DISSERTATION

Matched Weyl–Heisenberg Expansions of Nonstationary Environments

ausgeführt zum Zwecke der Erlangung des akademischen Grades
eines Doktors der technischen Wissenschaften

eingereicht an der Technischen Universität Wien
Fakultät für Elektrotechnik
von

Dipl.–Ing. Werner Kozek

Eipeldauerstr. 38/27/8, 1220 Wien
geboren am 26. Juli 1964
in Steyr, Oberösterreich
Mtr.-Nr. 8225078

Wien, im September 1996

Abstract

This thesis is about various aspects of linear time–varying systems and nonstationary processes (together nonstationary environments). Such nonstationary environments play an important role in modern communication engineering, particularly as models for natural signals or time–varying communication channels.

Emphasis is on time–frequency–parametrized representations of nonstationary environments, i.e., time–varying power spectra and time–varying transfer functions. Introduction of the generalized Weyl correspondence enables a unified formulation of classical, so far seemingly disparate definitions like Priestley's evolutionary spectrum, the Wigner–Ville spectrum, Zadeh's time–varying transfer function (Kohn–Nirenberg symbol) and the Weyl symbol. Nonstationary Wiener filtering provides an illustrative example for the limited applicability of these time–frequency concepts to a straightforward generalization of frequency domain solutions. We introduce a fundamental classification into underspread/overspread environments based on characterizing the underlying linear operator by the essential support of its spreading function. For underspread environments it is shown that the time–frequency—parametrized representations get essentially definition–independent and can be used in the same manner as the frequency–parametrized representations of stationary environments. Combining the practical efficiency of time–frequency–parametrized representations with the theoretical optimality of a diagonalizing transform leads to window matching criteria for the short–time Fourier transform/Gabor expansion (discrete/continuous Weyl–Heisenberg expansion) of signals and linear systems.

Zusammenfassung

Diese Dissertation behandelt die Theorie linearer zeitvarianter Systeme und nichtstationärer Prozesse (zusammen nichtstationäre Umgebungen). Nichtstationäre Umgebungen stellen ein wichtiges aktuelles Forschungsgebiet der modernen Nachrichtentechnik dar. Sie sind insbesonders bei der Modellierung natürlicher Signale oder zeitvarianter Nachrichtenkanäle von Bedeutung.

Der Schwerpunkt der Untersuchungen liegt auf zeit–frequenz–parametrisierten Darstellungen, also zeitvarianter Leistungsdichtespektren und zeitvarianter Übertragungsfunktionen. Die Einführung der verallgemeinerten Weyl–Korrespondenz ermöglicht eine einheitliche Formulierung bislang nur undurchsichtig zusammenhängender Konzepte wie dem evolutionären Spektrum nach Priestley, dem Wigner–Ville–Spektrum, der von Zadeh eingeführten zeitvarianten Übertragungsfunktion (Kohn–Nirenberg–Symbol) und dem Weyl–Symbol. Das nichtstationäre Wiener–Filter dient als illustratives Beispiel für die begrenzte Anwendbarkeit solcher Darstellungen. Basierend auf der Spreadingfunktion des die nichtstationäre Umgebung charakterisierenden Operators wird eine fundamentale Klassifikation in Underspread/Overspread–Umgebungen eingeführt. Für Underspread–Umgebungen werden eine Reihe von Eigenschaften bewiesen, die die Anwendbarkeit von Zeit–Frequenz–Konzepten analog zu den Frequenzbereichsdarstellungen stationärer Umgebungen verdeutlichen. Die Kombination von effizienten, zeit–frequenz–parametrisierten Darstellungen mit der theoretischen Optimalität diagonalisierender Transformationen führt zu Fensteroptimierungskriterien für die Kurzzeit–Fouriertransformation/-Gabor–Entwicklung (kontinuierliche/diskrete Weyl–Heisenberg–Entwicklung) von Signalen und linearen Systemen.

Die Begutachtung dieser Arbeit erfolgte durch:

1. Prof. Wolfgang Mecklenbräuker

   Institut für Nachrichtentechnik und Hochfrequenztechnik
   Technische Universität Wien

2. Prof. Hans Georg Feichtinger

   Institut für Mathematik
   Universität Wien

To my parents

Acknowledgment

I would like to thank the following persons who implicitly or explicitly contributed to the final success of this work:

- Prof. Wolfgang Mecklenbräuker for helpful discussions about linear systems, careful proof–reading and detailed suggestions that improved the clarity and readability of the presentation.
- Prof. Hans Georg Feichtinger for useful hints about the mathematical aspects of this work.
- Doz. Franz Hlawatsch for inviting me to take part in his research project "Time–Frequency Signal Processing" and for numerous stimulating discussion.
- Dipl.–Ing. Helmut Bölcskei for many useful discussions about Gabor analysis, frame theory and DFT filter banks.
- Dr. Gerhard Doblinger for helpful discussions about practical implementations of signal processing concepts.
- Dr. Kurt Riedel for sharing his experience in spectral estimation.
- Dr. Gernot Kubin for helpful discussions about system identification and speech signals.
- My high–school teacher Mag. Alfred Schuhmacher for initiating my unbroken joy in the mathematical modeling of physical phenomena.
- My wife Marion for endless patience and understanding.

Contents

Chapter 1

Introduction

This introduction motivates the present work followed by an overview of the original results and the basic structure of the text.

1.1 Motivation

Theoretically optimal signal processing procedures are almost always (implicitly or explicitly) based on the solution of an eigenproblem, i.e., the determination of eigenfunctions and eigenvalues of a linear system. Related to the context of this work we point out the role of eigenexpansions in three specific areas of communication engineering:

- In the theory of linear time–invariant systems the Fourier transform and the Laplace transform are standard tools [266, 267, 224]. The "basis"–functions of these integral transforms are generalized eigenfunctions of a linear time–invariant system and the transfer function plays the role of a continuous eigenvalue distribution[1].

- The Karhunen–Loeve (KL) transform[2] is based on the eigenexpansion of a covariance matrix (or covariance kernel in a continuous setting) [194, 225, 4]. The solution of many important statistical signal processing problems such as the design of optimum source coding schemes requires knowledge of the KL basis signals [355, 326].

- For the digital communication over linear channels it is highly desirable to employ eigensignals as transmission pulses since this means vanishing orthogonal distortion and in turn straightforward implementation of optimum detectors. The Nyquist criterion for the design of digital transmission pulses [227] or the theory of prolate spheroidal wave functions [331] are classical examples where eigen–theory is fundamental for the design of transmission pulses.

However, in the context of linear time–varying systems or nonstationary random processes (together nonstationary *environments*) exact eigenexpansions are often not applicable due to the following reasons:

[1] With this terminology we follow the usual mathematical physicists viewpoint [125, 5]. Mathematically precise (part of the spectral theory of linear operators) the LTI system's transfer function for some specific f_0, $H(f_0)$ is no eigenvalue of the underlying translation invariant operator acting on the Hilbert space $L_2(\mathbb{R})$, rather it is an element of the *continuous spectrum*. However, $H(f_0)$ is an approximate eigenvalue in the sense of the approximate point spectrum and it gets to a precise eigenvalue whenever $H(f) = \text{const}$ in a neighborhood of f_0 [311, p.115]. Alternatively, by virtue of Gelfand theory it is correct to speak about *generalized* eigenvalues whose corresponding eigen–"functions" are distributions, thus generalized functions.

[2] In a discrete context the KL transform is sometimes called *Hotelling* transform. Note, furthermore that the concept of *principal components* and *factor analysis* boil down to the same idea of covariance diagonalization.

- An exact eigenbasis typically does not possess the mathematical structure that admits efficient implementation or it does not establish an admissible set of transmission pulses for digital communication.

- In case of incomplete a priori knowledge about a nonstationary environment (as for example in mobile radio communication or practical Wiener filtering for signal enhancement) an approximate diagonalization is the best one can hope to achieve.

1.2 Why Weyl–Heisenberg?

Structured signal sets are fundamental in many signal processing applications. The basic concept is to generate the whole set out of one prototype signal (wavelet, window, pulse, atom, logon etc.) by applying groups of unitary operators (e.g. time shift, frequency shift, scaling). The group structure is fundamental for the fast discrete implementation of such transforms. The most prominent structures are the *affine* group leading to the *Wavelet transform* and the *Weyl–Heisenberg* group leading to the short–time Fourier transform and Gabor expansion [68]. As the title says, this thesis exclusively deals with the Weyl–Heisenberg group. This fact requires some words of justification particularly because the current mainstream of "time–varying signal processing" research puts the focus on the extremely popular Wavelet transform. We make the following points in favour of the Weyl–Heisenberg group:

- Time and frequency shifted versions of prototype signals are used in numerous *existing* applications such as digital communication, source coding, speech recognition, signal analysis and signal enhancement. Hence, one can try to improve performance without total redesign by an appropriate choice of the prototype signal and the time–frequency sampling grid.

- Practically important nonstationary environments such as the mobile radio channel are characterized by two energy–preserving effects: the narrowband Doppler shift and time delay, which are just the basic ingredients of the Weyl–Heisenberg group (the wideband Doppler effect leads to the affine group).

- For slowly time–varying linear systems a time–varying transfer function should make sense—it is time–frequency parametrized. The analog reasoning holds for slowly nonstationary random processes where one expects that the (time–frequency–parametrized) time–varying power spectrum achieves essentially the same properties as the power spectrum of wide–sense stationary random processes.

1.3 Addressed Problems

Time–frequency–parametrized representations of signals and systems have a comparatively long tradition. Most of the important concepts where introduced in two decades following the second world war, thus before the advent of powerful digital signal processing hardware. However, while short–time Fourier analysis became a standard signal analysis tool, the other prominent and mathematically even more sophisticated time–frequency concepts such as Zadeh's time–varying transfer function or Priestley's evolutionary spectrum hardly ever explicitly entered real–world signal processing solutions. This fact may be explained by a lack of mathematical justification in the sense of the following open problems:

- Given the usual (Zadeh's) definition of a time–varying transfer function, in how far does this function reflect algebraic properties of the LTV system's eigenvalues? In particular, does the minimum/maximum of the transfer function reliably reflect the maximum/minimum eigenvalue of the system? When does the cascade of two systems correspond to a multiplication of their transfer functions?

- Given the notion of a time–varying power spectrum, is there any connection to the process' KL eigenvalues? (For wide–sense stationary random processes the power spectrum establishes indeed the continuous eigenvalue distribution of the convolution–type covariance kernel.)
- The linear time–invariant filtering of stationary random processes is best characterized by a frequency domain multiplication of the process' power spectrum and the magnitude squared transfer function of the filter. Can one set up appropriate conditions on a linear time–varying filter and a nonstationary random process such that the well–known frequency domain multiplication relation carries over to an (approximate) time–frequency domain multiplication?
- What about the various classical definitions of a time–varying power spectrum related to the previous question, is anyone marked out by better theoretical behavior?
- For slowly time–varying systems it is an intuitive assumption that windowed sinusoids are approximate eigenfunctions of the system. Can this thought be formulated in a mathematically precise way? If yes, what is the optimum window given a (typically incomplete) a priori knowledge about the system?

Illustrative Example: Nonstationary Wiener Filter. Consider a linear time–invariant system with transfer function $H(f)$. If the input signal is stationary white noise (with normalized variance) then the output process $y(t)$ is wide–sense stationary with power spectrum $|H(f)|^2$. Assume, furthermore, that the output process is subject to additive white noise with spectral density σ_n^2. Then it is well–known that the minimum–mean squared error estimator of a realization of $y(t)$ given its noisy version is also an LTI system whose transfer function is given by [355]:

$$E(f) = \frac{|H(f)|^2}{|H(f)|^2 + \sigma_n^2},$$

this concept is known as *Wiener filter*.

Now, when we switch to a nonstationary environment, and the time–varying transfer function were a consistent generalization of its time–invariant counterpart then the time–varying transfer function of the nonstationary Wiener filter would be given by

$$E(t,f) \stackrel{?}{=} \frac{|H(t,f)|^2}{|H(t,f)|^2 + \sigma_n^2}. \tag{1.1}$$

The key questions associated with this concept are in the focus of the present work: (i) How can one characterize nonstationary environments where (1.1) works in good approximation? (ii) Can one realize the filter via efficient signal transforms?

1.4 Outline of this Thesis

In what follows we give an overview of the contents of the various chapters:

- The first chapter is devoted to the representation of nonstationary random processes via time–frequency–parametrized second order statistics. We give a brief review of the optimal representation of stationary and nonstationary processes via the Fourier transform and the Karhunen–Loeve transform. Particular emphasis is put on the fact that these transforms provide an eigenexpansion of the correlation operators.

  Then follows a critical review of the most prominent definitions of a time–varying power spectrum: The physical spectrum (expected spectrogram), the generalized Wigner–Ville spectrum and Priestley's evolutionary spectrum. We discuss the connection of the Wigner–Ville spectrum and the generalized Weyl correspondence. The Wiener filter provides an excellent example for pointing out the limited applicability of the time–frequency concepts.

We study the expected ambiguity function as an original concept for the correlative representation of nonstationary random processes. We introduce a fundamental classification of random processes (underspread/overspread). For underspread processes we show that the generalized Wigner–Ville spectrum and the evolutionary spectrum lead to an essentially identical result.

- In the second chapter we derive criteria for matching an STFT/Gabor window to the second order statistic of a nonstationary random process. The basic idea is to try to combine time–frequency–parametrization and correlation operator diagonalization by setting up off–diagonal norms as the cost function for the window optimization. The matching criteria are formulated in terms of the window's ambiguity function and expected ambiguity function (EAF) of the process.

  For rectangular and elliptical shape of the EAF support we derive an approximate, low cost matching rule that characterizes the optimum ratio of the temporal and the spectral moment of a window with arbitrary shape. For elliptical shape of the EAF support, the Gaussian window is shown to be optimal for the STFT. We furthermore prove that the physical spectrum with matched analysis window establishes a complete second order statistic of an underspread process. We briefly point out the natural extension to multiwindow methods for the representation of underspread processes.

- The third chapter is devoted to linear system theory. The concept of a time–varying transfer function and a time–varying power spectrum are shown to be mathematically equivalent in so far as in both concepts one maps a linear operator onto the time–frequency plane. In this sense, the spreading function of an LTV system corresponds to the expected ambiguity function of a nonstationary random process and the generalized Weyl symbol to the generalized Wigner–Ville spectrum. We give a critical review of these system representations and we introduce a slightly modified version of the classical underspread/overspread classification of LTV systems.

  We study STFT–based system analysis and design. The window matching theory of the previous chapter carries over to system theory. Specifically, we show that underspread systems can be analyzed and realized via the short–time Fourier transform.

  Finally, we study the practically important WSSUS class of stochastic time–varying systems and derive criteria for optimum distortion free transmission pulses both considering single pulse transmission and intersymbol interference in a time–frequency–division multiple access setup. The mathematical structure is shown to be equivalent with the STFT/Gabor window matching to a nonstationary process.

- The Chapter 4 is written in a more mathematical style. It is shown that underspread operators form an approximate commutative operator algebra and the generalized Weyl correspondence establishes an approximate homomorphism (i.e., the generalized Weyl symbol of the product operator is approximately equal to product of the symbols). For each of the presented theorems we point out its relevance in signal processing applications.

- Finally, in Chapter 5 the main results of this thesis are summarized and various future research problems are pointed out.

- Appendix A gives a very brief review of the basic facts of linear operator theory matched to the scope of this thesis. In Appendices B and C we summarize the definition and property of the generalized spreading function and the generalized Weyl symbol. In the Appendices D and E we derive the minimum–variance unbiased estimator for the generalized Wigner–Ville spectrum of an underspread process and the generalized Weyl symbol of an underspread system. Appendix F summarizes the mathematical properties of various time–frequency signal representations used in this work.

Chapter 2

Time–Frequency Representation and Classification of Random Processes

This chapter starts with a brief review of the theoretically optimal representation of stationary and nonstationary random processes via the Fourier transform and the Karhunen–Loeve transform, respectively. Particular emphasis is put on the fundamental "diagonalizing" property of these transforms. We discuss classical definitions of a time–varying power spectrum, Priestley's evolutionary spectrum and the generalized Wigner–Ville spectrum, and we point out their limited applicability. We study the properties and relations of the expected ambiguity function (EAF). Based on the support of the EAF, we introduce a fundamental classification (underspread/overspread) of nonstationary random processes. For underspread processes it is shown that i) any of the considered definitions of a time–varying spectrum leads to a two–dimensional lowpass function with the bandlimits given by the maximum temporal/spectral correlation width, and ii) Priestley's evolutionary spectrum and the generalized Wigner–Ville spectrum are essentially equivalent.

2.1 Stationary Environments and the Fourier Transform

The Fourier transform plays a fundamental role in the theory of signals and linear systems. This is mainly due to the fact that it gives a diagonalization of translation–invariant operators. Such operators appear either as a linear time–invariant (LTI) system or they correspond to the correlation kernel of a wide–sense stationary process. In the first case the Fourier transform describes the action of an LTI system $\mathbf{H}$ with kernel (impulse response) $h(\tau)$ as frequency domain multiplication

$$\begin{aligned} (\mathbf{H}x)(t) &= \int_\tau h(t-\tau)x(\tau)d\tau \\ &\Updownarrow \\ (\mathcal{F}\mathbf{H}x)(f) &= H(f)(\mathcal{F}x)(f), \end{aligned} \tag{2.1}$$

where $H(f) = (\mathcal{F}h)(f)$ is the transfer function (frequency response) of the LTI system and $\mathcal{F}$ denotes the Fourier transform. More generally, the Fourier transform helps to treat abstract operator calculus such as inversion or composition of LTI systems as simple, pointwise scalar operations of the corresponding transfer functions.

In the case of a wide–sense stationary process $x(t)$, with correlation kernel $r(\tau) = \mathrm{E}\{x(t)x^*(t-\tau)\}$, the Fourier transform leads to uncorrelated increments [268, 287, 85]:

$$\mathrm{E}\left\{d(\mathcal{F}x)(f)d(\mathcal{F}x)^*(\nu)\right\} = S_x(f)\delta(f-\nu)df\,d\nu, \tag{2.2}$$

where $S_x(f)$ denotes the *power spectrum* as defined by the *Wiener–Khintchine relation*:

$$S_x(f) = (\mathcal{F}r)(f). \tag{2.3}$$

This property forms the basic motivation of spectral analysis as a means for extracting the relevant information from a realization of a stationary process. In the statistical context the usual terminology says that the Fourier "basis" $e^{j2\pi ft}$ is *doubly orthogonal*, i.e., orthogonal in the deterministic sense of

$$\int_t e^{j2\pi ft} e^{-j2\pi\nu t} dt = \delta(f-\nu),$$

and orthogonal in the stochastic sense according to (2.2).

The conceptual significance of operator diagonalization is well illustrated by the example of *minimum mean–squared error filtering.*

2.1.1 Illustrative Example: Wiener Filter

Consider the classical setup for statistically optimal linear filters: Given is a noisy observation $y(t)$ of a signal process $x(t)$:

$$y(t) = x(t) + n(t),$$

where $n(t)$ is a statistically independent noise process. The minimum mean–squared error filter (linear estimator) is defined as [267, 326, 355, 199]

$$\mathbf{H}_{MMSE} \stackrel{\text{def}}{=} \arg\min_{\mathbf{H}} \mathrm{E}\left\{ \|x - \mathbf{H}y\|^2 \right\}.$$

Unique solution of this problem requires at least knowledge of the correlation operators of the signal process and the noise, $\mathbf{R}_x$ and $\mathbf{R}_n$, respectively[1]. Then, the formal solution to this optimization problem can be derived by the statistical orthogonality principle [267, 326, 355]. It requires addition, multiplication and inversion of linear operators:

$$\mathbf{H}_{MMSE} = \mathbf{R}_x(\mathbf{R}_x + \mathbf{R}_n)^{-1}. \tag{2.4}$$

If both the signal and the noise are (wide–sense) stationary one is in the lucky situation that $\mathbf{R}_x$ and $\mathbf{R}_n$ and in turn $\mathbf{H}_{MMSE}$ have a common set of generalized eigensignals and the operator calculus of (2.4) carries over to a scalar multiplication in terms of the eigenvalue distributions:

$$H_{MMSE}(f) = \frac{S_x(f)}{S_x(f) + S_n(f)}, \tag{2.5}$$

where $H_{MMSE}(f)$ is the transfer function of the optimum filter (in the stationary case it is indeed an LTI system), $S_x(f)$ is the power spectrum of the signal, and $S_n(f)$ is the power spectrum of the noise. The linear operator formulation (2.4) remains valid for nonstationary processes with finite expected energy, but then $\mathbf{H}_{MMSE}$ corresponds to a linear time–varying system. In the following section we shall investigate this issue in more detail.

2.2 Nonstationary Environments and the KL Transform

In many applications the assumption of stationarity is well justified and the Fourier transform is at least implicitly used for the signal and system analysis. However, in some of the classical applications and particularly in current research areas such as mobile communication or waveform coding a stationarity assumption can not be maintained without significant performance penalty.

[1] We define the correlation operator $\mathbf{R}_x$ of a random process $x(t)$ as the integral operator whose kernel is the correlation function $r_x(t,s) = \mathrm{E}\{x(t)x^*(s)\}$, i.e., one has

$$(\mathbf{R}_x y)(t) = \int_s r_x(t,s)y(s)ds.$$

Nonstationary environments are characterized by time–varying operators with a general two–dimensional kernel $h(t,s)$. Background information about linear operators is summarized in Appendix A, for the following discussion we presuppose familiarity with basic linear operator terminology.

For a square integrable kernel the associated operator is of Hilbert–Schmidt type. If we furthermore assume a normal operator[2], i. e., $\mathbf{H}^*\mathbf{H} = \mathbf{H}\mathbf{H}^*$[3], then the following spectral decomposition holds (i.e., the sum always converges at least in the weak sense of quadratic forms, for more details see Appendix A),

$$\begin{aligned} h(t,s) &= \sum_{k=1}^{\infty} \lambda_k u_k(t) u_k^*(s) \\ &\Updownarrow \\ \mathbf{H} &= \sum_{k=1}^{\infty} \lambda_k \mathbf{P}_{u_k}, \end{aligned} \tag{2.6}$$

where λ_k is the (complex–valued) eigenvalue distribution and $\{u_k(t)\}_{k=1,2,\ldots\infty}$ is the *orthonormal* basis of eigenfunctions,

$$\langle u_k, u_l \rangle \stackrel{\text{def}}{=} \int_t u_k(t) u_l^*(t) dt = \delta_{kl}, \tag{2.7}$$

and $\mathbf{P}_{u_k}$ denotes the rank–one projection operator onto $u_k(t)$:

$$\left(\mathbf{P}_{u_k} x\right)(t) \stackrel{\text{def}}{=} \langle x, u_k \rangle \, u_k(t). \tag{2.8}$$

The eigenvalues can be formally written as a HS operator inner product:

$$\lambda_k = \langle \mathbf{H}, \mathbf{P}_{u_k} \rangle \stackrel{\text{def}}{=} \int_t \int_s h(t,s) u_k^*(t) u_k(s) dt \, ds. \tag{2.9}$$

For the sake of a compact notation we introduce the unitary operator $\mathcal{U}$ that maps an $x \in L_2(\mathbb{R})$ onto a coefficient vector $\langle x, u_k \rangle \in l_2(\mathbb{Z})$ such that

$$x(t) = \sum_{k=1}^{\infty} \left(\mathcal{U}x\right)(k) u_k(t), \qquad \text{with} \qquad \left(\mathcal{U}x\right)(k) \stackrel{\text{def}}{=} \langle x, u_k \rangle.$$

The discrete analogue to (2.1) is given by

$$\begin{aligned} \left(\mathbf{H}x\right)(t) &= \int_s h(t,s) x(s) ds \\ &\Updownarrow \\ \left(\mathcal{U}\mathbf{H}x\right)(k) &= \lambda_k \left(\mathcal{U}x\right)(k). \end{aligned}$$

Replacing the impulse response $h(t,s)$ by a still square–integrable but now self–adjoint[4] correlation kernel $r_x(t,s)$ of a nonstationary, zero–mean random process we obtain the nonstationary analogue to (2.2) [167]

$$\mathrm{E}\left\{\left(\mathcal{U}x\right)(k) \left(\mathcal{U}x\right)^*(l)\right\} = \lambda_k \delta_{kl}. \tag{2.10}$$

[2] Practically important LTV systems such as, e.g., the mobile radio channel are usually not normal in the precise mathematical sense. In Chapter 5 we shall however prove a theorem that determines the *approximate normality* of underspread LTV systems, i.e., systems with appropriately limited memory and limited velocity of the time–variation. We define approximate normality in a HS sense meaning that there exists a normal LTV operator which deviates from the given operator with small HS norm.

[3] $\mathbf{H}^*$ denotes the adjoint operator defined by $(\mathbf{H}^*)(t,s) = (\mathbf{H})^*(s,t)$.

[4] A self–adjoint operator is defined by $\mathbf{H} = \mathbf{H}^*$, it is normal with real–valued eigenvalues.

Here, the unitary map $\mathcal{U}$ is well–known as *Karhunen–Loeve transform* [194, 225]. As a map of a nonstationary process onto uncorrelated coefficients the Karhunen–Loeve transform is of fundamental importance in statistical signal processing, specifically in the design of efficient source coding schemes [4, 326, 183]. It is the consistent generalization of the Fourier transform as it is based on *doubly orthogonal* basis signals (orthogonality in the deterministic sense of (2.7) and the stochastic sense (2.10)).

2.2.1 Formal Derivations of the KL Transform

The Karhunen–Loeve transform is matched to a normal Hilbert–Schmidt (HS) operator since it is based on the specific eigenbasis of the operator. This matching aspect can be formalized by setting up optimization criteria which lead to the KL transform as their optimum.

In this section we denote the KL transform by $\mathcal{U}_{opt}$ to stress its optimality while $\mathcal{U}$ denotes a general unitary transform (mapping from $L_2(\mathbb{R})$ onto $l_2(\mathbb{Z})$). We furthermore introduce a matrix representation of the normal HS operator in the following form

$$H_{\mathcal{U}}(k,k') \stackrel{\text{def}}{=} \langle \mathbf{H}u_{k'}, u_k \rangle .$$

Minimization of an Off–Diagonal Norm. As a diagonalizing transform one can formally derive the Karhunen–Loeve transform by the minimization of an off–diagonal norm $M_{\mathcal{U}}$ which is defined as:

$$M_{\mathcal{U}} \stackrel{\text{def}}{=} \sum_{k=1}^{\infty}\sum_{k'=1}^{\infty} \left|H_{\mathcal{U}}(k,k')\right|^2 (1-\delta_{kk'}). \tag{2.11}$$

It is straightforward to show that:

$$M_{\mathcal{U}} = \|\mathbf{H}\|^2 - \sum_{k=1}^{\infty} |H_{\mathcal{U}}(k,k)|^2,$$

where the operator norm is the Hilbert–Schmidt norm (see p. 129). The optimization problem thus amounts to maximizing the norm of the diagonal:

$$\mathcal{U}_{opt} = \arg\min_{\mathcal{U}} M_{\mathcal{U}} = \arg\max_{\mathcal{U}} \sum_{k=1}^{\infty} |H_{\mathcal{U}}(k,k)|^2, \tag{2.12}$$

subject to the constraint that $\mathcal{U}$ be an orthogonal transform from $L_2(\mathbb{R})$ onto $l_2(\mathbb{Z})$.

Optimal Concentration of Coefficient Power Distribution. We mention the formal derivation (2.11)—(2.12) for two reasons: i) it motivates the choice of the cost function in the subsequently discussed window optimization theory, and ii) it shows that optimum diagonalization of a correlation operator $\mathbf{H} = \mathbf{R}_x$ is equivalent with optimum power concentration in the distribution of the expansion coefficients. Since one has an invariance of the sum of the diagonal entries ($H_{\mathcal{U}}(k,k)$ is real–valued and positive since $\mathbf{R}_x$ is positive self–adjoint per definition.):

$$\sum_{k=1}^{\infty} H_{\mathcal{U}}(k,k) = \mathrm{tr}\mathbf{H}, \qquad \text{for any choice of} \quad \mathcal{U}, \tag{2.13}$$

we could equally well derive the Karhunen–Loeve transform by merely optimizing a concentration measure of the diagonal entries:

$$\mathcal{U}_{opt} = \arg\max_{\mathcal{U}} \sum_{k=1}^{\infty} \left\{ H_{\mathcal{U}}^2(k,k) - H_{\mathcal{U}}(k,k) \right\}, \tag{2.14}$$

where we assume $H_{\mathcal{U}}(k,k) < 1$, which assures $H_{\mathcal{U}}(k,k) > H^2_{\mathcal{U}}(k,k)$ ($H_{\mathcal{U}}(k,k) < 1$ is satisfied for the usual assumption of a process with normalized expected energy $\mathrm{tr}\mathbf{R}_x = 1$). For $\mathbf{H} = \mathbf{R}_x$ the diagonal entries of $H_{\mathcal{U}}(k,k')$ are the variance of the kth coefficient in the process representation,

$$\langle \mathbf{R}_x u_k, u_k \rangle = \mathrm{E}\left\{ |\langle x, u_k \rangle|^2 \right\} = \langle \mathbf{R}_x, \mathbf{P}_{u_k} \rangle .$$

In the representation of a nonstationary process the KL transform thus provides both uncorrelated coefficients, i.e., optimum linear representation via $x(t) \mapsto (\mathcal{U}_{opt}x)(k)$ and optimally concentrated power distribution of the coefficients, i.e., optimum quadratic representation in the sense of $x(t) \mapsto |(\mathcal{U}_{opt}x)(k)|^2$.

2.2.2 Nonstationary Wiener Filter

Under the assumption of nonstationary signal and noise processes with identical Karhunen–Loeve eigenbasis[5] one can formulate a KL based Wiener filter analog to (2.5):

$$\lambda_k^{(MMSE)} = \frac{\lambda_k^{(x)}}{\lambda_k^{(x)} + \lambda_k^{(n)}} = \frac{\langle \mathbf{R}_x, \mathbf{P}_{u_k} \rangle}{\langle \mathbf{R}_x, \mathbf{P}_{u_k} \rangle + \langle \mathbf{R}_n, \mathbf{P}_{u_k} \rangle}. \tag{2.15}$$

The nonstationary Wiener filter is thus a linear time–varying system with impulse response

$$h_{MMSE}(t,s) = \sum_{k=1}^{\infty} \lambda_k^{(MMSE)} u_k(t) u_k^*(s),$$

and input–output relation as follows:

$$(\mathbf{H}_{MMSE}y)(t) = \int_s h_{MMSE}(t,s)y(s)ds = \sum_{k=1}^{\infty} \lambda_k^{(MMSE)} \langle y, u_k \rangle u_k(t).$$

2.2.3 Practical Limitations of the KL Transform

From a strictly mathematical point of view, the Karhunen–Loeve transform (and its system theoretic pendant) is the ultimate tool for nonstationary environments. However, in practical engineering exact eigenexpansions are of restricted applicability due to the following reasons:

- Most often the starting point is an *incomplete a priori knowledge* of an underlying kernel. Instead of solving a usual eigenproblem one has to encounter the conceptually difficult task of matching a signal transform to the *a priori* knowledge at hand. Prominent examples for such an incomplete *a priori* knowledge include the *scattering function* of a stochastic time–varying channel or a *quasistationarity condition* for a random waveform.

- Even if one assumes exact knowledge of a correlation kernel or of a system's impulse response, the eigensignals in general are not equipped with a specific mathematical structure that leads to fast implementations. In practice, one thus uses more or less coarse approximations to the Karhunen–Loeve transform that allow efficient computation.

- When important "natural" signals like speech or music are analyzed by a human observer, ease of interpretation is more important than mathematical optimality. An intuitive interpretation of the transform domain parametrization is given by the notion of a time–varying signal spectrum. Using the general Karhunen–Loeve transform we have no justification to apply any physical interpretation to the transform domain index k.

[5] Although this is a severe constraint it is satisfied in the practically important case of stationary white noise. (The correlation operator of stationary white noise is the identity operator which trivially assures the required assumption of identical KL eigenbases of the signal and noise processes.)

