Gabor Wavelets and the Heisenberg Group: Gabor Expansions and Short Time Fourier Transform from the Group Theoretical Point of View

Hans G. Feichtinger and Karlheinz Gröchenig

Abstract. We study series expansions of signals with respect to Gabor wavelets and the equivalent problem of (irregular) sampling of the short time Fourier transform. Using Heisenberg group techniques rather than traditional Fourier analysis allows to design stable iterative algorithms for signal analysis and synthesis. These algorithms converge for a variety of norms and are compatible with the time-frequency localization of signals.

§1. Introduction and Notations

Among scientist and mathematicians it is a well known that the Fourier transform (FT) is a perfect tool to analyze periodic functions, to represent them as a superposition of pure frequencies and to consider the behavior of the FT as indication about the contribution of certain frequencies to the total signal. This makes sense for stationary signals, but even for such nice signals as a piece of classical music it would not make sense to consider the global FT. Instead, the function has to be considered locally, and some local Fourier analysis has to be carried out (much in the same way as a musician would describe a piece of music as a sequence of pure tones at certain times). Thus, the idea of time-frequency representations of signals came up. Among the pioneers of this approach were Wigner and Gabor. Wigner proposed a function, now named Wigner distribution, over the time frequency plane (TF-plane for short) which shows many properties that one would like to have for a joint time-frequency energy distribution (which actually does not exist). For many purposes the Short Time Fourier Transform (STFT) of a signal is easier to handle, because it depends in a linear way on the analyzed signal. It has become a standard tool in signal analysis (see [1], [3], [40] and in particular [46] for survey notes). As we shall see it requires the choice of a window function used to localize the analyzed signal in a decent way. The concept of the TF-plane was also
motivating for D. Gabor's [34] suggestion to use a coherent family of Gauss functions, i.e., a family of functions generated from the ordinary Gauss function by time-frequency shifts along the integer lattice in $\mathbb{Z}^2$ (the standard lattice in $\mathbb{R}^2$) as building blocks for a series representation (now called Gabor series) for arbitrary signals. Although his suggestion dates back to 1946 it has found serious interest by mathematicians only in the last decade. Meanwhile it is known that it is not possible to expand arbitrary $L^2$-functions exactly in the way suggested by Gabor if one wants to have square summable coefficients, or at least convergence of the series in $L^2$ (see [43] or [35]).

On the other hand it turned out that much more is true. Following the current trend in the mathematical community, we call a collection of functions of the form $M_y T_x g$, obtained from a single function $g$ by TF-shifts, a collection of Gabor wavelets. The established terminology in the engineering is literature is Gabor type functions and in quantum physics these functions are referred to as (generalized) coherent states (see [49] and [45]). The great interest in Gabor wavelet expansions may be documented by a selection of recent articles, mostly in the applied literature, due to Bastiaans [4], Cenker [8], Daubechies [11], [12], [14], Daugman [16], Dufaux and Kunt [17], Gertner and Zeevi [36], Porat and Zeevi [50], [51], Super an Bovik [58], or Walnut [62], [63].

As we will point out in detail in this note Gabor wavelets are a special case of the general approach described by the authors in [9], [22] and [25]--[28].

In this article we treat a general theory of Gabor wavelets. This theory emerges as a special case from the group theoretical approach ([25]--[27],[38]). It is our intention to give it a formulation concrete enough so that this very general theory can be useful to applied mathematicians. Besides this we shall discuss in some detail the connections between various concepts related to the TF-plane and those using the Heisenberg group. In addition to the known results we also state several new results (such as Theorems 15, 18 and 25 and most of section 7) which have not yet appeared in print.

Our exposition will focus on the following main problems:

A) Atomic representation problem (Gabor wavelet expansions)

We are looking for sufficient conditions on the basic building block $g$ and the discrete family $(z_i)_{i \in I} = (x_i, y_i)_{i \in I}$ in $\mathbb{R}^{2d}$ such that any $f$ in the Hilbert space $L^2(\mathbb{R}^d)$ allows a norm convergent series representation of the form (also called Gabor type expansion or Gabor series expansion): 

$$f(t) = \sum_{i \in I} c_i e^{2 \pi i y_i(t-x_i)} g(t-x_i),$$

(1.1)

Of course we would like to have stability in the following sense:

$$\left( \sum_{i \in I} |c_i|^2 \right)^{1/2} \leq C \cdot \|f\|_2 \quad \forall \ f \in L^2(\mathbb{R}^d).$$

(1.2)

One cannot expect uniqueness of the representation because coherent families are in general linear dependent. Therefore a constructive method of choosing the coefficients $(c_i)_{i \in I}$, depending linearly on $f$, is asked for. Finally good
locality and stability of the expansion are desirable (i.e., small local changes should effect only few coefficients; small errors in the coefficients should only result in a minimal distortion of the corresponding series,...).

B) **Irregular sampling of the Short Time Fourier Transform**

There are formulas (see Bastiaans [4],[5], Heil/Walnut [39], Daubechies et al. [13]) which allow to recover the signal or equivalently the full STFT from its sampling values over a sufficiently fine regular grid in the TF-plane. For this case the Zak-transform is a very useful tool (see [44], [39]). If the signal $f$ belongs to the Hilbert space of $L^2$-functions (signals of finite energy) this question is equivalent to the question, whether a coherent family is a so-called frame.

Again we are interested in efficient iterative reconstruction algorithms, which should be stable as well. Thus, replacing the window by 'similar' one, subsequent reconstruction of the signal from samples of the STFT should still give a good approximation to the signal in discussion. As we will show, the algorithms derived by our group theoretical approach satisfy all these requirements.

The plan of this paper is as follows. First we recall some notations, mainly motivated by applications in signal analysis in Chap. 2. In particular, concepts that are related to time-frequency representations of signal such as the STFT, are explained. We define two families of function spaces (or spaces of tempered distributions) which are formed according to the behavior of the STFT of their elements, and which are called generalized modulation spaces and formulate two main questions. In Section 3 we describe our solution of the two main problems. In Chap. 4 we explain the role of the reduced Heisenberg group and how the concepts of signal analysis are directly related to the representation coefficients for the Schrödinger representation. This point of view also allows us to identify modulation spaces with coorbit spaces. Section 5 is devoted to coorbit spaces related to $L^1$-spaces, which are important as spaces of test functions. The kernel-theorem for the Segal algebra $S_0(\mathbb{R}^d)$ is the main result of this section. In Section 6 we show how the main results can be obtained from the general theory, and some further applications of the group theoretical point of view are given. Recent results concerning Wilson bases are mentioned among others. The iterative methods used to solve the two problems is compared to the standard Hilbert space approach using frames, and the method of adaptive weights presented. The final chapter contains results on the stability of both the atomic representation method and the reconstruction methods for the STFT. Invariance properties of modulation spaces (e.g., under chirp operators) are proved, but it is also shown how seemingly different spaces (such as Bargmann-Fock spaces) can be identified in a very natural way with modulation spaces.
§2. Signal Analysis, Function Spaces Defined by Means of the STFT

Let the $\mathcal{H}$ be the Hilbert space $L^2(\mathbb{R}^d)$ of square integrable functions ("signals of finite energy") over the $d$-dimensional Euclidean space $\mathbb{R}^d$. Following the tradition of physicists and consistent with [25]–[28] we describe the inner product by the formula $\langle g, f \rangle := \int_{\mathbb{R}^d} g(x) f(x) dx$, which is linear with respect to the second argument. This will be convenient later. For the FT $\mathcal{F}$ we choose the following normalization:

$$\mathcal{F} f(y) = \hat{f}(y) := \int_{\mathbb{R}^d} f(x) e^{-2\pi i x \cdot y} dx,$$  \hspace{1cm} (2.1)

where $x \cdot y$ denotes the inner product on $\mathbb{R}^d$ given as $x \cdot y := \sum_{i=1}^{d} x_i y_i$. By Plancherel's theorem $\mathcal{F}$ is a unitary mapping on $L^2(\mathbb{R}^d)$, in particular, one has the fundamental relation $\langle f, g \rangle = \langle \hat{f}, \hat{g} \rangle$. The inversion formula is given by $\mathcal{F}^{-1} g(x) = \int_{\mathbb{R}^d} \hat{g}(y) e^{2\pi i x \cdot y} dy$.

Time-frequency shifts are denoted by $T_x$ for the translation operators and $M_y$ for the modulation operator with $x, y \in \mathbb{R}^d$, i.e.,

$$T_x f(z) := f(z - x), \hspace{1cm} M_y f(z) := e^{2\pi i y \cdot z} f(z),$$  \hspace{1cm} (2.2)

and we have

$$T_x \hat{f} = M_y \hat{f}, \hspace{1cm} M_y f = T_y f,$$  \hspace{1cm} (2.3)

and the important commutation relation

$$M_y T_x = e^{2\pi i x \cdot y} T_x M_y.$$  \hspace{1cm} (2.4)

By $\mathcal{S}(\mathbb{R}^d)$ we denote the Schwartz space of rapidly decreasing functions on $\mathbb{R}^d$. It is invariant under the FT. The dual space $\mathcal{S}'(\mathbb{R}^d)$ consists of tempered distributions and the FT can be extended to $\mathcal{S}'(\mathbb{R}^d)$ by $\hat{\sigma}(f) := \sigma(\hat{f})$.

$(L^1(\mathbb{R}^d), \| \cdot \|_1)$ denotes the Banach space of all Lebesgue integrable functions with the natural norm $\|f\|_1 := \int_{\mathbb{R}^d} |f(x)| dx$. It contains $\mathcal{K}(\mathbb{R}^d)$, the space of all continuous, complex-valued functions $K$ with compact support (supp$(k)$) as a dense subspace. $(L^1(\mathbb{R}^d), \| \cdot \|_1)$ is a Banach algebra with respect to convolution, which is defined as the integral:

$$f \ast g(x) := \int_{\mathbb{R}^d} g(x - y)f(y)dy = \left( \int_{\mathbb{R}^d} T_y g f(y) dy \right)(x).$$  \hspace{1cm} (2.5)

For $1 \leq p < \infty$ set $L^p_{m}(\mathbb{R}^d) := \{f \mid \|f\|_{p,m} := (\int_{\mathbb{R}^d} |f(y)|^p m(y)^p dy)^{1/p} < \infty \}$. For the trivial weight $m(y) \equiv 1$ we just write $(L^p, \| \cdot \|_p)$. If $m$ is a moderate weight (see below) then $L^p_{m}$ is translation invariant. Using the fact that $\|T_x f\|_p = \|f\|_p$ for all $x \in \mathbb{R}^d$ and $\|T_x f - f\|_p \to 0$ for $|x| \to 0$ and all $f \in L^p(\mathbb{R}^d)$ is also follows (by vector-valued integration)

$$L^1 \ast L^p \subseteq L^p,$$  \hspace{1cm} and $\|g \ast f\|_p \leq \|g\|_1 \|f\|_p \forall g \in L^1(\mathbb{R}^d), f \in L^p(\mathbb{R}^d).$  \hspace{1cm} (2.6)
The involution in $L^1$, given by $g^*(x) = \overline{g(-x)}$, satisfies $(g * f)^* = f^* * g^*$.

The Short Time Fourier Transform (or Sliding Window Fourier Transform), for short STFT, is a well-known tool for the time-frequency representation of signals, especially of non-stationary signals (see Allen-Rabiner [1], Nawab and Quatieri [46], Papoulis [47], Flandrin [32], Hlawatsch [40] for the use of STFT in signal analysis). It is also found under the names of continuous Gabor transform in [39] and holographic transform (see Schempp [54], [55]).

By means of a window function $g$, usually a plateau-like real-valued functions with compact support centered near the origin a piece of the signal $f$ is "cut out" and its frequencies are then analyzed with the Fourier transform.

The STFT of $f$ with respect to the window $g$ is given by

$$S_g f(x, y) := \int_{\mathbb{R}^d} e^{-2\pi i y \cdot z} \overline{g(z-x)} f(z) dz = (M_y T_x g, f) \quad \text{for} \quad (x, y) \in \mathbb{R}^{2d}. \quad (2.7)$$

The STFT is symmetric with respect to $f$ and $g$ in the following sense:

$$S_g f(x, y) = e^{-\pi i x \cdot y} \overline{S_f g(-x, -y)}. \quad (2.8)$$

This is important as it shows that decay properties of $S_g f$ are joint properties of $f$ and $g$. In particular, if $g$ is not smooth, one cannot expect good decay of $S_g f$ in the frequency direction, even if $f$ is very smooth (or even constant). On the other hand much smoothness of $g$ requires a large support which destroys the locality properties of $S_g f$. This dilemma makes "window design" so interesting. However, any window $g$ which is sufficiently well localized in time and frequency will allow to identify a "piece of music", i.e., a superposition of a couple of pure tones with smooth envelopes by its time-frequency behavior.

Due to its good time-frequency concentration (it gives equality in the Heisenberg uncertainty relation) and its invariance under the FT the Gauss kernel $g_0, g_0(x) := e^{-\pi x^2}$ is a very good choice. The formula $S_g f(x, y) = e^{-2\pi i x \cdot y} S_g \hat{f}(y, -x)$ implies

$$|S_g f(x, y)| = |S_{g_0} \hat{f}(y, -x)| \quad \text{for} \quad f \in L^2(\mathbb{R}^d), \quad (2.9)$$

showing that the behavior of $S_{g_0} \hat{f}$ is exactly the same as that of $S_g f$, rotated by $90^\circ$ in the TF-plane.

