwindow Gabor-type schemes by frame methods," *Appl. Comput. Harmon. Anal.*, vol. 4, no. 2, pp. 188–221, 1997. DOI: 10.1006/acha.1997.0209.

ISBN 978-82-326-9068-8 (printed ver.)
ISBN 978-82-326-9067-1 (electronic ver.)
ISSN 1503-8181 (printed ver.)
ISSN 2703-8084 (online ver.)

NTNU
Norwegian University of
Science and Technology