2.3 Time–Varying Power Spectra

Notwithstanding the practical limitations of the KL transform we emphasize that the KL transform as a diagonalizing transform of time–variant operators is in unique correspondence to the Fourier transform as the diagonalizing transform of time–invariant operators. This abstract correspondence gets physical life by "slowing down" the time–variations of an operator; one can show that the KL basis signals then converge to the Fourier "basis" $e^{-j2\pi ft}$ [287, 353]. Hence it is a classical idea to extend the powerful concept of the Fourier transform from strictly stationary environments to "slowly" nonstationary environments. Since this is no precise mathematical problem, researchers in statistics and signal processing suggested various different definitions of a time–varying power spectrum. In the following sections we give a brief review of the most important definitions.

2.3.1 The Physical Spectrum

The short–time Fourier transform (STFT)

$$STFT_x^{(\gamma)}(t,f) \stackrel{\text{def}}{=} \int_s x(s)\gamma^*(s-t)e^{-j2\pi fs}ds,$$

is a linear time–frequency representation of a signal $x(t)$. The STFT depends on the analysis window $\gamma(t)$. (The properties of the STFT are discussed in Appendix F.)

Perhaps the most natural way to define a time–varying power spectrum is via the expectation of the spectrogram, the magnitude–squared STFT:

$$ESPEC_x^{(\gamma)}(t,f) = \mathrm{E}\left\{\left|STFT_x^{(\gamma)}(t,f)\right|^2\right\}. \tag{2.16}$$

The so–defined spectrum has been originally introduced as *physical spectrum* [230]. The most obvious disadvantage of this definition is its window dependence, i.e., different windows may lead to quite different spectra.

Stationary Process and Nonstationary White Noise. For a stationary process with time–invariant power spectrum $S_x(f)$ the physical spectrum can be shown to be time–invariant and its frequency characteristic is given by a smoothed version of the true power spectrum:

$$ESPEC_x^{(\gamma)}(t,f) = S_x(f) * |\Gamma(-f)|^2, \tag{2.17}$$

in the dual case of nonstationary white noise with correlation $r_x(t,s) = m_x(t)\delta(t-s)$, we have a dual result:

$$ESPEC_x^{(\gamma)}(t,f) = m_x(t) * |\gamma(-t)|^2. \tag{2.18}$$

Operator Theoretic Formulation. For later use we introduce a more abstract, formal definition of the physical spectrum as a HS operator inner product

$$\begin{aligned} ESPEC_x^{(\gamma)}(t,f) &= \mathrm{E}\left\{STFT_x^{(\gamma)}(t,f)STFT_x^{(\gamma)*}(t,f)\right\} \\ &= \mathrm{E}\left\{\int_{t_1}\int_{t_2} x(t_1)x^*(t_2)\gamma(t_2-t)\gamma^*(t_1-t)e^{-j2\pi(t_1-t_2)f}dt_1\,dt_2\right\} \\ &= \int_{t_1}\int_{t_2} r_x(t_1,t_2)\gamma(t_2-t)\gamma^*(t_1-t)e^{-j2\pi(t_1-t_2)f}dt_1\,dt_2 \\ &= \left\langle \mathbf{R}_x, \mathbf{P}_\gamma^{(t,f)}\right\rangle, \end{aligned} \tag{2.19}$$

where $\mathbf{R}_x$ is the above introduced correlation operator, $\mathbf{P}_\gamma$ is the rank–one projection onto γ, and time–frequency shifting of operators is defined as:

$$\begin{aligned} \mathbf{P}^{(t,f)} &\stackrel{\text{def}}{=} \mathbf{M}_f \mathbf{T}_t \mathbf{P} \mathbf{T}_{-t} \mathbf{M}_{-f} = \mathbf{M}_f \mathbf{T}_t \mathbf{P} \left(\mathbf{M}_f \mathbf{T}_t\right)^{-1} && (2.20) \\ &\Updownarrow \\ \left(\mathbf{P}^{(t,f)}\right)(t_1,t_2) &= (\mathbf{P})(t_1 - t, t_2 - t) e^{j2\pi(t_1-t_2)f}. && (2.21) \end{aligned}$$

Here, $\mathbf{M}_f$ is the frequency–shift (modulation) operator defined by

$$(\mathbf{M}_f x)(t) \stackrel{\text{def}}{=} x(t) e^{j2\pi f t},$$

and $\mathbf{T}_t$ denotes the time–shift operator defined by

$$(\mathbf{T}_t x)(s) \stackrel{\text{def}}{=} x(s - t).$$

Observe the formal equivalence of (2.19) with the previously introduced expression for KL eigenvalues:

$$\lambda_k = \langle \mathbf{R}_x, \mathbf{P}_{u_k} \rangle .$$

We shall make use of this abstract correspondence later on for the statistically matched window design.

Definition (2.20) will often be used in this work and we urge the reader to become familiar with its meaning: When $\mathbf{P}$ is a time–frequency–localization operator which selects signals centered about the origin of the time–frequency plane, then $\mathbf{P}^{(t,f)}$ is unitarily equivalent and selects signals centered about (t,f)[6].

Relation to KL Transform. It is helpful to view the expected spectrogram as a distribution of the KL eigenvalues over the time–frequency plane. This is supported in a global way by the fact that

$$\int_t \int_f ESPEC_x^{(\gamma)}(t,f) dt\, df = \sum_{k=1}^{\infty} \lambda_k = \mathrm{tr}\mathbf{R}_x. \tag{2.22}$$

Proof: Our proof is based on the fundamental fact that the integral over time–frequency shifted operators gives the identity times the trace of the shifted operator:

$$\int_t \int_f \mathbf{P}^{(t,f)} dt\, df = \mathrm{tr}\left\{\mathbf{P}\right\} \mathbf{I}. \tag{2.23}$$

This formula is immediately obtained by integrating over the kernel of $\mathbf{P}^{(t,f)}$ (cf.(2.21)) or via the trace formula of the Weyl correspondence (cf. (C.14)) Now, the proof of (2.22) is straightforward:

$$\begin{aligned} \int_t \int_f ESPEC_x^{(\gamma)}(t,f) dt\, df &= \int_t \int_f \left\langle \mathbf{R}_x, \mathbf{P}_\gamma^{(t,f)} \right\rangle dt\, df \\ &= \left\langle \mathbf{R}_x, \int_t \int_f \mathbf{P}_\gamma^{(t,f)} dt\, df \right\rangle \\ &= \langle \mathbf{R}_x, \mathrm{tr}\left\{\mathbf{P}_\gamma\right\} \mathbf{I} \rangle \\ &= \int_{t_1} \int_{t_2} r_x(t_1,t_2) \delta(t_1 - t_2) dt_1\, dt_2 \\ &= \mathrm{tr}\mathbf{R}_x. \end{aligned}$$

[6] A more conventional, special case of (2.20) is the frequency shifting of a lowpass filter $\mathbf{H}$ with transfer function $H(f)$. The transfer function of the frequency–shifted version $\mathbf{H}^{(0,\nu)}$ is given by

$$H^{(0,\nu)}(f) = H(f - \nu),$$

i.e., the shifted lowpass filter gets to a bandpass filter with center frequency ν.

We still have to check whether the KL eigenvalues are distributed in a locally meaningful way. One has:

$$ESPEC_x^{(\gamma)}(t,f) = \sum_{k=1}^{\infty} \lambda_k SPEC_{u_k}^{(\gamma)}(t,f). \tag{2.24}$$

This would be useful only for essentially non–overlapping spectrograms $SPEC_{u_k}^{(\gamma)}(t,f)$, such that each KL eigenvalue would correspond to a different "Weyl–Heisenberg cell"[7]. But orthonormal signals may have totally overlapping spectrograms such that different eigenvalues would be superposed over the very same time–frequency–localization which ruins the idea of a meaningful time–varying power spectrum. At this point we do not further elaborate this issue (this is the topic of the following chapter) rather we continue with our review of prominent time–varying power spectra.

2.3.2 The Generalized Wigner–Ville Spectrum

The theoretically unsatisfactory window–dependence of the spectrogram has led researchers to define time–varying power spectra directly in the quadratic domain without any underlying linear process representation.

The temporal correlation in its "kernel" form

$$r_x(t,s) = \mathrm{E}\{x(t)x^*(s)\}$$

does not appropriately reflect the absolute time–variations of a process' second order statistics (the degree of nonstationarity). Hence, it is often useful to switch to an absolute time/time lag (t,τ) form. However, such a map is not unique, a general form is given by

$$\hat{r}_x^{(\alpha)}(t,\tau) \stackrel{\text{def}}{=} r_x\left(t + \left(\frac{1}{2} - \alpha\right)\tau, t - \left(\frac{1}{2} + \alpha\right)\tau\right),$$

where α is a real–valued parameter. The natural members of this family are the symmetrical version

$$\hat{r}_x^{(0)}(t,\tau) = r_x(t+\tau/2, t-\tau/2), \tag{2.25}$$

and the asymmetrical versions

$$\hat{r}_x^{(1/2)}(t,\tau) = r_x(t,t-\tau), \tag{2.26}$$

$$\hat{r}_x^{(-1/2)}(t,\tau) = r_x(t+\tau,t). \tag{2.27}$$

For a stationary process, $\hat{r}_x^{(\alpha)}(t,\tau)$ gets α–invariant and depends only on the lag variable:

$$\hat{r}_x^{(\alpha)}(t,\tau) = \bar{r}_x(\tau). \tag{2.28}$$

Now, one can write the Wiener–Khintchine relation for stationary processes formally as

$$S_x(f) = \mathcal{F}_{\tau\to f}\hat{r}_x^{(\alpha)}(t,\tau) \qquad \text{for} \qquad r_x(t,s) = \bar{r}_x(t-s)$$

such that it is near at hand to define a time–varying power spectrum by

$$EW_x^{(\alpha)}(t,f) = \mathcal{F}_{\tau\to f}\hat{r}_x^{(\alpha)}(t,\tau),$$

which is known as *generalized Wigner–Ville spectrum* [121]. For $\alpha = 0$ we have specifically the *Wigner–Ville spectrum* (expected Wigner distribution) and for $\alpha = 1/2$ we have the *Rihaczek spectrum* (expected Rihaczek distribution).

[7] For the simplicity of the discussion we do not admit identical eigenvalues since they lead to nonunique eigensignals.

The consistency with the usual power spectrum of stationary processes follows immediately from (2.28):

$$\begin{aligned} EW_x^{(\alpha)}(t,f) &= S_x(f) \\ &\Updownarrow \\ r_x(t,s) &= \bar{r}_x(t-s). \end{aligned}$$

In the dual case of nonstationary white noise the generalized Wigner–Ville spectrum is again α–invariant and leads to the natural result

$$\begin{aligned} EW_x^{(\alpha)}(t,f) &= m_x(t) \\ &\Updownarrow \\ r_x(t,s) &= m_x(t)\delta(t-s). \end{aligned}$$

Operator Formulation. Similar to the physical spectrum (see (2.19)) we can define the generalized Wigner–Ville spectrum $EW_x^{(\alpha)}(t,f)$ by a formal operator inner product[8]

$$EW_x^{(\alpha)}(t,f) = \left\langle \mathbf{R}_x, \mathbf{P}^{(t,f)}(\alpha) \right\rangle, \tag{2.29}$$

where the prototype operator is defined by

$$\left(\mathbf{P}^{(t_0,f_0)}(\alpha)\right)(t,s) \stackrel{\text{def}}{=} \delta\left(\left(\frac{1}{2}+\alpha\right)t + \left(\frac{1}{2}-\alpha\right)s - t_0\right) e^{j2\pi f_0(t-s)}. \tag{2.30}$$

This operator is indeed the infinitesimal building block underlying a time–frequency–parametrized expansion of the correlation operator,

$$\mathbf{R}_x = \int_t \int_f EW_x^{(\alpha)}(t,f)\mathbf{P}^{(t,f)}(\alpha)\, dt\, df,$$

we call such an operator decomposition a *continuous Weyl–Heisenberg expansion* of $\mathbf{R}_x$.

The prototype operator $\mathbf{P}^{(t,f)}(\alpha)$ is trace–normalized:

$$\mathrm{tr}\mathbf{P}(\alpha) = 1,$$

which, by virtue of (2.23), shows that the KL eigenvalues are distributed over the time–frequency plane in a globally correct way (a well–known fact [121]):

$$\int_t \int_f EW_x^{(\alpha)}(t,f) dt\, df = \sum_{k=1}^{\infty} \lambda_k = \mathrm{tr}\mathbf{R}_x. \tag{2.31}$$

However, a point–wise interpretation is in obvious conflict with Heisenberg's uncertainty principle, which in a statistical formulation says that there can not exist ideally time–frequency concentrated processes[9].

Yet, it is interesting to study the infinitesimal prototypes for the most prominent choices of α:

[8] In contrast to the inner product formulation of the physical spectrum, this inner product is no well–defined HS operator inner product, since $\mathbf{P}$ is definitely not HS. A mathematically rigorous development would require the use of linear functionals and the associated concept of dual topological vector spaces. However, such a rigorous treatment is beyond the scope of this work, and it blurs the interrelation between various classical definitions of a power spectrum.

[9] An *optimally* time–frequency concentrated process is obviously given by a Gaussian signal $\gamma(t)$ with random amplitude. For normalized variance the correlation operator of this ideally localized process is just the above introduced rank–one projection operator $\mathbf{P}_\gamma$. This point of view leads back to the physical spectrum. However, the physical spectrum is a nonunitary representation of the correlation operator which means in particular that we do not have validity of a continuous Weyl–Heisenberg expansion in the form of:

$$\mathbf{R}_x \stackrel{!}{\neq} \int_t \int_f ESPEC_x^{(\gamma)}(t,f)\mathbf{P}_\gamma^{(t,f)} dt\, df.$$

- For $\alpha = 1/2$, $\mathbf{P}^{(t_0,f_0)}$ is ideally frequency selective and the output is ideally temporally localized:

$$\left(\mathbf{P}^{(t_0,f_0)}x\right)(t) = \delta(t-t_0)X(f_0)e^{j2\pi f_0 t_0}.$$

- For $\alpha = -1/2$, we have the dual behavior: ideal frequency concentration on the range and ideal temporal selectivity on the domain:

$$\left(\mathbf{P}^{(t_0,f_0)}x\right)(t) = e^{j2\pi f_0 t}x(t_0)e^{-j2\pi f_0 t_0}.$$

- For the specific case $\alpha = 0$ the prototype operator acts in a totally symmetric way as a "time–frequency point mirror":

$$\begin{aligned}\left(\mathbf{P}^{(t_0,f_0)}x\right)(t) &= 2x(2t_0-t)e^{j2\pi 2f_0(t-t_0)},\\ \left(\mathcal{F}\mathbf{P}^{(t_0,f_0)}x\right)(f) &= 2X(2f_0-f)e^{j2\pi 2t_0(f_0-f)}.\end{aligned}$$

From this operator decomposition point of view it may be expected that the generalized Wigner–Ville spectrum of a *general* nonstationary process does not lead to a practically useful representation.

Relation to the Weyl Symbol and the Wigner Distribution. The generalized Wigner–Ville spectrum is the generalized Weyl symbol of the correlation operator

$$L_{R_x}^{(\alpha)}(t,f) = EW_x^{(\alpha)}(t,f),$$

the generalized Weyl symbol will be specifically studied in Chapter 4, it can be interpreted as a time–varying transfer function of a linear time–varying system (the properties and relations of the generalized Weyl correspondence are summarized in Appendix B,C).

The generalized Wigner–Ville spectrum is the expectation of the generalized Wigner distribution (see (F.11))

$$EW_x^{(\alpha)}(t,f) = \mathrm{E}\left\{W_x^{(\alpha)}(t,f)\right\}.$$

Based on the KL expansion of $\mathbf{R}_x$ the generalized Wigner–Ville spectrum can be written as a KL eigenvalue weighted sum of the generalized Wigner distribution of the KL eigensignals:

$$EW_x^{(\alpha)}(t,f) = \sum_{k=1}^{\infty}\lambda_k W_{u_k}^{(\alpha)}(t,f).$$

Hence, we argue that the process representation via $EW_x^{(\alpha)}(t,f)$ is useful only in that cases where the KL eigensignals (corresponding to different eigenvalues) are non–overlapping in their essential support. Then, the generalized Wigner–Ville spectrum does indeed give a reliable power distribution over the time–frequency plane, one has:

$$\lambda_k = \mathrm{E}\left\{|\langle x,u_k\rangle|^2\right\} = \left\langle EW_x^{(\alpha)}, W_{u_k}^{(\alpha)}\right\rangle. \tag{2.32}$$

When the KL eigensignal $u_k(t)$ is localized at (t_0,f_0) then the range of the Wigner–Ville spectrum about that very location reflects the power of the process. In the converse case of KL eigensignals with overlapping Wigner distributions (2.32) remains valid but $EW_x(t,f)$ will generically show an oscillatory behavior, different KL eigenvalues are superimposed over the same time–frequency location and the range of $EW_x(t,f)$ does not allow to say anything about the power of the process about some time–frequency localization. Later on in this work we shall see that this problem is related to the underspread/overspread classification of nonstationary processes.

Interrelation to Physical Spectrum. The physical spectrum is a smoothed version of the generalized Wigner–Ville spectrum [230] ($**$ denotes 2D convolution),

$$ESPEC_x^{(\gamma)}(t,f) = EW_x^{(\alpha)}(t,f) ** W_\gamma^{(\alpha)*}(-t,-f), \tag{2.33}$$

where the smoothing kernel is the generalized Wigner distribution of the window. This relation shows in an obvious way how the choice of the window blurs the spectral and temporal characteristic of a nonstationary process. In contrast to the physical spectrum which is based on a noninvertible map from $\mathbf{R}_x$ onto the time–frequency plane, the generalized Wigner–Ville spectrum is a unique second order statistic of a nonstationary process (for the map $EW_x^{(\alpha)}(t,f) \mapsto r_x(t,s)$ see (C.2)).

2.3.3 The Evolutionary Spectrum

The evolutionary spectrum is a classical definition of a time–varying power spectrum introduced by Priestley [286]. Priestley's definition relies on the idea of a doubly orthogonal process expansion. Instead of a detailed review of the original motivation (the interested reader is referred to the work of Priestley [286, 287] and to experimental and theoretical comparisons with the generalized Wigner–Ville spectrum [149, 138, 121, 236, 237]) we motivate Priestley's definition via the time–varying transfer function of the innovations system. This approach is particularly matched to the context of this work.

Any nonstationary process can be viewed as the output of a linear time–varying system, the so–called *innovations system* $\mathbf{H}$ excited by stationary white noise [268]. The correlation operator of the process is then given by:

$$\mathbf{R}_x = \mathbf{H}\mathbf{H}^*, \tag{2.34}$$

i.e., in terms of the kernels:

$$r_x(t,s) = \int_r h(t,r)h^*(s,r)\,dr,$$

where the input–output relation of the LTV system is defined by

$$(\mathbf{H}x)(t) = \int_s h(t,s)x(s)ds.$$

Eq.(2.34) shows that the innovations system is not uniquely defined since based on an arbitrary unitary operator $\mathbf{U}$ (with $\mathbf{U}\mathbf{U}^* = \mathbf{I}$) one can take an alternative innovations filter $\mathbf{G} = \mathbf{H}\mathbf{U}$ such that

$$\mathbf{G}\mathbf{G}^* = \mathbf{H}\mathbf{U}\mathbf{U}^*\mathbf{H}^* = \mathbf{H}\mathbf{H}^*.$$

For stationary processes the innovations representation can be easily formulated in the frequency domain:

$$S_x(f) = |H(f)|^2\,, \tag{2.35}$$

where $S_x(f)$ is the power spectrum of the process and $H(f)$ is the transfer function of the time–invariant innovations system

In order to define a straightforward time–varying generalization of (2.35) we need a time–varying generalization of the LTI system's transfer function $H(f)$. When we consider Zadeh's time–varying transfer function

$$Z_H(t,f) = \int_\tau h(t,t-\tau)e^{-j2\pi f\tau}d\tau$$

as the time–varying generalization of the time–invariant transfer function $H(f)$, then we can axiomatically define the evolutionary spectrum as the time–varying generalization of (2.35)

$$ES_x(t,f) \stackrel{\text{def}}{=} |Z_H(t,f)|^2\,. \tag{2.36}$$

For a stationary process the evolutionary spectrum is consistent with the usual power spectrum,

$$\begin{array}{rcl} ES_x(t,f) & = & S_x(f) \\ & \Updownarrow & \\ r_x(t,s) & = & \bar{r}_x(t-s). \end{array}$$

For nonstationary white noise it leads to the theoretically expected result just as the Wigner–Ville spectrum:

$$\begin{array}{rcl} ES_x(t,f) & = & m_x(t) \\ & \Updownarrow & \\ r_x(t,s) & = & m_x(t)\delta(t-s). \end{array}$$

However, for a general nonstationary process Priestley's spectrum is not unique since the innovations system is not unique.

Operator Theoretic Definition. Zadeh's time–varying transfer function is equivalent to the generalized Weyl symbol with $\alpha = 1/2$ (we shall discuss the concept of a time–varying transfer function of an LTV system in more detail in Chapter 4). It is the theory of the generalized Weyl correspondence which connects Priestley's definition with the generalized Wigner–Ville spectrum.

In order to give a compact operator formulation of the evolutionary spectrum we formally define a (non–unique) operator square–root by:

$$\mathbf{H} \stackrel{\text{def}}{=} \sqrt{\mathbf{R}_x} \qquad \Longleftrightarrow \qquad \mathbf{R}_x = \mathbf{H}\mathbf{H}^*.$$

With this notational convention we can write the evolutionary spectrum as a magnitude–squared operator inner product:

$$ES_x(t,f) \stackrel{\text{def}}{=} \left| \left\langle \sqrt{\mathbf{R}_x}, \mathbf{P}^{(t,f)}(1/2) \right\rangle \right|^2, \tag{2.37}$$

where $\mathbf{P}(1/2)$ denotes a prototype operator as defined in (2.30) with $\alpha = 1/2$. From this point of view, it is natural to define a *generalized evolutionary spectrum* [237] in terms of the α–dependent prototype operator which underlies the generalized Weyl symbol:

$$ES_x^{(\alpha)}(t,f) \stackrel{\text{def}}{=} \left| L_{\sqrt{R_x}}(t,f) \right|^2 = \left| \left\langle \sqrt{\mathbf{R}_x}, \mathbf{P}^{(t,f)}(\alpha) \right\rangle \right|^2. \tag{2.38}$$

Hence, compared to the Wigner–Ville spectrum

$$EW_x^{(\alpha)}(t,f) = L_{R_x}(t,f) = \left\langle \mathbf{R}_x, \mathbf{P}^{(t,f)}(\alpha) \right\rangle,$$

the evolutionary spectrum is always real–valued and positive but it does not give a unique second–order statistic of a general nonstationary process since the map $\mathbf{R}_x \mapsto ES_x(t,f)$ is not invertible. Note, moreover, that if the generalized Weyl symbol would satisfy the so–called "perfect symbol calculus" then the evolutionary spectrum and the Wigner–Ville spectrum would be equivalent:

$$\begin{array}{rcl} L_{HH^*}^{(\alpha)}(t,f) & \stackrel{?}{=} & L_H^{(\alpha)}(t,f) L_{H^*}^{(\alpha)}(t,f) \stackrel{?}{=} \left| L_H^{(\alpha)}(t,f) \right|^2 \\ & \Updownarrow & \\ ES_x^{(\alpha)}(t,f) & \stackrel{?}{=} & EW_x^{(\alpha)}(t,f). \end{array} \tag{2.39}$$

2.3.4 Applicability to Nonstationary Wiener Filtering

None of the previously discussed definitions of a time–varying power spectrum satisfies a useful local property that would justify its application. As an illustration for this fact we again consider the Wiener filter: if a time–varying transfer function $H(t,f)$ and a time–varying signal spectrum $E(t,f)$ were consistent generalizations of the corresponding time–invariant entities (the transfer function of an LTI system and the power spectrum of a stationary process) then one would have a unique representation of the nonstationary Wiener filter (corresponding to (2.4)):

$$H_{MMSE}(t,f) \stackrel{?}{=} \frac{E_x(t,f)}{E_x(t,f) + E_n(t,f)}. \tag{2.40}$$

Without even specifically considering the prominent examples for $H(t,f)$ (Zadeh's function, Weyl symbol) and $E_x(t,f)$ (Physical spectrum, Evolutionary spectrum, generalized Wigner–Ville spectrum) we know that (2.40) can never work in general simply because there does not exist a unique diagonalizing linear transform for a time–varying operator. Moreover we guess that it is even very hard to find one specific (nondegenerate) example where (2.40) works *exactly*. As one of the main results of Chapter 5 we give precise conditions under which (2.40) works in an approximate sense, i.e., the validity of an approximate symbol calculus. Note, moreover, that validity of an approximate symbol calculus also implies the approximate equivalence of the Wigner–Ville spectrum and the evolutionary spectrum (see (2.39)). Hence, without detailed theoretical investigations we can already expect that for those processes where the notion of a time–varying power spectrum makes sense such a power spectrum must be essentially definition independent.

2.4 The Expected Ambiguity Function

The (radar) ambiguity function of a signal $x(t)$,

$$A_x(\tau,\nu) \stackrel{def}{=} \int_t x(t)x^*(t-\tau)e^{-j2\pi\nu t}dt,$$

can be interpreted as *deterministic time–frequency correlation function* (see Appendix F, Section F.0.4). A stochastic time–frequency correlation function of a nonstationary process can be defined as the *expected ambiguity function* (EAF)[211]:

$$EA_x(\tau,\nu) \stackrel{\text{def}}{=} \mathrm{E}\left\{A_x(\tau,\nu)\right\} = \int_t r_x(t,t-\tau)e^{-j2\pi\nu t}dt. \tag{2.41}$$

The so defined EAF is mathematically equivalent to the asymmetrical spreading function (delay Doppler spread function) of the correlation operator (see (B.8)):

$$S_{R_x}(\tau,\nu) = EA_x(\tau,\nu). \tag{2.42}$$

The mathematical properties of the spreading function as a unitary operator representation are discussed in Appendix B, p. 134. The correspondence (2.42) gets physical life if we think of the process $x(t)$ as the output signal of an LTV system $\mathbf{H}$ driven by stationary white noise. Then, since $\mathbf{R}_x = \mathbf{H}^*\mathbf{H}$, the spreading function of the composite system is the EAF of $x(t)$.

The EAF $EA_x(\tau,\nu)$ is in one–to–one correspondence with the usual correlation function,

$$r_x(t,s) = \int_\nu EA_x(t-s,\nu)e^{j2\pi\nu t}d\nu, \tag{2.43}$$

and gives an alternative viewpoint of the second order statistic of a process.

2.4.1 Heuristic Derivation Based on STFT Correlation.

Applying an expectation operator to the ambiguity function of a nonstationary process is certainly not a very convincing way of introducing a useful concept. However, as it will turn out in the following discussion it is not so easy to come up with a useful point–wise interpretation of the EAF. Yet we can promise to develop formulas for later use in this work. We start with the question: How to define a time–frequency correlation function of a nonstationary process?

In order to find an answer, recall the well established concepts of a temporal and spectral correlation function. The temporal correlation function is defined as

$$r_x(t,s) = \mathrm{E}\{x(t)x^*(s)\},$$

which degenerates to a one–dimensional function in the case of wide–sense stationarity, $r_x(t,s) = \bar{r}_x(t-s)$. Whereas, the spectral correlation function is defined as

$$B_x(f,\nu) = \mathrm{E}\{X(f)X^*(\nu)\},$$

which degenerates to a one–dimensional function in the case of nonstationary white noise, which is the "spectrally stationary" process, in Fourier dualism to the wide–sense stationarity in the usual (temporal) sense.

In order to combine these concepts in the sense of a time–frequency correlation function we need a linear time–frequency signal representation. The only linear time–frequency representation is the short–time Fourier transform (STFT) (see Appendix F, Section F.0.1),

$$STFT_x^{(\gamma)}(t,f) \stackrel{\text{def}}{=} \int_{t'} x(t')\gamma^*(t'-t)e^{-j2\pi f t'}dt',$$

and its correlation is defined as:

$$R_{STFT}^{(x,\gamma)}(t,f,t',f') \stackrel{\text{def}}{=} \mathrm{E}\{STFT_x^{(\gamma)}(t,f)STFT_x^{(\gamma)*}(t',f')\}. \tag{2.44}$$

The so–defined STFT correlation is a linear function of the correlation of $x(t)$:

$$R_{STFT}^{(x,\gamma)}(t,f,t',f') = \int_{t_1}\int_{t_2} r_x(t_1,t_2)\gamma(t_2-t')\gamma^*(t_1-t)e^{-j2\pi(t_1 f - t_2 f')}dt_1\,dt_2. \tag{2.45}$$

However, a more illuminating form is obtained by introducing both the EAF of the process and the ambiguity function of the window. When we insert (2.43) and (F.17) in the expression of the STFT correlation (2.45) we have

$$\begin{aligned} R_{STFT}^{(x,\gamma)}(t,f,t',f') &= \int_{t_1}\int_{t_2}\int_{\nu_1}\int_{\nu_2} EA_x(t_1-t_2,\nu_1)A_\gamma(t_2-t'-t_1+t,\nu_2) \\ &\quad \cdot e^{j2\pi[\nu_2(t_2-t')+\nu_1 t_1 - f t_1 + f' t_2]}dt_1\,dt_2\,d\nu_1\,d\nu_2. \end{aligned}$$

Using the substitution $t_1 - t_2 = \tau$ and $\frac{t_1+t_2}{2} = \tau'$ the four–dimensional integral collapses to a two–dimensional integral and one arrives at (for notational convenience we now replace t' by $t-\tau$ and f' by $f-\nu$):

$$R_{STFT}^{(x,\gamma)}(t,f,t-\tau,f-\nu) = \int_{\tau'}\int_{\nu'} EA_x(\tau',\nu')A_\gamma(\tau-\tau',\nu-\nu')\,e^{j2\pi[(\nu'-\nu)(t-\tau)+(\nu'-f)\tau']}d\tau' d\nu'. \tag{2.46}$$

This relation implies a remarkable upper bound on the magnitude of the STFT correlation:

$$\left|R_{STFT}^{(x,\gamma)}(t,f,t-\tau,f-\nu)\right| \le |EA_x(\tau,\nu)| * *|A_\gamma(\tau,\nu)|, \tag{2.47}$$

i.e., this bound is "time–frequency stationary" in the sense that it depends only on the time lag τ and the frequency lag ν. Hence, we have the intuitively appealing picture that the STFT correlation is bounded by the inherent time–frequency correlation of the process as determined by $|EA_x(\tau,\nu)|$, smoothed by the time–frequency correlation of the window as expressed by $|A_\gamma(\tau,\nu)|$.

STFT–Correlation for Stationary White Noise. Stationary white noise with correlation function $r_x(t,s) = \delta(t-s)$ is the unique "time–frequency–stationary" process. The EAF of stationary white noise is ideally concentrated in the origin:

$$EA_x(\tau,\nu) = \delta(\tau)\delta(\nu).$$

Here, the magnitude of the STFT correlation depends only on the time–frequency lag variables τ, ν and is given by the magnitude of the ambiguity function of the window:

$$\left|R^{(x,\gamma)}_{STFT}(t, f, t-\tau, f-\nu)\right|^2 = |A_\gamma(\tau,\nu)|^2 . \tag{2.48}$$

Formal Deconvolution. Now, we have two fundamental problems: (i) The time–frequency correlation is obviously four–dimensional except for the case of white noise which is the only time–frequency stationary process, and (ii) the correlation depends on the choice of a window since there does not exist a window–independent linear time–frequency signal representation. This is where the rigorously inductive method ends and we have to switch to a heuristic reasoning.

As a formal way out of this twofold dilemma we start with the (idealized) assumption of the nonexisting window with ideally concentrated ambiguity function (denoted by $\delta\delta$)

$$A_{\delta\delta}(\tau,\nu) = \delta(\tau)\delta(\nu).$$

Formally inserting this (idealized) ambiguity function in the STFT correlation formula (2.46) leads to a *two–dimensional quadratic correlation function* that determines the magnitude of the STFT correlation:

$$|R^{(x,\delta\delta)}_{STFT}(t, f, t', f')|^2 = |EA_x(t-t', f-f')|^2.$$

That is, by the formal trick of an idealized window we have found a way out of both sides of our dilemma: (i) Our so defined time–frequency correlation function depends merely on the second order statistic of the process and (ii) it is two–dimensional because the magnitude of the correlation of the idealized STFT depends only on the lag variables. Alternatively, one may view this derivation as the formal deconvolution of the upper bound (2.47)[10]:

$$\left|R^{(x,\gamma)}_{STFT}(t, f, t-\tau, f-\nu)\right| \leq |EA_x(\tau,\nu)| **|A_\gamma(\tau,\nu)|.$$

We finally note that this heuristic reasoning suggests that it is the (squared) magnitude of $EA_x(\tau,\nu)$ which will be of practical interest, an observation that turns out to be particularly true as far as the results of this thesis are concerned. Of course, to have a unique second order statistic of a nonstationary process also requires knowledge of the phase of $EA_x(\tau,\nu)$ (it is the absolute time–frequency localization of the process that is "encoded" in the phase). Moreover, in the course of this chapter we shall consider a whole family of alternative definitions of a time–frequency correlation function that — fortunately — have equal magnitude and differ only in the phase.