**Lemma 1. (Inversion Formula for the STFT)**

a. Given two non-orthogonal elements $g, g_1 \in L^2(\mathbb{R}^d)$ any signal $f$ can be recovered (for some $C_1 \neq 0$) by

$$f(z) = C_1 \int_{\mathbb{R}^d} T_x (\overline{g g_1})(z) f(x) dx = C_1 \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} S_g f(x, y) M_y T_x g_1(z) dy dx, \quad (2.10)$$

or in vector-valued form

$$f = C_1 \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} (M_y T_x g, f) M_y T_x g_1 dy dx. \quad (2.11)$$
b. The mapping \( f \mapsto S_g f \) defines an isometry from \( L^2(\mathbb{R}^d) \) onto \( L^2(\mathbb{R}^{2d}) \), if \( g \in L^2(\mathbb{R}^d) \) satisfies \( \|g\|_2 = 1 \), i.e.,
\[
\|S_g(f)\|_2^2 = C\|f\|_2^2 \quad \forall \ f \in L^2(\mathbb{R}^d).
\] (2.12)

**Sketch of Proof:** The Fourier inversion formula allows to recover \((T_x \tilde{g}) f\) from \(S_g f(x,.)\). If \( \sum_{n \in \mathbb{Z}^d} T_n g \equiv 1 \) one recovers \( f \) as \( \sum_{n \in \mathbb{Z}^d} T_n g \cdot f \). For general \( g \in L^1 \cap L^2 \) summation is replaced by an integral. Since
\[
\int_{\mathbb{R}^d} \tilde{g}(z-x)dx = \int_{\mathbb{R}^d} g^*(x-z)dx = \int_{\mathbb{R}^d} g(u)du = \tilde{g}(0) ,
\] (2.13)
one derives
\[
\tilde{g}(0) \cdot f(z) = \int_{\mathbb{R}^d} (T_x \tilde{g}(z)) f(z)dx = \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} e^{2\pi iy \cdot z} S_g f(x,y)dxdy .
\] (2.14)

Since the ***same*** argument holds if we replace \( \tilde{g} \) above by some other function \( \tilde{g} g_1 \), with \( \int_{\mathbb{R}^d} \tilde{g}(z)g_1(z)dz =: (C_1)/1 \neq 0 \), the result is proved.

b) Take \( g = g_1 \) and \( \|g\|_2 = 1 \) and take the inner product with \( f \) on both sides of (2.10).

The lemma indicates how the signal \( f \) can be recovered from complete knowledge of the STFT, as a (continuous) superposition of copies of \( g_1 \), shifted in the time-frequency sense. If we want to write \( f \) as such a superposition, with basic building block \( g \), we just have to interchange the roles of \( g_1 \) and \( g \). From this point of view the two main questions are: Whether reconstruction is possible if the STFT is only given over a discrete set of points in the TF-plane, and whether the continuous collection \( \{M_y T_x g\} \) of "atoms" can be replaced by a discrete (but sufficiently rich) subcollection of Gabor wavelets.

Following Folland [33] or Heil/Walnut [39] (see more detailed references there, see also Auslander/Tolimieri [2]) we also introduce the ***radar ambiguity function*** \( A(f,g) \)
\[
A(f,g)(x,y) := \int_{\mathbb{R}^d} e^{2\pi iy \cdot z} f(z + x/2)\tilde{g}(z - x/2)dz ,
\] (2.15)
It has the same absolute value as the STFT. For the description of good window functions we also need the (cross) ***Wigner distribution*** \( W(f,g) \)
\[
W(f,g)(x,y) := \int_{\mathbb{R}^d} e^{2\pi iy \cdot z} f(z + 2/2)\tilde{g}(x - z/2)dz.
\] (2.16)

Despite their deceptive similarity these two functions can be shown to be mapped into each other via the Fourier transform. For most applications in signal analysis the quadratic expressions \( A_f := A(f,f) \) and \( W_f := W(f,f) \) are used (see Hlawatsch [40] and references there).

Lemma 1.b) suggests to think of \( \|S_g f\|^2 \) as an indicator for the "energy distribution" of the signal \( f \) on the TF-plane. The behavior of \( S_g f \) also reflects
local smoothness of \( f \) in some domain \( U \in \mathbb{R}^d \), by good decay of \( S_g f(x,y) \), for \( |y| \rightarrow \infty \), for all \( x \in U \) (at least if \( g \) is smooth enough). If \( g \) is well concentrated near the origin then \( S_g f \) will reflect decay (or growth) of \( f \) itself. It is therefore a natural idea to use the STFT as a tool to measure decay and smoothness properties of a signal as well as its time-frequency concentration. We do this by considering the following class of generalized modulation spaces. But first we need an appropriate class of weight functions.

**Definition 2.** A strictly positive and continuous function \( m \) on the TF-plane will be called \( a \)-moderate if for some \( a \geq 0 \) there exists \( C > 0 \) such that

\[
(M1) \quad m(z + v) \leq C(1 + |v|)^a m(z) \quad \forall z, v \in \mathbb{R}^d ,
\]

(2.17)

For any such weight we define the generalized modulation space \( M^{m}_{p,q}(\mathbb{R}^d) \):

**Definition 3.** Let \( g \in S \) and \( m \) be an \( a \)-moderate function, then we define the generalized modulation space \( M^{m}_{p,q}(\mathbb{R}^d) \) as

\[
M^{m}_{p,q}(\mathbb{R}^d) = \{ f \in S'(\mathbb{R}^d) \mid ||f| M^{m}_{p,q}(\mathbb{R}^d)|| :=
\]

\[
= \left[ \int_{\mathbb{R}^d} \left( \int_{\mathbb{R}^d} |S_g f(x,y)|^p m^q(x,y) dx \right)^{2/p} dy \right]^{1/q} < \infty .
\]

(2.18)

For the special case

\[
m(x,y) = (1 + |x|)^r (1 + |y|)^s \quad \forall x, y \in \mathbb{R}^d ,
\]

(2.19)

with \( r, s \in \mathbb{R} \) we shall use the symbol \( M^{r,s}_{p,q} \). The "classical" modulation spaces \( M^{s}_{p,q} \) (see [22] and Triebel [61]) are identical with the spaces \( M^{0,s}_{p,q} \) in the new terminology. On the other hand the spaces \( M^{r,s}_{p,q} \) may be considered as a weighted version of \( M^{s}_{p,q} \).

By reversing the order of integration we define another family of spaces, to be called \( W^{m}_{p,q} \). We have chosen the letter \( W \) because this family contains the Wiener-amalgam spaces \( W(\mathcal{F}L^p, L^q) \) as special cases, just as the classical modulation spaces are contained in the \( M \)-family. See [7], [20] and [21] for results about these spaces and their role in the description of the generalized FT (GFT).

**Definition 4.**

\[
W^{m}_{p,q}(\mathbb{R}^d) = \{ f \in S'(\mathbb{R}^d) \mid ||f| W^{m}_{p,q}(\mathbb{R}^d)|| :=
\]

\[
= \left[ \int_{\mathbb{R}^d} \left( \int_{\mathbb{R}^d} |S_g f(x,y)|^p m^{q}(x,y) dy \right)^{2/p} dx \right]^{1/p} < \infty .
\]

(2.20)

Obvious modifications take place if \( p = \infty \) or \( q = \infty \).

The definition of both families of spaces is inspired by the theory of Besov-Triebel-Lizorkin spaces \( B^{s}_{p,q} \) and \( F^{s}_{p,q} \) (see Stein [57], Peetre [48], and Triebel [60]). While the theory of modulation spaces is very similar to the theory of BTL-spaces, there are some fundamental differences. The \( B \)- and \( F \)-spaces are (with few exceptions) not isomorphic as Banach spaces, on the other hand, the Fourier transform is an isomorphism between \( M^{m}_{p,q}(\mathbb{R}^d) \) and \( W^{m}_{p,q} \), where \( \hat{m}(x,y) = m(-y,x) \), as follows immediately from the definitions.
Lemma 5. For $1 \leq p, q \leq \infty$ the $M_{p,q}^s(\mathbb{R}^d)$-spaces are Banach spaces of distribution on $\mathbb{R}^d$ with respect to their natural norms; they are invariant under translation and modulation operators and these operations are continuous on $M_{p,q}^s(\mathbb{R}^d)$ for $1 \leq p, q < \infty$ in the following sense (writing $B$ for any of them)

$$\|T_x f - f\|_B \to 0 \text{ for } |x| \to 0, \forall f \in B;$$

(2.21)

as well as

$$\|M_y f - f\|_B \to 0 \text{ for } |y| \to 0, \forall f \in B;$$

(2.22)

The submultiplicativity of the weight $w_a$, given by $w_a(v) := (1 + |v|)^a$, for $a \geq 0$, i.e., the inequality $w_a(x + y) \leq w_a(x)w_a(y)$ for $x, y \in \mathbb{R}^d$, implies

$$\|T_x f\|_B \leq w_1(x)\|f\|_B \quad \text{and} \quad \|M_y f\|_B \leq w_2(y)\|f\|_B$$

(2.23)

for all $x, y \in \mathbb{R}^d$, where $w_1(x) = w(x, 0)$ and $w_2(y) = w(0, y)$. It follows by vector-valued integration that $L_{w_1}^1 * B \subseteq B$, and

$$\|g * f\|_B \leq \|g\|_{1, w_1} \|f\|_B \quad \forall g \in L_{w_1}^1, f \in B.$$  

(2.24)

and by a similar argument $\mathcal{F}(L_{w_2}^1) \cdot B \subseteq B$, and

$$\|h \cdot f\|_B \leq \|\hat{h}\|_{1, w_2} \|f\|_B \text{ for } h \in \mathcal{F}L_{w_2}^1, f \in B.$$  

(2.25)

It is also true for the same reason that for any Dirac sequence $(e_n)_{n=1}^\infty$ in $L_{w_1 + w_2}^1(\mathbb{R}^d)$ one has $\|e_n * f - f\|_B \to 0$. More precisely, for any sequence of normalized functions in $L^1(\mathbb{R}^d)$ with supports shrinking to zero this is true. One also has $\|e_n \cdot f - f\|_B \to 0$ by a similar argument. Combining both facts and using the inclusion $(S * S') \cdot S \subseteq S$ we derive also that $S$ (or even $\mathcal{D} := S \cap \mathcal{K}$) is dense in any of these spaces (for $1 \leq p, q < \infty$).

The above definitions raise a couple of questions. First of all, are these spaces well defined, i.e., actually independent of the choice of the window function $g$? If so, how can a sufficiently large class of windows be described, such that any non-trivial choice of a window from that class gives the same spaces (with equivalent norms). This questions will be answered in detail in Lemma 26.

The following lemma gives a list of familiar spaces that are contained in the family of modulation spaces.

Lemma 6. $L^2$ is contained in the family of modulation spaces as $M_{2,2}^0$. More generally the $L^2$ Bessel potential spaces $H^s = \{f \in S' : \int |\hat{f}(t)|^2(1 + |t|)^{2s}dt < \infty\}$ (or $L^2_s$ in Stein's notation, see Stein [57]) coincide with the modulation spaces $M_{2,2}^s$. The Segal algebra $S_0$ (see section 6) arises as $M_{1,1}^0$ (see [19]). Its algebra of multipliers (operators commuting with translations) can be identified with $M_{1,\infty}^0$, and the dual space $S'_0(\mathbb{R}^d)$ is just $M_{\infty,\infty}^0$ (with appropriate interpretations of the symbols for $p$ or $q = \infty$).

For details we refer to [20]–[22], [26], [27]. Atomic decompositions of these spaces are studied in [22], and [25] – [28].
§3. Gabor Wavelets and Irregular Sampling of the STFT

In this section we describe expansions of signals in the modulation spaces by means of Gabor wavelets. These results are special cases of the general theory in our papers [9] and [25]–[29], obtained by choosing as group the reduced Heisenberg group with the Schrödinger representation. The relevance of the Heisenberg group for a wide range of problems in mathematical analysis and also for applications in signal analysis is meanwhile well documented (see Auslander and Tolimieri [2], Folland [33], Howe [41], Schempp [54],[55], and Taylor [59]). However, since the translation from the abstract results into the concrete form may not be obvious for "users", and since Gabor wavelet expansions can be stated independently and directly, we postpone the discussion of the background to section 4 and state the results in the language of signal analysis. We start with the characterization of modulation spaces by means of Gabor wavelet expansions.

In order to describe the 'density' of a discrete family \((z_i)_{i \in I}\) in \(IR^d\) we use the following concept: Given some set \(W\) with nonvoid interior in \(IR^d\) we call \((z_i)_{i \in I}\) \(W\)-dense, if \(\bigcup_{i \in I} z_i + W = IR^d\).

**Theorem 7.** (Gabor wavelet expansions for modulation spaces). Let us consider the family of all modulation spaces \(M^m_{p,q}(IR^d)\), with \(m\) satisfying (M1) for the same \(a \geq 0\) and \(C > 0\). Then for any \(g \in L^2(IR^d)\) with Wigner distribution \(W_g\) (or equivalently with ambiguity function \(A_g\) in \(L^1_{w_a}(IR^d)\) there exists a neighborhood \(W\) of \((0,0)\) in \(IR^{2d}\) such that for any \(W\)-dense family \((z_i)_{i \in I} = (x_i, y_i)_{i \in I}\) in \(IR^{2d}\) it is possible to write any \(f \in M^m_{p,q}(IR^d)\) as

\[
f = \sum_{i \in I} c_i T_{x_i} M_{y_i} g ,
\]

where the family \((c_i)_{i \in I}\) belonging to a sequence space connected with \(M^m_{p,q}(IR^d)\). The mapping \(f \mapsto (c_i)_{i \in I}\) is linear. If \((z_i)_{i \in I}\) is a subset of some lattice \(aZZ^d \times \beta ZZ^d \subseteq IR^{2d}\) the coefficient mapping is bounded from \(M^m_{p,q}(IR^d)\) into the sequence space

\[
(\ell^q(\ell^p))_m := \{(c_{i,n})_{i,n \in Z^d} : \left[ \sum_{i \in Z^d} \left( \sum_{n \in Z^d} |c_{i,n}|^p m_{i,n}^p \right)^{q/p} \right]^{1/q} < \infty \}
\]

where \(m_{i,n} := m(\alpha l, \beta n)\), for \(k, n \in ZZ^d\). Conversely, for any set of coefficients \((c_{i,n})_{i,n \in Z^d} \in (\ell^q(\ell^p))_m\) the series

\[
f = \sum_{n,l} c_{n,l} T_{\alpha l} M_{\beta n} g
\]

is unconditionally convergent in \(M^m_{p,q}(IR^d)\), since for every \(\epsilon > 0\) there exists a finite subset \(F_0 \subseteq I\) such that for any finite set \(F \supseteq F_0\) one has

\[
\|f - \sum_{i \in F} c_{i,n} T_{\alpha l} M_{\beta n} g |M^m_{p,q}(IR^d)\| \leq \epsilon .
\]
In particular
\[ f = \sum_n \sum_i c_{n,i} T_{\alpha i} M_{\beta n} g = \sum_i \sum_n c_{n,i} T_{\alpha i} M_{\beta n} g. \]  
(3.5)

A version of the above result which gives more quantitative information about the set \( W \) reads as follows (details are to be given in a forthcoming note by Gröchenig). For the case of regular sampling along lattices in the TF-plane results of this kind are given by Walnut [62].

**Theorem 7.b.** Suppose that \( \text{supp}(g) \subseteq [-a, a] \) for some \( g \in L^2(R) \). Let \( (x_n)_{n \in \mathbb{N}} \) be any increasing sequence such that for some \( A, B > 0 \)
\[ 0 < A_1 \leq \sum_{n \in \mathbb{N}} |g(t - x_n)|^2 \leq B_1 < \infty \quad \text{a.e.} \]  
(3.6)
Assume that \( (y_{n,k})_{k \in \mathbb{Z}} \) are increasing sequences with \( \sup_{k,n} (y_{k+1,n} - y_{k,n}) < 1/2a \). Then every \( f \in L^2(R) \) has a stable expansions
\[ f = \sum_{n,k} c_{k,n} (y_{k+1,n} - y_{k,n}) M_{y_{k,n}} T_{x_n} g \]  
(3.7)
with
\[ (c_{k,n})_{n,k \in \mathbb{Z}} \in \ell^2 \quad \text{and} \quad \| (c_{k,n}) \|_2 \leq C \cdot \| f \|_2. \]  
(3.8)
Actually, the collection \( \{(y_{k+1,n} - y_{k,n})^{1/2} M_{y_{k,n}} T_{x_n} g\}_{n,k \in \mathbb{Z}} \) is a frame with frame constants \( A = A_1 \times (1 - 2\delta) \) and \( B = B_1 \times (1 + 2\delta) \).

**Remark 8.** Note that condition (3.6) implies \( \sup_{k \in \mathbb{Z}} |x_{k+1} - x_k| < 2a \).

**Remark 9.** Of course it is possible to characterize the \( W \)-spaces in the same way, just reversing the order of summation compared to the \( M \)-spaces.

The results in [38] can be interpreted as statements, that sufficiently rich coherent families) are suitable as building blocks for atomic representations for large classes of function spaces if and only if they are Banach frames for the same class.

As answer to the second main question we present the following result.

**Theorem 10.** Consider the same family of spaces as in Theorem 7, and let \( g \) be as above. Then there exists a neighborhood \( \tilde{W} \) of zero in \( \mathbb{R}^{2d} \) such that for any \( \tilde{W} \)-dense family \( (z_i)_{i \in I} \) in \( \mathbb{R}^{2d} \) there is a sequence of adaptive weights \( (w_i)_{i \in I} \) such that the sequence \( (f^{(n)})_{n=1}^{\infty} \), given recursively by
\[ f^{(0)} := A f := \sum_{i \in I} w_i S_g f(x_i, y_i) M_{y_i} T_{x_i} g \]  
(3.9)
and
\[ f^{(n+1)} := f^{(n)} + A(f - f^{(n)}) \quad \text{for } n \geq 1 \]  
(3.10)
is convergent in $M_{p,q}^m(\mathbb{R}^d)$ to $f$ at a geometric rate, i.e., there is $\gamma, 0 < \gamma < 1$ and $C > 0$, so that
\[
\|f - f^{(n)}\|_{M_{p,q}^m(\mathbb{R}^d)} \leq C \cdot \gamma^n \cdot \|f\|_{M_{p,q}^m(\mathbb{R}^d)}.
\]  
(3.11)

Remark 11. The input in this algorithm are the values $S_g f(z_i)$. The theorem is thus an irregular sampling theorem for the STFT and provides an easy reconstruction method for all modulation spaces. The proposed method will be called the ADAPTIVE WEIGHTS METHOD for the irregular sampling problem of the STFT. The actual rate of convergence depends mainly on the density of the sampling family in the TF-plane.

Corollary 12. If in addition to the assumptions of Theorem 7 the sampling set belongs to a (sufficiently fine) lattice in $\mathbb{R}^{2d}$, then a function in $L^2(\mathbb{R}^d)$ belongs to $M_{p,q}^m(\mathbb{R}^d)$ if the sampling family $(S_g f(x_i, y_i))_{i \in I}$ belongs to the corresponding sequence space $(\ell^q(\ell^p))_m$.

There are several immediate questions related to these results. If the coefficients are not uniquely determined, to what extent does the choice of the specific method, and in particular the choice of auxiliary parameters influence the result. Let us assume we want to replace the building block (atom) by another one which is close in some sense. Will the corresponding family of coefficients be similar to the original one? What is a good way to measure this similarity of atoms? For the STFT we have similar questions. What can be said about the reconstruction, if the window that has been used for the STFT is only known approximately (i.e., close to some known window)? Can we still hope for approximate reconstruction? What, if there is jitter error, i.e., if we do not have exact informations about the actual sampling positions in the time-frequency plane for the STFT? In the best case we can hope that the reconstruction error is small if the maximal jitter deviation from a family of sampling points is sufficiently small. Such results and many related questions can be answered affirmatively, based on our group theoretic point of view.

From a practical point of view one will have to look for a combined optimization of $\alpha$ and $\beta$ simultaneously, given $g$, if one has the freedom to chose $\alpha$ and $\beta$ independently. Also, both the speed of convergence as well as the degree of stability of the scheme will depend very much on the choice of $g$, $\alpha$ and $\beta$.

The following corollary is of great practical use, especially if the same sampling geometry and the same window occur for many different signals.

Corollary 13. There are functions $e_i \in M_{w_\alpha}^{1,1}$, so that
\[
f = \sum_i S_g f(x_i, y_i) e_i,
\]  
(3.12)

with convergence in $M_{p,q}^m(\mathbb{R}^d)$, for any $f \in M_{p,q}^m(\mathbb{R}^d)$.

The practical relevance of this Corollary is based on the fact that one may calculate the collection $(e_i)_{i \in I}$ offline, perhaps on a faster computer, and recombine them with the sampling values according to (3.12).
§4. The Group Theoretical Background, Correspondence Principle

Gabor wavelets are obtained from a single function — sometimes called mother wavelet — by means of time and frequency shifts $T_x$ and $M_y$. These operators do not commute, but the family

$$\{\tau T_x M_y \mid \tau \in C, |\tau| = 1; x, y \in \mathbb{R}^d \}$$

forms a natural group of unitary operators on $L^2(\mathbb{R}^d)$, which also acts continuously on a variety of other function spaces. Identifying these operators with their parameters, which are in $\mathcal{H}^d := \mathbb{R}^d \times \mathbb{R}^d \times \mathcal{T}$ we recognize the following group law for the reduced Heisenberg group $\mathcal{H}^d$:

$$h_1 \cdot h_2 := (x_1, y_1, \tau_1) \cdot (x_2, y_2, \tau_2) = (x_1 + x_2, y_1 + y_2, \tau_1 \cdot \tau_2 e^{2\pi i x_2 y_1}) \cdot (4.2)$$

The inverse can be calculated as $h^{-1} = (x, y, \tau)^{-1} = (-x, -y, \tau^{-1} e^{2\pi i xy})$.

NOTE: Since we use the factor $2\pi$ in the definition of the modulation operators (because of our definition of $F$) this law differs slightly from the description in our earlier notes [25]–[27]. We refer to Folland [33], Taylor [59], and Heil/Walnut [40] and Reiter [53] for technical details concerning following facts: The Lebesgue measure $dh := dxdydr$ is the Haar measure for $\mathcal{H}^d$. It is both left and right invariant, i.e., $\mathcal{H}^d$ is unimodular. This means that for $F \in \mathcal{K}(\mathcal{H}^d)$

$$\int_{\mathcal{H}^d} F(h_0^{-1}h)dh = \int_{\mathcal{H}^d} F(h)dh = \int_{\mathcal{H}^d} F(hh_0)dh \quad \forall \ h_0 \in \mathcal{H}^d. \quad (4.3)$$

Unimodularity also implies that the left and right translation operators, given by $L_h F(v) := F(h^{-1}v)$ and $R_h F(v) := F(vh)$ respectively, act isometrically on $L^p(\mathcal{H}^d)$, for $1 \leq p \leq \infty$. For $F, G \in \mathcal{K}(\mathcal{H}^d)$ convolution is given (as point-wise or vector-valued integral)

$$F \ast G(h_0) := \int_{\mathcal{H}^d} G(h^{-1}h_0) F(h)dh \quad \text{or} \quad \left( \int_{\mathcal{H}^d} F(h)L_h Gdh \right)(h_0). \quad (4.4)$$

Convolution on $\mathcal{H}^d$ is associative, but not commutative. In the following we summarize basic facts concerning convolution. (1) Translation and convolution:

$$L_h(F \ast G) = L_h F \ast G, \quad R_h(F \ast G) = F \ast R_h G. \quad (4.5)$$

(2) $(L^1(\mathcal{H}^d), \|\cdot\|_1)$ is a Banach algebra with respect to convolution.

(3) Young's inequality: since $\mathcal{H}^d$ is unimodular, the convolution relations

$$L^p(\mathcal{H}^d) \ast L^q(\mathcal{H}^d) \subseteq L^{p} (\mathcal{H}^d), \quad \text{and} \quad L^2(\mathcal{H}^d) \ast L^2(\mathcal{H}^d) \subseteq C^0(\mathcal{H}^d), \quad (4.6)$$

hold, where $C^0(\mathcal{H}^d)$ is the space of continuous functions vanishing at infinity, endowed with the sup-norm $\|F\|_\infty := sup_{x \in \mathcal{H}^d} |F(x)|$. 

370 H. Feichtinger and K. Gröchenig
(4) If $F^*$ denotes the involution $F^*(h) := \hat{F}(h^{-1})$, then $(F*H)^* = H^*F^*$. Convolution by $G^*$ (from the right) is the adjoint operator of the convolution operator $F \mapsto F*G$, i.e.,

$$
\langle H, F*G \rangle = \langle H*G^*, F \rangle \quad \text{for } F, H \in L^2(\mathbb{H}^d), G \in L^1(\mathbb{H}^d).
$$

(4.7)

(5)

$$
H*G(h) = \langle L_h G^*, H \rangle \quad \forall \, h \in \mathbb{H}^d, \quad H*G^*(0) = \langle G, H \rangle.
$$

(4.8)

(6) Convolution inequalities for weighted $L^p$-spaces on $\mathbb{H}^d$ are derived as on $\mathbb{R}^d$. If $w$ is a submultiplicative (Beurling) weight $w$, i.e.,

$$
w(hh') \leq w(h)w(h') \quad \forall \, h, h' \in \mathbb{H}^d,
$$

(4.9)

then $L^1_w(\mathbb{H}^d)$ is a Banach convolution algebra. An important example of weight functions on $\mathbb{H}^d$ is $w(h) := w_a(h) := (1 + |(x,y)|)^{\alpha}$, with $\alpha \geq 0$.

(7) A strictly positive and continuous function on $\mathbb{H}^d$ is called $a$-moderate if $m(hh') \leq m(h)w_a(h') \quad \forall \, h, h' \in \mathbb{H}^d$. Then the pointwise estimate

$$
|F*G(h)| \cdot m(h) \leq (|F|m*|G|w)(h) \quad \forall \, h \in \mathbb{H}^d
$$

(4.10)

holds and Young's inequality implies that $L^p_w(\mathbb{H}^d) := \{ F | Fm \in L^p(\mathbb{H}^d) \}$ satisfies $L^p_m * L^1_w \subseteq L^p_m$ for $1 \leq p \leq \infty$.

In order to apply the general theory of [26] to Gabor wavelet expansions, we proceed along the following steps:

(A) Check that the assumptions of the general theory are satisfied in the particular case of the Heisenberg group and the Schrödinger representation.

(B) Relate the abstract concepts of [25] and [26] to the language of signal analysis, as developed in Section 2.

(C) Translate the main result on series expansions in [26] into the concrete situation of Gabor wavelet expansions.

To verify the assumptions made in [26], we have to check that the Schrödinger representation is irreducible, one-to-one, and integrable with respect to all weights $w_a$, $a > 0$ on $\mathbb{H}^d$, i.e., has matrix coefficients in $L^1_w(\mathbb{H}^d)$.

Recall that the matrix (or representation) coefficients or the generalized wavelet transform (with respect to the Schrödinger representation) are functions on $\mathbb{H}^d$ given by

$$
V_g f(h) := \langle \pi(h)g, f \rangle \quad \text{for } h \in \mathbb{H}^d.
$$

(4.11)

We speak of $g$ as analyzing window (since $g$ may be any bump function we avoid the word “wavelet” at this point), and $f$ is called the analyzed signal. We will call $f \mapsto V_g f$ the Gabor-(Heisenberg) transform with respect to $g$ in this note in order to indicate the intimate connection to the Heisenberg group.
The Gabor transform is a linear mapping from the Hilbert space $L^2(\mathbb{R}^d)$ into a space of bounded and continuous functions on $\mathbb{H}^d$, since for $f, g \in L^2(\mathbb{R}^d)$

$$|V_g f(h)| \leq \|g\|_2 \|f\|_2 \ \forall \ h \in \mathbb{H}^d, \text{ by Cauchy-Schwartz} \quad (4.12)$$

$$|V_g f(h) - V_g f(h \cdot h')| \leq \|((\pi(h') - 1)g, \pi(h^{-1})f)\| \to 0 \text{ for } |h'| \to 0. \quad (4.13)$$

It will be crucial that $V_g$ also satisfies the intertwining property

$$V_g(\pi(h)f) = L_h(V_g(f)) \quad \text{for } h \in \mathbb{H}^d. \quad (4.14)$$

**Proof:**

$$V_g(\pi(h)f)(\tilde{h}) = \langle \pi(\tilde{h})g, \pi(h)f \rangle =$$

$$= \langle \pi(h^{-1}\tilde{h})g, f \rangle = V_g f(h^{-1}\tilde{h}) = L_h(V_g(f))(\tilde{h}). \quad \blacklozenge$$

Changing the role of $f$ and $g$ is the same as applying the involution to $V_g f$

$$(V_g f)^* = V_f(g). \quad (4.15)$$

**Proof:**

$$V_g f^*(h) = \overline{V_g f(h^{-1})} = \overline{(\pi(h^{-1})g, f)} = \langle \pi(h)f, g \rangle = V_f g(h). \quad \blacksquare$$

The existence of admissible vectors $g$, i.e., of non-zero functions $g \in L^2(\mathbb{R}^d)$ for which $V_g g \in L^1(\mathbb{H}^d)$, is by definition the same as integrability of the representation $\pi$. Note that by the boundedness of $V_g g$ square integrability is a consequence of integrability. Since we have

$$|V_g f(x, y, \tau)| = |V_g f(x, y, 1)| \ \forall \ \tau \in T^d \quad (4.16)$$

we may use the symbol $V_g f(x, y)$ for $V_g f(x, y, 1)$ and check integrability of $V_g f$ over $R^{2d}$ only. To this end let us derive two pointwise estimates. One has

$$|V_g f(x, y)| = |\langle T_x M_y g, f \rangle| = |M_y g^* \ast f(x)| \leq (\|g^*\| \ast |f|)(x); \quad (4.17)$$

for $f, g \in L^2(\mathbb{R}^d)$, but also by Plancherel’s theorem

$$|V_g f(x, y)| = |\langle M_{-x} T_y \hat{g}, \hat{f} \rangle| = |M_{-x} \hat{g}^* \ast \hat{f}(y)| \leq (\|\hat{g}\| \ast |\hat{f}|)(y). \quad (4.18)$$

Combining these two estimates one obtains

$$|V_g g(x, y)| \leq \min[|g^*| \ast |g|](x), (|\hat{g}^*| \ast |\hat{g}|)(y)]. \quad (4.19)$$

Since $g, g^*, \hat{g}, \hat{g}^*$ and therefore both convolution products are faster decaying than the inverse of any polynomial for any $g \in S(\mathbb{R}^d)$, it follows that $V_g g \in L^1(\mathbb{H}^d)$ (or $L^1(\mathbb{R}^{2d})$). Actually, we have shown that $g \in \mathcal{H}_{wa}^1 = \{g \mid V_g g \in L^1_{wa} \}$ for any $a \geq 0$. 