2.4.2 Local Interpretation

In a non–point–wise manner one can interpret the EAF via a pair of (now realistically) well time–frequency localized signals. The deterministic ambiguity function satisfies the fundamental property (see (F.22))

$$\langle A_x, A_{h,g}\rangle = \langle x, h\rangle\langle x, g\rangle^*,$$

which leads to a corresponding property of the EAF:

$$\langle EA_x, A_{h,g}\rangle = \mathrm{E}\left\{\langle x, h\rangle\langle x, g\rangle^*\right\}. \tag{2.49}$$

This provides a useful (local) interpretation when we assume that the basis signals $g(t)$ and $h(t)$ are well time–frequency localized and centered at different spots of the time–frequency plane. We

[10] This formal deconvolution is similar to the heuristic interpretation of the Wigner distribution as a spectrogram with the (nonexisting) window that has an ideally concentrated Wigner distribution $W_\gamma(t,f) = \delta(t)\delta(f)$, since the spectrogram can be written as

$$SPEC^{(\gamma)}_x(t,f) = W_x(t,f) **W_\gamma(-t,-f).$$

furthermore presuppose that the distance between these locations is large enough such that the cross ambiguity function $A_{h,g}(\tau,\nu)$ is concentrated about a certain point (τ_0,ν_0) that is far away from the origin. Then (2.49) allows to conclude that whenever $EA_x(\tau,\nu)$ is concentrated about the origin then the coefficients $\langle x,h\rangle$ and $\langle x,g\rangle$ are uncorrelated since $EA_x(\tau,\nu)$ and $A_{h,g}(\tau,\nu)$ do not overlap in their essential supports. Figure 2.1 shows a schematic sketch of this situation.

With regard to the converse statement one has to be careful: Given an "overspread" process with large spread of $EA_x(\tau,\nu)$ such that $EA_x(\tau,\nu)$ and $A_{h,g}(\tau,\nu)$ overlap, the two coefficients $\langle x,h\rangle$ and $\langle x,g\rangle$ may still be uncorrelated as both functions are typically highly oscillating and the inner product may still vanish[11].

In conclusion, we can say that the magnitude of the EAF allows to estimate the potential correlation between different locations of the time–frequency plane only in the form of an upper bound. With other words, the EAF predicts stochastic orthogonality but it does not allow to predict correlation. Note that this is fundamentally different from the characterization of stationary processes via the one–dimensional temporal correlation that allows to predict correlation in a precise sense.

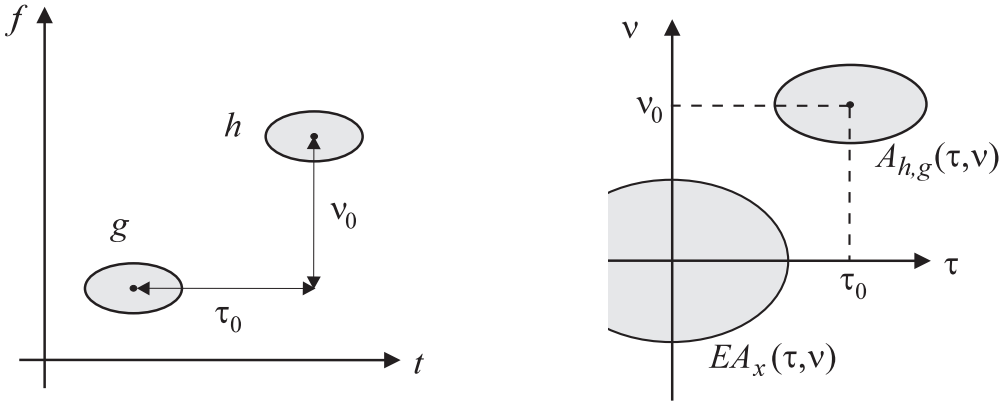

Figure 2.1: On the interpretation of the EAF: (a) Time–frequency localization of h and g, (b) essential support of $EA_x(\tau,\nu)$ and $A_{h,g}(\tau,\nu)$

2.4.3 Spectral Decomposition and Related Properties

Based upon the process' KL expansion one can write the EAF as a KL eigenvalue–weighted sum of the deterministic ambiguity functions of the KL eigensignals:

$$EA_x(\tau,\nu)=\sum_{k=1}^{\infty}\lambda_k A_{u_k}(\tau,\nu). \tag{2.50}$$

This shows that when all KL eigensignals $u_k(t)$ are well time–frequency concentrated then $EA_x(\tau,\nu)$ will be well concentrated about the origin of the (τ,ν)–plane. Here, again, the converse statement is generally not true.

The EAF takes on its maximum magnitude in the origin (as may be postulated for a correlation function). This maximum is real–valued and equal to the trace of the correlation kernel (sum of the Karhunen–Loeve eigenvalues λ_k which is equal to the process' expected energy):

$$EA_x(0,0)=\int_t r_x(t,t)dt=\sum_{k=1}^{\infty}\lambda_k=\mathrm{tr}\mathbf{R}_x\geq\left|EA_x(\tau,\nu)\right|. \tag{2.51}$$

[11] Later on in this chapter, we shall define "overspread" processes in a precise manner.

The total integral of the squared magnitude of the EAF is equal to the squared Hilbert–Schmidt norm of the correlation operator:

$$\int_{\tau}\int_{\nu} \left|EA_x(\tau,\nu)\right|^2 d\tau\, d\nu = \sum_{k=1}^{\infty} \lambda_k^2 = \|\mathbf{R}_x\|^2. \tag{2.52}$$

2.4.4 Symmetry

The magnitude of the EAF is symmetric w.r.t. the origin:

$$\left|EA_x(\tau,\nu)\right| = \left|EA_x(-\tau,-\nu)\right|. \tag{2.53}$$

For a *real–valued process*, $|EA_x(\tau,\nu)|$ is furthermore symmetric w.r.t. the ν–axis:

$$\left|EA_x(\tau,\nu)\right| = \left|EA_x(\tau,-\nu)\right|, \qquad \text{for} \qquad x(t) \in \mathbb{R}. \tag{2.54}$$

2.4.5 EAF of Important Processes

Wide–Sense Stationary Process. For a (wide–sense) stationary process with correlation function

$$r_x(t,s) = \bar{r}_x(t-s),$$

the EAF is ideally concentrated on the τ–axis

$$EA_x(\tau,\nu) = \bar{r}_x(\tau)\delta(\nu).$$

This indicates that different spectral components are uncorrelated.

Nonstationary White Noise. In the dual case of nonstationary white noise with correlation function

$$r_x(t,s) = m_x(t)\delta(t-s),$$

the EAF is ideally concentrated on the ν–axis:

$$EA_x(\tau,\nu) = \delta(\tau)M_x(\nu).$$

Here, $M_x(\nu)$ denotes the Fourier transform of $m_x(t)$

$$M_x(\nu) = \int_t m_x(t)e^{-j2\pi\nu t}.$$

Cyclostationary Process. A cyclostationary process is characterized by

$$r_x(t,s) = r_x(t+lT, s+kT),$$

where $k \in \mathbb{Z}$ and T is the period [137]. The EAF is ideally concentrated on equally spaced lines parallel to the τ–axis (for an illustration see Fig. 2.2):

$$EA_x(\tau,\nu) = \sum_k \rho_k(\tau)\, \delta\left(\nu - \frac{k}{T}\right),$$

where $\rho_k(\tau)$ are usually defined Fourier coefficients [137] :

$$\rho_k(\tau) = \frac{1}{T}\int_0^T r_x(t, t-\tau)e^{-j2\pi\frac{k}{T}t}dt.$$

Tapped Delay Line Process. The Fourier dual of the cyclostationary process is given by the output of a tapped delay line defined as

$$x(t) = \sum_{n=1}^{N} a_k y(t - nT),$$

driven by nonstationary white noise with correlation

$$r_y(t, s) = m(t)\delta(t - s).$$

The EAF is ideally concentrated on lines parallel to the τ–axis:

$$EA_x(\tau, \nu) = \sum_{k=-N+1}^{N-1} \rho_k(\nu)\, \delta(\tau - kT).$$

The Fourier coefficients $\rho_k(\nu)$ are determined by the tap weights a_k as follows ($M(\nu)$ is the Fourier transform of $m(t)$):

$$\rho_k(\nu) = M(\nu) \cdot \begin{cases} \displaystyle\sum_{l=\frac{|k|}{2}+1}^{N-\frac{|k|}{2}} a_{l+k/2} a^*_{l-k/2} e^{-j2\pi\nu[l+k/2]T} & \text{for } k \text{ even}, \\ \displaystyle\sum_{l=\frac{|k|+1}{2}}^{N-\frac{|k|+1}{2}} a_{l+(k+1)/2} a^*_{l-(k+1)/2+1} e^{-j2\pi\nu[l+(1+k)/2]T} & \text{for } k \text{ odd}. \end{cases}$$

Processes with Finite Temporal Correlation Width. A process with finite temporal correlation width is characterized by a finite τ–support of

$$\tilde{r}_x(t, \tau) = r_x(t, t - \tau).$$

We denote the correlation width by τ_0. We characterize the finite support by

$$\tilde{r}_x(t, \tau) = \tilde{r}_x(t, \tau)\, \chi_{[-\tau_0, \tau_0]}(\tau),$$

where $\chi_{[-\tau_0,\tau_0]}(\tau)$ denotes the indicator function of the interval $[-\tau_0, \tau_0]$.

A zero–mean, normal process with finite correlation width is a special case of a so–called *a–dependent* process (with $a = \tau_0$), which is defined by requiring statistical independence of the random variables $x(t_1)$ and $x(t_2)$ whenever $|t_1 - t_2| > a$ [267, p.302].

Finite correlation width means in particular that two windowed versions of $x(t)$, defined as

$$\begin{aligned} w_1(t) &= x(t)\, \chi_{[-T,T]}(t - t_1), \\ w_2(t) &= x(t)\, \chi_{[-T,T]}(t - t_2), \end{aligned}$$

are uncorrelated whenever the windows cover two separate intervals with the gap greater than τ_0:

$$\mathrm{E}\{w_1(t) w_2^*(s)\} = 0 \qquad \text{for} \qquad |t - s| - 2T > \tau_0.$$

The finite correlation width is reflected in the τ–support of the EAF:

$$EA_x(\tau, \nu) = EA_x(\tau, \nu)\, \chi_{[-\tau_0,\tau_0]}(\tau), \tag{2.55}$$

since one has

$$EA_x(\tau, \nu) = \mathcal{F}_{t\to\nu} \tilde{r}_x(t, \tau).$$

Processes with Finite Spectral Correlation Width. The Fourier dual of the finite correlation width process may be defined by a finite spectral correlation width. The spectral correlation is defined as

$$B_x(f, f') = \mathrm{E}\left\{X(f)X^*(f')\right\},$$

finite spectral correlation width can thus be characterized by:

$$B_x(f, f-\nu) = B_x(f, f-\nu)\chi_{[-\nu_0,\nu_0]}(\nu).$$

The EAF can be written as:

$$EA_x(\tau,\nu) = \int_f B_x(f, f-\nu)e^{j2\pi(f-\nu)\tau}df,$$

such that one has finite ν–support (dual to (2.55))

$$EA_x(\tau,\nu) = EA_x(\tau,\nu)\,\chi_{[-\nu_0,\nu_0]}(\nu).$$

However, when we recall the definition of the EAF

$$EA_x(\tau,\nu) = \mathcal{F}_{t\to\nu}\tilde{r}_x(t,\tau),$$

we have the fact that finite spectral correlation width is equivalent to the requirement that $r_x(t, t-\tau)$ be a lowpass function of t.

Quasistationary Process. Quasistationary processes are often only vaguely defined by requiring slow time–variation. This can be made precise by the above discussed finite spectral correlation width. However, in the context of this work it is remarkable how Papoulis defines quasistationarity [267, p. 302]: *A process with finite temporal correlation width τ_0 is quasistationary whenever the following approximation holds:*

$$\tilde{r}_x(t,\tau) \approx \tilde{r}_x(t+\Delta t,\tau) \qquad \text{for} \qquad \Delta t < \tau_0. \tag{2.56}$$

We now show that a (finite expected energy) process with appropriately limited product of temporal and spectral correlation width is quasistationary in the sense of Papoulis.

First, we reformulate the requirement (2.56) by selecting a small constant ϵ:

$$|\tilde{r}_x(t+\Delta t,\tau) - \tilde{r}_x(t,\tau)| \le \epsilon \qquad \text{for} \qquad \Delta t < \tau_0. \tag{2.57}$$

Hence, we have to show that to any given ϵ we can find constants τ_0 (temporal correlation width) and ν_0 (spectral correlation width) such that (2.57) holds true.

Bernstein's inequality for bandlimited functions [267, p. 144] allows to conclude that

$$\left|\frac{\partial\tilde{r}_x(t,\tau)}{\partial t}\right| \le 2\pi\nu_0 \max_t |\tilde{r}_x(t,\tau)|\,. \tag{2.58}$$

Based on (2.51) we have

$$|\tilde{r}_x(t,\tau)| \le \int_{-\nu_0}^{\nu_0} |EA_x(\tau,\nu)|\,d\nu \le 2\nu_0 \mathrm{tr}\mathbf{R}_x,$$

such that we have a bound in terms of the expected energy $\mathrm{tr}\mathbf{R}_x$:

$$\left|\frac{\partial\tilde{r}_x(t,\tau)}{\partial t}\right| \le 4\pi\nu_0^2 \mathrm{tr}\mathbf{R}_x.$$

With the mean–value theorem of differential calculus we have:

$$|\tilde{r}_x(t+\Delta t,\tau)-\tilde{r}_x(t,\tau)| \leq 4\pi\nu_0^2 \mathrm{tr}\mathbf{R}_x|\Delta t|. \tag{2.59}$$

Hence, we can easily compute the condition

$$\nu_0\sqrt{\tau_0} < \sqrt{\frac{\epsilon}{4\pi \mathrm{tr}\mathbf{R}_x}},$$

which finishes our proof. We have thus shown that a process with properly limited product of temporal and spectral correlation width is quasistationary in the sense of Papoulis.

Locally Stationary Process. A *locally stationary process* has been precisely defined in [329] via a separable correlation function of the form:

$$r_x(t,s) = m\left(\frac{t+s}{2}\right) r(t-s).$$

The magnitude of the EAF is given by:

$$|EA_x(\tau,\nu)| = |r(\tau)M(\nu)|,$$

which shows that whenever i) $m(t)$ is bandlimited

$$M(\nu) = M(\nu)\,\chi_{[-\nu_0,\nu_0]}(\nu),$$

and ii) $r(\tau)$ has compact support

$$r(\tau) = r(\tau)\,\chi_{[-\tau_0,\tau_0]}(\tau),$$

then the EAF support is restricted to a centered rectangular domain,

$$EA_x(\tau,\nu) = EA_x(\tau,\nu)\,\chi_{[-\tau_0,\tau_0]}(\tau)\,\chi_{[-\nu_0,\nu_0]}(\nu).$$

Uniformly Modulated Process. A uniformly modulated process is a nonstationary random process $x(t)$ that has been defined by Priestley [287] as the product of a stationary, zero–mean random process $x_s(t)$ and a modulation function $m(t)$:

$$x(t) = m(t)x_s(t).$$

(Priestley furthermore requires a modulation function with $m(0)=1$ and $|(\mathcal{F}_{t\to f}m)(0)| \geq |(\mathcal{F}_{t\to f}m)(f)|$, which is satisfied for a usual window function.) The autocorrelation function of such a process is given by:

$$r_x(t,s) = m(t)m^*(s)r_s(t-s), \qquad \text{with} \qquad r_s(t-s) = \mathrm{E}\{x_s(t)x_s^*(s)\}.$$

The EAF is given by the product of the correlation function of the stationary process $x_s(t)$ and the ambiguity function of the modulation function $m(t)$,

$$EA_x(\tau,\nu) = r_s(\tau)A_m(\tau,\nu).$$

Whenever (i) the modulation function is strictly lowpass and (ii) the correlation of the stationary process has finite support, we can conclude that the EAF lies inside a rectangle about the origin of the (τ,ν)–plane.

Figure 2.2 shows the support of the time–frequency correlation function for the above considered processes.

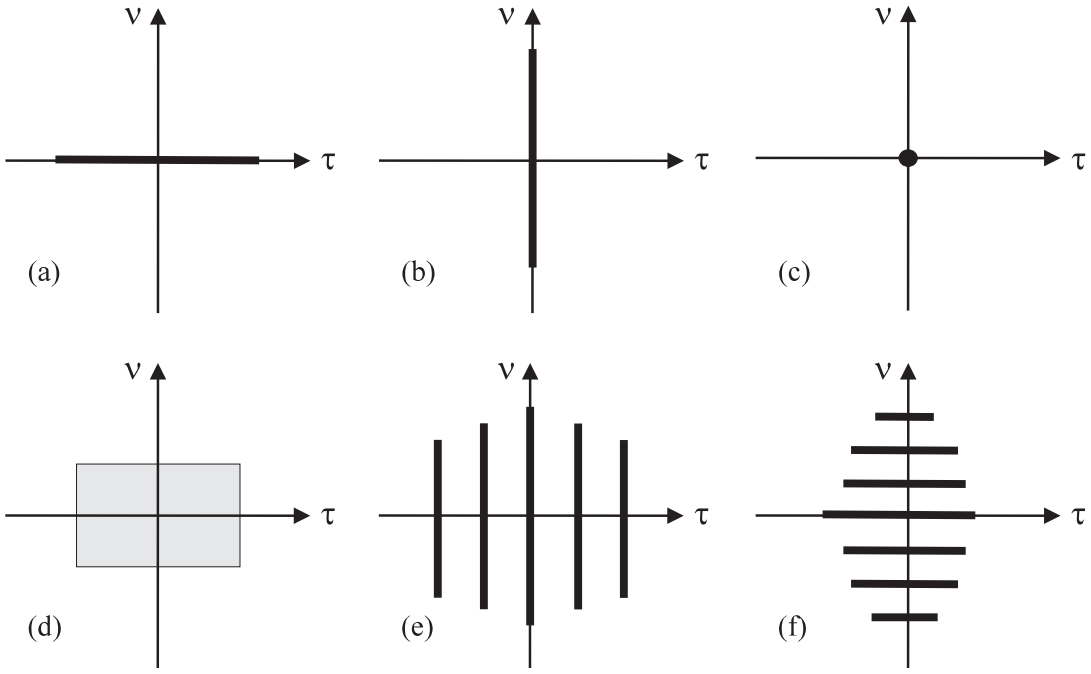

Figure 2.2: Support of the EAF for important classes of processes: (a) Stationary process, (b) nonstationary white process, (c) stationary white process, (d) locally stationary process, (e) "tapped delay line" process, (f) cyclostationary process.

2.4.6 Generalized EAF

In order to relate the EAF to the generalized Wigner–Ville spectrum we now switch to the *generalized* EAF, defined as the expectation of the generalized ambiguity function (the stochastic pendant to (F.19))

$$EA_x^{(\alpha)}(\tau,\nu) \stackrel{def}{=} \mathrm{E}\{A_x^{(\alpha)}(\tau,\nu)\} = \int_t r_x\left(t+\left(\frac{1}{2}-\alpha\right)\tau, t-\left(\frac{1}{2}+\alpha\right)\tau\right) e^{-j2\pi\nu t}dt. \tag{2.60}$$

the previously discussed EAF in particular is obtained by setting $\alpha = 1/2$. The mutual interrelation of the generalized EAF members is given by an unimodular factor:

$$EA_x^{(\alpha)}(\tau,\nu) = EA_x^{(0)}(\tau,\nu)e^{-j2\pi\nu\tau\alpha}. \tag{2.61}$$

Hence, the magnitude of $EA_x^{(\alpha)}(\tau,\nu)$ is α–invariant:

$$\left|EA_x^{(\alpha_1)}(\tau,\nu)\right| = \left|EA_x^{(\alpha_2)}(\tau,\nu)\right|.$$

2.4.7 Interrelation of EAF and Time–Varying Spectra

Generalized Wigner–Ville Spectrum. The generalized Wigner–Ville spectrum is the symplectic 2D Fourier transform of the EAF

$$EW_x^{(\alpha)}(t,f) = \int_\tau \int_\nu EA_x^{(\alpha)}(\tau,\nu)e^{-j2\pi(\tau f-\nu t)}d\tau\, d\nu = \mathcal{F}_{\tau\to f}\mathcal{F}^{-1}_{\nu\to t}\left\{EA_x^{(\alpha)}(\tau,\nu)\right\}, \tag{2.62}$$

a relation that may be seen as the (consistent) generalization of the Wiener–Khintchine relation (2.3) to nonstationary processes (or as the stochastic counterpart of the interrelation of the Wigner

distribution and the ambiguity function (F.23)).

Physical Spectrum. The physical spectrum can be written as a smoothed generalized Wigner–Ville spectrum (see (2.33)). Using the just discussed "nonstationary Wiener–Khintchine" relation (2.62) one has

$$ESPEC_x^{(\gamma)}(t,f) = \mathcal{F}_{\tau\to f}\mathcal{F}^{-1}_{\nu\to t}\left\{EA_x^{(\alpha)}(\tau,\nu)A_\gamma^{(\alpha)*}(\tau,\nu)\right\},$$

where $A_\gamma^{(\alpha)}(\tau,\nu)$ is the generalized ambiguity function of the window.

Evolutionary Spectrum. In order to relate the evolutionary spectrum to the EAF we have to use the spreading function of the innovations system[12]. The generalized EAF is formally equivalent to the generalized spreading function of the correlation operator:

$$EA_x^{(\alpha)}(\tau,\nu) = S_{R_x}^{(\alpha)}(\tau,\nu).$$

The evolutionary spectrum can be written as

$$ES_x(t,f) = \mathcal{F}_{\tau\to f}\mathcal{F}^{-1}_{\nu\to t}\left\{S_H^{(1/2)}(\tau,\nu) **S_H^{(1/2)*}(\tau,\nu)\right\}, \tag{2.63}$$

where $S_H^{(1/2)}(\tau,\nu)$ is the asymmetrical spreading function of the innovations system and $**$ denotes convolution:

$$S_H^{(1/2)}(\tau,\nu) **S_H^{(1/2)*}(\tau,\nu) \stackrel{\text{def}}{=} \int_{\tau'}\int_{\nu'} S_H^{(1/2)}(\tau-\tau',\nu-\nu')S_H^{(1/2)*}(\tau',\nu')d\tau'\,d\nu'.$$

Formulating the generalized EAF in terms of the spreading function of the innovations system leads to the so–called *twisted convolution*:

$$EA_x^{(\alpha)}(\tau,\nu) = \int_{\tau'}\int_{\nu'} S_H^{(\alpha)}(\tau-\tau',\nu-\nu')S_H^{(\alpha)*}(\tau',\nu')e^{-j2\pi\phi(\tau,\nu,\tau',\nu',\alpha)}d\tau'\,d\nu' \tag{2.64}$$

with

$$\phi(\tau,\nu,\tau',\nu',\alpha) = \tau'\nu(\alpha+1/2) + \tau\nu'(\alpha-1/2) - 2\tau'\nu'\alpha.$$

Observe that up to the unimodular factor $e^{-j2\pi\phi(\tau,\nu,\tau',\nu',\alpha)}$, the twisted convolution (2.64) looks similar to a usual convolution. As will be used later, we already mention that with regard to support relations of $EA_x^{(\alpha)}(\tau,\nu)$ and $S_H^{(\alpha)}(\tau,\nu)$ the twisted convolution behaves just like a regular convolution. For a discrete setting ($\tau,\nu\in\mathbb{Z}$) it is trivial to see that for self–adjoint $\mathbf{H}$ (where one has $|S_H^{(\alpha)}(\tau,\nu)| = |S_H^{(\alpha)}(-\tau,-\nu)|$) the twisted convolution and the usual convolution enlarge the support by a factor two in both directions[13]. For the continuous–time setting a precise proof may be based on Titchmarsh's convolution theorem [195, p.178].

2.5 Processes with Compactly Supported EAF

In previous sections we have seen that three classical definitions of mildly nonstationary processes, i.e., Silverman's locally stationary, Papoulis' quasistationary, and Priestley's uniformly modulated process, are all essentially characterized by a compactly supported EAF. This suggests to study such processes in more detail.

[12] The spreading function of a linear system will be explained in more detail in Chapter 4, the properties and relations are summarized in Appendix B.

[13] For many practical applications it is sufficient to study processes on a limited rectangle of the time–frequency–plane $[-T/2,T/2]\times[-F/2,F/2]$, of course, with $TF\gg 1$. Formal periodization of the Wigner–Ville spectrum with periods T,F leads to a *time–frequency periodic* process whose expected ambiguity function is given by a superposition of delta distributions on the reciprocal grid $(k/F,l/T)$ $(k,l\in\mathbb{Z})$. Hence, in this case one can work with a discrete (τ,ν)–domain.

We characterize the support of the EAF by a separable 0/1–valued indicator function with minimum area such that

$$EA_x^{(\alpha)}(\tau,\nu) = EA_x^{(\alpha)}(\tau,\nu)\chi_{[-\tau_0,\tau_0]}(\tau)\chi_{[-\nu_0,\nu_0]}(\nu), \tag{2.65}$$

and we call

$$\sigma_x \stackrel{\text{def}}{=} 4\tau_0\nu_0, \tag{2.66}$$

total spread. Based upon this characterization we classify nonstationary processes in *underspread* for $\sigma_x \ll 1$ and *overspread* for $\sigma_x \gg 1$ (see Figure 2.3) [209]. Moreover, we call two underspread processes with identical (τ_0,ν_0) *jointly underspread*.

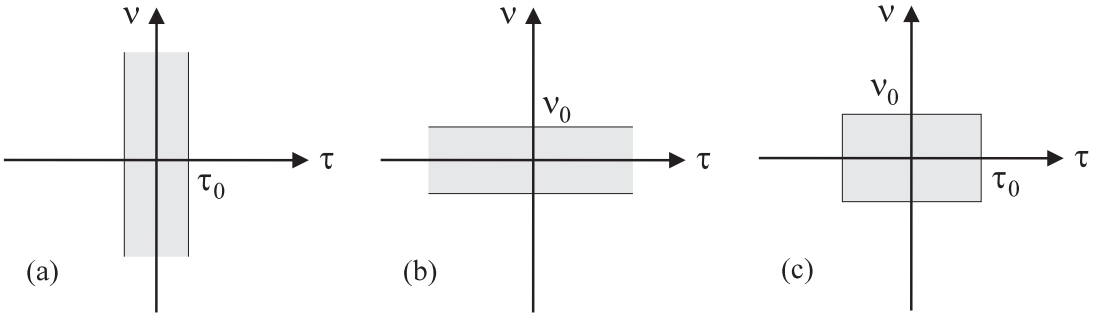

Figure 2.3: Support of the expected ambiguity function for the following processes: (a) finite temporal correlation, (b) finite spectral correlation, (c) underspread process (assuming $4\tau_0\nu_0 \ll 1$).

Here, we have employed the system theoretic terminology [197], where linear time–varying systems are characterized through the essential support of their spreading function or its stochastic pendant, the scattering function, respectively. Mathematically, the EAF characterization of nonstationary process and the spreading characterization of linear systems are equivalent due to the equivalence of the time–frequency correlation function and the spreading function of the correlation operator (2.42).

Note that the underspread/overspread classification is not merely based on the degree of nonstationarity (temporal variation of statistics), rather it takes into account the variation of the process second order statistic in both time and frequency direction. This aspect will become clearer by studying the impact of the underspread/overspread classification on classical definitions of a time–varying power spectrum of a nonstationary process.

2.5.1 Canonical Reformulation of the Wigner–Ville Spectrum

Based on the "nonstationary Wiener–Khintchine relation" (2.62) the (τ,ν)–domain support constraint (2.65) carries over to an idempotent (t,f)–domain convolution:

$$\begin{aligned} EA_x^{(\alpha)}(\tau,\nu) &= EA_x^{(\alpha)}(\tau,\nu)\chi_{[-\tau_0,\tau_0]}(\tau)\chi_{[-\nu_0,\nu_0]}(\nu) \\ &\Updownarrow \\ EW_x^{(\alpha)}(t,f) &= EW_x^{(\alpha)}(t,f) **\left(\mathcal{F}_{\tau\to f}\mathcal{F}^{-1}_{\nu\to t}\chi_{[-\tau_0,\tau_0]}(\tau)\chi_{[-\nu_0,\nu_0]}(\nu)\right). \end{aligned}$$

By using the shift–invariance property of the generalized Weyl symbol ((C.15)) we can interpret the convolution kernel as symbol of an α–dependent prototype operator $\mathbf{P}(\alpha)$:

$$\begin{aligned} EW_x^{(\alpha)}(t,f) **L^{(\alpha)}_{P(\alpha)}(t,f) &= \int_\tau\int_\nu EW_x^{(\alpha)}(\tau,\nu)L^{(\alpha)}_{P(\alpha)}(t-\tau,f-\nu)d\tau\,d\nu \\ &= \int_\tau\int_\nu EW_x^{(\alpha)}(\tau,\nu)L^{(\alpha)}_{P^{(\tau,\nu)}(\alpha)}(t,f)d\tau\,d\nu. \end{aligned}$$

Switching from symbols to operators gives an alternative *continuous Weyl–Heisenberg expansion* of the correlation operator:

$$\mathbf{R}_x = \int_t \int_f EW_x^{(\alpha)}(t,f)\mathbf{P}^{(t,f)}(\alpha)dt\,df, \tag{2.67}$$

where the prototype operator is characterized by the spreading constraint as follows:

$$\begin{aligned} S_{P(\alpha)}^{(\alpha)}(\tau,\nu) &= \chi_{[-\tau_0,\tau_0]}(\tau)\chi_{[-\nu_0,\nu_0]}(\nu) \\ &\Updownarrow \\ L_{P(\alpha)}^{(\alpha)}(t,f) &= \frac{\sin(2\pi\nu_0 t)\sin(2\pi\tau_0 f)}{\pi^2 tf} \\ &\Updownarrow \\ (\mathbf{P}(\alpha))\,(t,s) &= \frac{1}{4\pi\tau_0\nu_0}\frac{\sin\left(2\pi\nu_0\left(\frac{1}{2}+\alpha\right)t+\left(\frac{1}{2}-\alpha\right)s\right)}{\left(\frac{1}{2}+\alpha\right)t+\left(\frac{1}{2}-\alpha\right)s}\chi_{[-\tau_0,\tau_0]}(t-s). \end{aligned} \tag{2.68}$$

The so–defined prototype operator is quite different from that introduced in Section 2.3.2 (it approaches that operator asymptotically for $\tau_0, \nu_0 \to \infty$). It is Hilbert–Schmidt with the norm given by the total spread:

$$\|\mathbf{P}(\alpha)\|^2 = \sigma_x.$$

Note, moreover, that we can redefine the Wigner–Ville spectrum for a process that satisfies (2.65) by an operator inner product (analogous to (2.29)):

$$EW_x^{(\alpha)}(t,f) = \left\langle \mathbf{R}_x, \mathbf{P}^{(t,f)}(\alpha)\right\rangle \tag{2.69}$$

It should be emphasized that the continuous Weyl–Heisenberg expansion (2.67) and the operator inner product formulation (2.69) are not uniquely defined. In fact when we replace $\mathbf{P}(\alpha)$ by an operator $\mathbf{P}'(\alpha)$ whose spreading function satisfies

$$S_{P'(\alpha)}^{(\alpha)}(\tau,\nu) = \begin{cases} 1, & \text{for} \quad (\tau,\nu) \in [-\tau_0,\tau_0]\times[-\nu_0,\nu_0], \\ \text{arbitrary}, & \text{else}, \end{cases} \tag{2.70}$$

both (2.67) and (2.69) hold true. However, with the unitarity of the spreading function we know that

$$\|S_H(\tau,\nu)\| = \|\mathbf{H}\|.$$

Hence, among all possible candidate prototype operators that satisfy (2.70) the specific prototype operator defined by (2.68) is canonical in so far as it achieves minimum Hilbert–Schmidt norm. Moreover, in Appendix D we show that this specific prototype operator establishes an optimum time–frequency shift–invariant estimator of real–valued spectra in the following form

$$\widehat{EW}_x^{(0)}(t,f) = \left\langle \mathbf{P}^{(t,f)}(0)y, y\right\rangle,$$

where $y(t)$ is a noisy observation of $x(t)$.

2.5.2 Stochastic Sampling Principle

In view of the Fourier correspondence between the generalized Wigner–Ville spectrum and the EAF (2.62) we see that the previously defined underspread condition implies a smoothness condition of the time–varying spectrum of the process.

Using the theory of the generalized Weyl correspondence (see Appendix C) and the sampling principle for 2D signals leads to a *discrete Weyl–Heisenberg expansion* of the correlation operator:

$$\begin{aligned} \mathbf{R}_x &= TF \sum_l \sum_m EW_x^{(\alpha)}(lT, mF)\mathbf{P}^{(lT,mF)}(\alpha) \\ &\Updownarrow \\ r_x(t,s) &= \sum_l \sum_m EW_x^{(\alpha)}(lT, mF)p_\alpha(t - lT, s - lT)e^{j2\pi mF(t-s)}, \end{aligned} \tag{2.71}$$

valid for a sampling grid with

$$T \leq \frac{1}{2\nu_0} \qquad \text{and} \qquad F \leq \frac{1}{2\tau_0}. \tag{2.72}$$

Without going into details we note that it is obvious that the critical spread $\sigma_x = 1$ plays a theoretically interesting role as it corresponds to the critical grid

$$TF = 1,$$

which in turn means that in this case the rate of innovation in the second order statistic of the process is equal to the critical sampling rate of one realization (considering bandlimited processes). Moreover, the theoretical study of underspread operators in Chapter 5 will show the fact that, from a mathematical point of view, it is the class of underspread processes where the terminology of a time–varying spectrum makes sense[14].

Recall that the power spectrum of a stationary process is the eigenvalue distribution of its correlation operator. Hence in the fundamental interpretation of the generalized Wigner–Ville spectrum as a time–varying power spectrum we interpret the samples $EW_x^{(\alpha)}(lT, mF)$ as (approximate) eigenvalues of the correlation operator such that (2.71) should be a Weyl–Heisenberg structured version of the KL based spectral decomposition:

$$\begin{aligned} \mathbf{R}_x &= \sum_{k=1}^{\infty} \lambda_k \mathbf{P}_{u_k} \\ &\Updownarrow \\ r_x(t,s) &= \sum_{k=1}^{\infty} \lambda_k u_k(t) u_k^*(s). \end{aligned} \tag{2.73}$$

The conceptual correspondence between (2.71) and (2.73) does not work in a precise sense since the prototype operator $\mathbf{P}(\alpha)$ is not an orthogonal projection operator. In Chapter 5 we shall discuss various mathematical theorems that support the eigenvalue interpretation of the GWS samples. It is important to emphasize that the (approximate) eigenvalue interpretation of $EW_x^{(\alpha)}(lT, mF)$ implies an (approximate) multiplicity of the KL eigenvalues of $1/\sigma_x$ as can be seen from the discrete trace–formula

$$\sum_l \sum_m EW_x^{(\alpha)}(lT, mF) = \frac{1}{TF} \sum_{k=1}^{\infty} \lambda_k = \sigma_x \mathrm{tr}\mathbf{R}_x. \tag{2.74}$$

Multiplicity of the KL eigenvalues means that underspread processes are time—frequency locally white in the same sense as stationary processes are approximately white over frequency bands with small variation of the power spectrum. However, one should be aware that mathematically precise

[14] To be precise, we should additionally admit any (centered) symplectic relatives of the rectangular domain in our underspread definition (sheared or rotated versions of the rectangular domain). However, in view of the classical definitions of limited nonstationarity, rectangular symmetry seems to be of predominant practical interest. Moreover, for real–valued processes a sheared or rotated underspread support is excluded due to the symmetry w.r.t. to the ν–axis (see (2.54)).

the $EW_x^{(\alpha)}(t,f)$–samples depend *both on the KL eigenvalues and the KL eigensignals* of the process. Which means that two underspread processes with identical (τ_0, ν_0) do not necessarily have a common KL eigenbasis. We also emphasize that one has to be cautious when using the discrete trace formula (2.74) for a characterization of the approximate rank of time–frequency–localized KL subspaces[15]. In Chapter 5 we shall see that it is the ratio T/F which establishes a grid matching rule which remains valid whatever goodness of approximation of the KL subspaces one requires. Whereas the rank of the subspaces depends strongly on the required approximation threshold. Moreover, much in this spirit, we shall return to the matched grid ratio in the statistical optimization of the Gabor expansion, where TF is determined by pragmatical considerations while T/F has to be matched to the a priori knowledge.

The physical spectrum is a smoothed version of the GWS which means that the sampling grid for the GWS is also sufficient for sampling the physical spectrum without loss of information.

With regard to Priestley's evolutionary spectrum we note that, under the assumption of an underspread innovations system, using the "two–step" relation to the EAF (2.63), (2.64), it easy to see that for a self–adjoint innovations system (whose generalized spreading function has a magnitude which is symmetrical w.r.t. the origin, see (B.25)) the evolutionary spectrum is a 2D lowpass function with exactly the same bandlimits as the generalized Wigner–Ville spectrum.

Hence, we have shown that any of the classical definitions of a time–varying power spectrum of a process with compactly supported EAF can be sampled without loss of information on the grid as determined by (2.72). One has a general *stochastic sampling principle* (formulation of (2.72) in a more intuitively appealing form):

$$\frac{T}{F} = \frac{\tau_0}{\nu_0} \qquad \text{and} \qquad TF = \frac{1}{4\tau_0\nu_0} = \frac{1}{\sigma_x}. \tag{2.75}$$

2.5.3 Asymptotic Equivalence of Time–Varying Spectra

For stationary processes all of the above mentioned window–independent spectrum definitions yield the conventional power spectrum. The stationary process may be seen as a limit case of an underspread process, and in fact, *for underspread processes with underspread innovations representation the evolutionary spectrum is essentially equal to the generalized Wigner–Ville spectrum* [211][16]. The proof of this fact can be based on the theory of the symbolic calculus of underspread operators as follows (detailed proofs can be found in Chapter 5).

As pointed out in the previous section, we know that given a self–adjoint underspread innovations system $\mathbf{H}$ with spreading constraints $(\tau_0/2, \nu_0/2)$, white noise excitation leads to an underspread process whose correlation operator has spreading constants (τ_0, ν_0). Then, Theorem 5.1 allows to conclude that

$$\left|L_H^{(0)}(t,f)\right|^2 = L_H^{(0)}(t,f) L_{H^*}^{(0)}(t,f) \approx L_{HH^*}^{(0)}(t,f),$$

and, since

$$L_{HH^*}^{(1/2)}(t,f) = L_{R_x}^{(1/2)}(t,f) = EW_x^{(1/2)}(t,f),$$

we have the approximate equivalence of the evolutionary spectrum defined by (2.36) and the generalized Wigner–Ville spectrum with $\alpha = 1/2$:

$$ES_x(t,f) \approx EW_x^{(1/2)}(t,f).$$

[15] Subspaces spanned by KL eigensignals with identical KL eigenvalues.

[16] Numerical experiments indicate that the existence of an underspread innovations representation seems not necessary. However, for our proof this assumption is crucial.

Due to (2.36), Priestley's definition stands conceptually closer to the Rihaczek spectrum (generalized Wigner–Ville spectrum with $\alpha = 1/2$) . With the approximate α–invariance of the generalized Weyl symbol of underspread operators — as will be quantified in Theorem 5.5, Chapter 5 — we can furthermore show that the evolutionary spectrum approaches any member of the $EW_x^{(\alpha)}$–family with decreasing spread of the process. Theorem 5.5 and 5.1 yield the following bounds

$$\begin{aligned}
\left| |L_H^{(\frac{1}{2})}(t,f)|^2 - |L_H^{(0)}(t,f)|^2 \right| &< 32\sigma_x^2 \sin\left(\pi\sigma_x^2\right) \mathrm{tr}^2\mathbf{H}, \\
\left| L_{R_x}^{(\alpha)}(t,f) - L_{R_x}^{(0)}(t,f) \right| &< 8\sigma_x \sin\left(\pi\sigma_x|\alpha|\right) \mathrm{tr}\mathbf{R}_x, \\
\left| L_{HH^*}^{(0)}(t,f) - |L_H^{(0)}(t,f)|^2 \right| &< 32\sigma_x^2 \sin\left(\frac{1}{4}\pi\sigma_x\right) \mathrm{tr}^2\mathbf{H}.
\end{aligned}$$

Combining these inequalities leads to the following result, valid for $\sigma_x < 1/2$:

$$\left| EW_x^{(\alpha)}(t,f) - ES_x(t,f) \right| < 8\pi\sigma_x^2 \left(8\sigma_x \mathrm{tr}^2\sqrt{\mathbf{R}_x} + |\alpha|\, \mathrm{tr}\mathbf{R}_x \right), \tag{2.76}$$

where for notational simplicity we have made the bound coarser by utilizing $\sin(x) < x$ for $x \in [0, \pi/2]$. This formula relates all of the classical window independent definitions of a time–varying signal spectrum in the form of an L_∞–bound valid for underspread processes.

2.6 Numerical Experiments

As an illustration for the above discussed theoretical results we consider two numerical experiments about zero–mean, finite–length, discrete–time nonstationary processes synthetically defined by a covariance matrix, i.e., the computation of the various spectra is based on complete knowledge of the second–order statistics. The discrete implementation of the generalized Weyl correspondence is the key problem of this experiment.

The map that underlies the Weyl symbol ($\alpha = 0$), $(t,\tau) \mapsto (t + \tau/2, t - \tau/2)$, is not directly discretizable while the map underlying the Kohn–Nirenberg correspondence ($\alpha = 1/2$), $(t,\tau) \mapsto (t, t-\tau)$, can be trivially discretized. In particular one can formulate a discrete Kohn–Nirenberg correspondence as a unitary one–to–one mapping of $N \times N$ matrices by

$$L_H^{(1/2)}[n,k] = \sum_{m=0}^{N-1} H\left[n, (n-m)_{\mathrm{mod}N}\right] e^{-j2\pi\frac{mk}{N}}, \qquad n,k \in [0, N-1],$$

which is impossible for the Weyl correspondence. One pragmatical way out of this dilemma is to compress the matrix at hand onto the space $\mathcal{H}$ of halfband signals (complex signals whose spectrum vanishes over its half length)

$$\mathbf{H}_\mathcal{H} = \mathbf{P}_\mathcal{H}\mathbf{H}\mathbf{P}_\mathcal{H},$$

where $\mathbf{P}_\mathcal{H}$ denotes the halfband projection operator and $\mathbf{H}_\mathcal{H}$ is the compressed matrix. This procedure maps an $N \times N$ matrix on a $2N \times 2N$ matrix. For such a half–band compressed matrix, one can define a discrete $N \times N$ Weyl symbol as:

$$L_H^{(0)}[n,k] = \sum_{m=1}^{N} 2H\left[(n+m)_{\mathrm{mod}2N}, (n-m)_{\mathrm{mod}2N}\right] e^{-j4\pi\frac{mk}{N}}, \qquad n,k \in [0, N-1].$$

Such an ad hoc implementation keeps the desirable behavior of the Weyl correspondence for operators with chirp–like eigenfunctions. However, this approach comes at the cost of (i) loss of unitarity and (ii) considerably increased memory and computational expense.

Figure 2.4 shows contour plots of various spectra of an underspread process. The process has been synthesized by defining an innovations system in the form of a moderately time–variant bandpass filter. The total spread has been chosen near to the critical threshold, see Fig. 2.4(j). This choice leads to visible differences between the various window independent spectrum definitions. In the middle (chirpy) part of the process, one can observe the better concentration of the $\alpha = 0$, Weyl–type spectra Fig. 2.4(f) and Fig. 2.4(i) compared to the $\alpha = 1/2$, Kohn–Nirenberg type spectra Fig. 2.4(g) and Fig. 2.4(h). Another observation that can be predicted from theory is the dramatic influence of the window on the appearance of the physical spectrum for one and the same process. The matched window has been computed according to the theory presented in the following chapter. The matched window physical spectrum gives a satisfactory process representation, see Fig. 2.4(e).

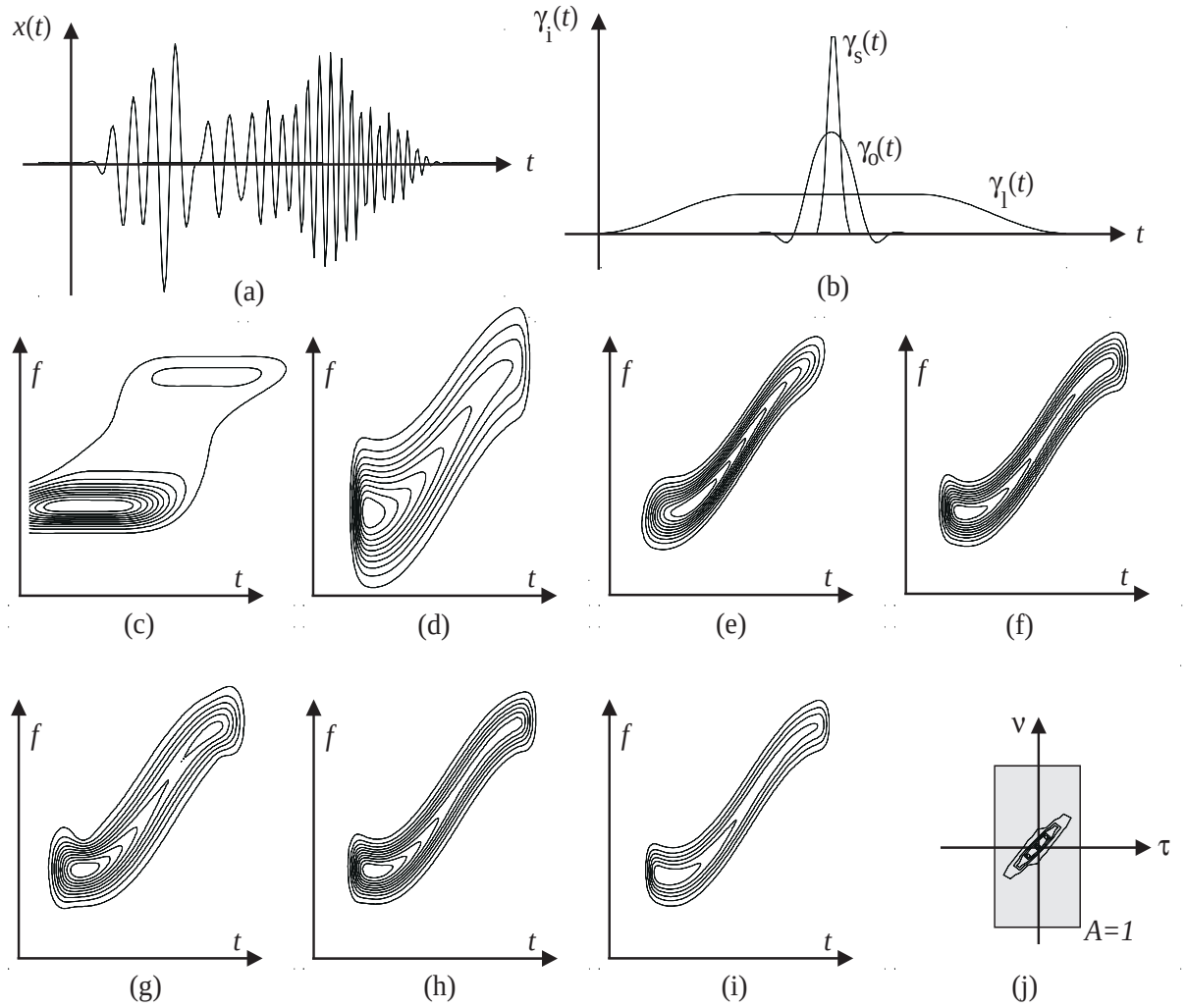

Figure 2.4: Various time–varying power spectra of an underspread process: (a) one realization of the process, (b) three different window functions: $\gamma_l(t) \cdots$ "long" window, $\gamma_s(t) \cdots$ "short" window, $\gamma_o(t) \cdots$ matched window, (c) physical spectrum using $\gamma_l(t)$, (d) physical spectrum using $\gamma_s(t)$, (e) physical spectrum using $\gamma_o(t)$, (f) Wigner–Ville spectrum, (g) Rihaczek spectrum (generalized Wigner–Ville spectrum with $\alpha = 1/2$), (h) evolutionary spectrum (generalized evolutionary spectrum with $\alpha = 1/2$), (i) generalized evolutionary spectrum with $\alpha = 0$, (j) expected ambiguity function.

The dramatic divergence of the various power spectra for an overspread process is the topic of Figure 2.5. We defined an overspread innovations system in a piecemeal fashion by putting together highly temporally localized and highly frequency localized eigen–components. Although the corresponding

process certainly has some specific time–frequency–structure, its overspread behavior leads to the theoretically expected result: None of the classical definitions of a time–varying power spectrum allows to learn about the basic structure of the process. It is only the *combined* observation of a "short–window" (wideband) physical spectrum and a "long–window" (narrowband) physical spectrum that elucidates the structure of the process. This fact can be explained by the window matching theory of the following chapter: The considered overspread process is a combination of two underspread processes (one of them is almost nonstationary white and the other almost stationary) which are not *jointly underspread*. Hence, the short window is matched to the almost nonstationary white process and the long window is matched to the stationary process. It should be emphasized that the combined consideration of narrowband/wideband spectrograms is in common use by speech analysts.

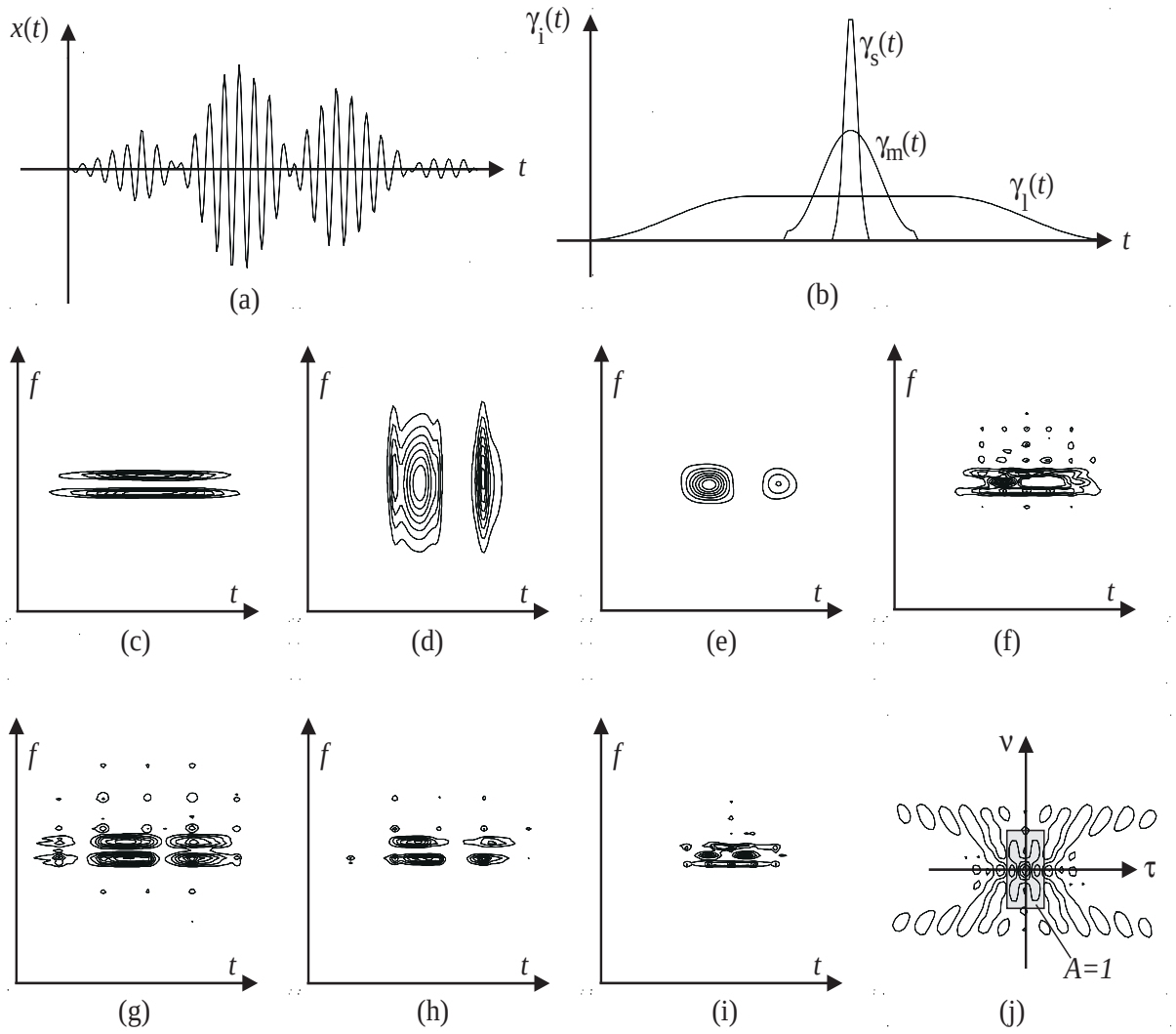

Figure 2.5: Various time–varying power spectra of an overspread process: (a) one realization of the process, (b) three different window functions: $\gamma_l(t)\cdots$"long" window, $\gamma_s(t)\cdots$"short" window, $\gamma_o(t)\cdots$"middle duration" window, (c) physical spectrum using $\gamma_l(t)$, (d) physical spectrum using $\gamma_s(t)$, (e) physical spectrum using $\gamma_m(t)$, (f) Wigner–Ville spectrum, (g) Rihaczek spectrum (generalized Wigner–Ville spectrum with $\alpha = 1/2$), (h) evolutionary spectrum (generalized evolutionary spectrum with $\alpha = 1/2$), (i) generalized evolutionary spectrum with $\alpha = 0$, (j) expected ambiguity function.

2.7 Summary

From an operator diagonalization point of view the Fourier transform is matched to wide–sense stationary processes and the Karhunen–Loeve (KL) transform to nonstationary processes with finite expected energy. We have pointed out two different ways of deriving the KL transform given the correlation operator of a zero–mean random process: (i) The KL transform can be obtained by diagonalizing the correlation operator. (ii) Alternatively, the KL transform yields optimum concentration of the diagonal coefficients.

The Wiener filter served as an example for the relevance of diagonalizing transforms. Using a diagonalizing transform, the abstract Wiener filter, $\mathbf{H}_{MMSE} = \mathbf{R}_x(\mathbf{R}_x + \mathbf{R}_n)^{-1}$, can be reformulated by scalar operations. In the stationary case one has in terms of the power spectra of the signal and noise processes:

$$H_{MMSE}(f) = \frac{S_x(f)}{S_x(f) + S_n(f)},$$

while in the nonstationary case (with commuting signal and noise correlation operators) the Wiener filter is determined by the KL eigenvalues of the signal and noise process: $\lambda_k^{(MMSE)} = \frac{\lambda_k^{(x)}}{\lambda_k^{(x)} + \lambda_k^{(n)}}$.

We have reviewed the most prominent definitions of a stochastic time–varying spectrum, the (window–dependent) physical spectrum, the generalized Wigner–Ville spectrum and the generalized evolutionary spectrum. We have introduced a compact notation for time–frequency shifting of linear operators in the form

$$\mathbf{P}^{(t,f)} \stackrel{\text{def}}{=} \mathbf{M}_f\mathbf{T}_t\mathbf{P}\left(\mathbf{M}_f\mathbf{T}_t\right)^{-1},$$

where $\mathbf{M}_f\mathbf{T}_t$ denotes a time–frequency shift operator. With this notation, one can formulate all of the considered definitions of a time–varying spectrum by a (formal) operator inner product. The physical spectrum can be written as

$$ESPEC_x^{(\gamma)}(t,f) = \left\langle \mathbf{R}_x, \mathbf{P}_\gamma^{(t,f)} \right\rangle,$$

where $\mathbf{P}_\gamma$ is the rank–one projection operator onto the analysis window.

The generalized Wigner–Ville spectrum,

$$EW_x^{(\alpha)}(t,f) = \left\langle \mathbf{R}_x, \mathbf{P}^{(t,f)}(\alpha) \right\rangle,$$

and the generalized evolutionary spectrum

$$ES_x^{(\alpha)}(t,f) = \left| \left\langle \sqrt{\mathbf{R}_x}, \mathbf{P}^{(t,f)}(\alpha) \right\rangle \right|^2$$

are based on the same (α–dependent) infinitesimal prototype operator which can be defined via the generalized Weyl correspondence as $L_{P(\alpha)}^{(\alpha)}(t,f) = \delta(t)\delta(f)$. We have pointed out the limited relevance of the time–varying spectra for general nonstationary processes due to a lack of a point–wise interpretation. This becomes obvious when asking about the time–frequency formulation of the Wiener filter:

$$H_{MMSE}(t,f) \stackrel{?}{=} \frac{E_x(t,f)}{E_x(t,f) + E_n(t,f)}.$$

We have discussed the *expected ambiguity function* EAF as an alternative second–order process representation. Based on the EAF we have introduced a fundamental classification of nonstationary random processes: Underspread processes are defined by a small product of spectral and temporal correlation width and, conversely, overspread processes by a large product. It has been shown that all of the considered time–varying spectra are 2D lowpass functions with identical bandlimits and are asymptotically equivalent (to include the physical spectrum we have to assume a matched window). For underspread processes the time–frequency formulation of the nonstationary Wiener filter indeed makes sense as we shall learn from the following chapters.

Chapter 3

Matched Weyl–Heisenberg Expansions of Nonstationary Processes

We derive the statistically optimum STFT/Gabor window matched to the second order statistic of a nonstationary process in the sense of an optimum approximate diagonalization via a Weyl–Heisenberg structured signal set. The STFT window criterion is compared to other criteria aiming at minimum global/local bias in spectrogram based estimation of the Wigner–Ville spectrum. For underspread processes, the matched window physical spectrum (expected spectrogram) is shown to provide a complete second order statistic. We discuss a simple window adaptation rule based on optimizing the duration for a given window shape. Finally, the extension to multiwindow methods is briefly pointed out.

3.1 Optimally Uncorrelated STFT Expansion

We assume a zero–mean, nonstationary, finite energy process $x(t)$ characterized by its correlation function $r_x(t,s) = \mathrm{E}\{x(t)x^*(s)\}$. When applying the short–time Fourier transform (STFT) to $x(t)$ it is our implicit hope that the STFT does quite the same job for a nonstationary process which the Fourier transform does for a stationary process. In particular, when the process $x(t)$ is assumed to be slowly nonstationary one may argue that the windowed portion of $x(t)$ is in good approximation second–order time–invariant ("locally stationary") which in turn justifies to decorrelate this portion via the Fourier transform. Although we know that precise decorrelation requires the Karhunen–Loeve transform it is still worthwhile to study the statistics of the STFT.

Statistics of the STFT. The STFT of the zero–mean, nonstationary process $x(t)$ is a two–dimensional (nonstationary) process that is also zero–mean,

$$\mathrm{E}\left\{ STFT_x^{(\gamma)}(t,f)\right\} = 0,$$

and its four–dimensional correlation has been obtained in the previous chapter as

$$R_{STFT}^{(x,\gamma)}(t,f,t-\tau,f-\nu) = \int_{\tau'}\int_{\nu'} EA_x(\tau',\nu')A_\gamma(\tau-\tau',\nu-\nu')\, e^{j2\pi[(\nu'-\nu)(t-\tau)+(\nu'-f)\tau']}d\tau' d\nu',$$

where $EA_x(\tau,\nu)$ is the expected ambiguity function of the process and $A_\gamma(\tau,\nu)$ is the ambiguity function of the window.

Continuous Off–Diagonal Norm. Exact decorrelation, i.e.,

$$R_{STFT}^{(x,\gamma)}(t,f,t',f') \stackrel{?}{=} 0 \qquad \text{for} \qquad (t\neq t') \vee (f\neq f'), \tag{3.1}$$

is impossible because the underlying expansion set $\gamma^{(t,f)}(s) = \gamma(s-t)e^{j2\pi fs}$ is highly linear dependent and statistically orthogonal coefficients require as a first prerequisite deterministic orthogonality.

However, we expect that depending on the choice of the window function the decorrelating property (3.1) holds in an approximate sense.

An obvious way to measure the deviation from optimum decorrelation (i.e. optimum diagonalization of a correlation operator) is a continuous off–diagonal norm (similar to the discrete case off–diagonal norm (2.11) that formally led to the KL transform):

$$M_K(x,\gamma) \stackrel{\text{def}}{=} \int_t \int_f \int_\tau \int_\nu \left| R_{STFT}^{(x,\gamma)}(t,f,t-\tau,f-\nu) \right|^2 (1-K(\tau,\nu))\, dt\, df\, d\tau\, d\nu, \tag{3.2}$$

where $K(\tau,\nu)$ is a radially non–increasing smoothing function satisfying

$$K(0,0)=1, \qquad \frac{\partial K(\tau,\nu)}{\partial \tau} \le 0, \qquad \frac{\partial K(\tau,\nu)}{\partial \nu} \le 0, \qquad 0 \le K \le 1, \tag{3.3}$$

such that $1-K(\tau,\nu)$ penalizes the off–diagonal spread of $R_{STFT}^{(x,\gamma)}(t,f,t',f')$.

It must be additionally noted that the L_2–norm of $R_{STFT}^{(x,\gamma)}(t,f,t',f')$ is equal to the L_2–norm of the correlation function (we always assume $\|\gamma\|=1$)

$$\begin{aligned}
\|R_{STFT}^{(x,\gamma)}\|^2 &= \int_t \int_f \int_{t'} \int_{f'} \left| R_{STFT}^{(x,\gamma)}(t,f,t',f') \right|^2 dt\, df\, dt'\, df' \\
&= \int_t \int_f \int_{t'} \int_{f'} \int_{t_1} \int_{t_2} \int_{t_3} \int_{t_4} r_x(t_1,t_2)\gamma^{(t',f')}(t_2)\gamma^{(t,f)*}(t_1) \\
&\quad \cdot r_x^*(t_3,t_4)\gamma^{(t',f')*}(t_4)\gamma^{(t,f)}(t_3) dt\, df\, dt'\, df'\, dt_1\, dt_2\, dt_3\, dt_4 \\
&= \int_{t_1} \int_{t_2} |r_x(t_1,t_2)|^2 dt_1\, dt_2 \\
&= \|\mathbf{R}_x\|^2,
\end{aligned} \tag{3.4}$$

where we have used the STFT–based resolution of the identity (F.2)

$$\int_t \int_f \gamma^{(t,f)}(t_1)\gamma^{(t,f)*}(t_2) dt\, df = \delta(t_1-t_2).$$

Hence, the global correlation measure $M_K(x,\gamma)$ is well defined for any process with square–integrable correlation function.

The Optimum Window. The correlation of the STFT coefficients depends on the process $x(t)$ and the analysis window $\gamma(t)$. Matching the STFT to the process in the sense of minimum global correlation is thus equivalent to the following window optimization problem:

$$\gamma_{opt,1}(t) = \arg\min_\gamma M_K^{(x,\gamma)} \qquad \text{subject to} \quad \|\gamma\|=1. \tag{3.5}$$

The expression for the STFT correlation (2.46) and in particular its special form for white noise (2.48):

$$\left| R_{STFT}^{(n,\gamma)}(t,f,t-\tau,f-\nu) \right|^2 = |A_\gamma(\tau,\nu)|^2$$

suggests an ideal window with $A_\gamma(\tau,\nu) \stackrel{?}{=} \delta(\tau)\delta(\nu)$ for which the magnitude of the STFT would indeed be given by the magnitude of the process' time–frequency correlation:

$$A_\gamma(\tau,\nu) \stackrel{?}{=} \delta(\tau)\delta(\nu) \qquad \Longrightarrow \qquad \left| R_{STFT}^{(x,\gamma)}(t,f,t-\tau,f-\nu) \right| = |EA_x(\tau,\nu)|.$$

However, the radar uncertainty principle ((F.29) for a normalized signal, $\|\gamma\| = 1$)

$$\int_{\tau}\int_{\nu} |A_\gamma(\tau,\nu)|^2 d\tau\, d\nu = 1, \qquad \text{with} \qquad A_\gamma(0,0) = 1,$$

forbids the existence of such an ideal window. Notwithstanding this fundamental limitation we know that different windows may yield different amount of additional, undesired time–frequency correlation. This is what we quantify by the global STFT correlation $M_K(x,\gamma)$ defined in (3.2).