Among the basic consequences of the (square) integrability of the Schrödinger representation are the orthogonality relations, which are also known as Moyal’s formulas in this case.

\[
\langle V_{g_1}f_1, V_{g_2}f_2 \rangle_{L^2(\mathbb{H}^d)} = \langle f_1, f_2 \rangle_{L^2(\mathbb{R}^d)} \cdot \langle g_2, g_1 \rangle_{L^2(\mathbb{R}^d)}. \tag{4.20}
\]

Taking \( f_1 = f_2 = f \) and \( g_1 = g_2 = g \), then we obtain

\[
\|V_g f\|_2 = \|f\|_2 \|g\|_2 \quad \forall \ f, g \in L^2(\mathbb{R}^d). \tag{4.21}
\]

Both formulas follow immediately from the inversion formula of the STFT (3.10), since for the representation coefficients the integration over the torus is trivial.

The orthogonality relations have several important consequences. In order to show that \( \pi \) is irreducible we have to verify that for any \( g \neq 0 \) in \( L^2(\mathbb{R}^d) \) the finite linear combinations of elements of the form \( T_x M_y g \) are dense in \( L^2(\mathbb{R}^d) \).

**Proof:** Assume \( f \perp \pi(h)g \) for all \( h \in \mathbb{H}^d \). Then we have \( V_g f = 0 \), hence \( \|V_g f\|_2 = \|f\|_2 \|g\|_2 = 0 \), which in turn implies that \( f = 0 \). \( \blacksquare \)

Next we verify the **convolution formula** for the representation coefficients:

\[
V_{g_1}f_1 * V_{g_2}f_2 = \langle g_1, f_2 \rangle \cdot V_{g_2}f_1 \quad \text{for} \ f_1, f_2, g_1, g_2 \in L^2(\mathbb{R}^d) \tag{4.22}
\]

**Proof:**

\[
V_{g_1}f_1 * V_{g_2}f_2(h) = \langle L_h(V_{g_2}f_2)^*, V_{g_1}f_1 \rangle = \\
= \langle L_h(V_{f_2}g_2), V_{g_1}f_1 \rangle = \langle V_{f_2}(\pi(h)g_2), V_{g_1}f_1 \rangle = \\
= \langle g_1, f_2 \rangle \langle \pi(h)g_2, f_1 \rangle = \langle g_1, f_2 \rangle V_{g_2}f_1(h). \quad \blacklozenge
\]

As a special case of (4.22) we observe the reproducing convolution equation, (assuming only that \( \|g\|_2 = 1 \)):

\[
V_g f * V_g g = V_g f \quad \forall \ f \in L^2(\mathbb{R}^d). \tag{4.23}
\]

Writing \( G := V_g g \) it follows that \( G * G = G \) and \( G = G^* \), i.e., \( F \mapsto F * G \) is an (orthogonal) projection operator on \( L^2(\mathbb{H}^d) \). It also can be shown that \( L^2(\mathbb{H}^d) * G \) is exactly \( V_g(L^2(\mathbb{R}^d)) \), the range of \( V_g \).

The spaces \( \mathcal{H}_w^1 \) play an important role as spaces of test functions: For any functional \( f \) on one of these Banach spaces it is possible to define the Gabor transform \( V_g f \) for any analyzing signal \( g \in \mathcal{H}_w^1 \). For reasons of convenience and consistency with our papers we work with the so-called anti-dual \((\mathcal{H}_w^1)^\sim\) of all additive and continuous mappings \( \sigma \) from \( \mathcal{H}_w^1 \) into the complex numbers satisfying \( \sigma(\lambda g) = \lambda \sigma(g) \). This allows us to use the symbols \( \langle \cdot, \cdot \rangle \), both for the new duality and the Hilbert space duality. For the definition of **coorbit spaces**
we need also a solid Banach space \((Y, \| \cdot \|_Y)\) of locally integrable functions on \(\mathbb{H}^d\) with

\[
(Y1) \ |F(h)| \leq |G(h)| \text{ a.e. and } G \in Y \rightarrow F \in Y \text{ and } \|F\|_Y \leq \|G\|_Y. \tag{4.24}
\]

\[
(Y2) \ Y \text{ is left and right translation invariant and for some } a \geq 0
\]

\[
\|L_h F\|_Y \leq w_a(h)\|F\|_Y, \text{ and } \|R_h F\|_Y \leq w_a(h)\|F\|_Y. \tag{4.25}
\]

\[
(Y2') \text{ Furthermore we assume that for the same } a \geq 0 \text{ we have}
\]

\[
Y * L^1_a \subseteq Y \text{ with } \|F * G\|_Y \leq \|F\|_Y \|G\|_{1,a} \forall F \in Y, g \in L^1_a, \tag{4.26}
\]

where convolution is in the sense of the Heisenberg group.

With these assumption made we are ready to define the \textit{coorbit spaces with respect to the Schr"odinger representation}:

\textbf{Definition 14.}

\[
Co(Y) = \{f | f \in (\mathcal{H}^d_w)^{-'} , V_g f \in Y\}, \text{ with } \|f\|_Y := \|V_g f\|_Y. \tag{4.27}
\]

Given the definition of generalized modulation spaces and \(W\)-spaces we have to establish the link between these spaces and coorbit spaces next. First a general remark. The above definition raises the question why we replace the handy and concrete spaces using only the time frequency plane and the STFT by objects related to the Heisenberg group and the matrix coefficients of the Schrödinger representation. This seems like an unnecessary abstraction and complication. The answer lies in the reproducing formula (4.22). It is a simple and ordinary convolution equation and can be treated by methods of non-commutative harmonic analysis. Staying in the TF-plane would lead to a "twisted" convolution (see [33]) and to a rather "twisted" analysis of the problem. Adding the trivial and almost redundant parameter of the torus component provides a fundamental group structure and thus makes life much easier. In fact, it makes the powerful machinery of general atomic representations (see [26]), which also covers many other "wavelet-theories", applicable to \(M\) and \(W\)-spaces. The link consists of two main observations:

1) The representation coefficients \(V_g f\), which a priori "live" on the group \(\mathbb{H}^d\), are uniquely determined by their values on \(\mathbb{R}^d \times \mathbb{R}^d \times \{1\}\), since

\[
V_g f(x,y,\tau) = \tau V_g f(x,y,1) \quad \forall f, g \in L^2(\mathbb{R}^d). \tag{4.28}
\]

The connection to the STFT (with window \(g\)) is given trough

\[
V_g f(x,y,\tau) = \tau e^{2\pi i x \cdot y} S_g f(x,y), \forall f, g \in L^2(\mathbb{R}^d) \tag{4.29}
\]

2) There are several natural ways to relate a Banach space \((V, \| \cdot \|_V)\) on \(\mathbb{R}^d\) satisfying \((Y1)-(Y2')\) (and ordinary addition as group multiplication)
to a Banach space \((Y, \| \cdot \|_Y)\) on the reduced Heisenberg group \(H^d\) satisfying 
\((Y1) - (Y2')\) in such a way that for continuous functions \(F\) on \(H^d\) satisfying \(F(x, y, \tau) = \tau F(x, y, 1)\) membership of \(F_{\text{red}}: (x, y) \mapsto F(x, y, 1)\) in \((V, \| \cdot \|_V)\) is exactly the same as membership of \(F\) in \((Y, \| \cdot \|_Y)\) (with equivalence of norms).

Despite the possibility of giving a "general recipe" it seems to be better to treat this questions individually in order to preserve "natural" identifications. Thus, it is clear that \(V = L^p(R^{2d})\) matches with the space \(Y = L^p(H^d)\), for \(1 \leq p < \infty\). There are also no serious problems with moderate weight functions \(m\), because any moderate weight can only vary to a certain amount over the compact torus, i.e., there are positive constants \(C_1\) and \(C_2\) such that

\[
C_1 \cdot m(x, y, 1) \leq m(x, y, \tau) \leq C_2 \cdot m(x, y, 1) \quad \forall \tau \in T. \tag{4.30}
\]

Also moderateness of \(m\) with respect to the weight \(w_a\) on \(H^d\) is equivalent to moderateness of its restriction to \(R^{2d}\) with respect to the weight \(w_a\) (on \(R^{2d}\)).

For more general spaces, such as the weighted, mixed norm spaces used for the definition of M-spaces, a good general recipe to generate \((Y, \| \cdot \|_Y)\) from \((V, \| \cdot \|_V)\) is to use the sublinear mapping

\[
P_T F := \int_T |F(x, y, \tau)| d\tau \tag{4.31}
\]

and to define

\[
Y := Y_V := \{ F | F \text{ locally integrable on } H^d, \ P_T F \in V \} \tag{4.32}
\]

with the natural norm \(\| F \|_Y := \| P_T F \|_V\). Note that whenever it is possible to give an estimate \(\| T_{(x,y)} F \|_V \leq C \cdot w(x,y) \| F \|_V \forall F \in V\) it also follows that

\[
\| L_{(x,y,\tau)} F \|_Y \leq C \cdot w(x,y) \| F \|_Y \forall F \in Y. \tag{4.33}
\]

This is obvious for weighted \(L^p\)-spaces and follows directly for the spaces as defined in (4.32), using the identity

\[
|P_T (L_{(x,y,\tau)} F) | = | T_{(x,y)} P_T F |. \tag{4.34}
\]

Furthermore, \(K(H^d)\) is dense in \(Y\) if \(K(R^{2d})\) is dense in \(V\). In this case also continuity of \(h \mapsto L_h F\) (or \(R_h F\)), from \(H^d\) to \(Y\), and \(\| L_h F \|_Y \leq w_a(h) \| F \|_Y\) follows, hence by vector-valued integration \(Y \ast L_w Y \subseteq Y\) (now with Heisenberg convolution), i.e., conditions \((Y1) - (Y2')\) are satisfied.

There remains one last (formal) difference in the definition of generalized modulation spaces and the corresponding coorbit spaces to be discussed. Whereas \(M_{p,q}^m(R^d)\) is defined as a subset of \(S'(R^d)\), the abstract definition selects the elements from the antidual space \((H^d_w)^{-1'}\) \((w = w_a\) in our case). The reasons choosing the definition in this way have been the following ones:

i) For the abstract approach there is no such "natural" reservoir of "distributions" such as \(S'(R^d)\) in the present case.
ii) The generality of coorbit spaces using the spaces \( (H_w^1)^{-r} \) allows to go beyond tempered distributions, and opens the way to the definition of Banach spaces of ultra-distributions.

Fortunately this ambiguity has been settled already by Theorem 4.2.iii in [26], because it is easy to check that \( S'(\mathbb{R}^d) \) is continuously embedded into \( (H_w^1)^{-r} \) for sufficiently strong weights, such as exponential weights on \( \mathbb{H}^d \).

Altogether we have now verified that part (A) is OK. Let us come to part (B). The key to this the intertwining property \( V_g(\pi(h)f) = L_h(V_g f) \) (4.14) of the Gabor transform. It allows to translate the two main problems into questions about functions on \( \mathbb{H}^d \) which satisfy the reproducing convolution equation (4.23). We call this fact the correspondence or transference principle. As we shall see both problems involve coherent systems of functions (see Peremelov [49] for this notation), i.e., families of the form \( (\pi(h_i)g)_{i \in I} \), where \( (h_i)_{i \in I} \) is some discrete family in \( \mathbb{H}^d \) (or the TF-plane \( \mathbb{R}^d \times \mathbb{R}^d \times \{1\} \subseteq \mathbb{H}^d \)). They correspond to families \( (L_{h_i}V_gg)_{i \in I} \) under the Gabor transform. Thus, we may reformulate the main questions in the group theoretical language:

i) Atomic decomposition problem: We are looking for sufficient conditions on \( g \) and \( (h_i)_{i \in I} \) such that any \( F = V_gf \) , with \( f \in L^2(\mathbb{R}^d) \) allows a series representation of the form (writing \( G := V_gg) : F = \sum_{i \in I} c_i L_{h_i}G \).

ii) The irregular sampling problem for the STFT translates by means of (4.29) into the question of reconstructing \( F \) from the family of coefficients

\[
(F(h_i))_{i \in I} = (\langle \pi(h_i)g, f \rangle)_{i \in I} = (\tau e^{2\pi \tau y_i} S_g f(x_i, y_i))_{i \in I}.
\]  

(4.35)

The inversion formula (2.14) then allows to recover \( f \) from \( F \). Of course we shall look for constructive ways of obtaining \( f \) more directly from the given data.

§5. \( S_0(\mathbb{R}^d) \) and Harmonic Analysis

In this section we investigate the space \( M_{1,1}^0(\mathbb{R}^d) \), traditionally denoted as \( S_0(\mathbb{R}^d) \), in more detail. This space is of great interest in window design and can serve as a substitute for the Schwartz space \( S_0(\mathbb{R}^d) \). Since the Banach space \( S_0(\mathbb{R}^d) \) has a much simpler structure than the Frechet space \( S(\mathbb{R}^d) \), an alternative approach to abstract harmonic analysis can be based on \( S_0(\mathbb{R}^d) \) [23]. The space \( S_0(\mathbb{R}^d) \) has a great number equivalent characterizations. We present here mainly those which can be well formulated using notions from signal analysis.