Based on (2.46) the global STFT correlation is given by:

$$M_K(x,\gamma) = \int_t\int_f\int_\tau\int_\nu\int_{\tau_1}\int_{\nu_1}\int_{\tau_2}\int_{\nu_2} EA_x(\tau_1,\nu_1)A_\gamma(\tau-\tau_1,\nu-\nu_1)EA_x^*(\tau_2,\nu_2)A_\gamma^*(\tau-\tau_2,\nu-\nu_2)$$

$$e^{-j2\pi[(\nu_2-\nu_1)t+(\tau_2-\tau_1)f-\nu_1\tau_1+\nu_2\tau_2]}\,(1-K(\tau,\nu))\,dt\,df\,d\tau\,d\nu\,d\tau_1\,d\nu_1\,d\tau_2\,d\nu_2,$$

where four integrals collapse and we obtain the following expression for the off–diagonal norm

$$M_K(x,\gamma) = \|\mathbf{R}_x\|^2 - \int_\tau\int_\nu |A_\gamma(\tau,\nu)|^2 \left[\int_{\tau'}\int_{\nu'} |EA_x(\tau',\nu')|^2 K(\tau-\tau',\nu-\nu')d\tau'\,d\nu'\right] d\tau\,d\nu.$$

Minimizing $M_K(x,\gamma)$ by an optimal choice of $\gamma(t)$ is thus equivalent to the maximization of an $L_2(\mathbb{R}^2)$–inner product of the smoothed magnitude–squared expected ambiguity function of the process and the ambiguity function of the window:

$$\gamma_{opt,1} = \arg\min_\gamma M_K(x,\gamma) = \arg\max_\gamma \left\langle |EA_x|^2 * *K\,,\, |A_\gamma|^2 \right\rangle, \qquad \|\gamma\| = 1. \tag{3.6}$$

The limit case smoothing kernel $K(\tau,\nu) = \delta(\tau)\delta(\nu)$ does not correspond to an admissible penalty function (i.e., it does not satisfy (3.3)). In the following section we show however that — from a slightly different view of measuring the eigenfunction deviation — $K(\tau,\nu) = \delta(\tau)\delta(\nu)$ makes sense in (3.6).

3.1.1 Matching via Orthogonality Principle

Recall that optimum decorrelation of a nonstationary process is equivalent to the diagonalization of its correlation operator. If we restrict ourselves to a set of time–frequency shifted versions of the analysis window, $\{\gamma^{(\tau,\nu)}\}$, we hope that any member of this set is an approximate eigensignal of $\mathbf{R}_x$. For each individual time–frequency shifted version $\gamma^{(\tau,\nu)}(t)$ we can measure the deviation from an eigenfunction of $\mathbf{R}_x$ by splitting up $(\mathbf{R}_x\gamma^{(\tau,\nu)})(t)$ into a scalar multiple of $\gamma^{(\tau,\nu)}$ and an error function

$$\left(\mathbf{R}_x\gamma^{(\tau,\nu)}\right)(t) = c^{(\tau,\nu)}\gamma^{(\tau,\nu)}(t) + \epsilon^{(\tau,\nu)}(t),$$

where $c^{(\tau,\nu)}$ is a constant factor. This split–up gets unique by minimizing the L_2–norm of the error function via the orthogonality principle

$$\left\langle \epsilon_{min}^{(\tau,\nu)}, \gamma^{(\tau,\nu)} \right\rangle = 0.$$

Since we always presuppose $\|\gamma\| = 1$ the minimum error norm is given by

$$\|\epsilon_{min}^{(\tau,\nu)}\|^2 = \left\|\mathbf{R}_x\gamma^{(\tau,\nu)}\right\|^2 - \left|\left\langle \mathbf{R}_x\gamma^{(\tau,\nu)}, \gamma^{(\tau,\nu)} \right\rangle\right|^2$$

In order to match the entire STFT–expansion set $\{\gamma^{(\tau,\nu)}\}$ we integrate $\|\epsilon_{min}^{(\tau,\nu)}\|^2$ over the total range of τ and ν:

$$M(x,\gamma) = \int_\tau \int_\nu \|\epsilon_{min}^{(\tau,\nu)}\|^2 d\tau\, d\nu.$$

Using the STFT–based resolution of the identity (F.2) the first integral can be evaluated as follows

$$\begin{aligned}\int_\tau \int_\nu \left\|\mathbf{R}_x\gamma^{(\tau,\nu)}\right\|^2 d\tau\, d\nu &= \int_\tau \int_\nu \int_t \int_{t_1} \int_{t_2} r_x(t,t_1)\gamma^{(\tau,\nu)}(t_1) r_x^*(t,t_2)\gamma^{(\tau,\nu)*}(t_2) d\tau\, d\nu\, dt\, dt_1\, dt_2 \\ &= \int_t \int_{t'} |r_x(t,t')|^2 dt\, dt',\end{aligned}$$

for the second integration we introduce the STFT correlation in the form of (2.46):

$$\begin{aligned}\int_\tau \int_\nu \left|\left\langle \mathbf{R}_x\gamma^{(\tau,\nu)}, \gamma^{(\tau,\nu)}\right\rangle\right|^2 d\tau\, d\nu &= \int_\tau \int_\nu \left|R_{STFT}^{(x,\gamma)}(\tau,\nu,\tau,\nu)\right|^2 d\tau\, d\nu \qquad (3.7)\\ &= \int_\tau \int_\nu \int_{\tau_1} \int_{\nu_1} \int_{\tau_2} \int_{\nu_2} EA_x(\tau_1,\nu_1) EA_x^*(\tau_2,\nu_2) A_\gamma(-\tau_1,-\nu_1) \\ &\quad \cdot A_\gamma^*(-\tau_2,-\nu_2) e^{j2\pi[\tau(\nu_1-\nu_2)+(\nu_1-\nu)\tau_1-(\nu_2-\nu)\tau_2]} d\tau\, d\nu\, d\tau_1\, d\nu_1\, d\tau_2\, d\nu_2 \\ &= \int_\tau \int_\nu |EA_x(\tau,\nu)|^2 |A_\gamma(\tau,\nu)|^2 d\tau\, d\nu,\end{aligned}$$

combining these results we have

$$M(x,\gamma) = \|\mathbf{R}_x\|^2 - \left\langle |EA_x|^2, |A_\gamma|^2 \right\rangle, \qquad \|\gamma\| = 1.$$

Hence, the window matching via orthogonality principle leads to the following optimization criterion:

$$\gamma_{opt,2} = \arg\min_\gamma M(x,\gamma) = \arg\max_\gamma \left\langle |EA_x|^2, |A_\gamma|^2 \right\rangle, \qquad \|\gamma\| = 1. \qquad (3.8)$$

Comparing with (3.6), we see that the result of window matching via orthogonality principle (3.8) is equivalent to the use of $K(\tau,\nu) = \delta(\nu)\delta(\tau)$ in the definition of the off–diagonal norm $M_K(x,\gamma)$ (3.2). Then however, $M_K(x,\gamma)$ is no more an off—diagonal norm rather it reflects the norm of the diagonal itself. Before we discuss the optimum windows in more detail we point out another intuitively appealing derivation of (3.8).

3.1.2 From Optimum STFT to Optimum Spectrogram

Analogously to the quadratic optimality criterion that leads to the Karhunen–Loeve transform (2.14), the "sharp off–diagonal measure" $M(x,\gamma)$ as defined in (3.8) admits an interpretation in terms of optimum power concentration in the coefficient distribution. The diagonal of the STFT correlation equals the physical spectrum $ESPEC_x^{(\gamma)}(t,f)$ (see (2.16)):

$$R_{STFT}^{(x,\gamma)}(t,f,t,f) = \mathrm{E}\left\{\left|STFT_x^{(\gamma)}(t,f)\right|\right\}^2 = ESPEC_x^{(\gamma)}(t,f). \qquad (3.9)$$

We define a concentration measure of the expected spectrogram analog to the concentration measure that formally led to the KL transform (see (2.14)) presupposing a process with normalized expected energy $\mathrm{tr}\mathbf{R}_x = 1$

$$\widetilde{M}(x,\gamma) = \int_t \int_f \left\{ESPEC_x^{(\gamma)^2}(t,f) - ESPEC_x^{(\gamma)}(t,f)\right\} dt\, df.$$

However, since using (3.9) and (2.46) we have:

$$\begin{aligned}\int_t \int_f ESPEC_x^{(\gamma)}(t,f)dt\,df &= \int_t \int_f \int_\tau \int_\nu EA_x(\tau,\nu A_\gamma(-\tau,-\nu)e^{j2\pi[\nu t+(\nu-f)\tau]}dt\,df\,d\tau\,d\nu \\ &= EA_x(0,0)A_\gamma(0,0) \\ &= \mathrm{tr}\mathbf{R}_x,\end{aligned}$$

and with (3.7)

$$\begin{aligned}\int_t \int_f {ESPEC_x^{(\gamma)}}^2(t,f)dt\,df &= \int_t \int_f \left|R_{STFT}^{(x,\gamma)}(t,f,t,f)\right|^2 dt\,df \\ &= \left\langle |EA_x|^2, |A_\gamma|^2 \right\rangle\end{aligned}$$

we can conclude that the maximization of the concentration measure $\widetilde{M}(x,\gamma)$ is equivalent to the matching via orthogonality principle:

$$\arg\max_\gamma \widetilde{M}(x,\gamma) = \arg\min_\gamma M(x,\gamma).$$

Hence we see that *matching the STFT expansion set via the orthogonality principle is equivalent to optimizing the concentration of the expected spectrogram.* We remind the reader that this is just what we expect by generalizing the concept of the Fourier and KL transform: statistically optimum linear representation leads to a corresponding statistically optimum quadratic representation.

Moreover, whenever both $EA_x(\tau,\nu)$ and $A_\gamma(\tau,\nu)$ are smooth and well–concentrated about the origin (which is usually satisfied in cases of practical interest) then $\left|R_{STFT}^{(x,\gamma)}(t,f,t',f')\right|^2$ is a well–behaved function. Then it is clear that maximizing the diagonal $\left|R_{STFT}^{(x,\gamma)}(t,f,t,f)\right|^2$ must diminish the off–diagonal spread of $\left|R_{STFT}^{(x,\gamma)}(t,f,t',f')\right|^2$ since the total integral is independent w. r. t. a variation of $\gamma(t)$ on the unit sphere (see (3.4)).

3.2 Discussion of the Matched STFT Window

3.2.1 Twofold Ambiguity

In the window optimization criterion (3.6) the ambiguity functions of the window and of the process both appear magnitude squared, a fact that entails a twofold ambiguity.

First, the optimum window is not uniquely determined. Rather there exists at least[1] a whole Weyl–Heisenberg family of optimum windows, $\gamma_{opt}^{(\tau,\nu)}(t)$, since the cost function of the window optimization criterion is invariant w.r.t. a time–frequency shift of the window. However, in most practical applications a window function that is not spectrally and temporally centered about zero is not acceptable as it falsifies the (absolute) time–frequency localization of the signal. Fortunately, in the iterative numerical solution of the window optimization [206] it suffices to start with a time–frequency centered window in order to prevent convergence to "time–frequency–biased" solutions.

Second, there exist processes with different correlation functions but identical optimum window. A trivial example is a Weyl–Heisenberg family of processes $x^{(\tau,\nu)}(t)$. This ambiguity makes our window optimization theory even more useful since it allows *to match the window to the incomplete a priori knowledge of a correlation kernel.* It should be stressed that such incomplete a priori knowledge prevents the application of the KL transform.

[1] Apart from other potential, less obvious ambiguities of the optimum window.

3.2.2 Choice of Smoothing Kernel

The smoothing kernel $K(\tau,\nu)$ has been introduced for the design of the penalty (weight) function in the continuous off–diagonal norm (3.2). It is obvious that with increasing spread of $K(\tau,\nu)$ and decreasing spread of $EA_x(\tau,\nu)$ the choice of the smoothing kernel has considerable influence on the result. For the limit case of stationary white noise the window optimization theory is no more valid (it is restricted to trace class correlation kernels), it is however illuminating to look at this case in an approximate way:

$$|EA_x(\tau,\nu)|^2 \approx \delta(\tau)\delta(\nu) \qquad \Longrightarrow \qquad \gamma_{opt} \approx \arg\max_{\gamma} \left\langle |A_\gamma|^2, K \right\rangle, \qquad \|\gamma\| = 1.$$

We see that for approximate stationary white processes the optimum window tends to depend entirely on the choice of the smoothing kernel. This is not astonishing if one is aware of the fact that in this case the KL basis itself is highly ambiguous, since any orthonormal basis leads to uncorrelated coefficients. It is only our desire to penalize temporal and spectral correlation of the STFT coefficients that determines the optimum window. Moreover, in Section 3.1.1 we have seen that even the ideally concentrated kernel $K(\tau,\nu) = \delta(\tau)\delta(\nu)$ makes sense although it is not admissible in the sense of measuring the off–diagonal spread of the STFT correlation[2].

Hence, in practice one could completely abandon the smoothing kernel apart from two exceptions: (i) When there is a need to penalize temporal and spectral correlation in an unequal manner, this can be expressed by the choice of a radially non–symmetric $K(\tau,\nu)$. (ii) In practice, the window optimization will be based on an *estimate of* $EA_x(\tau,\nu)$. Whenever this estimate appears to be highly oscillatory then a slight smoothing may help to improve convergence in the iterative numerical optimization.

3.2.3 Residual Correlation

Given the optimized window we now ask about the residual global correlation of the STFT coefficients. Using (3.8) and (2.52) one can write the global correlation in the following form (for simplicity we assume negligible smoothing, i.e. $K(\tau,\nu) \approx \delta(\tau)\delta(\nu)$):

$$M(x,\gamma) = \int_\tau \int_\nu |EA_x(\tau,\nu)|^2 \left(1 - |A_\gamma(\tau,\nu)|^2\right) d\tau\, d\nu.$$

It is obvious that the optimum window should have an ambiguity function that is close to one in that area of the (τ,ν)–plane where $|EA_x(\tau,\nu)| > 0$. Restricting the discussion to typical time–frequency–localized window functions, the radar uncertainty principle for normalized signals,

$$\int_\tau \int_\nu |A_\gamma(\tau,\nu)|^2 d\tau\, d\nu = 1 \qquad \text{and} \qquad |A_\gamma(0,0)|^2 = 1 \geq |A_\gamma(\tau,\nu)|^2,$$

shows that the underspread/overspread classification plays an important role as explained in what follows.

Impact of Underspread/Overspread Classification. A usual window is a well time–frequency–localized signal with an ambiguity function that can be coarsely approximated by a rectangular indicator function with area one. Then it is clear that for an underspread process the optimum window indeed achieves

$$|A_\gamma(\tau,\nu)| \approx 1 \qquad \text{where} \qquad |EA_x(\tau,\nu)| > 0,$$

i.e., we thus have small global correlation. On the other hand, for an overspread process, we cannot find a window that achieves such a perfect matching since due to the radar uncertainty principle the area with $|A_\gamma(\tau,\nu)| \approx 1$ cannot be larger than one.

[2] In a discrete setting this problem disappears, since the delta distributions carry over to Kronecker functions and the ideally concentrated kernel establishes a valid off–diagonal norm.

3.3 Time–Varying Spectral Estimation and the Spectrogram

In Section 3.1.1 we have seen that the optimum STFT of a nonstationary process establishes an expected spectrogram (physical spectrum) that is optimum in the sense of concentration. The natural question is to ask about the relation of this optimum physical spectrum to classical window–independent definitions of a time–varying spectrum. One may suppose that the optimally concentrated spectrogram is in some sense also optimum as an estimator of the generalized Wigner–Ville spectrum. In this section we study various window optimization criteria which aim at minimum bias of the spectrogram as an estimator of the generalized Wigner–Ville spectrum.

The above–discussed "nonstationary Wiener–Khintchine relation"

$$EA_x^{(\alpha)}(\tau,\nu) = \int_t \int_f EW_x^{(\alpha)}(t,f) e^{-j2\pi(\nu t-\tau f)} dt df = \mathcal{F}_{t\to\nu}\mathcal{F}^{-1}_{f\to\tau}\left\{EW_x^{(\alpha)}(t,f)\right\},$$

provides the starting point in investigating this issue. Observe that due to the α–dependence of the expected generalized Wigner distribution we can a priori exclude a simple relation to the α–independent STFT window optimization criterion. However, we hope that in cases of practical interest all window optimization criteria lead to essentially the same optimum.

3.3.1 Bias of Spectrogram

The expected spectrogram is a smoothed version of the expected generalized Wigner distribution (GWD), where the smoothing kernel is the GWD of the analysis window $\gamma(t)$:

$$ESPEC_x^{(\gamma)}(t,f) = EW_x^{(\alpha)}(t,f) * *W_\gamma^{(\alpha)*}(-t,-f).$$

Considering the expected spectrogram as an estimate of the expected GWD, the bias clearly depends on the choice of the analysis window[3]:

$$B_{SPEC}^{(\gamma)}(t,f) \stackrel{\text{def}}{=} EW_x^{(\alpha)}(t,f) - ESPEC_x^{(\gamma)}(t,f).$$

For a general nonstationary process it is hard to set up a useful criterion for the choice of $\gamma(t)$. However, by considering processes with compactly supported EAF we are able to obtain concrete window optimization criteria.

We henceforth presuppose a priori knowledge of the 0/1–valued characteristic function $\chi_x(\tau,\nu)$ of the support of $EA_x^{(\alpha)}(\tau,\nu)$, defined by

$$EA_x^{(\alpha)}(\tau,\nu) = EA_x^{(\alpha)}(\tau,\nu)\,\chi_x(\tau,\nu) \qquad \text{and} \qquad \int_\tau \int_\nu \chi_x(\tau,\nu) d\tau\, d\nu \to \min. \tag{3.10}$$

The minimum area side constraint is intuitively clear but requires further theoretical reasoning. In Appendix D, it is shown that the indicator function of the support of $EA_x(\tau,\nu)$ establishes the *minimum variance unbiased estimator* of the Wigner–Ville spectrum. This fact gives the motivation to take the smallest indicator function among those which contain the support of $EA_x(\tau,\nu)$.

For the following derivations we mention two properties of the indicator function $\chi_x(\tau,\nu)$: (i) it does not depend on α as the magnitude of $EA_x(\tau,\nu)$ is α–invariant, (ii) one has

$$\chi_x(\tau,\nu) = \chi_x(-\tau,-\nu),$$

since

$$|EA_x^{(\alpha)}(\tau,\nu)| = |EA_x^{(\alpha)}(-\tau,-\nu)|.$$

[3] For simplicity, we here do not consider more general estimators of the Wigner–Ville spectrum and refer to Appendix D. There, it turns out that the window matching is indeed a mere bias problem, i.e. it suffices to compare the *expected* spectrogram with the *expected* generalized Wigner distribution. The spectogram with its specific structure does not allow a bias versus variance trading, for a typical underspread process it is usually a suboptimal estimator. The natural step to multi–window estimators will be discussed in Section 3.5.

3.3.2 Minimum Bias Window

Based on (3.10) one can easily obtain a tight bound on the bias magnitude via triangle and Schwarz inequality:

$$\begin{aligned}\left|B_{SPEC}^{(\gamma)}(t,f)\right| &= \left|EW_x^{(\alpha)}(t,f) - ESPEC_x^{(\gamma)}(t,f)\right| \\ &= \left|\int_\tau\int_\nu \left\{EA_x^{(\alpha)}(\tau,\nu) - EA_x^{(\alpha)}(\tau,\nu)A_\gamma^{(\alpha)*}(\tau,\nu)\right\} e^{-j2\pi(f\tau-t\nu)}d\tau\,d\nu\right| \\ &\leq \int_\tau\int_\nu \left|EA_x^{(\alpha)}(\tau,\nu)\right|\left|\chi_x(\tau,\nu) - A_\gamma^{(\alpha)*}(\tau,\nu)\right| d\tau\,d\nu \\ &\leq \|\mathbf{R}_x\|\left\|\chi_x - A_\gamma^{(\alpha)}\right\|. \end{aligned} \tag{3.11}$$

Minimization of the upper bound leads to a classical signal synthesis problem [371, 339],

$$\gamma_{opt,3} = \arg\min_\gamma \left\|\chi_x - A_\gamma^{(\alpha)}\right\|^2 = \arg\max_\gamma \left\langle \chi_x, A_\gamma^{(\alpha)}\right\rangle \qquad \text{subject to} \qquad \|\gamma\|^2 = 1, \tag{3.12}$$

i.e., one has to find that normalized signal $\gamma(t)$ whose ambiguity function comes closest to a given target function. (This optimization problem is slightly less complicated than (3.5), it leads to a (partial) eigenvalue problem.) In contrast to the statistically optimal STFT window which is invariant w.r.t. the choice of α, the minimum bias window will be α–dependent. This fact is not at all astonishing, since $\gamma_{opt,3}$ minimizes the bias of the spectrogram as an estimator of the α–dependent GWD.

3.3.3 Minimum Bias Window for Real–Part of the GWVS

The generalized Wigner–Ville spectrum (GWVS) is generally complex–valued. By considering the real–part of the GWVS one has to sacrifice some of the usually required properties of a time–varying spectrum [159].

By the employment of (F.21) and (2.61) one can easily parallel the derivation in (3.11):

$$\begin{aligned}\left|B_{SPEC,R}^{(\gamma)}(t,f)\right| &= \left|\mathrm{Re}\left\{EW_x^{(\alpha)}(t,f)\right\} - ESPEC_x^{(\gamma)}(t,f)\right| \\ &\leq \int_\tau\int_\nu \left|EA_x^{(0)}(\tau,\nu)\right|\left|\chi_x(\tau,\nu)\cos 2\pi\tau\nu\alpha - A_\gamma^{(0)*}(\tau,\nu)\right| d\tau\,d\nu \\ &\leq \|\mathbf{R}_x\|\left\|\tilde{\chi}_x^{(\alpha)} - A_\gamma^{(0)}\right\|. \end{aligned} \tag{3.13}$$

where we have introduced a modification of $\chi_x(\tau,\nu)$ which is no more an indicator function:

$$\tilde{\chi}_x^{(\alpha)}(\tau,\nu) \stackrel{\text{def}}{=} \chi_x(\tau,\nu)\cos 2\pi\tau\nu\alpha.$$

(For underspread processes this modification will be negligible since $\cos 2\pi\tau\nu\alpha \approx 1$ for $\tau_0\nu_0 \ll 1$.) Equation (3.13) results in a slightly modified version of the minimum bias window criterion (3.12): we have the structurally equivalent optimization problem now expressed as a signal synthesis problem in terms of the symmetrical ambiguity function $A_\gamma^{(0)}(\tau,\nu)$ and an α–dependent target function:

$$\gamma_{opt,4} = \arg\max_\gamma \left\langle \tilde{\chi}_x^{(\alpha)}, A_\gamma^{(0)}\right\rangle \qquad \text{subject to} \qquad \|\gamma\|^2 = 1. \tag{3.14}$$

3.3.4 Matching via Symbolic Calculus

The physical spectrum can be viewed as a linear time–frequency representation ("Weyl–Heisenberg symbol") of the correlation operator with the inner product formulation (see (2.19) and (2.20)):

$$ESPEC_x^{(\gamma)}(t,f) = \left\langle \mathbf{R}_x, \mathbf{P}_\gamma^{(t,f)} \right\rangle,$$

where $\mathbf{P}_\gamma$ denotes the rank–one projection operator onto the analysis window γ. The generalized Wigner–Ville spectrum establishes a different "Weyl–Heisenberg symbol" of the correlation operator (see Section 2.5.1):

$$EW_x^{(\alpha)}(t,f) = \left\langle \mathbf{R}_x, \mathbf{P}^{(t,f)}(\alpha) \right\rangle,$$

where the prototype operator $\mathbf{P}(\alpha)$ is given by the support constraint of the EAF:

$$S_{P(\alpha)}^{(\alpha)}(\tau,\nu) = \chi_x(\tau,\nu).$$

Now it is near at hand to define the optimum spectrogram window by matching the rank–one prototype operator $\mathbf{P}_\gamma$ to the general prototype operator $\mathbf{P}(\alpha)$. This is indeed just the previously discussed window $\gamma_{opt,3}$ that achieves minimum squared bias:

$$\gamma_{opt,3}(t) = \arg\min_\gamma \|\mathbf{P}(\alpha) - \mathbf{P}_\gamma\|^2 = \arg\max_\gamma \left\langle \chi_x, A_\gamma^{(\alpha)} \right\rangle \qquad \text{subject to} \qquad \|\gamma\|^2 = 1.$$

The same reasoning holds when studying the Weyl–Heisenberg expansion via the real–valued generalized Wigner distribution; the symbolic calculus leads to $\gamma_{opt,4}$. Hence, we see that the abstract mathematical view leads to formal short–cut derivations.

3.3.5 Approximate Equivalence

The introduced window optimization criteria lead to complicated nonlinear constrained optimization problems. In practice the numerical expense for the iterative solution may prohibit the application of the optimum window. This motivates the derivation of approximate window criteria aiming at low–cost design. Fortunately, we will see that, for underspread processes, all of the previously discussed window optimization criteria lead to a unique, intuitively appealing window matching rule.

Assume that $EA_x(\tau,\nu)$ is approximately given by an 0/1 valued indicator function $\chi_x(\tau,\nu)$ with area much smaller than one:

$$EA_x(\tau,\nu) = \chi_x(\tau,\nu) \qquad \text{with} \qquad \int_\tau \int_\nu \chi_x(\tau,\nu) d\tau\, d\nu \ll 1. \tag{3.15}$$

We can replace the ambiguity function of a (real–valued) window by its Taylor approximation (see Section F.0.4):

$$A_\gamma^{(\alpha)}(\tau,\nu) \approx 1 - \frac{1}{2\pi}\left(F_\gamma^2\tau^2 + T_\gamma^2\nu^2\right), \tag{3.16}$$

$$\left|A_\gamma^{(\alpha)}(\tau,\nu)\right|^2 \approx 1 - \frac{1}{\pi}\left(F_\gamma^2\tau^2 + T_\gamma^2\nu^2\right), \tag{3.17}$$

where T_γ^2 and F_γ^2 are the temporal and spectral moments of order two defined as:

$$\begin{aligned} T_\gamma^2 &= 4\pi \int_t t^2 |\gamma(t)|^2 dt, \\ F_\gamma^2 &= 4\pi \int_f f^2 |\Gamma(f)|^2 df, \end{aligned}$$

where $\Gamma(f) = (\mathcal{F}\gamma)(f)$. The square–roots of these moments, T_γ and F_γ are one natural, mathematically convenient way to define the *duration* and *bandwidth* of a window function. In particular, the uncertainty principle for $\gamma(t)$ can be formulated as

$$T_\gamma F_\gamma \geq 1, \tag{3.18}$$

where equality holds only for the Gaussian signal, which is exactly determined by its second order moments:

$$g(t) = \sqrt[4]{2F_\gamma/T_\gamma}\, e^{-\pi(F_\gamma/T_\gamma)t^2}.$$

More generally, for most applications good windows are always near to a Gaussian function in the sense that

$$T_\gamma F_\gamma \approx 1,$$

and the main problem is to determine the optimum duration/bandwidth ratio T_γ/F_γ.

To this end, we consider a family of (unit energy) windows $\gamma_a(t)$ which are dilated versions of a (energy normalized) reference window $\gamma_0(t)$:

$$\gamma_a(t) = \sqrt{a}\gamma_0(at). \tag{3.19}$$

The dilation factor a characterizes the duration–bandwidth ratio of the window $\gamma_a(t)$. With the well–known facts

$$T_\gamma = \frac{T_0}{a}$$

and

$$F_\gamma = aF_0,$$

one has

$$\frac{T_\gamma}{F_\gamma} = \frac{1}{a^2}\frac{T_0}{F_0},$$

where T_0^2 and F_0^2 denote the temporal and spectral moments of the reference window $\gamma_0(t)$ and T_γ^2 and F_γ^2 are the respective moments of the dilated version $\gamma_a(t)$.

The approximations (3.15), (3.16), and (3.17) provide the basis for replacing the general (multivariate) window optimization criteria by the optimization of the single parameter $\frac{T_\gamma}{F_\gamma}$.

For the window optimization criteria (3.5) (with negligible smoothing $K(\tau,\nu) \approx \delta(\tau)\delta(\nu)$) and (3.8) we obtain the following approximation ($i \in \{1,2\}$):

$$\left(\frac{T_\gamma}{F_\gamma}\right)_{opt,i} = \arg\max_{\left(\frac{T_\gamma}{F_\gamma}\right)} \left\langle \chi_x, |A_\gamma|^2 \right\rangle \approx \arg\max_{\left(\frac{T_\gamma}{F_\gamma}\right)} \int_\tau \int_\nu \chi_x(\tau,\nu) \left(T_\gamma^2\nu^2 + F_\gamma^2\tau^2\right) d\tau\, d\nu.$$

The α–dependent design criteria (3.12), (3.14) get approximately α–independent since one has:

$$e^{j2\pi\pi\nu\tau\alpha} \approx 1,$$

for $\tau\nu \ll 1$. (Note that we always consider $|\alpha| \leq 1/2$.) Hence, the window optimization criteria (3.12), (3.14) lead to an α–independent approximate cost function ($i \in \{3,4\}$):

$$\left(\frac{T_\gamma}{F_\gamma}\right)_{opt,i} = \arg\max_{\left(\frac{T_\gamma}{F_\gamma}\right)} \langle \chi_x, A_\gamma \rangle \approx \arg\max_{\left(\frac{T_\gamma}{F_\gamma}\right)} \int_\tau \int_\nu \chi_x(\tau,\nu) \left(T_\gamma^2\nu^2 + F_\gamma^2\tau^2\right) d\tau\, d\nu.$$

It is equivalent to the approximate optimization of $\gamma_{opt,i}, i \in \{1,2\}$. That is, with regard to matching the window to the support of the EAF (in case of incomplete a priori knowledge) we have already established the approximate equivalence of all previously defined window optimization criteria

$$\gamma_{opt,1}(t) \approx \gamma_{opt,2}(t) \approx \gamma_{opt,3}(t) \approx \gamma_{opt,4}(t) \qquad \text{for} \qquad \tau_0\nu_0 \ll 1$$

In order to obtain a simple window matching rule we restrict the discussion to specific shapes of $\chi_x(\tau,\nu)$. Both for rectangular shape

$$\chi_x(\tau,\nu) = \chi_{[-\tau_0,\tau_0]}(\tau)\chi_{[-\nu_0,\nu_0]}(\nu)$$

and elliptical symmetry of $\chi_x(\tau,\nu)$

$$\chi_x(\tau,\nu) = \begin{cases} 1, & \left(\frac{\tau}{\tau_0}\right)^2 + \left(\frac{\nu}{\nu_0}\right)^2 \leq 1, \\ 0, & \text{else} \end{cases}$$

one obtains an intuitively appealing solution for the optimum duration–bandwidth–ratio:

$$\boxed{\left(\frac{T_\gamma}{F_\gamma}\right)_{opt} = \frac{\tau_0}{\nu_0}.} \tag{3.20}$$

Proof: We introduce the following notation:

$$\begin{array}{lcl} p \stackrel{\text{def}}{=} \dfrac{T_\gamma}{F_\gamma}, & & q \stackrel{\text{def}}{=} T_\gamma F_\gamma, \\ & \Updownarrow & \\ pq = T_\gamma^2, & & \dfrac{q}{p} = F_\gamma^2. \end{array}$$

The cost function for rectangular shape of $\chi_x(\tau,\nu)$ can be evaluated straightforward:

$$\begin{aligned} Q_1(p) &= \int_{-\tau_0}^{\tau_0}\int_{-\nu_0}^{\nu_0}\left(pq\nu^2 + \frac{q}{p}\tau^2\right)d\tau\, d\nu \\ &= \frac{4}{3}\tau_0\nu_0 q\left(p\nu_0^2 + \frac{1}{p}\tau_0^2\right). \end{aligned}$$

For the elliptical case we introduce another short notation:

$$s(\tau) \stackrel{\text{def}}{=} \nu_0\sqrt{1-\left(\frac{\tau}{\tau_0}\right)^2}.$$

The integration shows that the stationary points for rectangular and elliptical shape are identical:

$$\begin{aligned} Q_2(p) &= \int_{-\tau_0}^{\tau_0}\int_{\nu=-s(\tau)}^{s(\tau)}\left(pq\nu^2 + \frac{q}{p}\tau^2\right)d\tau\, d\nu \\ &= 2\int_{-\tau_0}^{\tau_0}\left(pq\frac{s^3(\tau)}{3} + \frac{q}{p}\tau^2 s(\tau)\right)d\tau \\ &= 2\int_{-\pi/2}^{\pi/2}\left(pq\frac{\nu_0^3}{3}\tau_0\cos^4 x + \frac{q}{p}\tau_0^3\nu_0\sin^2 x\cos^2 x\right)dx \\ &= \frac{\pi}{4}\tau_0\nu_0\left(p\nu_0^2 + \frac{1}{p}\tau_0^2\right). \end{aligned}$$

Here, we have used the substitution:

$$\frac{\tau}{\tau_0} = \sin x,$$

and the following integrals

$$\int_{-\pi/2}^{\pi/2}\cos^4 x dx = \frac{3\pi}{8},$$

$$\int_{-\pi/2}^{\pi/2} \sin^2 x \cos^2 x dx = \frac{\pi}{8}.$$

Differentiation of the p–dependent term of the cost functions $Q_1(p), Q_2(p)$ leads to

$$\frac{\partial Q_i}{\partial p} = K_i \left(\nu_0^2 - \frac{1}{p^2} \tau_0^2 \right), \qquad i \in \{1, 2\}.$$

These functions vanish for $p = \frac{\tau_0}{\nu_0}$, which finishes our proof. ∎

The matching rule can be alternatively written in terms of an optimum dilation factor (see (3.19)):

$$a_{opt} = \sqrt{\frac{T_0}{F_0} \frac{\nu_0}{\tau_0}},$$

where T_0^2 and F_0^2 are the temporal and spectral moments of the reference window $\gamma(t)$ (the optimum window is then given by $\sqrt{a_{opt}}\gamma(a_{opt}t)$). It should be emphasized that according to our fundamental matching rule the optimum window length depends *both on the spectral and temporal variations of the nonstationary process* in a symmetrical manner. This fact is in obvious contradiction with the usual rule of thumb where one selects the longest possible window such that the windowed part of the process is approximately stationary. In this rule one completely disregards the spectral variations (f–dependence of the time–varying spectrum) due to a temporal correlation of the process.

3.3.6 General Elliptical Symmetry and the Gaussian Window

We consider two positive, real–valued 2D functions $C_i(\tau, \nu)$ with elliptical symmetry as defined by

$$C_i(\tau, \nu) = \bar{C}_i \left(\left(\frac{\tau}{\tau_0} \right)^2 + \left(\frac{\nu}{\nu_0} \right)^2 \right), \qquad i \in \{1, 2\}. \tag{3.21}$$

When we insert such functions into the window optimization criteria (3.8),

$$\gamma_{opt,2} = \arg\max_{\gamma} \left\langle C_1, |A_\gamma^{(\alpha)}|^2 \right\rangle \qquad \text{subject to} \qquad \|\gamma\| = 1,$$

and (3.12) specialized to $\alpha = 0$:

$$\gamma_{opt,3} = \arg\max_{\gamma} \left\langle C_2, A_\gamma^{(0)} \right\rangle \qquad \text{subject to} \qquad \|\gamma\| = 1,$$

then, one can find the Gaussian pulse with matched duration as an — at least local — optimum of the window optimization problem (other local optima are the Hermite functions of order $k \geq 2$ which do not establish a useful STFT window):

$$\gamma_{opt,2}(t) = \gamma_{opt,3}(t)\Big|_{\alpha=0} = \sqrt[4]{2F_\gamma/T_\gamma}\, e^{-\pi(F_\gamma/T_\gamma)t^2},$$

where T_γ^2 and F_γ^2 are the spectral and temporal moments defined according to (F.30), (F.31). Here, the matching rule (3.20) holds exactly and remains valid for the case of overspread processes

$$\frac{T_\gamma}{F_\gamma} = \frac{\tau_0}{\nu_0}.$$

For the case of $\gamma_{opt,3}$ with $\alpha = 0$ this is essentially well–known [371, 339]. However, for the (α–independent) $\gamma_{opt,2}$ we are not aware of existing analytical results. We now show that the Hermite

functions with matched scale satisfy the necessary condition equation of a Lagrange function corresponding to the constrained optimization problem leading to $\gamma_{opt,2}$.

Proof: The proof is based on the fact that [371, 125] any function $C(\tau,\nu) \in L_2(\mathbb{R}^2)$ with elliptical symmetry defined as in (3.21) can be expanded into a weighted sum of the (auto–)ambiguity functions

$$C(\tau,\nu) = \sum_{k=1}^{\infty} a_k A_{h_k}(\tau,\nu), \tag{3.22}$$

where $h_k(t)$ denotes the orthonormal Hermite function of order k with matched duration [1]

$$h_k(t) = \sqrt[4]{\frac{2\nu_0}{\tau_0}} \frac{1}{\sqrt{n!}} \left(\frac{-1}{2\sqrt{\pi}}\right)^n e^{\pi\left(\frac{\nu_0}{\tau_0}\right)t^2} \frac{d^n}{dt^n} e^{-2\pi\left(\frac{\nu_0}{\tau_0}\right)t^2}. \tag{3.23}$$

We now perform a variational calculus using the orthonormal basis of the matched Hermite functions. We can write the prototype signal $\gamma(t) \in L_2(\mathbb{R})$ via the coefficients $g(k) \in l_2(\mathbb{Z})$:

$$\gamma(t) = \sum_{k=1}^{\infty} g(k) h_k(t) \qquad \text{with} \qquad g(k) = \langle g, h_k \rangle .$$

The squared magnitude of the ambiguity function of $g(t)$ can be written in terms of the cross–ambiguity functions of the Hermite functions $h_k(t)$ and the coefficients $g(k)$

$$|A_\gamma(\tau,\nu)|^2 = \sum_{k=1}^{\infty}\sum_{l=1}^{\infty}\sum_{k'=1}^{\infty}\sum_{l'=1}^{\infty} g(k)g^*(l)g(k')g^*(l') A_{h_k,h_l}(\tau,\nu) A^*_{h_{k'},h_{l'}}(\tau,\nu). \tag{3.24}$$

The autoambiguity functions $A_{h_k,h_k}(\tau,\nu) \stackrel{\text{def}}{=} A_{h_k}(\tau,\nu)$ of the Hermite functions are real–valued and given by Laguerre functions of $(\tau^2+\nu^2)$:

$$A_{h_k}(\tau,\nu) = e^{-(\pi/2)r^2} L_k(\pi r^2), \qquad \text{with} \qquad r^2 \stackrel{\text{def}}{=} \tau^2+\nu^2, \tag{3.25}$$

where $L_k(x)$ are the Laguerre polynomials [1]:

$$L_k(x) \stackrel{\text{def}}{=} \sum_{l=0}^{k} \binom{k}{l} \frac{(-x)^l}{l!}.$$

(The general cross–ambiguity functions of the Hermite functions are complex–valued and given by the generalized Laguerre functions.) We incorporate the side constraint using the method of Lagrange (ρ is the Lagrange multiplier)

$$J = \sum_{k=1}^{\infty}\sum_{l=1}^{\infty}\sum_{k'=1}^{\infty}\sum_{l'=1}^{\infty} g(k)g^*(l)g^*(k')g(l')Q(k,l,k',l') - \rho\left(\sum_{k=1}^{\infty} g(k)g^*(k) - 1\right), \tag{3.26}$$

where we have introduced the short notation

$$Q(k,l,k',l') \stackrel{\text{def}}{=} \int_\tau \int_\nu C(\tau,\nu) A_{h_k,h_l}(\tau,\nu) A^*_{h_{k'},h_{l'}}(\tau,\nu) d\tau\, d\nu. \tag{3.27}$$

The gradient w.r.t. $g^*(m)$ is then given by

$$\frac{\partial J}{\partial g^*(m)} = \sum_{k=1}^{\infty}\sum_{k'=1}^{\infty}\sum_{l'=1}^{\infty} g(k)g^*(k')g(l')Q(k,m,k',l') + \sum_{k=1}^{\infty}\sum_{l=1}^{\infty}\sum_{l'=1}^{\infty} g(k)g^*(l)g(l')Q(k,l,m,l') - \rho g(m)$$

We have to show that $g(m) = \delta_{mn}$ gives a stationary point of J, i.e., $\frac{\partial J}{\partial g^*(m)} = 0$, we have the necessary condition

$$Q(n, m, n, n) + Q(n, n, m, n) = \rho\delta_{mn}.$$

Since we know that the auto–ambiguity functions are real–valued (see (3.25)) and with

$$A_{x,y}(\tau, \nu) = A^*_{y,x}(\tau, \nu)$$

we have immediately

$$Q(n, m, n, n) = Q(n, n, m, n).$$

Hence, it remains to show that (with arbitrary ρ)

$$Q(n, m, n, n) = 2\rho\delta_{mn}.$$

To see that this is indeed true, we have to analyze $Q(n, m, n, n)$ in more detail, one has (see (3.27))

$$Q(n, m, n, n) = \int_\tau \int_\nu C(\tau, \nu) A_{h_n,h_m}(\tau, \nu) A_{h_n}(\tau, \nu) d\tau d\nu.$$

The function $C(\tau, \nu)A_{h_n}(\tau, \nu)$ obviously satisfies the elliptical symmetry condition (3.21) such according to (3.22) we can expand this function into auto–ambiguity functions:

$$C(\tau, \nu)A_{h_n}(\tau, \nu) = \sum_{k=1}^{\infty} b_k A_{h_k}(\tau, \nu).$$

Now, by using the well–known fact [371]

$$\langle A_{x,y}, A_{g,h} \rangle = \langle x, g \rangle \langle h, y \rangle.$$

we can finish our proof:

$$Q(n, m, n, n) = \sum_{k=1}^{\infty} b_k \langle A_{h_k}, A_{h_n,h_m} \rangle = \sum_{k=1}^{\infty} b_k \langle h_k, h_n \rangle \langle h_m, h_k \rangle = b_n \delta_{mn}.$$

∎

3.3.7 Discussion

In the previous chapter we have seen that, for underspread processes, the prominent window–independent definitions of a time–varying power spectrum lead to essentially equivalent results (see (2.76)). Concluding this section on time–varying spectrum estimation we can state that *for underspread processes the matched window spectrogram provides a satisfactory estimate of this essentially unique spectrum.* Here, the qualification *satisfactory* is to be understood in a twofold sense:

- First, considering the spectrogram as an estimator of the generalized Wigner distribution, the bias using the matched window is comparatively small. As a quantitative example consider a strongly underspread process with rectangular support of the expected ambiguity function. Here, the maximum bias can be bounded as (see (3.11))

$$\left| B^{(\gamma_{opt})}_{SPEC} \right| \leq \|\mathbf{R}_x\| \left\| \chi_x - A^{(\alpha)}_\gamma \right\| < \|\mathbf{R}_x\| \frac{2}{\pi} (\tau_0\nu_0)^{\frac{3}{2}} \ll 1$$

  approximately valid for $\tau_0\nu_0 \ll 1$ with the approximations as discussed in Section 3.3.5.

Proof: We assume a matched, time–bandwidth efficient window with temporal moment T_γ^2 and spectral moment F_γ^2 such that

$$T_\gamma F_\gamma \approx 1 \qquad \text{and} \qquad \frac{T_\gamma}{F_\gamma} = \frac{\tau_0}{\nu_0}$$

$$\Longrightarrow \qquad T_\gamma^2 = \frac{\tau_0}{\nu_0} \qquad \text{and} \qquad F_\gamma^2 = \frac{\nu_0}{\tau_0}$$

and hence using the Taylor expansion of the ambiguity function:

$$\left\| \chi_x - A_\gamma^{(\alpha)} \right\|^2 \approx \int_{-\tau_0}^{\tau_0} \int_{-\nu_0}^{\nu_0} \left| \frac{1}{2\pi} \left(F^2 \tau^2 + T^2 \nu^2 \right) \right|^2 d\tau d\nu < \left(\frac{2}{\pi} \right)^2 (\tau_0 \nu_0)^3.$$

∎

- Second, presupposing knowledge of a spreading constraint, the physical spectrum (expected spectogram) is a complete second order statistic of an underspread process (which is not true for a general nonstationary process). One can switch to a unitary representation of $\mathbf{R}_x$ via a minimum–norm deconvolution in the (t, f)–domain that is equivalent to a stable division in the (τ, ν)–domain:

$$EA_x^{(\alpha)}(\tau, \nu) = \begin{cases} \dfrac{\left(\mathcal{F}_{t \to \nu} \mathcal{F}_{f \to \tau}^{-1} ESPEC_x \right)(\tau, \nu)}{A_\gamma^{(\alpha)*}(\tau, \nu)}, & |A_\gamma(\tau, \nu)|^2 \geq \epsilon, \\ 0, & |A_\gamma(\tau, \nu)|^2 < \epsilon. \end{cases} \tag{3.28}$$

Note that based on an even only coarsely matched window one has

$$|A_\gamma(\tau, \nu)|^2 \geq \epsilon \qquad \text{wherever} \qquad EA_x(\tau, \nu) > 0,$$

such that the minimum–norm deconvolution (3.28) works without error. This fact is illustrated in Figure 3.1.

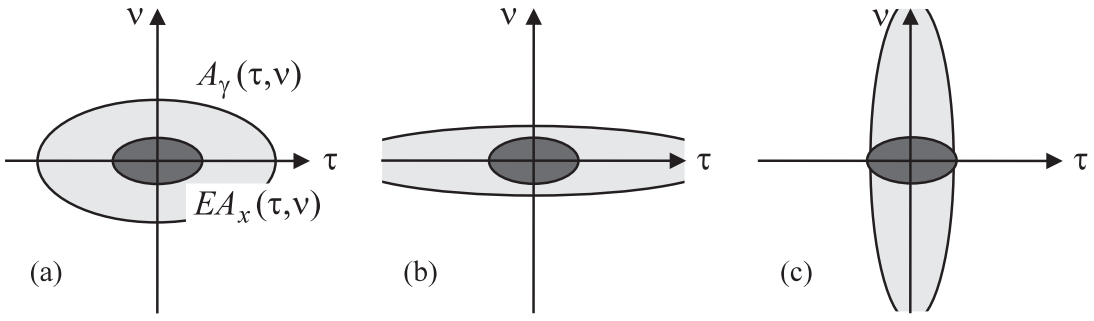

Figure 3.1: Schematic illustration of $EA_x(\tau, \nu)$ and $A_\gamma(\tau, \nu)$ for matched/mismatched situations: (a) matched γ, (b) "long" window, (c) "short" window.

As a last remark we remind the reader that this discussion was devoted to a mere consideration of the *bias* of spectrogram based spectrum estimation. For strongly underspread processes, variance reduction in the spectrum estimation leads to multi–window spectrograms as discussed in Section 3.5.

3.4 Optimal Gabor Expansion

The short–time Fourier transform is a highly redundant signal representation and with our main objective — obtaining uncorrelated coefficients — time–frequency discretization comes up naturally.

The Gabor coefficients can be defined by sampling the STFT on a rectangular grid. Hence, we expect that the statistically optimum STFT window is a good Gabor analysis window too; but in general it is certainly not the optimum window. Moreover, in the discrete case the analysis window $\gamma(t)$ is not the only free parameter, rather we have to select the sampling periods T, F and the sampling offsets τ_p, ν_p. The consideration of the sampling offsets will be a crucial point in our derivation of the optimum Gabor expansion. We define the Gabor coefficients as:

$$G_x^{(\gamma)}(l,m) \stackrel{\text{def}}{=} \int_s x(s)\gamma^*(s - lT + \tau_p)e^{-j2\pi(mF-\nu_p)s}ds, \tag{3.29}$$

with

$$\|\gamma\| = 1, \qquad l,m \in \mathbb{Z}, \qquad 0 \le \tau_p < T, \qquad 0 \le \nu_p < F.$$

The reconstruction is given by

$$x(t) = \sum_l \sum_m G_x^{(\gamma)}(l,m)g^{(lT,mF)}(t),$$

where $g(t)$ is the so–called Gabor *synthesis* window. In this chapter, the focus is on the design of the analysis window $\gamma(t)$. The problem of finding the synthesis window to a given analysis window is equivalent to the design problem of the analysis window for a given synthesis window. This problem has attracted considerable interest in recent years [366, 108, 37, 181, 72, 114], we shall return to this point in the following chapter, Section 4.5.2.

3.4.1 Statistics of the Gabor Expansion

For a zero–mean process $x(t)$ the Gabor coefficients form a two–dimensional discrete random process with zero mean

$$\mathrm{E}\left\{G_x^{(\gamma)}(l,m)\right\} = 0 \qquad \text{for} \qquad \mathrm{E}\{x(t)\} = 0.$$

The correlation of the Gabor coefficients is a four–dimensional discrete function:

$$R_G^{(x,\gamma)}(l,m,l',m') \stackrel{\text{def}}{=} \mathrm{E}\left\{G_x^{(\gamma)}(l,m)\, G_x^{(\gamma)*}(l',m')\right\}.$$

Just as in the STFT case we express the Gabor coefficient correlation in terms of the ambiguity function of the window and the expected ambiguity function (EAF) of the process:

$$\begin{aligned} R_G^{(x,\gamma)}(l,m,l',m') = \int_\tau \int_\nu & EA_x(\tau,\nu)A_\gamma^*\left(\tau - (l-l')T, \nu - (m-m')F\right) \\ & \cdot e^{j2\pi[(\nu-(m-m')F)(lT-\tau_p)-\tau(m'F-\nu_p)]}d\tau\, d\nu. \end{aligned} \tag{3.30}$$

Global Coefficient Correlation. Similar to (3.2) we measure the global correlation of the Gabor expansion via an off–diagonal norm

$$M_{Gabor,S}^{(x,\gamma)}(T,F,\tau_p,\nu_p) = \sum_l \sum_m \sum_{l'} \sum_{m'} \left|R_G^{(x,\gamma)}(l,m,l',m')\right|^2 S(l-l',m-m'),$$

where $S(l,m)$ is a radially nondecreasing weight function with

$$S(0,0) = 0, \qquad 0 \le S(l,m) \le 1.$$

Based on (3.30) we start with a complicated expression for the global Gabor correlation:

$$M_{Gabor,S}^{(x,\gamma)}(T,F,\tau_p,\nu_p) = \sum_l \sum_m \sum_{l_1} \sum_{m_1} S(l_1,m_1) \int_{\tau_1} \int_{\nu_1} \int_{\tau_2} \int_{\nu_2} EA_x(\tau_1,\nu_1) EA_x^*(\tau_2,\nu_2)$$

$$A_\gamma\left(\tau_1 - l_1 T, \nu_1 - m_1 F\right) A_\gamma^*\left(\tau_2 - l_1 T, \nu_1 - m_1 F\right)$$

$$e^{j2\pi[(\nu_1-\nu_2)(lT-\tau_p)-(\tau_1-\tau_2)((m+m_1)F-\nu_p)]} d\tau_1\, d\nu_1\, d\tau_2\, d\nu_2.$$

Using twice Poisson's sum formula:

$$\sum_l e^{j2\pi(\nu_1-\nu_2)lT} = \frac{1}{T} \sum_{m_2} \delta\left(\nu_1 - \nu_2 - \frac{m_2}{T}\right),$$

and

$$\sum_m e^{j2\pi(\tau_1-\tau_2)mF} = \frac{1}{F} \sum_{l_2} \delta\left(\tau_1 - \tau_2 - \frac{l_2}{F}\right),$$

leads to

$$\begin{aligned} M_{Gabor,S}^{(x,\gamma)}(T,F,\tau_p,\nu_p) = \tfrac{1}{TF} \sum_{l_1} \sum_{m_1} \sum_{l_2} \sum_{m_2} S(l_1,m_1) \int_{\tau} \int_{\nu} EA_x(\tau,\nu) EA_x^*(\tau - \tfrac{l_2}{F}, \nu - \tfrac{m_2}{T}) \\ A_\gamma\left(\tau - l_1 T, \nu - m_1 F\right) A_\gamma^*\left(\tau - l_1 T - \tfrac{l_2}{F}, \nu - m_1 F - \tfrac{m_2}{T}\right) e^{-j2\pi(\frac{m_2}{T}\tau_p + \frac{l_2}{F}\nu_p)} d\tau d\nu. \end{aligned} \tag{3.31}$$

The general statistical optimization of the Gabor expansion in the form of

$$(\gamma, T, F, \tau_p, \nu_p)_{opt} = \arg\min_{(\gamma,T,F,\tau_p,\nu_p)} M_{Gabor,S}^{(x,\gamma)}(T,F,\tau_p,\nu_p) \qquad \text{subject to} \qquad \|\gamma\| = 1, \tag{3.32}$$

is obviously too complicated for a numerical solution. However, the dependence on the sampling offsets is undesirable as in the on–line implementation of the Gabor expansion as a multirate filter bank it is not realistic to match the sampling offsets to the statistics of long natural signals. We thus perform a usual "randomization" step in order to match the window w.r.t. the whole ensemble of possible Gabor grid positions.

3.4.2 Random Offset Averaging

We assume that both the time and frequency sampling offset are uniformly distributed over their admissible intervals, i.e.,

$$p_{\tau_p}(\xi) = \frac{1}{T}\chi_{[0,T]}(\xi) \qquad \text{and} \qquad p_{\nu_p}(\xi) = \frac{1}{F}\chi_{[0,F]}(\xi), \tag{3.33}$$

where $p_{\tau_p}(\xi)$ and $p_{\nu_p}(\xi)$ are the probability density functions of the random variables τ_p and ν_p. Averaging the global coefficient correlation w.r.t. the uniformly distributed sampling offsets yields a considerable simplification:

$$\begin{aligned} \overline{M}_{Gabor,S}^{(x,\gamma)}(T,F) &\stackrel{\text{def}}{=} \mathrm{E}_{(\tau_p,\nu_p)}\left\{M_{Gabor,S}^{(x,\gamma)}(T,F,\tau_p,\nu_p)\right\} \\ &= \frac{1}{TF} \int_{\xi_1=0}^{T} \int_{\xi_2=0}^{F} M_{Gabor,S}^{(x,\gamma)}(T,F,\xi_1,\xi_2) d\xi_1\, d\xi_2 \\ &= \frac{1}{TF} \sum_l \sum_m S(l,m) \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |EA_x(\tau,\nu)|^2 |A_\gamma(\tau - lT, \nu - mF)|^2 d\tau\, d\nu, \end{aligned} \tag{3.34}$$

where we have used the fact that

$$\frac{1}{T}\int_{\xi_1=0}^{T} e^{j2\pi(\frac{m'}{T}\xi_1)}d\xi_1 = \delta_{m'0} \qquad \text{and} \qquad \frac{1}{F}\int_{\xi_2=0}^{F} e^{j2\pi(\frac{l'}{F}\xi_2)}d\xi_2 = \delta_{l'0}.$$

Based on the offset–averaged version of the global Gabor correlation the remaining optimization problem can be formulated as:

$$(\gamma,T,F)_{opt} = \arg\min_{(\gamma,T,F)} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |A_\gamma(\tau,\nu)|^2 \left|\widehat{E}A_x^{(T,F)}(\tau,\nu)\right|^2 d\tau d\nu, \qquad \text{subject to} \qquad \|\gamma\| = 1, \tag{3.35}$$

where we have introduced a weighted–periodized version of the magnitude–squared EAF:

$$|\widehat{E}A_x^{(T,F)}(\tau,\nu)|^2 \stackrel{\text{def}}{=} \sum_l \sum_m S(l,m)|EA_x(\tau - lT, \nu - mF)|^2.$$

Although the optimization problem is now considerably simplified compared to its original setup (3.32) it is still too complicated for numerical optimization with reasonable expense. As a further step towards a tractable window optimization problem we restrict ourselves to underspread processes and we suggest the heuristic choice of a *matched grid* as discussed in the following section. That is, *first* we determine T and F and *then* we optimize $\gamma(t)$.

3.4.3 Matched Sampling Grid

The choice of an appropriate sampling grid (as determined by T and F) is basically a two–parameter problem. In our context, it is best divided into the determination of TF and T/F.

The product TF has to be considered as an *a priori* side constraint because within the statistical optimization framework the product should be as large as possible (a higher sampling density always increases the global correlation), while from a signal reconstruction point of view it should be small with the upper bound $TF = 1$ [173]. In order to fulfill the reconstruction condition we henceforth assume $TF = 1 - \epsilon$ with small ϵ.

In the statistical optimization of the Gabor expansion it is the grid ratio T/F which has to be matched to the process. Given an underspread process with τ_0 and ν_0 we propose a *matched sampling grid ratio* in the form

$$\boxed{\frac{T}{F} = \frac{\tau_0}{\nu_0},} \tag{3.36}$$

but we have no mathematical proof for its optimality. Rather, we provide various heuristic lines of argumentation that all lead to this specific grid ratio (3.36).

Via Frame Theory. In our first approach, we start with the matched Gaussian window for STFT–based representation of an underspread process, as discussed in Section 3.3.6 one has:

$$\gamma_{opt,STFT}(t) = \sqrt[4]{2F_\gamma/T_\gamma}\, e^{-\pi(F_\gamma/T_\gamma)t^2}, \qquad \text{with} \qquad \frac{T_\gamma}{F_\gamma} = \frac{\tau_0}{\nu_0}, \tag{3.37}$$

where T_γ^2,F_γ^2 are the spectral and temporal moments of the Gaussian window and τ_0, ν_0 are the temporal and spectral correlation width of the given underspread process. Now, one way of deriving the matched grid ratio is to match the Gabor grid to the matched Gaussian window (3.37) via frame theory.

A brief review of frame theory can be found in [68, 37]. We here just mention the basic definition of a Weyl–Heisenberg frame.

Definition. *The set* $\gamma^{(lT,mF)}(t) = \gamma(t - lT)e^{j2\pi mFt}$ *establishes a Weyl–Heisenberg frame whenever the following inequality holds with* $0 < A \leq B < \infty$,

$$A\|x\|^2 \leq \sum_l \sum_m \left|\langle x, \gamma^{(lT,mF)}\rangle\right|^2 \leq B\|x\|^2. \tag{3.38}$$

A and B are called lower and upper frame bound.

A measure for the goodness of the frame is $B/A - 1$, which should be as small as possible. Clearly, for a given window the frame bounds depend on the choice of the grid ratio. It has been shown both numerically [68] and theoretically [347, 348] that, given the Gaussian window (3.37), the optimum grid ratio is given by:

$$\frac{T}{F} = \frac{T_\gamma}{F_\gamma},$$

which in turn corroborates our definition (3.36).

Via Symbolic Calculus. Another motivation for our matched grid ratio definition comes directly from the stochastic sampling principle as discussed in Section 2.5.2. There, we have shown that for an underspread process any of the previously discussed classical definitions of a time–varying spectrum (generalized Wigner–Ville, physical and Priestley's evolutionary spectrum) leads to a 2D lowpass function which is uniquely characterized by its samples on a rectangular grid, with the grid ratio given by the matching rule (3.36). This clearly shows that *the matched grid ratio is a natural, window–independent quantity of an underspread process.*

Via Residual Global Correlation. A more intuitive, but coarser way of reasoning starts with the assumption of a rectangular support of the EAF with critical spread $\tau_0\nu_0 = 1/4$. Recall that the (offset averaged) global coefficient correlation is given by the inner product of the "aliased" version of the EAF and the magnitude squared ambiguity function of the Gabor coefficients:

$$\overline{M}^{(x,\gamma)}_{Gabor,S}(T,F) = \left\langle \left|\widehat{E}A_x^{(T,F)}\right|^2, |A_\gamma|^2 \right\rangle.$$

The ambiguity function of a useful window is well concentrated about the origin. With this fact in mind it is obvious that one has to prefer a grid ratio such that $\left|\widehat{E}A_x^{(T,F)}\right|^2$ leaves enough space for the ambiguity function of the window which in turn causes less global correlation. The situation is illustrated in Figure 3.2. The area of the largest rectangle that fits into the centered gap between the nearest terms of $\left|\widehat{E}A_x^{(T,F)}\right|^2$ is given by

$$A = 4(F - \nu_0)(T - \tau_0).$$

It is simple to show that maximizing the area of the gap rectangle leads again to the matched grid ratio:

$$\left(\frac{T}{F}\right)_{opt} = \arg\min_{\left(\frac{T}{F}\right)} A = \frac{\tau_0}{\nu_0}.$$

3.4.4 Alternative Derivation for Underspread Processes

It is interesting to note that based upon the three practically reasonable assumptions of :

1. an underspread process,

$$\tau_0\nu_0 \leq 1/4,$$

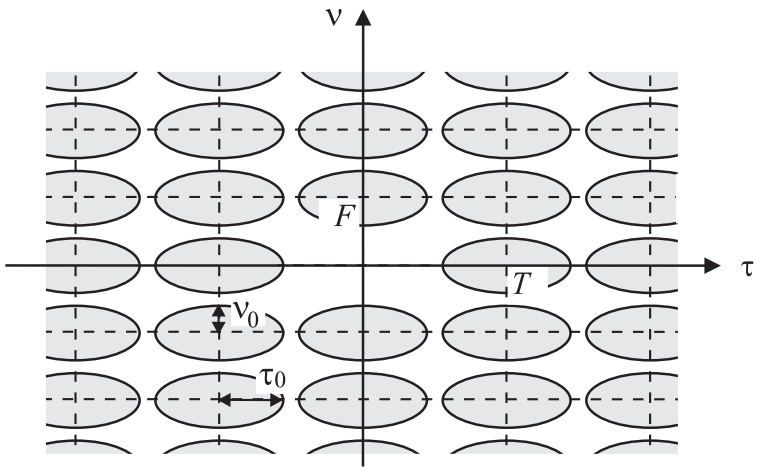

Figure 3.2: *Schematic illustration of the modified magnitude–squared EAF,* $|\widehat{E}A_x^{(T,F)}(\tau,\nu)|^2$ *of a process with compactly supported EAF.*

2. overcritical sampling, i.e.,

$$TF \leq 1,$$

3. the choice of the matched sampling grid

$$\frac{T}{F} = \frac{\tau_0}{\nu_0}.$$

one can avoid the offsets averaging argument (3.33)—(3.35) by recognizing that[4]

$$EA_x(\tau,\nu)EA_x\left(\tau - \frac{l'}{F}, \nu - \frac{m'}{T}\right) = |EA_x(\tau,\nu)|^2 \delta_{l'0}\delta_{m'0}. \tag{3.39}$$

This is exactly the "anti–aliasing condition" for sampling the generalized Wigner–Ville spectrum on a grid with constants T and F. Using (3.39) in Equation (3.31) leads exactly to the simplified expression for the global correlation (3.34) which has been derived in Section 3.4.2 via the offset averaging argument.

3.4.5 The Optimum Window

With the choice of a matched sampling grid according to (3.36) the statistical optimization of the Gabor expansion amounts to the following window optimization problem:

$$\gamma_{opt,G} = \arg\min_{\gamma} \left\langle \left|\widehat{E}A_x^{(T,F)}\right|^2, |A_\gamma|^2 \right\rangle, \qquad \text{subject to} \qquad \|\gamma\| = 1.$$

However, since

$$\left\langle 1, |A_\gamma(\tau,\nu)|^2 \right\rangle = 1 \qquad \text{for} \qquad \|\gamma\| = 1,$$

[4] This is the already published original approach of the author [206].

one has (we slightly abuse the inner product notation)

$$\arg\min_{\gamma} \left\langle \left| \widehat{E} A_x^{(T,F)}(\tau,\nu) \right|^2, |A_\gamma(\tau,\nu)|^2 \right\rangle = \arg\max_{\gamma} \left\langle 1 - \left| \widehat{E} A_x^{(T,F)}(\tau,\nu) \right|^2, |A_\gamma(\tau,\nu)|^2 \right\rangle$$
$$\text{subject to} \qquad \|\gamma\| = 1$$

which finally allows to formulate the Gabor analysis window optimization problem as follows:

$$\gamma_{opt,G} = \arg\min_{\gamma} \left\langle 1 - \left| \widehat{E} A_x^{(T,F)}(\tau,\nu) \right|^2, |A_\gamma(\tau,\nu)|^2 \right\rangle, \qquad \text{subject to} \qquad \|\gamma\|^2 = 1. \tag{3.40}$$

In this form the optimization problem is structurally equivalent to the statistical window optimization for the STFT Eq. (3.6).