Theorem 15. For \( f \in L^2(\mathbb{R}^d) \) the following conditions are equivalent:

1. For some \( g \in S(\mathbb{R}^d) \), \( g \neq 0 \): \( S_gf \in L^1(\mathbb{R}^{2d}) \), or \( V_gf \in L^1(\mathbb{H}^d) \), i.e., \( g \in Co(L^1) \).
2. For some (or any) non-zero \( g \in S_0(\mathbb{R}^d) \): \( S_gf \in L^1(\mathbb{R}^{2d}) \);
3. \( S_f \in L^1(\mathbb{R}^{2d}) \) (or equivalently \( V_f \in L^1(\mathbb{H}^d) \));
4. The Wigner distribution \( W_f \) is integrable (over the TF-plane \( \mathbb{R}^{2d} \));
5. The ambiguity function \( A_f \) of \( f \) is integrable (over the TF-plane \( \mathbb{R}^{2d} \));
6. For some/any non-zero \( g \in S_0(\mathbb{R}^d) \), \( f \) has a Gabor series representation

\[
 f = \sum_{i \in I} a_i T_{x_i} M_y g, \text{ with } \sum_{i \in I} |a_i| < \infty. \tag{5.1}
\]

7. \( f = \sum_{j \in J} f_j \), with \( f_j \in L^1(\mathbb{R}^d) \) and \( \text{spec}(f_j) \subseteq y_j + Q \) for some fixed set \( Q \) (bounded with nonvoid interior), and \( \sum_{j \in J} \|f_j\|_1 < \infty \).

**Proof:** Since \( |V_\alpha f(x, y, \tau)| = |S_\alpha f(x, y)| \) for \( (x, y) \in \mathbb{R}^{2d} \) the first two characterizations are more or less direct reformulations of the membership of \( f \in \mathcal{H} = L^2(\mathbb{R}^d) \) in \( Co(L^1(\mathbb{H}^d)) \). Since \( L^1(\mathbb{H}^d) \) is isometrically left and right translation invariant the weight function \( w \) required in the general theory can be taken as the trivial one \( (w(h) \equiv 1 \text{ for all } h \in \mathbb{H}^d) \), and therefore the set of admissible vectors \( A_w \) (described by 3.) coincides with \( S_0 = Co(Y) \).

That \( S_f f \in L^1(\mathbb{R}^{2d}) \) if and only if \( W_f \in L^1(\mathbb{R}^{2d}) \) follows from the identity (recall that \( \hat{g}(z) = g(-z) \))

\[
 W(f, g)(x, y) = 2^d S \hat{g}(-2x, -2y)e^{-\pi i(2x)(2y)}, \tag{5.2}
\]

which implies (choosing \( f = g \)) that \( W_f \) is obtained by dilation from

\[
 S_f f(-x, -y)e^{-\pi ixy}.
\]

Therefore integrability of \( W_f \) is equivalent to integrability of \( S_f \) or \( V_f \).

Characterization 6. is just a reformulation of the atomic characterization of coorbit spaces. Note however, that in the \( \ell^1 \) case under discussion it is possible to admit arbitrary point sets \( (x_i, y_i)_{i \in I} \), without assuming that they are separated. On the other hand it is sufficient (according to the general theory) to use sufficiently dense discrete families.

In order to check 7. let \( f \) be given as described, and choose \( g \in \mathcal{S}(\mathbb{R}^d) \subseteq Co(L^1) \) such that \( \hat{g}(y) \equiv 1 \) on \( Q \). Using \( L^1(\mathbb{R}^d) \ast S_0(\mathbb{R}^d) \subseteq S_0(\mathbb{R}^d) \) and

\[
 \|f_j\|_{S_0} = \|M_{y_j}(M_{-y_j} f_j \ast g)\|_{S_0} \leq \|M_{-y_j} f_j\|_1 \|g\|_{S_0} \leq \|f_j\|_1 \|g\|_{S_0}, \tag{5.3}
\]

it follows that the functions as described in 7. can be represented as absolutely convergent series in \( S_0(\mathbb{R}^d) \), and therefore belong to \( S_0(\mathbb{R}^d) \), since

\[
 \|f\|_{S_0} \leq \sum_j \|f_j\|_{S_0} \leq \|g\|_{S_0} \sum_j \|f_j\|_1 < \infty. \tag{5.4}
\]

For the converse let \( f \in S_0(\mathbb{R}^d) \) be given, and choose some \( g \in \mathcal{S}(\mathbb{R}^d) \) with compact spectrum satisfying \( \sum_{n \in \mathbb{Z}^d} T_n \hat{g} \equiv 1 \). It follows for \( f \in S_0(\mathbb{R}^d) \)

\[
 f = \sum_{n \in \mathbb{Z}^d} f \ast M_n g, \text{ with } \sum_{n \in \mathbb{Z}^d} \|f \ast M_n g\|_1 \leq C_2 \cdot \|f\|_{S_0}. \tag{5.5}
\]
In fact, if \( g_1 \in S(\mathbb{R}^d) \) satisfies \( g_1 \equiv 1 \) on \( \text{supp}(\hat{g}) + B_1(0) \) (the ball of radius one around zero), then \( T_{s+n} \hat{g} \cdot T_n \hat{g} = T_n \hat{g} \) for \( |s| \leq 1 \), and therefore

\[
\|M_n g * f\|_1 \leq \int_{n+Q} \|M_s g_1 * M_n g * f\|_1 ds \leq \|M_n g\|_1 \int_{n+Q} \|M_s g_1 * f\|_1 ds
\]

(5.6)

and consequently

\[
\sum_{n \in \mathbb{Z}^d} \|f * M_n g\|_1 \leq \|g\|_1 \sum_{n \in \mathbb{Z}^d} \int_{n+Q} \|M_s g_1 * f\|_1 ds = \|g\|_1 \|S_{g_1, f}\|_1 = C_2 \|f\|_{S_0},
\]

(5.7)

as claimed. \( \blacksquare \)

**Remark 16.** Due to its special properties the Segal algebra \( S_0(\mathbb{R}^d) \) plays a special role among modulation spaces. It shares many properties with the Schwartz space \( S(\mathbb{R}^d) \). Of rapidly decreasing functions, in particular the invariance under the Fourier transform. It is a Banach space, isometrically invariant under time-frequency shifts, and actually the smallest Banach space with this property. Applying the usual duality argument it is possible to define a *generalized Fourier transform* on \( S'_0(\mathbb{R}^d) \). Although \( S'_0(\mathbb{R}^d) \) is a proper subspace of \( S'(\mathbb{R}^d) \) it is large enough to contain all kinds of functions and measures (or distributions) that are of interest in signal analysis, such as Dirac functions, Dirac pulse trains (also called Shah-distributions), or functions in \( L^p \)-spaces on \( \mathbb{R}^d \), for any \( p \geq 1 \). An elementary introduction to harmonic analysis (i.e., an approach which is not based on Lebesgue integration theory) based on the pair \( S_0(\mathbb{R}^d) \) and \( S'_0(\mathbb{R}^d) \) has been proposed in [23]. \( S_0(\mathbb{R}^d) \) (or more generally \( S_0(\mathcal{G}) \), for arbitrary locally compact Abelian groups \( \mathcal{G} \)) has been discovered independently by J.P. Bertrandias [7] and the first author [19], who also was the first to point out the minimality properties of \( S_0(\mathcal{G}) \) and its consequences. The existence of a very weak form of atomic decompositions for \( S_0(\mathcal{G}) \) together with a hint concerning Gabor’s classical expansions for \( L^2 \)-signals (see Gabor [34] ) motivated the search for more general Gabor-type expansions of distributions, which resulted in a characterization of modulation spaces in terms of Gabor expansions (see [22] ). The observed similarity to the construction of Y.Meyers orthogonal (affine) wavelets sparked the development of the general group theoretical approach to atomic decompositions.

An easy consequence of the atomic characterization of \( S_0(\mathbb{R}^d) \) is the “tensor-product stability”. In order to describe it we need the space

\[
S_0(\mathbb{R}^d) \otimes S_0(\mathbb{R}^d) := \{ f \in L^1(\mathbb{R}^{2d}) \mid f(x,y) = \sum_{n=1}^{\infty} f_n(x)g_n(y), \quad \text{with} \quad \sum_{n=1}^{\infty} \|f_n\|_{S_0} \|g_n\|_{S_0} < \infty \}. \quad (5.8)
\]

The infimum over these series expressions defines the natural norm on this space, which will be denoted by \( \|f\|_{S_0 \otimes S_0} \).
Corollary 17. The spaces $S_0(\mathbb{R}^d) \hat{\otimes} S_0(\mathbb{R}^d)$ and $S_0(\mathbb{R}^{2d})$ coincide, and their natural norms are equivalent.

The most direct argument (we leave it to the reader to check details, or to compare with [19]) uses atomic building blocks $g = g_1 \otimes g_2$, with $g_1, g_2 \in S_0(\mathbb{R}^d)$ for the characterization of $S_0(\mathbb{R}^{2d})$, where we use the notation $g(x, y) = g_1 \otimes g_2(x, y) = g_1(x)g_2(y)$.

As the main consequence of this property we shall verify the so-called kernel theorem for $S_0(\mathbb{R}^d)$. The possibility of proving a kernel theorem (announced already in [20]) is remarkable, because usually the "nuclearity" of a topological vector space is the basis for such a result.

Theorem 18. (Kernel Theorem) For every bounded linear operator $T$ from $S_0(\mathbb{R}^d)$ into $S'_0(\mathbb{R}^d)$ there exists exactly one (distributional kernel) $\sigma \in S'_0(\mathbb{R}^{2d})$ such that $Tf(g) = \sigma(f \otimes g)$, for all $f, g \in S_0(\mathbb{R}^d)$.

Proof:

i) Given $\sigma$ it is clear that for fixed $f$ the mapping $\sigma_f: g \mapsto \sigma(f \otimes g)$ defines a bounded linear functional, i.e., some element in $S'_0(\mathbb{R}^d)$. Moreover, the mapping $T_\sigma : f \mapsto \sigma_f$ is linear and satisfies $|\sigma(f \otimes g)| \leq \|\sigma\|s_0 \|f\|s_0 \|g\|s_0$. This allows to estimate the operator norm of $T_\sigma$ by $\|T\| \leq \|\sigma\|s'_0$.

ii) In order obtain $\sigma$ from the operator $T$ it seems to be sufficient to use the tensor product property of $S_0(\mathbb{R}^d)$ and to define $\sigma$ on $f = \sum_{n=1}^\infty f_n \otimes g_n$ through $\sigma(f) := \sum_{n=1}^\infty [T(f_n)](g_n)$. However, it is hard to verify directly that this definition makes sense, i.e., that it is independent from the representation of $f$. It is therefore helpful to remember that a 'kernel theorem' is a distributional analogue of a matrix representation. If $T$ is an integral operator of the form

$$T_h(f)(x) = \int_{\mathbb{R}^d} h(x, y)f(y)dy,$$

for some continuous function $h$, it is possible to recover $h(x, y)$ by calculating $T(\delta_x)(y) = \delta_y(T(\delta_x))$ (in the same way as the columns of a matrix are the images of the unit vectors of the induced linear mapping). Of course this does not work in the general setting, because Dirac measures are not in the domain of $T$ and also the range of $T$ consists in general of distributions (i.e., elements of $S'_0(\mathbb{R}^d)$), so it does not make sense to evaluate them pointwise. However, it is possible to use impulse-like instead of Dirac measures elements from $S_0(\mathbb{R}^d)$, and to go to the limit afterwards (in $S'_0(\mathbb{R}^d)$). This motivates the following procedure:

Fix some Dirac sequence $(\varepsilon_k)_{k=1}^\infty$ in $S_0(\mathbb{R}^d)$. For simplicity we may assume that $\varepsilon_k(x) = 2^{-k}\varepsilon_0(2^{-k}x)$ for $k \geq 1$, with $\varepsilon_0 \in S(\mathbb{R}^d)$ and compact support. We may also assume that $\varepsilon_0(x) \geq 0$ for all $x \in \mathbb{R}^d$ and that $\|\varepsilon_0\|_1 = 1$, and therefore $\|\varepsilon_k\|_1 = 1$ for $k \geq 1$. Using this sequence we are able to define a sequence (of bounded, continuous functions) $h_k$ on $\mathbb{R}^{2d}$ by

$$h_k(u, v) := [T(T_u\varepsilon_k)](T_v(\varepsilon_k))$$

for $k \geq 1$. (5.10)
Since the translation operators $T_x$, $x \in \mathbb{R}^d$, are isometric on $S_0(\mathbb{R}^d)$ and since $x \mapsto T_x(f)$ is a continuous mapping from $\mathbb{R}^d$ into $S_0(\mathbb{R}^d)$, both, boundedness and continuity of each single $h_k$, follows easily from the definition. In general this family of functions on $\mathbb{R}^{2d}$ will be unbounded with respect to the sup-norm. However, we show next that it is bounded with respect to the norm of $S_0(\mathbb{R}^d)$. To this end we denote the functional generated by $h_k$ by $\sigma_k$. Then

$$\sigma_k(f \otimes g) = [T(e_k * f)](e_k * g). \quad (5.11)$$

Since for any fixed $k$ the functions $u \mapsto T_u \tilde{e}_k f(u)$ and $v \mapsto T_v \tilde{e}_k g(v)$ are bounded, continuous and integrable (even in the sense of a vector-valued Riemannian integral), and because both integration and the operator $T$ are continuous mappings, this is verified as follows, for $f, g \in S_0(\mathbb{R}^d)$:

$$\sigma_k(f \otimes g) = \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} h_k(u, v)(f \otimes g)(u, v) =$$

$$= T \left( \int_{\mathbb{R}^d} T_u \tilde{e}_k f(u) du \right) \left( \int_{\mathbb{R}^d} T_v \tilde{e}_k g(v) dv \right) = [T(e_k * f)](e_k * g).$$

Using $\|e_k * f - f\|_{S_0} \to 0$ for $k \to \infty$ it follows therefrom that

$$\sigma_k(f \otimes g) \to T f(g) \text{ for } f, g \in S_0(\mathbb{R}^d). \quad (5.12)$$

Using the tensor product representation we may write any $h \in S_0(\mathbb{R}^{2d})$ as $h = \sum_{n=1}^{\infty} f_n \otimes g_n$, with the appropriate norm estimate. It follows therefrom that

$$|\sigma_k(h)| \leq \sum_{n=1}^{\infty} \|T(e_k * f_n)\|_{S_0(\mathbb{R}^d)} \|e_k * g\|_{S_0(\mathbb{R}^d)} \leq \|T\| \cdot \sum_{n=1}^{\infty} \|f_n\|_{S_0} \|g_n\|_{S_0},$$

for any admissible representation of $h$, and thus for some $C > 0$:

$$|\sigma_k(h)| \leq C \cdot \|T\| \cdot \|h\|_{S_0(\mathbb{R}^{2d})}. \quad (5.14)$$

Furthermore it follows that $\sigma_k(h)$ is a Cauchy-sequence of complex numbers for any $h \in S_0(\mathbb{R}^{2d})$ (this is obvious for functions $h$ which are finite linear combinations of tensor products and follows therefrom for general elements of $S_0(\mathbb{R}^d)$ by approximation). We denote the pointwise (or $w^*$ in the functional analytic description) limit by $\sigma$, i.e.

$$\sigma(h) := \lim_{k \to \infty} \sigma_k(h) \text{ for } h \in S_0(\mathbb{R}^{2d}). \quad (5.15)$$

It now follows from (5.12) that

$$\sigma(h) = \sum_{n=1}^{\infty} [T f_n](g_n) \text{ for } h \in S_0(\mathbb{R}^{2d}). \quad (5.16)$$

Thus, this expression is independent of the representation of $h$ and

$$\|\sigma(h)\|_{S_0(\mathbb{R}^{2d})} \leq C \cdot \|T\|, \quad (5.17)$$
i.e., \( \sigma \) meets all the requirements. □

**Remark 19.** Looking once more at the sequence \((\sigma_k)_{k=1}^{\infty}\) we can relate it to \(\sigma\) more directly by verifying that

\[
\sigma_k(f \otimes g) = [T(e_k * f)](e_k * g) = \sigma((e_k * f) \otimes (e_k * g)) = \\
= \sigma((e_k \otimes e_k) * (f \otimes g)) = [(e_k \otimes e_k) * \sigma](f \otimes g) \text{ for } f, g \in S_0(\mathbb{R}^d),
\]

or in other words

\[
\sigma_k = (e_k \otimes e_k) * \sigma, \quad (5.18)
\]

where the convolution of \(\sigma\) is taken in the sense of \(\mathbb{R}^{2d}\).