3.4.6 Approximate Solutions

One may expect that the optimum STFT window is also a good Gabor analysis window. We show this fact by approximate reasoning. For the simplicity of the following discussion we assume the sharp off–diagonal weight function $S(l,m) = 1 - \delta_{l0}\delta_{m0}$. Note that for approximately critical sampling $TF = 1 - \epsilon$ and underspread processes, it is the nearest neighbor correlation ($R_g^{(x,\gamma)}(l,m,l',m')$ with $|l-l'||m-m'| \leq 1$) that will be dominant in the off–diagonal norm such that a more general weight function would not make considerable difference.

Matched Oversampling. The existence of simple, approximate results for the optimum Gabor analysis window is restricted to underspread processes with rectangular constraint, i.e.,

$$EA_x(\tau,\nu) \approx \chi_x(\tau,\nu) = \chi_{[-\tau_0,\tau_0]}(\tau)\chi_{[-\nu_0,\nu_0]}(\nu) \qquad \text{and} \qquad \tau_0\nu_0 < \frac{1}{4}. \tag{3.41}$$

Under the assumption of "matched oversampling" in the form of

$$\begin{aligned} T &= 2\tau_0, & F &= 2\nu_0, \\ &\Updownarrow & & \\ \frac{T}{F} &= \frac{\tau_0}{\nu_0}, & TF &= 4\tau_0\nu_0 < 1. \end{aligned} \tag{3.42}$$

one has simply (see Fig. 3.2):

$$1 - \sum_l \sum_m \chi_x(\tau - lT, \nu - mF) = \chi_x(\tau,\nu)$$

such that the optimum STFT window is equivalent to the optimum Gabor window:

$$\begin{aligned} \gamma_{opt,STFT} &\approx \arg\max_{\gamma} \left\langle \chi_x, |A_\gamma|^2 \right\rangle \\ &= \left\langle 1 - \sum_l \sum_m \chi_x(\tau - lT, \nu - mF), |A_\gamma(\tau,\nu|^2 \right\rangle \\ &= \gamma_{opt,G} \end{aligned}$$

We emphasize that this case is theoretically remarkable but it does not give any reason to prefer the specific product $TF = 4\tau_0\nu_0$. In fact, it is only for the critical spread $\tau_0\nu_0 = 1/4$ that the grid of (3.42)

is identical with the critical grid of the stochastic sampling principle. From our original motivation of obtaining uncorrelated coefficients the product TF should be kept as large as possible.

Note that once we omit the assumption of "matched oversampling" we still have the fact that due to (3.41) one has

$$\left|\widehat{E}A_x^{(T,F)}(\tau,\nu)\right|^2 \approx \sum_l \sum_m \chi_x(\tau - lT, \nu - mF)(1 - \delta_{l0}\delta_{m0}),$$

such that the largest rectangle that fits into the centered gap of $\left|\widehat{E}A_x^{(T,F)}(\tau,\nu)\right|^2$ has a length ratio

$$\frac{T_{gap}}{F_{gap}} = \frac{\tau_0}{\nu_0}$$

(see Figure 3.2, p. 54). The optimum window should have a maximum part of volume of $|A_\gamma(\tau,\nu)|^2$ in such a centered rectangle. This allows to conclude that just as in the case of the STFT a coarse matching of the window in the sense of keeping a constant shape and optimizing the scale (as discussed in Section 3.3.5) may again be defined by

$$\frac{T_\gamma}{F_\gamma} = \frac{\tau_0}{\nu_0},$$

where T_γ^2 and F_γ^2 are the temporal and spectral moments of the window.

Approximate Matching Via Symbolic Calculus. Similar to the STFT window optimization, an abstract mathematical view gives an approximate low–cost window design criterion.

The stochastic sampling principle (as discussed in Section 2.5.2) leads to a discrete Weyl–Heisenberg expansion of a correlation operator

$$\mathbf{R}_x = \sum_l \sum_m EW_x^{(\alpha)}(lT, mF)\mathbf{P}^{(lT,mF)}(\alpha), \tag{3.43}$$

where the prototype operator is uniquely defined for the choice of the critical sampling grid $T = \frac{1}{2\nu_0}, F = \frac{1}{2\tau_0}$. On the other hand, the Karhunen–Loeve based operator decomposition is given by

$$\mathbf{R}_x = \sum_k \lambda_k \mathbf{P}_{u_k}, \tag{3.44}$$

where $\mathbf{P}_{u_k}$ is the rank–one projection operator onto the KL basis signals $u_k(t)$.

Now, an ideally decorrelating Gabor expansion must be equivalent to the KL expansion such that one has a structured version of (3.44) in the form of:

$$\mathbf{R}_x \stackrel{?}{=} \sum_l \sum_m \mathrm{E}\left\{\left|G_x^{(\gamma)}(l,m)\right|^2\right\}\mathbf{P}_\gamma^{(lT,mF)}, \tag{3.45}$$

where the expectation of the magnitude squared Gabor coefficients would play the role of the KL eigenvalues.

The structural equivalence of (3.43) and (3.45) suggests a heuristic matching by selecting that γ whose rank–one projection is closest to the prototype operator. This is just the already mentioned α–dependent window optimization criterion $\gamma_{opt,2}$ defined as:

$$\gamma_{opt,2} = \arg\min_\gamma \|\mathbf{P}(\alpha) - \mathbf{P}_\gamma\|^2 = \arg\max_\gamma \left\langle \chi_x, A_\gamma^{(\alpha)} \right\rangle \qquad \text{subject to} \qquad \|\gamma\|^2 = 1.$$

3.4.7 Discussion

The adaptation of the Gabor expansion to the second order statistic of a nonstationary process shows quite similar aspects as discussed for the corresponding STFT problem.

We have a twofold ambiguity, both the window and the process appear in the form of their magnitude–squared ambiguity function. The ambiguity of the optimum window is only a minor practical problem while the ambiguity w.r.t. the process is one of the fundamental practical advantages of the Gabor expansion compared to the KL transform.

To see this, recall that incomplete a priori knowledge prevents the application of the Karhunen–Loeve transform. But the magnitude–squared EAF of the process (as it appears in the Gabor window optimization criterion) is in fact an incomplete a priori knowledge of a correlation kernel. Moreover, even if we assume the mere knowledge of a *quasistationarity condition* as can be made mathematically precise by an indicator function of the essential support of $EA_x(\tau,\nu)$, it is natural to replace $|EA_x(\tau,\nu)|$ by the indicator function $\chi_x(\tau,\nu)$ in the Gabor window optimization problem. Thus, one can obtain an approximate diagonalization of any underspread correlation operator via the matched Gabor transform in a case where lack of a priori knowledge completely prevents the use of the Karhunen–Loeve transform.

It is beyond the scope of this work to give an experimental verification of the notion of a slowly time–varying process as a model for natural signals. We emphasize however that *existing source coding concepts whether they are linear prediction based or transform coding rely on (or are explained by) the concept of slow nonstationarity.* The present work may thus be seen as a contribution towards a better understanding of the existing concepts. Clearly, the window optimization theory shows that a precise description of slow nonstationarity requires knowledge of both the temporal and spectral fluctuations present in the process' second order statistic.

3.5 Extension to Matched Multi–Window Expansions

In the statistical optimization of the Gabor expansion we have encountered a *deterministic and stochastic* (Weyl–Heisenberg) sampling principle:

- The signal can be uniquely reconstructed from the samples of the STFT $STFT_x^{(\gamma)}(t,f)$ on a rectangular grid with
$$TF < 1.$$
See [125, 68, 37] for mathematical details about this problem.

- The second order statistic of a process with restricted spreading is characterized by the samples of the generalized Wigner–Ville spectrum $EW_x^{(\alpha)}(t,f)$ on a rectangular lattice with $T < \frac{1}{2\nu_0}$ and $F < \frac{1}{2\tau_0}$, thus for an underspread process one has
$$TF \geq 1.$$

The obvious incompatibility of these sampling principles is not only theoretically unsatisfactory but also a problem in potential applications of time–varying spectra. In a *source coding* application, the estimate of the time–varying spectrum should control the bit allocation for the quantization of the Gabor coefficients. Another example is nonstationary minimum mean–squared error (Wiener) filtering, where it would be natural to use one and the same time–frequency grid for both the estimation and the filtering. The most convenient way out of this fundamental incompatibility is the employment of multi–window expansions as originally introduced by Thompson (for stationary spectrum estimation) [345].

Given a family of orthonormal *analysis* windows $\{\gamma_k(t)\}_{k=1,\ldots,N}$ we define the multi–window Gabor coefficients by:
$$G_x(l,m,k) = \left\langle x, \gamma_k^{(lT,mF)} \right\rangle \qquad k = 1,...,N; l,m \in \mathbb{Z},$$

where we assume that the expansion set $\gamma_k^{(lT,mF)}$ is complete which seems to be possible for

$$TF \leq N.$$

The signal expansion is then based on a corresponding family of *synthesis* windows $\{g_k(t)\}_{k=1,\ldots,N}$

$$x(t) = \sum_l \sum_m \sum_{k=1}^{N} G_x(l,m,k) g_k^{(lT,mF)}(t).$$

As to the choice of the analysis window family, it is obvious that a reformulation of the window matching theory via an off–diagonal norm minimization is too complicated to lead to practically useful results. However, by considering the multi–window Gabor coefficients as point estimates of the generalized Wigner–Ville spectrum we can obtain a tractable window matching criterion just as a natural generalization of the STFT/Gabor window matching via "symbolic calculus" (see Sections 3.3.4 and 3.4.6).

One can write

$$\widehat{EW}_x^{(\alpha)}(lT,mF) = \sum_{k=1}^{N} \left| G_x^{\{\gamma_k\}}(l,m,k) \right|^2 = \left\| \mathbf{P}_N^{(lT,mF)} x \right\|^2 = \left\langle \mathbf{P}_N^{(lT,mF)} x, x \right\rangle,$$

where $\mathbf{P}_N$ is the rank–N projection onto the span of the windows $\{\gamma_k\}$:

$$\left(\mathbf{P}_N\right)(t,s) \stackrel{\text{def}}{=} \sum_{k=1}^{N} \gamma_k(t) \gamma_k^*(s).$$

For the statistical derivation we refer the reader to the Appendix D, we here discuss the matching via symbolic calculus for underspread processes. Due to the stochastic sampling principle we have a discrete Weyl–Heisenberg expansion in the form of:

$$\mathbf{R}_x = \sum_l \sum_m EW_x^{(\alpha)}(lT,mF) \mathbf{P}^{(lT,mF)}(\alpha).$$

It can be shown (Appendix D) that $\mathbf{P}(\alpha)$ is essentially the characteristic operator of the MVUB estimator of $EW_x^{(\alpha)}(t,f)$:

$$\widehat{EW}_{x,MVUB}^{(\alpha)}(t,f) = \left\langle \mathbf{P}^{(t,f)}(\alpha) x, x \right\rangle.$$

For the sake of simplicity we narrow the discussion to the case $\alpha = 0$, here the prototype operator is self–adjoint and admits the usual eigendecomposition:

$$\mathbf{P}(0) = \sum_{k=1}^{\infty} \tilde{\lambda}_k \mathbf{P}_{\tilde{u}_k},$$

where $\mathbf{P}_{\tilde{u}_k}$ is the rank–one projection onto the eigensignals $\tilde{u}_k(t)$ and $\tilde{\lambda}_k$ are the eigenvalues. We furthermore assume that the eigenvalues are arranged in order of non–increasing magnitude. It should be emphasized that these eigenexpansion is in general completely different from the process' KL expansion (eigenexpansion of the correlation operator). The optimum window set is given by the first N eigensignals of the prototype operator, which is equivalent to saying that the optimum rank–N projection estimator of the Wigner–Ville spectrum is based on the following prototype operator:

$$\mathbf{P}_{N,opt} = \sum_{k=1}^{N} \mathbf{P}_{\tilde{u}_k}$$

The rank is selected such that it exceeds the area of the stochastic sampling grid

$$N \geq \frac{1}{4\tau_0\nu_0} = TF.$$

We see at least two potential applications for such multi–window methods:

- *Multiwindow Subband Coding of Underspread Processes.* The basic structure is similar to the usual single window transform coder consisting of an analysis step, followed by quantization and synthesis.

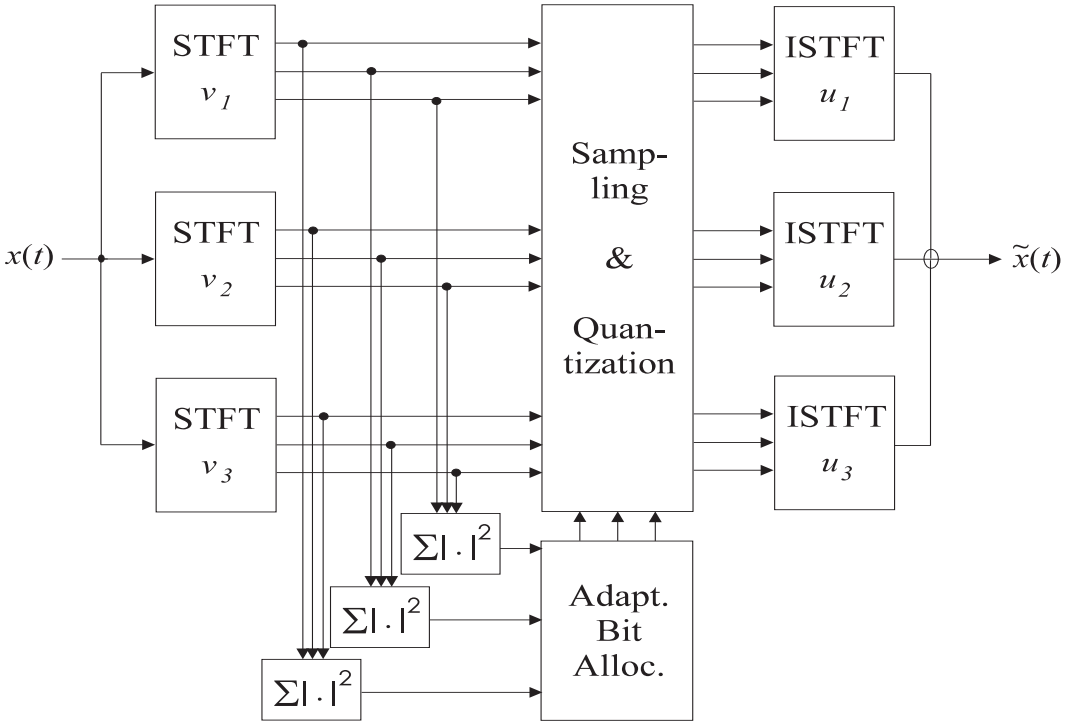

Figure 3.3: Multiwindow realization of subband coder.

The main difference lies in the bit allocation which is performed according to the estimate of the generalized Wigner–Ville spectrum thus on a considerably coarser grid compared to the usual Gabor based transform coder. This means reduced overhead information in adaptive bit allocation schemes.

We emphasize that the classical bit allocation scheme (averaging the coefficient energy over a number of successive blocks in order to obtain an estimate of the power in each channel) may be reinterpreted in terms of the proposed multi–window scheme by regarding the rectangular windows of succesive blocks as a trivial (time–disjoint) family of orthogonal windows.

- *Minimum Mean–Squared Error Filtering.* The nonstationary MMSE filter for jointly underspread signal and noise leads to an underspread LTV system [199]. As discussed in the next chapter one can use one and the same set of windows for the *estimation* and the *filtering*. The filter itself is a multirate–multiwindow STFT filter as has been originally proposed by the author [205] for the realization of underspread filters.

3.6 Numerical Experiments

In the following numerical experiments we consider the STFT window matching to a given deterministic signal. For comparison we have computed spectrograms using windows with deliberately long/short duration and a smoothed Wigner distribution based on a radially Gaussian smoothing kernel.

Figure 3.4 considers a combination of three chirp signals with equal chirp rate, but different amplitudes, contaminated by white Gaussian noise with resulting signal–to–noise ratio 0dB (the signal and the noise have equal energy). In this specific case, the optimum window is a chirp which is no traditional window function but matched to the signal at hand. Clearly, this is a best–case example

for the window matching concept, because the signal consists of time–frequency shifted versions of a well–concentrated "prototype" signal. However, the experiment demonstrates the noise robustness of the window matching concept. Note that the window matching is based on the noisy observation without further a priori knowledge. The noise robustness can be explained by the fact, that the noise causes merely a (large peak) in the origin of the signal's ambiguity function (see Fig. 3.4(j)) which does not influence the window optimization procedure (any normalized window is equal to one in the origin of the (τ, ν)–plane).

For the second experiment we have designed a signal that consists of components with quite different, specific time–frequency structure: a frequency modulated signal with parabolic frequency law, a "long" windowed sinusoid and a "short" Gaussian component centered between the other components. In contrast to the above considered three–chirp example, the various components of this signal would certainly require different window durations at different time–frequency localizations. Yet, the global window optimization procedure leads to a satisfactory compromise between the conflicting goals of representing the "short" Gaussian, the "long" sinusoidal component and the chirpy parts of the frequency–modulated signal.

Last, we consider a voiced part of a speech signal. Speech signals are one of the best studied "natural" signals. They feature quite complicated temporal and spectral structures. We consider a voiced part of the speech signal which is almost periodic with a small number of dominant spectral components. The period is called *pitch* and the spectral components are called *formants*. One of the key issues in speech analysis is to keep track of both structures which led to the combined consideration of "short–window" (wideband) and "long–window"(narrowband) spectrograms. The analyzed signal includes two pitch periods. Figure 3.6 illustrates the zooming–in procedure that led to the signal at hand. Using the "long" window blurs the pitch structure, see Fig. 3.6(c), while using the short window the formant structure gets lost: Fig. 3.6(d). The optimum window spectrogram Fig. 3.6(e) shows a good compromise between spectral and temporal resolution, it allows to recognize both the temporal and spectral fine structure of the speech signal.

3.7 Summary

The short–time Fourier transform is a classical tool for the analysis and processing of signals that can not be satisfactorily modeled as realizations of wide–sense stationary processes. In this chapter, we have studied the second–order statistic of the STFT of a nonstationary process. Minimization of a global STFT correlation measure leads to a window optimization criterion that is approximately equivalent to window optimization criteria that aim at minimum local/global bias of the spectrogram as an estimator of the generalized Wigner–Ville spectrum for underspread processes. The STFT window optimization can be most compactly formulated by maximizing an inner product of the smoothed expected ambiguity function (EAF) and the ambiguity function of the window:

$$\gamma_{opt} = \arg\max_{\gamma} \left\langle |EA_x|^2 * *K, |A_\gamma|^2 \right\rangle, \qquad \|\gamma\| = 1,$$

where the smoothing kernel $K(\tau, \nu)$ corresponds to the off—diagonal penalty function.

The Gabor window optimization can be formulated analogously:

$$\gamma_{opt,G} = \arg\min_{\gamma} \left\langle 1 - \left| \widehat{E}A_x^{(T,F)}(\tau, \nu) \right|^2, |A_\gamma(\tau, \nu)|^2 \right\rangle, \qquad \|\gamma\| = 1,$$

where the smoothing of the STFT criterion is replaced by a weighted periodization ($S(l, m)$ corresponds to the off–diagonal penalty function, and T, F are the sampling periods of the Gabor grid):

$$|\widehat{E}A_x^{(T,F)}(\tau, \nu)|^2 \stackrel{\text{def}}{=} \sum_l \sum_m S(l, m)|EA_x(\tau - lT, \nu - mF)|^2.$$

In the matching of a Gabor expansion we have also considered the freedom in the choice of a (rectangular) sampling grid.

For processes with finite temporal and spectral correlation,

$$EA_x(\tau,\nu) = EA_x(\tau,\nu)\,\chi_{[-\tau_0,\tau_0]}(\tau)\chi_{[-\nu_0,\nu_0]}(\nu)$$

we suggest the use of matched sampling grid whose ratio is given by

$$\frac{T}{F} = \frac{\tau_0}{\nu_0},$$

where τ_0, ν_0 denote the temporal/spectral correlation width (support constraints of the EAF) of the process.

We have specialized the above window matching criteria to the one–parameter optimization for specific (rectangular and elliptic) symmetry of the EAF, this leads to the intuitively appealing window matching rule:

$$\frac{T_\gamma}{F_\gamma} = \frac{\tau_0}{\nu_0},$$

where T_γ^2 and F_γ^2 are the temporal and spectral moments of the window. Moreover, for underspread processes ($\tau_0\nu_0 \ll 1$) we have seen that the matched window spectrogram achieves a low bias estimate of the, essentially definition–independent, classical time–varying power spectra. Specifically, we have shown that

$$\left\| EW_x^{(\alpha)} - ESPEC_x^{(\gamma)} \right\|_\infty^2 = \mathcal{O}\left((\tau_0\nu_0)^3\right).$$

Deterministic signals are always overspread processes. However, the numerical experiments of this chapter indicate that the matched window spectrogram gives a highly satisfactory, adaptive time–frequency representation, with no loss of structural information compared to a Wigner distribution with supervised smoothing.

We have pointed out the natural extension of the window matching theory to multiwindow methods for the representation of underspread processes. Loosely speaking, the main motivation for multiwindow methods can be seen in the fact that the time–varying power spectrum of an underspread process is so smooth that its reproducing kernel has essential rank greater than one. Hence, optimum estimation and representation requires more than one matched window. This idea is consistent with but more general than classical multiwindow methods for stationary spectral estimation.

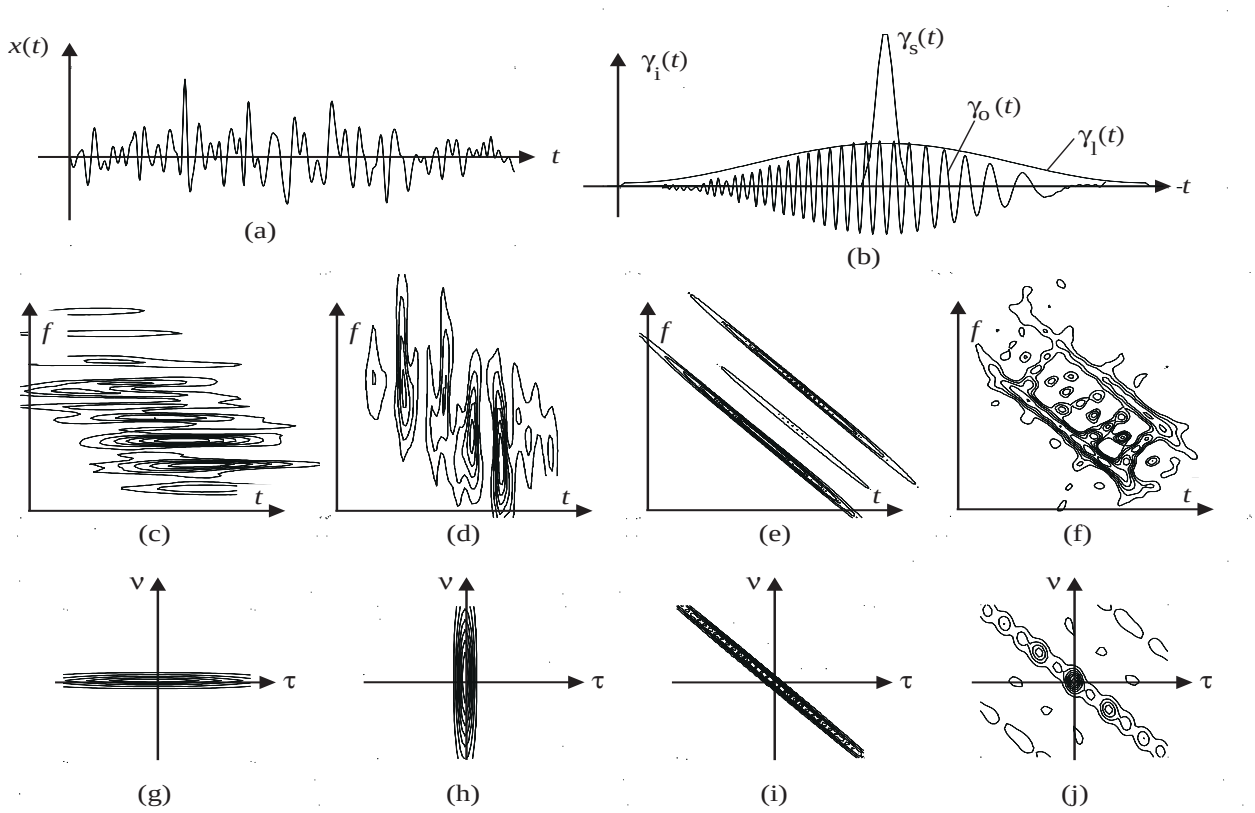

Figure 3.4: TF representations of a "three chirps + noise" signal. (a) signal $x(t)$, (b) windows: $\gamma_l(t) \cdots$ "long", $\gamma_s(t) \cdots$ "short", $\gamma_o(t) \cdots$ optimum, (c) spectrogram $SPEC_x(t,f)$ using $\gamma_l(t)$, (d) $SPEC_x(t,f)$ using $\gamma_s(t)$, (e) $SPEC_x(t,f)$ using $\gamma_o(t)$, (f) smoothed WD. (g–j) Magnitude of ambiguity functions: (g) $A_{\gamma_l}(\tau,\nu)$, (h) $A_{\gamma_s}(\tau,\nu)$, (i) $A_{\gamma_o}(\tau,\nu)$ (j) $A_x(\tau,\nu)$.

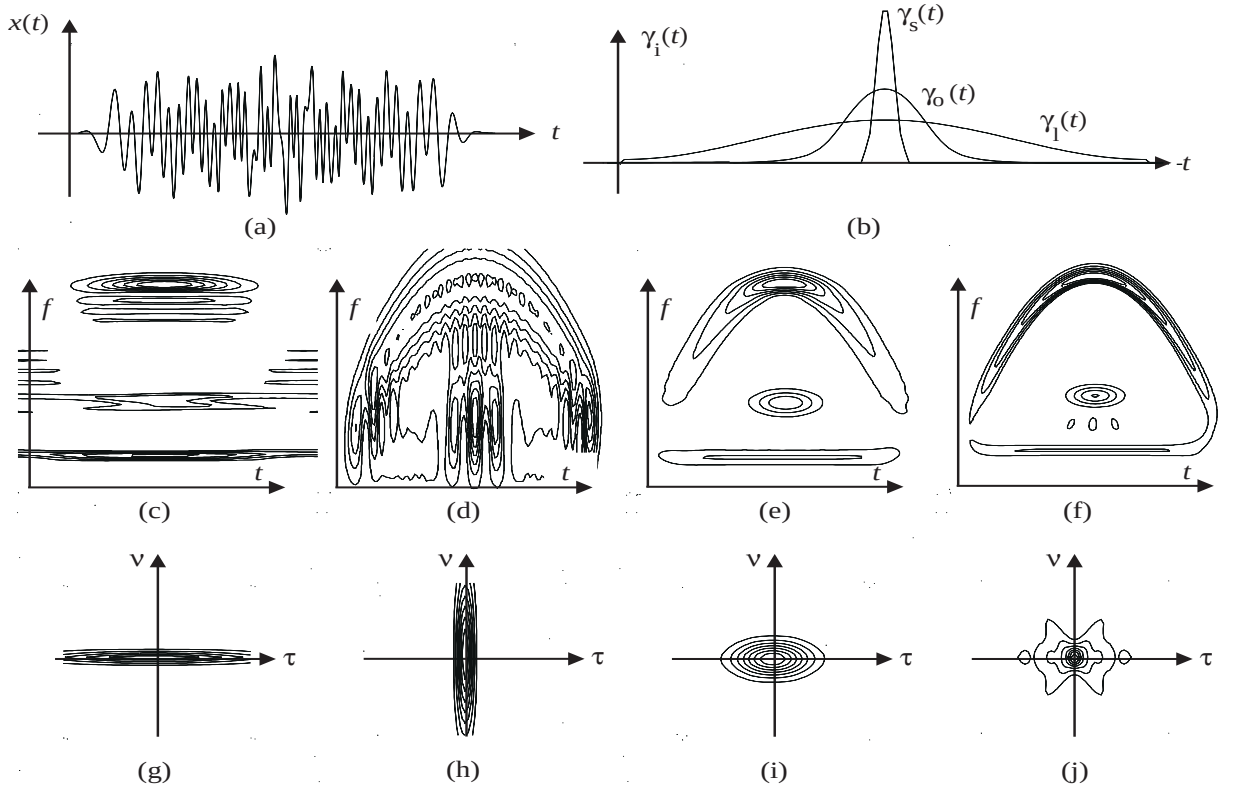

Figure 3.5: TF representations of "complicated" synthetic signal.

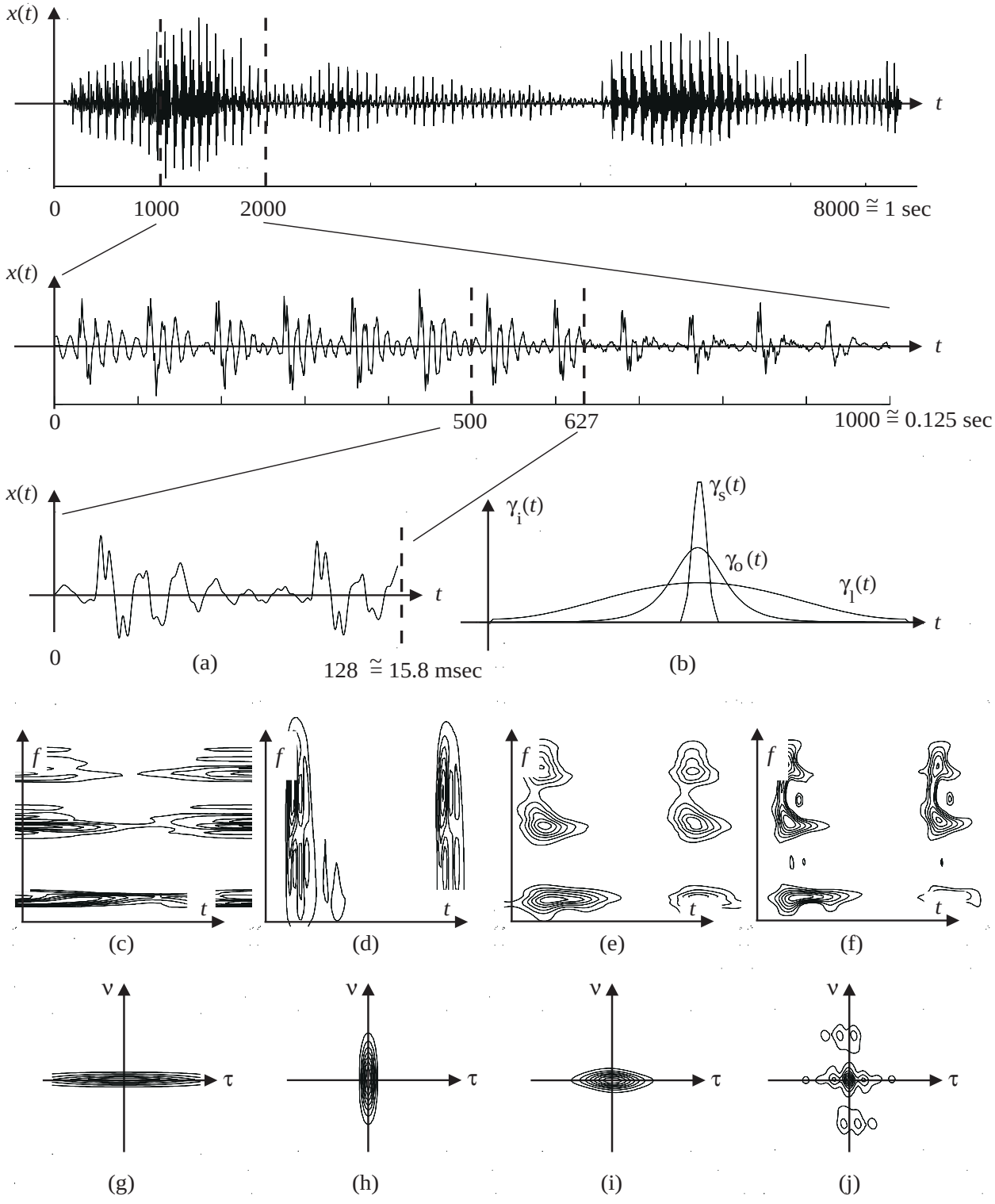

Figure 3.6: Optimum window spectrogram versus other representations for two pitch periods of a voiced speech signal. (a) The signal $x(t)$, (b) the considered window functions: $\gamma_l(t) \cdots$ "long" window, $\gamma_s(t) \cdots$ "short" window, $\gamma_o(t) \cdots$ optimum window, (c) spectrogram of $x(t)$ using $\gamma_l(t)$, (d) spectrogram of $x(t)$ using $\gamma_s(t)$, (e) spectrogram of $x(t)$ using $\gamma_o(t)$, (f) smoothed Wigner distribution of $x(t)$. (g–j) Ambiguity functions of: (g) the "long" window $\gamma_l(t)$, (h) the "short" window $\gamma_s(t)$, (i) the optimum window $\gamma_o(t)$, (j) the speech signal $x(t)$.