**Remark 20.** The above proof can be easily modified to give a proof for the case of general lca. groups.

**Remark 21.** Using Wilson bases (see below) it is possible to show that the space \(S_0(\mathbb{R}^d)\) is isomorphic as a Banach space to \(\ell^1\).

### §6. Proofs and further Results, Frames, Adaptive Weights

**Proof of Theorem 7.**

First observe that a weighted variant of Theorem 15 implies that \(W_g \in L^1_{w_a}(\mathbb{R}^{2d})\) (or the same for \(A_g\)) if and only if \(g \in \mathcal{H}^1_{w_a} = C_0(L^1_{w_a})\). As has been pointed out at the end of section 4 the modulation space \(M^n_{p,q}(\mathbb{R}^d)\) concides with a coorbit space \(C_0(Y)\), where \(Y\) is a suitable weighted mixed norm space on \(\mathbb{H}^d\), with moderate weight \(m\), hence satisfying \((Y1) - (Y2')\). Therefore the general theory applies (Theorem 6.1 in [26]), i.e., for a sufficiently small neighborhood \(U\) of the neutral element in \(\mathbb{H}^d\) any coherent family \(\pi(h_i)g\) can be used as a family of atoms for this coorbit space, i.e., allows a series representation \(f = \sum_{i \in I} \lambda_i \pi_i(g)\) with coefficients \((\lambda_i)\) in the appropriate sequence space.

In the present situation we only have to observe that without loss of generality we may assume that \(U\) is of the form \(W \times D\), where \(W\) and \(D\) are neighborhoods of the \((0, 0) \in \mathbb{R}^{2d}\) and \(1 \in \mathbb{T}\) respectively. If we fix a \(D\)-dense family \((\tau_i)_{i=1}^{\infty}\) in \(\mathbb{T}\), e.g. a collection of roots of unity of sufficiently high order it is clear that \((x_i, y_i, \tau_i)_{i \in I} \) is dense in \(\mathbb{H}^d\) if (and only if) the given family \((x_i, y_i)_{i \in I} \) is \(W\)-dense on \(\mathbb{R}^{2d}\).

Thus, by the general theory \(f\) can be represented as a series

\[
f = \sum_{i, l} \lambda_{i, l} \pi(z_i, \tau_l)g, \quad (6.1)
\]

with complex coefficients \((\lambda_{i, l})\) in a sequence space associated with the weighted mixed norm space \(Y\) generating the coorbit space.

If the additional partial lattice structure (or if more general conditions of a similar kind are fulfilled) this sequence space has the simple natural description used in the theorem. We have to leave this verification to the interested reader.
Observe, however, that we do not assume that the sampling set is a lattice by itself.

As a last argument in order to obtain the required series representation we add up over \( l \), i.e., we set \( c_i := \sum_{l=1}^{n} \tilde{\eta}_l \lambda_{i,l} \), in order to obtain

\[
f = \sum_i \sum_l \lambda_{i,l} \tilde{\eta}_l T_{x_l} M_{y_l} g = \sum_i c_i T_{x_i} M_{y_i} g,
\]

with \((c_i)_i \) being in the same weighted mixed norm sequence space. ■

Proof of Theorem 10.
Once more we observe that knowing the STFT on some \( \bar{W} \)-dense set is the same as knowing it on \( \bar{W} \times \mathcal{T} \), hence on a \( \bar{U} := \bar{W} \times \mathcal{T} \)-dense subset of \( \mathbb{H}^d \). By the general theory (see [38], Theorem S and Theorem U for details) the operator \( A := D^+_\psi \) (as well as other operators of a similar form) will do the reconstruction job. It is given by

\[
AF := \sum_{i \in I} F(h_i) w_i L_{h_i} G,
\]

where the sequence of weights \((w_i)_{i \in I}\) has to be chosen in an appropriate way. Since we choose them according to the local density of points, i.e., adaptively with respect to the distribution of sampling points, we call the resulting iterative method the adaptive weights method. The key estimate, which also decides about the size of \( \bar{W} \), is then of the following form (we have to omit the technical details here):

\[
\| D^+_\psi F * G - F * G \|_Y < \gamma \| F \|_Y
\]

for some \( \gamma < 1 \) and all \( F \in Y \), if \( \bar{W} \) is sufficiently small.

However, we would like to mention here that the formula

\[
S_g f(x, y) M_y T_y g = V_g f(z) \pi(z) g, \quad \text{for } z = (x, y) \in \mathbb{R}^{2d}.
\]

can be used to reformulate the reconstruction algorithm in a way avoiding the use of the torus component entirely in the description of the method. ■

At this point few remarks concerning the required sampling density, i.e., the size of the sets \( \bar{W} \) and \( \bar{W} \) above are in order.

For general \( g \) and general modulation spaces it is still an open problem to find good estimates. For the reader’s convenience we list some results in this direction:

- In Daubechies [11] (fairly complicated) estimates on \( \alpha, \beta > 0 \) are derived for a given \( g \in L^2 \), so that the family \( \{ e^{2\pi i k \alpha z} g(x - n \beta) \}_{k,n} \) constitutes a frame for \( L^2(\mathbb{R}^d) \).

- For the case of regular sampling sequences \( S_g(k\alpha, n\beta) \) of the STFT it is known that \( 0 < \alpha \beta \leq 1 \) is a necessary condition for stable reconstruction in \( L^2(\mathbb{R}^d) \).
If \( g(x) = e^{-\pi x^2} \), then \( \{ e^{2\pi i k \alpha x} g(x - n\beta) \}_{k, n \in \mathbb{Z}} \) is a frame for \( L^2(\mathbb{R}^d) \) if and only if \( 0 < \alpha \beta < 1 \), (see Seip/Wallsten [56]).

Let \( Z_g(\tau, \omega) = \sum_{k \in \mathbb{Z}} f(\tau + k)e^{2\pi i k \omega} \) be the Zak transform (or Weil-Brezin transform) of \( g \in S_0(\mathbb{R}) \). If for some \( g \in S \) there exists \( N \geq 2, A, B > 0 \), so that

\[
0 < A \leq \sum_{j=0}^{N-1} |Z_g(\tau, \omega - j/N)|^2 \leq B < \infty, \tag{6.6}
\]

then all modulation spaces \( M_{p,q}^m \) have Gabor wavelet expansions w.r.t. \( \{ e^{2\pi i k \alpha x} g(x - n\beta) \}_{k, n \in \mathbb{Z}} \), whenever \( \alpha \beta = 1/N \), see Walnut [62].

Daubechies/Jaffard/Journe [15] prove the existence of functions \( g \in S \) so that both \( g \) and \( \hat{g} \) are of exponential decay, i.e., satisfy

\[
|g(x)| \leq C e^{-a|x|}, |\hat{g}(\xi)| \leq C' e^{-b|\xi|},
\]

so that the following simple linear combinations of Gabor wavelets \( g_{l,n}, l \in \mathbb{N}, n \in \mathbb{Z} \) constitute an orthonormal basis for \( L^2(\mathbb{R}^d) \): \( g_{0,n}(x) = g(x - n) \) and

\[
g_{l,n}(x) = \sqrt{2} g(x - n/2) \cos(2\pi lx) \text{ if } l \neq 0, l + n \in 2\mathbb{Z}, \tag{6.7}
\]

\[
g_{l,n}(x) = \sqrt{2} g(x - n/2) \sin(2\pi lx) \text{ if } l \neq 0, l + n \in 2\mathbb{Z} + 1. \tag{6.8}
\]

In [31] it is show that \( \{ g_{l,n} \}_{l \geq 0, n \in \mathbb{Z}} \), also constitutes an unconditional basis for all modulation spaces \( M_{p,q}^m \) (and all spaces \( W_{p,q}^m \)). With a little more work it can be shown ([37]) that the collection \( \{ g_{l,n} \}_{l \geq 0, n \in \mathbb{Z}} \), where \( g(x) = e^{-\pi x^2} \), is also an unconditional basis for the modulation spaces. In this case, however, the biorthogonal system is hard to compute whereas in the former case the coefficient of \( g_{l,n} \) is just \( (f, g_{l,n}) \).

If the Gabor wavelet \( g \) has compact support or if \( g \) is band-limited then it is easy to give sufficient conditions on the sampling density for \( S_{g,f} \) for certain sampling sets:

Assume that \( \text{supp}(\hat{g}) \subseteq [-\omega, \omega] \). Let \( (y_n) \) be an increasing sequence so that for two constants \( A, B > 0 \)

\[
0 < A \leq \sum_{n} |\hat{g}(\xi - 2\pi y_n)|^2 \leq B < \infty. \tag{6.9}
\]

For each \( n \) let \( (x_{k,n})_{k \in \mathbb{Z}} \) be an increasing sequence so that \( \text{supp}_k (x_{k+1,n} - x_{k,n}) < \pi/\omega \) is true for all \( n \). Then

\[
\{ e^{2\pi i y_n x} g(x - x_{k,n}) \}_{k, n \in \mathbb{Z}} \tag{6.10}
\]

is a frame for \( L^2(\mathbb{R}^d) \).

We think that these known results give sufficient evidence that the required sampling density in Theorem 10 is close to the best possible value and that the condition "sufficiently dense" does not pose any problem in the application of these results. Having established the main results as special cases of
theorems which hold true for a much more general situation (described in group theoretical terms) we give next a short explanation of frames. We shall discuss the advantages of the adaptive weights method over the frame approach which works for Hilbert spaces only, and present the functional analytic foundation (omitting technical details) of the new approach as "Main Lemma".

We start with the observation that the reproducing equation (4.8) and (4.22) establish the connection between our two problems by

\[ F(h_i) = F \star G(h_i) = \langle L_{h_i} G, F \rangle \quad \text{since} \quad G = G^*. \quad (6.11) \]

Based on this fact one can solve both questions simultaneously using the theory of frames. The notion of frames goes back to the theory of non-harmonic Fourier series (see Duffin/Schaeffer [18], and Young [64]), but has become popular recently in connection with wavelet theory (e.g. [13],[14], ...). The family \((L_{h_i} G)_{i \in I}\) is a frame for the Hilbert space \(\mathcal{H} := \{ F \in L^2(\mathbb{R}^d) | F \star G = F \}\), i.e., that there are positive constants \(A, B > 0\) such that

\[ A\|f\|^2 \leq \sum_{i \in I} |\langle L_{h_i} G, F \rangle|^2 \leq B\|f\|^2 \quad \forall \, F \in \mathcal{H}, \quad (6.12) \]

The optimal values are called lower and upper frame constant. On \(\mathcal{H}\) the so-called frame-operator \(S\), given by

\[ S(F) := \sum_{i \in I} \langle L_{h_i} G, F \rangle \cdot L_{h_i} G = \sum_{i \in I} F(h_i) L_{h_i} G \quad (6.13) \]

is invertible because it satisfies the operator inequality \(A \cdot Id_{\mathcal{H}} \leq S \leq B \cdot Id_{\mathcal{H}},\) and thus \((\lambda S)^{-1} = \sum_{n=0}^{\infty} (Id - \lambda S)^n\) is well defined through Neumann's series for a certain range of values \(\lambda\), the optimal value being \(\lambda = 2/(A + B)\) (see Daubechies [14]). The constant \(\lambda\) may be considered as a relaxation parameter for the recursive description of the partial sums arising by this method, which starts with \(F^{(0)} := \lambda SF\) and is given as

\[ F^{(n+1)} = (Id - \lambda S)F^{(n)} + \lambda SF = F^{(n)} + \lambda S(F - F^{(n)}) \quad \text{for} \quad n \geq 1. \quad (6.14) \]

The rate of convergence can be estimated using the operator norm:

\[ \|F - F^{(n)}\| \leq C\|Id - \lambda S\|^n \quad \text{for} \quad n \geq 1. \quad (6.15) \]

Furthermore \(SF\) and therefore \(F^{(n)}\) for any \(n \geq 1\) can be obtained from the knowledge of the family \((F(h_i))_{i \in I}\) alone. Finally the formula

\[ F = S(S^{-1} F) = \sum_{i \in I} \langle L_{h_i} G, S^{-1} F \rangle L_{h_i} G, \quad (6.16) \]

shows that \(F\) can be represented as a series with atoms \(L_{h_i} G\).
Disadvantages of the Frame Approach

a) Only for special cases the frame bounds \( A, B \) have been estimated [11], but in general they are difficult to determine or even good estimates may not be available. As a consequence the relaxation parameter \( \lambda \) and the speed of convergence have to be determined by expensive trial and error experiments.

b) For very irregular sampling geometries, where the local density of the sampling points varies significantly within the TF-plane, the frame bounds are necessarily far apart. This entails extremely slow convergence. To improve the situation one might even throw away some information, but clearly this makes the reconstruction less stable and more sensitive to noise.

c) The frame approach works only for Hilbert spaces. In order to deal with more sensitive norms additional arguments, as given in [38] have to be given. Again they are based on group-theoretic concepts.

The Adaptive Weights Methods

The problems mentioned above suggests to look for alternatives with improved speed of convergence and stability. This also means that one should look for proofs which allow to check convergence with respect to other norms, not just the global \( L^2 \)-norm. This family of norms should include at least the weighted \( L^p \)-spaces (on the group level), i.e., convergence should be verified for the norm of coorbit spaces such as \( M \)- and \( W \)-spaces.

Convergence of several iterative methods which work for a variety of norms can be derived from the following Main Lemma.

**Lemma 22.** Assume that \( F \mapsto F \ast G \) is a bounded linear operator on a Banach space \((Y, \| \cdot \|_Y)\) of functions on a locally compact group \( G \). If \( A \) is a bounded, linear operator on \((Y, \| \cdot \|_Y)\) such that there exists some \( \gamma < 1 \) with

\[
\| F \ast G - AF \|_Y \leq \gamma \| F \|_Y \quad \forall \ F \in Y.
\]  

(6.17)

Then any \( F \) satisfying \( F = F \ast G \) can be recovered from \( AF \) by the recursion

\[
F^{(0)} = AF, \quad F^{(n+1)} = F^{(n)} \ast G + AF - AF^{(n)} \quad \forall \ n \geq 1.
\]  

(6.18)

Moreover, we have

\[
\| F - F^{(n)} \|_Y \leq \gamma^n \| F \|_Y \quad \forall \ n \geq 1.
\]  

(6.19)

Alternatively, one can show that there is a bounded linear operator \( D \) on \( Y \) such that \( F = D(AF) \) if \( F = F \ast G \). If furthermore \( G = G \ast G \) and \( AF = AF \ast G \) for all \( F \in Y \) then it also true that \( F = A(DF) \).

**Proof:** Write \( RF := F \ast G - AF \), or \( F \ast G = AF + RF \). By induction (replacing \( RF \) by \( R(F \ast G) = R(AF + RF) \) over and over again we derive therefrom

\[
F = F \ast G = \left( \sum_{k=0}^{n} R^k \right) AF + R^{n+1}F \quad \text{for} \ n \geq 1.
\]  

(6.20)
The sequence $F^{(n)} := (\sum_{k=0}^{n} R^k) AF$ can be described recursively through

$$
F^{(0)} = AF, \quad F^{(n+1)} = RF^{(n)} + AF = F^{(n)} * G + A(F - F^{(n)}). \quad (6.21)
$$

Since by assumption

$$
\| R^{n+1} F \|_Y \leq \gamma \| R^n F \|_Y \leq \gamma^{n+1} \| F \|_Y \quad \text{for } n \geq 1. \quad (6.22)
$$

the operator $D := \sum_{k=0}^{\infty} R^k$ is well defined and satisfies $D(AF) = F$ if $F = F * G \in B$.

Assuming now the additional condition one may apply the splitting operation $F = AF + RF$ also to $R^k F$ for any $k \geq 1$ and obtains by induction

$$
F = F * G = A \left( \sum_{k=0}^{n} R^k F \right) + R^{n+1}. \quad (6.23)
$$

Going with $n$ to infinity gives the result. $\blacksquare$

Remark 23. The proof also shows that $DF * G = DF$ under the additional assumption.

According to our experience the most simple and efficient approximation is obtained by (6.3). A simple recipe concerning the choice of the family of weights $(w_i)_{i \in I}$ is the following one: Think of a partition of the Heisenberg-group of the form $(P_i)_{i \in I}$, the sets $P_i$ being as small as possible, with $h_i \in P_i$. One might take for example as $P_i$ the set of nearest neighbors of $h_i$. Then a good choice is to take as weights $w_i$ the Haar measure (i.e., practically the 3D-volume) of $P_i$. In general it seems to be only important that the weights are small for areas of high density of $(h_i)_{i \in I}$ and larger elsewhere.

We do not have enough space here to demonstrate how one can obtain the required estimates in order to apply the Main Lemma. Let us just indicate that a very natural argument can be based on the fact that the approximation operator $A$ is a Riemannian sum to the vector-valued integral describing the convolution product $F * G$ (see (4.3)). Thus, convergence of those Riemannian sums with respect to a variety of norms stands in the background of our method. It is therefore not limited to the Hilbert space case. The group theoretical description also allows to make statements about uniformity with respect to the possible families $(h_i)_{i \in I} \in \mathbb{H}^d$. For example, one obtains uniform estimates on the speed of convergence for all families $(h_i)_{i \in I}$ which have the same density, i.e., for which the family $(h_i \cdot Q)_{i \in I}$ defines a covering of $\mathbb{H}^d$ the same (small) set $Q \subseteq \mathbb{H}^d$.

Next we indicate how the Lemma can be used to answer the main questions. Observe first that the sequence defined iteratively in the Main Lemma is convergent to $F$ at a geometric rate. Since $AF$ can be built if only the sampling values $(F(h_i))_{i \in I}$ are known the irregular sampling problem has been solved based on the first part of the lemma.
In order to verify the result concerning Gabor wavelet expansions we note that the choice of $A$ and the equation $G = G \ast G$ imply that the additional assumptions of the Lemma are fulfilled as well. Since any function of the form $AF$, for some $F \in Y$ has the required series representation the argument is complete.

We have to mention here that the convolution relations for the so-called Wiener amalgam spaces are a useful tool to verify that the sequence of coefficients in the series representation, which is just $(w_i F(h_i))_{i \in I}$, belongs to the sequence space corresponding the $Y$ in a natural way. This can be shown using once more the reproducing property $F \ast G = F$ and the fact that $G$ is a smooth function on $\mathbb{R}^{2d}$ for ‘nice’ analyzing Gabor wavelets, such as $g \in S_0(\mathbb{R}^d)$. However, we also have to omit a detailed discussion of the verification that the natural sequence spaces which correspond to a mixed norm space on $\mathbb{H}^d$ are the obvious ones.

We have to skip the discussion technical details of the crucial estimate: $\|F \ast G - AF\| \leq \gamma \|F\|$, for some $\gamma < 1$ and for all $F \in Y$, and refer to [26], etc. for details. each step produces a series in the term $L_{h_i} G$ which finally leads to the required series representation. It is also possible and sometimes convenient from a practical point of view that the recursion can also be described directly in terms of the coefficients:

$$f = \sum_{i \in I} \lambda_i \pi(z_i)g$$  \hspace{1cm} (6.24)

we only have to recall that we start with the family $\Lambda^1$, with $\Lambda^1_i = F(z_i)w_i$, for $i \in I$, and that the recursion delivers (for the coefficients):

$$\Lambda^{n+1} = \Lambda^n + ((F(z_i) - F^n(z_i)) \cdot w_i)_{i \in I}.$$  \hspace{1cm} (6.25)

The geometric decay of the error terms in the recursion guarantees that $\Lambda = \lim_{n \to \infty} \Lambda^n$ is an appropriate set of coefficients for (6.24) to hold true.

We have now shown that the general group theoretic approach can be used to give a positive answer to both of our main questions, not only for the Hilbert space case, but for a large family of norms. Now we want to indicate, how the convolution relations can be used to quickly answer a variety of related questions with almost no extra effort. First we will verify the independence of the definition of the $M^{m}_{p,q}(\mathbb{R}^d)$ spaces from the analyzing function $g$.

**Lemma 24.** The definition of $M^{m}_{p,q}(\mathbb{R}^d)$ is independent of the choice of the non-zero analyzing function $g$, i.e., two non-zero functions $g$ and $\tilde{g}$ with $W_g$ and $W_{\tilde{g}} \in L^1_{w_1}(\mathbb{R}^{2d})$ define the same spaces and equivalent norms on $M^{m}_{p,q}(\mathbb{R}^d)$.

**Proof:** We have seen that membership of $S_g f$ in the defining weighted, mixed norm space over the TF-plane is equivalent to membership of $V_{g \ast f}$ in some Banach function space $Y$ on $\mathbb{H}^d$, which satisfies (Y1) and $Y \ast L^1_{w_1} \subseteq Y$. 

Assuming that $V_g f \in Y$ we want to verify that $V_{\tilde{g}} f \in Y$. It follows directly from the convolution formula (4.22) that

$$V_g f \ast V_{\tilde{g}} \tilde{g} = \langle g, \tilde{g} \rangle \cdot V_{\tilde{g}} f \subseteq Y \ast L^1_{w_a} \subseteq Y.$$  \hspace{1cm} (6.26)

If $\langle g, \tilde{g} \rangle \neq 0$ this implies $V_{\tilde{g}} f \in Y$. Otherwise we may replace $\tilde{g}$ by $\pi(h_0)\tilde{g}$. Since $\langle \pi(h_0)\tilde{g}, g \rangle \neq 0$ for at least some $h_0 \in IH^d$ we obtain in any case that $V_{\pi(h_0)\tilde{g}} f \in Y$ for some $h_0 \in IH^d$. But it follows therefrom that

$$V_{\pi(h_0)\tilde{g}} f(h) = \langle \pi(h)\pi(h_0)g, \tilde{g} \rangle = \langle \pi(\pi(h_0)g), \tilde{g} \rangle = V_{\tilde{g}} f(hh_0),$$ \hspace{1cm} (6.27)

i.e., some right translate (in the sense of $IH^d$) of $V_{\tilde{g}} f$ belongs to $Y$, and therefore $V_{\tilde{g}} f$ itself, since $Y$ is left and right translation invariant.

The redundancy (linear dependence) of a coherent system of Gabor wavelets implies that the coefficients for a Gabor series representation of a signal is in general not uniquely determined. However, we will show that sometimes uniqueness can be achieved. Recall that we have shown in Theorem 7.3. [27] that it is possible to derive interpolation results for generalized wavelet transforms, if the sampling points are well separated. In order to motivate the name of the following theorem let us look at the following 1D-situation: Think of the signal $f$ as an acoustic signal, produced by a piano player, who is only able to produce tones at a well-defined (and known) scale of frequencies, and assume further for simplicity of the model that everything can be thought to happen on a discrete time-scale. Then it should be possible to reconstruct the scores from the acoustic signal, i.e., to obtain the unique Gabor expansion of the signal from the signal itself. Of course we have lost (with this restricted set of atoms) the ability to represent arbitrary signals, but nobody would try to produce an arbitrary sound by means of an ordinary piano.

**Theorem 25. (Piano Reconstruction Theorem)** Consider the family of $M^{m}_{p,q}(IR^d)$-spaces described in Theorem 7 (for fixed $C > 0$ and $a \geq 0$), and fix also $g \in \mathcal{H}^1_{w_a}$. Then there exists a compact set $S$, such that for any fixed family $(z_i)_{i \in I}$ in the TF-plane which is $S$-separated, i.e., which satisfies $(z_i + S) \cap (z_j + S) = \emptyset$ for $i \neq j$ one has: If $f \in M^{m}_{p,q}(IR^d)$ has a representation as

$$f = \sum_{i \in I} c_i T_{x_i} M_{y_i} g,$$ subject to $sup_{i \in I} |c_i|\omega_a^{-1}(z_i) < \infty$ \hspace{1cm} (6.28)

the sequence of coefficient is uniquely determined (assuming that the side-condition is satisfied).

**Proof:** The proof of Theorem 7.3 of [27] indicates how the correct coefficients for the atomic decomposition of $f$ can be obtained constructively. Starting with the sampling values of the STFT one just has to iterate in a way similar to the general atomic representation algorithm. The extra information that only certain building blocks will be required in the atomic representation can be used in the algorithm by simply setting the other coefficients to zero at each step.
§7. Stability and invariance properties

In this section we indicate that the algorithms derived from our approach can be shown to be stable with respect to minor perturbations. Later on in this section we shall discuss the invariance properties of coorbit spaces, in particular, of $M$- and $W$-spaces. We start showing that the STFT changes only a little bit if the window is not modified too much (in the sense of $S_0(\mathbb{R}^d)$). For simplicity we formulate the result for unweighted spaces $M^0_{p,p}(\mathbb{R}^d)$.

Lemma 26. Let $g \in S_0(\mathbb{R}^d)$ be given. Then for any $\varepsilon > 0$ there exists some $\delta > 0$ such that for all $\tilde{g} \in S_0(\mathbb{R}^d)$ with $\|g - \tilde{g}\|_{S_0} \leq \delta$ one has

$$\|S_g f - S_{\tilde{g}} f\|_{L^p(\mathbb{R}^{2d})} < \varepsilon \|S_g f\|_{L^p(\mathbb{R}^{2d})},$$

for any distribution $f$ with STFT $S_g f \in L^p(\mathbb{R}^{2d})$.

Proof: Since we are free to choose the analyzing wavelet we may suppose that $g_0 \in S(\mathbb{R}^d)$ is some non-zero Schwartz-function with $\|g_0\|_2 = 1$. The involution $F \mapsto F^*$ being isometric on $L^1(\mathbb{H}^d)$ and since $(V_g f)^* = V_f(g)$ (see (4.15)) our assumption is equivalent to

$$\|V_g g_0 - V_{g_0} g_0\|_1 \leq \delta. \quad (7.1)$$

It follows therefrom by means of (4.22)

$$\|V_g f - V_{g_0} f\|_p = \|V_{g_0} f * V_g g_0 - V_{g_0} f * V_{g_0} g_0\|_p$$

$$\leq \|V_{g_0} f\|_p \|V_g g_0 - V_{g_0} g_0\|_1 \leq \delta \|V_{g_0} f\|_p. \quad (7.2)$$

Since the $p$-norm of $S_g f$ over the TF-plane is a norm equivalent to the coorbit norm, i.e., the norm of $V_{g_0} f$ in $L^p(\mathbb{H}^d)$ the result is proved.

Modifying the window $g$ only a little bit in the $S_0$-sense has another consequence, which is more important. The $L^p$-norm of the corresponding STFT changes only a little bit, it is also true that for such function (i.e., for $f \in M^0_{p,p}(\mathbb{R}^d)$) the sampling sequence over any relatively separated set (e.g. on which is carried by some fine lattice as described in Theorem 10) will only undergo a small change with respect to the $\ell^p$-norm. Therefore imprecise knowledge concerning the window, which has been used for the calculation of the STFT, will not completely destroy the capability of our algorithms to reconstruct the signal from irregular sampling values of the STFT at least approximately. The same is true concerning the jitter error which arises, if sampling values $S_g f(\tilde{z}_i)_{i \in I}$ with $\tilde{z}_i = f(\tilde{x}_i, \tilde{y}_i)_{i \in I}$ are used instead of the sampling values at the points $(z_i)_{i \in I}$. This question comes up if only imprecise information about the sampling positions is available.

For simplicity we state the stability results (also indicating that such errors can be handled simultaneously) for both types of errors in one theorem (see [26], Theorem 6.5 for a proof, and [63] for a special case).
Theorem 27. (Stability Theorem) Consider the family of $M_{p,q}^m(\mathbb{R}^d)$-spaces as in Theorem 7, and let $\varepsilon > 0$ be given. Then there exists some $\delta > 0$ such that $\|g - \tilde{g}\|_{M_{p,q}^m(\mathbb{R}^d)} \leq \delta$ and $|z_i - \tilde{z}_i| \leq \delta$ implies that for any $f$ in $M_{p,q}^m(\mathbb{R}^d)$ the distribution $\tilde{f}$ reconstructed from the sampling sequence $S_\delta(z_i)$ satisfies

$$\|f - \tilde{f}\|_{M_{p,q}^m(\mathbb{R}^d)} \leq \varepsilon \|f\|_{M_{p,q}^m(\mathbb{R}^d)}.$$  \hfill (7.3)

Remark 28. For the purpose of better numerical treatment it is often desirable to replace a window $g$ with unbounded support by some compactly supported window $\tilde{g}$. That ordinary truncation is not a good idea will be clear to everyone familiar with Fourier analysis, as the gaps arising at the truncation points will introduce bad decay of the STFT $S_g f$, independently of the smoothness of $f$. However, if $\Psi = (\psi_i)_{i \in I}$ is a bounded partition of unity in the Fourier algebra $F L^1$, then the finite partial sums $\tilde{g} := \sum_{i \in F} g \psi_i$ will approximate $g \in S_0(\mathbb{R}^d)$ in the sense of $S_0(\mathbb{R}^d)$. In particular, this holds for a partition of unity built up by triangular functions over $\mathbb{R}$. For higher smoothness partitions of unity arising from B-splines of higher order are useful.

We have already shown that the coorbit spaces are invariant with respect to time/frequency shifts. For the subsequent discussion of further invariance properties of the modulation spaces it is more convenient to use the Heisenberg group with the symmetric multiplication

$$(x_1, y_1, \tau_1) \cdot (x_2, y_2, \tau_2) = (x_1 + x_2, y_1 + y_2, \tau_1 \tau_2 e^{i\pi(x_1 y_2 - x_2 y_1)})$$  \hfill (7.4)

and unitary representation $\pi'(x, y, \tau)f(t) = e^{i\pi xy} e^{2\pi i yt} f(t - x), f \in L^2(\mathbb{R})$.

It is then easily seen that $(x, y, \tau) \to (x, y, \tau e^{i\pi x y})$ defines an isomorphism between the old and the new multiplication. Since $< f, \pi'(x, y, \tau)g >$ and $V_g(f)(x, y, \tau)$ differ only by a phase factor, neither the definition of the modulation spaces nor any of the results do change.

The interesting part of the automorphism group of automorphism of $\mathbb{H}^d$ can now be described as follows (see [33] for details): Let $A$ be a $2n \times 2n$-matrix, which leaves the bilinear form $(x_1, y_1), (x_2, y_2) \to x_1 \cdot y_2 - x_2 \cdot y_1$, $x_i, y_i \in \mathbb{R}^{2d}$ invariant, i.e., a symplectic matrix. Then $h_A(x, y, \tau) = (A(x, y), \tau)$ for $(x, y, \tau) \in \mathbb{H}^d$ is an automorphism of $\mathbb{H}^d$, and $h_A$ leaves the center $0 \times 0 \times T$ of $\mathbb{H}^d$ invariant. Since the restrictions of the representations $\pi'(x, y, \tau)$ and $\pi' \circ h_A(x, y, \tau)$ to the center are identical, $\pi'$ and $\pi' \circ h_A$ are equivalent by the Stone-von Neumann uniqueness theorem. Thus, there exist unitary operators $\sigma(A)$, so that

$$\pi' \circ h_A(x, y, \tau) = \sigma(A) \pi'(x, y, \tau) \sigma(A)^{-1} \forall (x, y, \tau) \in \mathbb{H}^d.$$  \hfill (7.5)

The operators $\sigma(A)$ are called metaplectic operators and are determined only up to a phase factor. It can be shown that $A \to \sigma(A)$ can be defined in such a way that it yields a unitary representation of a two-fold covering group of the symplectic group (see Schempp [54] and Reiter [53] for facts concerning the metaplectic group).
For a large class of symplectic matrices the intertwining operators $\sigma(A)$ can be written out explicitly (see Folland [33] for an excellent exposition).

If $A = \begin{pmatrix} C & 0 \\ 0 & (C^*)^{-1} \end{pmatrix}$ for some $C \in GL(n, \mathbb{R})$, then

$$\sigma(A)f(x) = |\det A|^{{1/2}}f(A^{{1/2}}x).$$

(7.6)

For $A = \begin{pmatrix} 0 & I \\ I & 0 \end{pmatrix}$, one obtains $\sigma(A)f = \hat{f}$, the Fourier transform.

If $A = \begin{pmatrix} I & 0 \\ C & I \end{pmatrix}$ for some symmetric $n \times n$-matrix $C$, then

$$\sigma(A)f(x) = \pm e^{{i\pi x \cdot C x}} f(x),$$

(7.7)

i.e., multiplication with a chirp function.

Originally these operators are defined on $L^2(\mathbb{R}^d)$. It is an important consequence of a general principle ([26], Thm.4.8) that the metaplectic operators can be extended to modulation spaces and leave an important class of them invariant. On an intuitive level the invariance statements assert that the time-frequency concentration of a signal is not affected by the action of the metaplectic representation. In the following theorem we collect some mapping properties of the $\sigma(A)$.

**Theorem 29.**

(i) If the $a$-moderate function $m(x, y)$ satisfies $m(A(x, y)) \leq Cm(x, y)$ for some $C > 0$ and $\forall (x, y) \in \mathbb{R}^{2d}$, then $\sigma(A)$ extends to an isomorphism of $M_{pp}^m$ (onto $M_{p,p}^m$).

(ii) If in particular $w \equiv 1$ then $M_{pp}^0$, $1 \leq p \leq \infty$, is invariant under all $\sigma(A)$.

(iii) Let $\tilde{m}(x, y) = m(-y, x)$, then the Fourier transform establishes an isomorphism between the spaces $M_{p,q}^m$ and $W_{q,p}^{\tilde{w}}$.

(iv) If $m(x, y) = m_0(x)$, then $M_{p,q}^{m_0}$ is invariant under multiplication by chirps $e^{{2\pi i x \cdot C x}}$.

(v) In dimension $n = 1$, the operators $\sigma(A)$, where $A$ is an orthogonal matrix, are called Mehler transforms (They are diagonalized by the Hermite functions $h_n$, i.e., if $U_\vartheta = \begin{pmatrix} \cos \vartheta & \sin \vartheta \\ -\sin \vartheta & \cos \vartheta \end{pmatrix}$, then $\sigma(U_\vartheta)h_n = e^{i\vartheta n}h_n$, and form a one parameter group which contains the Fourier transform). If $w(x, y)$ is rotation invariant, then $M_{p,p}^m$ is invariant under $\sigma(U_\vartheta)$, $0 \leq \vartheta < 2\pi$.

**Remark 30.** Ad (i): A typical family of rotations invariant weights with polynomial growth of arbitrary order is given by $w_n(x, y) = (1 + |x| + |y|^\alpha$. For $\alpha \geq 0$ these weights satisfy $w(A(x, y)) \leq C_A w(x, y)$ for all $A \in GL(2n, \mathbb{R})$.

**Remark 31.** Besides the abstract proof contained in [26], Thm.4.6. there are several other, often involved proofs of the invariance of $S_0$ or $M_{1,1}^m$ under chirps.
In order to give the flavor of the proof of the general intertwining principle, we
give a short and direct proof for the chirp invariance of $S_0$ and $M_{1,1}^w$.

Let $M_g(x) = e^{i\pi x \cdot C x} g(x)$ be the "chirp operator" for a symmetric $n \times n$
matrix $C$. Then

$$S_{M_g}(M_g)(x, y) = \int e^{-2\pi i yt} e^{-i\pi (t-x) \cdot C (t-x) y} g(t-x) e^{-i\pi t \cdot C t} g(t) \, dt$$

$$= e^{i\pi x \cdot C x} \int e^{-2\pi i (y+C x) t} g(t-x) g(t) \, dt =$$

$$= e^{i\pi x \cdot C x} S_y(g)(x, y - C x)$$

Now using the characterization 3. of Theorem 15 for $S_0(\mathbb{R}^d)$ or $M_{1,1}^w(\mathbb{R}^d)$ with
$w(x, y) = w_0(x)$ and the above computation we obtain

$$e^{-i\pi x \cdot C x} g(x) \in S_0(\text{ resp. } M_{1,1}^w) \iff S_{M_g}(M_g)(x, y) L^1(\mathbb{R}^{2d}) (\text{ resp. } L^1_w(\mathbb{R}^{2d}))$$

$$\iff S_y(g) \in L^1(\mathbb{R}^{2d}) (\text{ resp. } L^1_w(\mathbb{R}^{2d}))$$

$$\iff g \in S_0 (\text{ resp. } M_{1,1}^w)).$$

Another instance of the intertwining theorem (Thm.4.8 in [26]) occurs for
the Bargmann–Fock spaces. These are spaces of entire functions in $\mathcal{G}^n$ defined by

$$\mathcal{F}_{p,q}^m(\mathcal{G}^n) = \{ F \text{ entire: } \int \left( \int |F(x+iy)|^p m(x, y)^p e^{-p|x|^2/2} \, dx \right)^{q/p} \, dy < \infty \}$$

(7.8)

where $1 \leq p, q \leq \infty$, $m$ is an $a$-moderate function and $z = x + iy \in \mathcal{G}^n$.

$\mathcal{F}^2$ is a Hilbert space on which the Heisenberg group $H^d$ acts by means
of the Fock representation $\beta$. Writing $w = r + is \in \mathcal{G}^n$, then $\beta$ is

$$[\beta(r, s, \tau) F](x) = \tau e^{i\pi \bar{w} \cdot z} e^{-\pi |w|^2/2} F(z - w)$$

(7.9)

for $F \in \mathcal{F}^2$. Since $\beta(0, 0, \tau)$ is just multiplication by $\tau$, by the Stone-von Neumann
uniqueness theorem $\beta$ and the Schrödinger representation $\pi$ must be equivalent, i.e., there is a unitary operator

$$B : L^2(\mathbb{R}) \rightarrow \mathcal{F}^2, \text{ so that } B \circ \pi(r, s, \tau) = \beta(r, s, \tau) \circ B.$$  

(7.10)

By the intertwining property (4.11) $B$ extends to the modulation spaces and
this extension is an isomorphism between $M_{p,q}^m$ and $\mathcal{F}_{p,q}^m$. Consequently all
results for modulation spaces have a corresponding version in the Bargmann–
Fock spaces.
Before we state the series expansions and sampling theorems for the $F_{p,q}^m$, let us look at the so-called Bargmann transform $B$ and its significance. $Bf(w) = 2^{n/4} \int f(x)e^{2\pi iwx - \pi x^2 - (\pi/2)x^2} dx = S_g(f)(r,s)e^{\pi |w|^2/2}$, where $g(x) = e^{-\pi x^2}$, is essentially the STFT with the Gaussian as the window. The study of Bargmann–Fock spaces with tools of complex analysis can be and have been used to obtain results about Gabor expansions in the strict sense (see Johnson/Peetre/Rochberg [42]). Theorem 7 translates into

**Theorem 32.** Let $m$ be a-moderate and $G \in F_{w_a}^{1,1}$. Then there exists a neighborhood $W \subseteq G^a$ of 0, such that for any $W$-dense family $(z_i) \subseteq G^a$ every $F \in F_{p,q}^m$, $1 \leq p, q < \infty$ can be written as

$$F(z) = \sum_i c_i \beta(z_i)G \sum_i c_i e^{-\pi |z_i|^2/2} e^{i \pi z_i \cdot z} G(z - z_i). \quad (7.11)$$

The coefficients depend linearly and stably on $F$ as in Theorem 7.

In particular, if $G(z) \equiv 1 \in F_{w_a}^{1,1} \forall a \geq 0$ is chosen, one obtains expansions in $F_{p,q}^m$ in terms of the exponential functions $e^{i \pi z_i \cdot z}$

$$F(z) = \sum c_i e^{-\pi |z_i|^2/2} e^{i \pi z_i \cdot z}. \quad (7.12)$$

Using a duality argument (see [38] for details), one can derive that $F \in F_{p,q}^m$ $B$ is uniquely determined by the coefficients $< \beta(z_i)G, F >$. Since for $G \equiv 1 < \beta(z_i)2, F > = e^{-\pi |z_i|^2/2} F(z_i)$, this yields a general sampling theorem for the Bargmann–Fock spaces (see [14] for the case of regular lattices).

**Theorem 33.** Given $1 \leq p, q \leq \infty$ and $m$, there exists a neighborhood $W \subseteq G^a$ of 0, such that $F \in F_{p,q}^m$ is uniquely determined by its sampled values $\{F(z_i), i \in I\}$ from any $W$-dense set $\{z_i\}$ in $G^a$.

We leave it to the reader to rephrase the reconstruction algorithms for Theorem 10 for this case.

**FINAL REMARKS.** There are close similarities between the reconstruction problem and the irregular sampling problem for band-limited functions as discussed in [29] and [30]. In particular, the adaptive weights for the 2D-band-limited reconstruction problem turn out to be the same as those for the STFT-problem. The analogy can be used to carry out an error analysis for the STFT-reconstruction algorithms in much the same way as in [30]. Theorem 27 above is just one of such statements. This analogy has been addressed in more detail in [24].

The above theory can be also formulated without serious changes for general locally compact Abelian groups, replacing the Euclidean space $\mathbb{R}^d$. This is not just of theoretical interest, but very useful for applications, because it readily provides a sound theoretical background for discrete implementations
of the algorithms used to solve the two main questions, just by replacing $\mathbb{R}^d$ by some finite (e.g., cyclic) group $G$.

The first named author wants to thank the Centre for Theoretical Physics in Marseille, in particular A. Grossman and B. Torresani, for their hospitality. The first draft of this paper was prepared during a one-month visit at Marseille in the summer of 1991. The second author acknowledges partial funding by the grant AFOSR-90-0311.

References


Hans G. Feichtinger
Institute für Mathematik
Universität Wien
Strudlhorgg.4, A-1090, Vienna, Austria
a8131dan@awiuni11

Karlheinz Gröchenig
Department of Mathematics
University of Connecticut
Storrs, CT 06269-3009