Chapter 4

Matched Weyl–Heisenberg Expansions of LTV Systems

Many aspects of the second order theory of nonstationary processes are mathematically equivalent to linear system theory. This chapter gives a compact review of the joint time–frequency representations of linear time–varying (LTV) systems: The generalized spreading function, Zadeh's time–varying transfer function and the Weyl symbol. Particular emphasis is put on underspread LTV systems and their analysis and realization via the short–time Fourier transform. The window matching theory as discussed in the previous chapter is shown to be directly applicable.

We review the practically important concept of wide–sense stationary uncorrelated scattering (WSSUS) and show its invariance w.r.t. the parametrization of the generalized Weyl correspondence. Finally, we discuss a criterion for a transmission pulse with minimum expected distortion for the communication over a Gaussian WSSUS channel. The mathematical structure turns out to be equivalent to the statistical window matching theory for the STFT/Gabor transform.

4.1 Nonstationary Processes and Linear Systems

In the previous chapter we have matched the STFT (continuous Weyl–Heisenberg signal expansion) and the Gabor expansion (discrete Weyl–Heisenberg expansion) to the second order statistics of a nonstationary process. Reduced to a mathematical viewpoint, the key idea was the optimum approximate diagonalization of a Hilbert–Schmidt operator via Weyl–Heisenberg structured bases. This suggests to apply the obtained optimization criteria to linear time–varying (LTV) systems that correspond to a linear operator $\mathbf{H}$ by just replacing the correlation operator $\mathbf{R}_x$ by $\mathbf{H}$. There are various different motivations for this extension of the window matching theory:

- Any nonstationary process $x(t)$ with given second order statistics $\mathbf{R}_x$ can be seen as the output of a linear time–varying *innovations* system $\mathbf{H}$ excited by white noise

$$x(t) = (\mathbf{H}n)(t) \qquad \text{with} \qquad \mathbf{R}_n = \mathbf{I}.$$

  The (non–unique) innovations system can be characterized by

$$\mathbf{R}_x = \mathbf{H}\mathbf{H}^*. \tag{4.1}$$

  The problem of defining an appropriate (time–frequency parametrized) *time–varying spectrum* of $x(t)$ as a linear representation of $\mathbf{R}_x$ may accordingly be traced back to the representation of $\mathbf{H}$ via a *time–varying transfer function*. Mathematically, these concepts are exactly equivalent. The only difference lies in the fact that $\mathbf{R}_x$ is an always positive operator while in practice $\mathbf{H}$ is typically not even normal.

- The short–time Fourier transform is a classical tool for both the representation *and filtering* of nonstationary processes. It may be expected that one and the same window is optimum for both the representation and the filtering of a process.

- In the digital communication over a linear channel it is natural to prefer transmission pulses that come close to eigensignals of the channel. Whenever we consider time and frequency shifted versions of a prototype transmission pulse (as e.g. in the practically predominant time/frequency division multiplex systems) this leads back to the matching of a discrete Weyl–Heisenberg set in the sense of optimum diagonalization of the channel.

Linear time–varying (LTV) systems appear in widespread applications. The choice of an appropriate mathematical representation of the system depends on the application. The vast literature on this topic can be coarsely split up into the predominant model based *state space approach* and model independent approaches via *kernel representations*. The kernel representation of deterministic and stochastic LTV systems as developed by Zadeh, Kailath, Bello and others [385, 386, 190, 191, 29, 197] provides the basis for our developments. A compact review of the fundamental theory can be found in [334].

In a system theoretical context it is usual to call the kernel $h(t,s)$ *impulse response*[1] (the formal explanation for this terminology is discussed in Appendix A); the input–output–relation based on $h(t,s)$ is given by

$$(\mathbf{H}x)(t) = \int_s h(t,s)x(s)ds. \tag{4.3}$$

The impulse response of an LTV system corresponds to the natural time–domain representation of the input signal and the output signal. The Fourier dual of the impulse response is the frequency domain kernel which is known as *bifrequency function* $B_H(f,s)$ [385] and is defined by

$$(\mathcal{F}\mathbf{H}x)(f) = \int_s B_H(f,s)X(s)ds. \tag{4.4}$$

The bifrequency function is related to the impulse response via a symplectic Fourier transform

$$B_H(f,s) = \int_t \int_r h(t,r)e^{-j2\pi(ft-sr)}dt\,dr. \tag{4.5}$$

From an operator representation point of view, the impulse response and the bifrequency function correspond to the temporal and spectral correlation function of a nonstationary process. For the specific classes of linear systems considered in this work the bifrequency function does not play a specifically helpful role. Rather, time–frequency parametrized representations are of key interest.

One can introduce time–frequency parametrization formally as kernels of integral operators essentially analog to (4.3) and (4.4) (by representing either the output or the input signal in a different domain). However, this point of view neither leads to theoretically useful system classification nor

[1] We remind the reader that the term impulse response is not uniquely defined in the LTV system literature. In many publications one considers $h_2(t,\tau) = h(t,t-\tau)$ as the impulse response of an LTV system. The input–output relation then reads

$$(\mathbf{H}x)(t) = \int_\tau h_2(t,\tau)x(t-\tau)d\tau, \tag{4.2}$$

which is advantageous in so far as it provides a split–up in absolute time t and time delay τ (causal or memoryless systems can be characterized merely by the τ–support of $h_2(t,\tau)$ which leads to more complicated conditions on $h(t,s)$). We use definition (4.3) as its corresponds to the usual mathematical definition for the kernel of a linear integral operator; the switch to $h_2(t,\tau)$ is trivial.

does it suggest practically efficient methods for system analysis and synthesis. Hence, we introduce these classical representations in a more physically intuitive way. Moreover, the physical motivation leads to an operator decomposition point of view in the sense of representing an LTV system by a weighted, parallel combination of simple, well–understood building blocks. This approach is particularly matched to the analysis and synthesis of underspread systems, to which the original results of this chapter are devoted.

4.2 Spreading Functions

4.2.1 Time–Frequency Shifting of Signals

A general linear time–varying (LTV) operator potentially causes both time shifts (time delay), i.e.,

$$\left(\mathbf{T}_\tau x\right)(t) = x(t-\tau),$$

and frequency shifts (modulation), i.e.,

$$\left(\mathbf{M}_\nu x\right)(t) = x(t)e^{j2\pi\nu t}.$$

Recall that these unitary operators correspond to the amplitude preserving physical effects of a time–varying multipath propagation: *time–delay* and *Doppler–shift*[2].

Whenever a linear operator does not cause any time shift then it is *memoryless* (linear, frequency–invariant, briefly LFI) and whenever there are no potential frequency shifts then the operator is linear, time–invariant (LTI). The only linear operator that does neither cause time nor frequency shifts is the identity.

The combination of a time shift τ and a frequency shift ν leads to a unitary time–frequency shift operator denoted by $\mathbf{S}^{(\tau,\nu)}$ with the input–output relation given by

$$\left(\mathbf{S}^{(\tau,\nu)}x\right)(t) \stackrel{\text{def}}{=} \left(\mathbf{M}_\nu\mathbf{T}_\tau x\right)(t) = x(t-\tau)e^{j2\pi\nu t}. \tag{4.6}$$

In the context of this work time and frequency play an equal role which becomes more obvious by studying the input–output mapping via the spectrogram:

$$SPEC^{(\gamma)}_{S^{(\tau,\nu)}x}(t,f) = SPEC^{(\gamma)}_x(t-\tau, f-\nu), \tag{4.7}$$

that is, the time–frequency shift corresponds to a translation of the signal's energetic time–frequency representation.

4.2.2 Asymmetrical Spreading Function

The amount of potential time and frequency shifts caused by an LTV system $\mathbf{H}$ can be specified by the *asymmetrical spreading function* (also known as delay–Doppler–spread function) [334]

$$S_H^{(1/2)}(\tau,\nu) \stackrel{\text{def}}{=} \int_t h(t,t-\tau)e^{-j2\pi\nu t}dt\,.$$

(The meaning of the superscript (1/2) will be clarified later.) The asymmetrical spreading function establishes an infinitesimal decomposition of $\mathbf{H}$ into time–frequency shift operators as defined in (4.6):

$$\begin{aligned}\left(\mathbf{H}x\right)(t) &= \int_\tau\int_\nu S_H^{(1/2)}(\tau,\nu)\left(\mathbf{S}^{(\tau,\nu)}x\right)(t)\;d\tau\,d\nu \\ &= \int_\tau\int_\nu S_H^{(1/2)}(\tau,\nu)x(t-\tau)e^{j2\pi\nu t}d\tau\,d\nu,\end{aligned} \tag{4.8}$$

[2] A general LTV system may cause frequency shift that is not due to a Doppler effect. In this work we refer to the time–varying multipath propagation as an illustrative example with important practical background.

which is well–defined whenever one has (i) $x(t) \in L_2(\mathbb{R})$ and (ii) $h(t,s) \in L_2(\mathbb{R}^2)$ (see [125, p.79,80])[3].

Hence, the spreading function admits a point–wise physical interpretation as the path loss and phase corresponding to a certain time–delay τ and Doppler–shift ν in a time–varying multipath environment [197].

4.2.3 Generalized Spreading Function

The definition of a time–frequency shift operator is not unique. Instead of (4.6) one can equally well put the frequency shift before the time shift which results in a different time–frequency shift operator that acts as

$$\left(\widetilde{\mathbf{S}}^{(\tau,\nu)}x\right)(t) \stackrel{\text{def}}{=} \left(\mathbf{T}_\tau \mathbf{M}_\nu x\right)(t) = x(t-\tau)e^{j2\pi\nu(t-\tau)}.$$

More general time–frequency shift operators with identical time shift and frequency shift can be obtained by arbitrarily splitting up and combining the time shift and the frequency shift. These time–frequency shift operators differ only w.r.t. a time–independent phase factor, and (4.7) holds for any version of a time–frequency shift operator.

Among this manifold of time–frequency shift operators there is one which is marked out by a certain symmetry[4] [125],

$$\left(\bar{\mathbf{S}}^{(\tau,\nu)}x\right)(t) = x(t-\tau)e^{j2\pi\nu t}e^{-j2\pi\frac{\nu\tau}{2}}. \tag{4.9}$$

The corresponding spreading function is

$$S_H^{(0)}(\tau,\nu) = \int_t h\left(t+\frac{\tau}{2}, t-\frac{\tau}{2}\right)e^{-j2\pi\nu t}dt, \tag{4.10}$$

and will be called *symmetrical spreading function*. Based on an α–parametrized time–frequency shift operator with kernel

$$\left(\mathbf{S}^{(\tau,\nu)}(\alpha)\right)(t,s) = \delta(t-s-\tau)e^{j2\pi\nu\{t+\tau(\alpha-1/2)\}} \tag{4.11}$$

the symmetrical and asymmetrical spreading function can be written in a unified way: the *generalized spreading function* is defined as

$$S_H^{(\alpha)}(\tau,\nu) \stackrel{\text{def}}{=} \int_t h\left(t+\left(\frac{1}{2}-\alpha\right)\tau, t-\left(\frac{1}{2}+\alpha\right)\tau\right)e^{-j2\pi\nu t}dt. \tag{4.12}$$

A lexical summary of the properties and relations of the generalized spreading function can be found in the Appendix B, we here mention only facts that are relevant for the specific context of this chapter.

System Decomposition. The generalized spreading function allows to characterize the action of a general LTV system as a superposition of time–frequency shifts (generalization of (4.8)):

$$\mathbf{H} = \int_\tau \int_\nu S_H^{(\alpha)}(\tau,\nu)\mathbf{S}^{(\tau,\nu)}(\alpha)\,d\tau\,d\nu, \tag{4.13}$$

where the kernel of $\mathbf{S}^{(\tau,\nu)}(\alpha)$ is defined in (4.11).

α–invariance of Magnitude. The magnitude of the generalized spreading function is α–invariant:

$$\left|S_H^{(\alpha_1)}(\tau,\nu)\right| = \left|S_H^{(\alpha_2)}(\tau,\nu)\right|,$$

[3] Specific systems with a non–square–integrable kernel can still be meaningfully characterized by the spreading function with the help of distributions, but then validity of (4.8) requires more severe mathematical restrictions on the input signal.

[4] This time–frequency shift operator has been originally defined by Weyl who was led by group theoretical considerations beyond the scope of this work. We mention this fact since later on one can observe the intimate relationship of the *symmetrical* spreading function to the Weyl correspondence.

this leads to a pleasing α–invariance of many important results.

Time–Frequency Correlation. The generalized spreading function of a correlation operator is equivalent to the time–frequency correlation function (expected ambiguity function) of the process:

$$S_{R_x}^{(\alpha)}(\tau,\nu) = EA_x^{(\alpha)}(\tau,\nu).$$

LTI System. An LTI system with kernel $h(t,s) = \bar{h}(t-s)$ does not cause any frequency shifts. Accordingly, the generalized spreading function is ideally concentrated on the time axis:

$$S_{H_{LTI}}^{(\alpha)}(\tau,\nu) = \bar{h}(\tau)\delta(\nu).$$

LFI System. In the dual case of an LFI system with kernel $h(t,s) = \delta(t-s)m(t)$ the generalized spreading function is ideally concentrated on the frequency axis

$$S_{H_{LFI}}^{(\alpha)}(\tau,\nu) = \delta(\tau)M(\nu),$$

where $M(\nu)$ is the Fourier transform of the multiplicator function $m(t)$.

Figure 4.1 shows the support of the generalized spreading function for various important classes of linear systems (underspread systems are the topic of the next section and time–frequency periodic systems will be discussed in Section 4.5.2).

4.2.4 Underspread Systems

Analog to the underspread/overspread classification of nonstationary processes as introduced in the previous chapter, we classify linear time–varying systems via a rectangular spreading constraint:

$$S_H^{(\alpha)}(\tau,\nu) = S_H^{(\alpha)}(\tau,\nu)\,\chi_{[-\tau_0,\tau_0]}(\tau)\,\chi_{[-\nu_0,\nu_0]}(\nu), \tag{4.14}$$

where we call

$$\sigma_H \stackrel{\text{def}}{=} 4\tau_0\nu_0$$

total spread. The total spread determines the conditioning of the least–squares identification problem and the conceptual value of frequency domain methods in general:

1. The value $\sigma_H = 1$ is a threshold for a variance–decreasing least–squares frequency–domain system identification from input/output observation. A thorough discussion of the system identification problem is the topic of Appendix E. There it is shown that the variance of the optimum transfer function estimator is proportional to σ_H. (For $\sigma_H = 1$ the variance of the coefficient estimate is equal to the SNR ratio of the output signal, i.e., loosely speaking, the system changes so fast that there is no possibility for averaging without introducing bias.)

2. With regard to the conceptual value of frequency domain methods in general, the threshold value $\sigma_H = 1$ itself establishes a merely formal mathematical side constraint for the quantitative results on underspread operators discussed in the next chapter. The derivation of the various approximation bounds often requires $\sigma_H < 1$ but in order to have small error terms the spread must be far below this threshold.

   Hence, we define *underspread* systems analogously to underspread processes by

   $$\sigma_H \ll 1.$$

   This asymptotic classification is largely consistent with the original, qualitative definition of underspread LTV systems in [197].

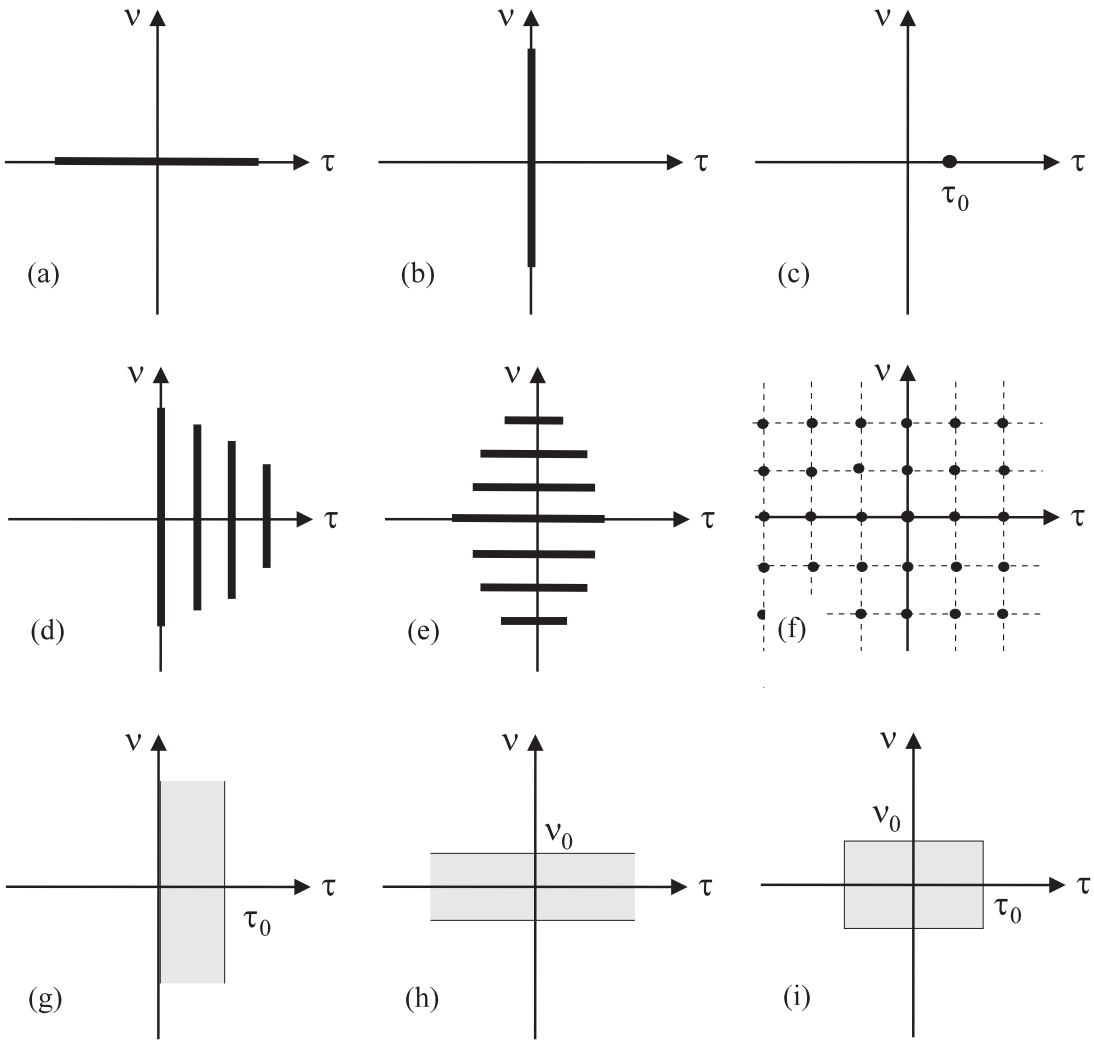

Figure 4.1: Support of the spreading function for important classes of linear systems: (a) LTI system, (b) LFI or memoryless system, (c) perfect reconstruction system (with time–delay τ_0), (d) time–varying tapped delay line, (e) periodically time–varying system, (f) time–frequency periodic system (e.g. Weyl–Heisenberg frame operator), (g) causal system with finite memory (limited time–delay), (h) quasistationary system, (i) underspread system.

Underspread LTV systems play an important role in this work, the practical relevance can be split up in two basic applications.

From a system *analysis* point of view we note that the time–varying communication channels as encountered in mobile communication are in good approximation underspread [269]. Here, the limited amount of Doppler shift is due to the limited velocity of the mobile radio station, while the maximum time–delay can be pragmatically motivated by the path loss of the (passive) propagation channel. Moreover, both from a practical and statistical point of view there is always an appropriate threshold value beyond which one has to disregard existing (τ, ν)–contributions (point scatterers) of the channel.

We emphasize, that many other practically important linear systems are slowly time–varying with limited memory (see [356, 240, 81] for control engineering examples). In fact, the key results of this work establish a theoretical framework which should contribute to a better understanding why frequency–domain concepts do work for underspread LTV systems although the frequency–domain theory is matched to time–invariant systems.

In the *synthesis* of filters for the enhancement of speech and audio signals it is natural to require limited potential time or frequency shift since such signals are characterized by a specific temporally localized frequency content, i.e., a noise reduction should take place in the form of *time–frequency–selective multiplication.*

From a purely deterministic signal separation point of view one is led to the concept of "time–frequency filtering" which can be performed most naturally by the modification of the signal's time–frequency representation followed by an appropriate signal synthesis [41]. Whenever the chosen signal representation is quadratic, the overall filter is highly nonlinear. It has been shown that such nonlinear analysis–modification–synthesis schemes perform remarkably bad as compared to linear time–frequency filters [205, 210]. Linear time–frequency filters can be defined by a time–varying transfer function (Zadeh's function or Weyl symbol, to be discussed in detail later on in this chapter). However, qualitatively speaking, such a time–frequency–parametrized transfer function must be smooth in order to keep the intuitive transfer function interpretation. Now, a canonical way to enforce smoothness of the system's transfer function is the underspread condition (4.14) (because the transfer function is the double Fourier transform of the spreading function).

Moreover, we shall see that any underspread system can be realized by multiplicative modification of the short–time Fourier transform (STFT) provided that one uses an appropriate (even only coarsely matched) window. This fact is in accordance with the experimental results presented in [210], which show that the classical STFT–based time–frequency filters can hardly be outperformed by much more complicated, off–line concepts for linear time–frequency filtering. It should be emphasized that the STFT is the conceptual basis for many modern multirate signal processing concepts such as modulated filter banks, generalized cosine transforms and lapped transforms [357]. From this point of view, the presented theory may help to improve existing concepts for signal enhancement.

Also from a theoretical point of view underspread systems are worthwhile to study, we here mention just two important qualitative properties (the mathematical details can be found in the following chapter, specifically Theorem 5.2).

Approximate Commutativity. We call two underspread systems $\mathbf{G}, \mathbf{H}$ with identical spreading constraints *jointly underspread.* Jointly underspread systems commute in an approximate sense:

$$\mathbf{HG} \approx \mathbf{GH}. \tag{4.15}$$

Approximate Normality. An underspread system is approximately normal:

$$\mathbf{HH}^* \approx \mathbf{H}^*\mathbf{H}. \tag{4.16}$$

Approximate normality means that we can write an underspread system with square–integrable impulse response (i.e., an underspread Hilbert–Schmidt operator) as

$$\mathbf{H} \approx \sum_k \lambda_k \mathbf{P}_{u_k}, \tag{4.17}$$

where $\mathbf{P}_{u_k}$ is the rank–one projection operator onto the eigensignal $u_k(t)$. These eigensignals establish an orthonormal basis for $L_2(\mathbb{R})$. The combined consideration of the commutativity and normality allows the conclusion that *jointly underspread operators possess a common (approximate) eigenbasis.* This thought leads back to the idea of approximate diagonalization as has been already elaborated for correlation operators in the previous chapter. A correlation operator is positive and self–adjoint thus a special case of a normal operator. This means that we can immediately apply our STFT/Gabor window matching theory to the approximate diagonalization of general underspread HS operators. We shall return to this point later on in this chapter.

4.3 Weyl–Heisenberg Symbols of LTV Systems

4.3.1 Transfer Function of LTI System

As discussed at the outset of the foregoing chapter, the Fourier transform is matched to linear time–invariant (LTI) operators. An LTI system acts as a convolution:

$$\left(\mathbf{H}_{LTI}x\right)(t) = \int_{\tau} \bar{h}(\tau)x(t-\tau)d\tau,$$

the complex sinusoids $e^{(\nu)}(t) = e^{j2\pi\nu t}$ are (generalized) eigensignals,

$$\left(\mathbf{H}_{LTI}e^{(\nu)}\right)(t) = H(\nu)e^{(\nu)}(t),$$

where the *transfer function* $H(f) = \mathcal{F}_{\tau\to f}\bar{h}(\tau)$ is the *continuous eigenvalue distribution*[5]. This transfer function establishes a powerful concept for the analysis and design of LTI systems. As an example we mention the cascade composition of LTI systems where we have the well–known correspondence:

$$\mathbf{H}_{12} = \mathbf{H}_1\mathbf{H}_2 \qquad \Longleftrightarrow \qquad H_{12}(f) = H_1(f)H_2(f). \tag{4.18}$$

We emphasize that LTI systems are in fact a limit case of underspread systems featuring exact commutativity and normality.

The transfer function of an LTI system can be formally written as,

$$H(f) = \left.\frac{\left(\mathbf{H}x\right)(t)}{x(t)}\right|_{x(t)\,=\,e^{j2\pi ft}}. \tag{4.19}$$

Of course, (4.19) is of limited practical relevance. In particular, the practical approximation of (4.19) establishes a very poor method for LTI system identification (transfer function measurement). Useful system identification methods are based on $L_2(\mathbb{R})$–signals and lead more naturally to a generalization towards LTV systems. We shall follow this idea in Section 4.4 where we study STFT–based system analysis.

4.3.2 Zadeh's Time–Varying Transfer Function

In a prominent work [385], Zadeh introduced a time–frequency–parametrized representation of an LTV system by a direct generalization of (4.19):

$$Z_H(t,f) \stackrel{\text{def}}{=} \left.\frac{\left(\mathbf{H}x\right)(t)}{x(t)}\right|_{x(t)\,=\,e^{j2\pi ft}}. \tag{4.20}$$

In electrical engineering this function is widely known as *time–varying transfer function* [386, 190, 191, 29, 54, 336, 269]. In the special case of an LTI system, definition (4.20) trivially assures consistency with the usual transfer function

$$Z_{H_{LTI}}(t,f) = H(f). \tag{4.21}$$

Zadeh's definition leads to an invertible map of the impulse response $h(t,s)$ onto a joint function of time and frequency[6]

$$Z_H(t,f) = \int_{\tau} h(t,t-\tau)e^{-j2\pi f\tau}d\tau. \tag{4.22}$$

[5] Mathematically precise: Approximate point spectrum, see the footnote on p. 1.

[6] Formally, Zadeh's time–varying transfer function may also be seen as kernel of an integral operator that maps the input spectrum onto the output signal (up to a complex factor):

$$\left(\mathbf{H}x\right)(t) = \int_{f} Z_H(t,f)X(f)e^{j2\pi ft}df.$$

However, the terminology of a time–varying "transfer function" is somewhat misleading as this function satisfies besides (4.20) practically none of the properties of the LTI system's transfer function. For example, one clearly has for general LTV systems:

$$Z_{H_{12}}(t,f) \stackrel{!}{\neq} Z_{H_1}(t,f) Z_{H_2}(t,f),$$

because LTV systems do not commute in general.

It is interesting to note that in mathematics (theory of pseudo–differential operators) this function was 15 years later independently introduced as *Kohn–Nirenberg symbol* of a linear operator [203, 125][7].

Recently, it has been recognized by the author [205] that Zadeh's time–varying transfer function can be related to quadratic time–frequency signal representations, namely the Rihaczek distribution [302]

$$R_x(t,f) \stackrel{\text{def}}{=} \int_\tau x(t) x^*(t-\tau) e^{-j2\pi f\tau} d\tau.$$

in the sense that

$$\langle \mathbf{H}x, x\rangle = \langle Z_H, R_x\rangle = \int_t \int_f Z_H(t,f) R_x^*(t,f)\, dt\, df. \tag{4.23}$$

This means, whenever $R_x(t,f)$ is concentrated about a certain point (t_0,f_0) and $x(t)$ is an eigensignal of the operator then the values of Zadeh's function in the region about (t_0,f_0) give a local information about the eigenvalue. This is in accordance with the intuitive idea of a time–frequency–parametrized eigenvalue distribution as a generalization of the frequency–parametrized transfer function of LTI systems.

However, to have a reliable local eigenvalue interpretation $R_x(t,f)$ should play the role of a positive 2D test function in (4.23). The quadratic form itself is always a reliable eigenvalue estimator in the sense that (i) it is real–valued for self–adjoint systems, (ii) for self–adjoint systems one has[8]

$$\lambda_{\min} \leq \langle \mathbf{H}x, x\rangle \leq \lambda_{\max} \qquad \text{for} \qquad \|x\| = 1.$$

Now, whenever the Rihaczek distribution of a (normalized) eigensignal shows oscillatory behavior (with large imaginary or large negative, real–valued components) (4.23) remains valid but $Z_H(t,f)$ can take on values which are far away from the true eigenvalue corresponding to this eigensignal[9]. For general LTV systems the consequence is that self–adjoint operators do not correspond to real–valued symbols and L_2–stable systems may well correspond to unbounded symbols and vice versa [125]. In particular, one can not study e.g. the (theoretical) L_2–invertibility of a general LTV system in terms of the minimum of Zadeh's function, because this minimum does not reflect the minimum eigenvalue of the system (which would be true in case of LTI systems).

It is curious that Zadeh's time–varying transfer function appears in all monographs on the mobile radio channel but to our best knowledge there is no single work about its limitations. We shall however show that in case of most practical time–varying communication channels one is in the lucky situation that Zadeh's function can be interpreted as an eigenvalue distribution with reasonable approximation, i.e., one can study such critical things as e.g. channel invertibility using $Z_H(t,f)$. Yet, one should be aware that the frequency domain characterization of general LTV systems is essentially mismatched and can only be justified by pragmatic reasoning.

Theoretically mismatched concepts always leave some mathematical freedom in the definitions. While this is well–known for the case of stochastic time–varying power spectra (as discussed in the

[7] Note, however, that the work of Kohn and Nirenberg goes far beyond Zadeh's results.

[8] These facts can be meaningfully generalized to normal non–selfadjoint systems but we concentrate on the self–adjoint case for simplicity of the discussion.

[9] Of course, one could equally well reverse this discussion in the sense of tracing back the inconsistency of the Rihaczek distribution (as an "energy" distribution) to the problems of Zadeh's function as an operator symbol. Rihaczek itself did not seem to be aware that his definition is Zadeh's function of the rank–one projection operator.

previous chapter), this essential freedom seems to be widely unknown in the LTV system literature[10]. Within the general context of this work it is appropriate to discuss a mathematically attractive alternative to Zadeh's function: The Weyl symbol.

4.3.3 Weyl Symbol

The Rihaczek distribution has received only moderate attention in the signal processing literature due to the inherent contradiction of a complex valued "energy" distribution. A more popular alternative is the *Wigner distribution* [51, 159] (see Appendix F)

$$W_x(t,f) \stackrel{\text{def}}{=} \int_\tau x\left(t+\frac{\tau}{2}\right) x^*\left(t-\frac{\tau}{2}\right) e^{-j2\pi f\tau} d\tau,$$

which is real–valued and leads to a much more satisfactory time–frequency representation of chirp signals compared to the Rihaczek distribution. Reformulation of the quadratic form (4.23) based on the signal's Wigner distribution leads to the *Weyl symbol* $L_H(t,f)$ [369, 277, 125, 327, 205, 293] of an LTV system:

$$\begin{aligned} \langle \mathbf{H}x, x\rangle &\stackrel{!}{=} \langle L_H, W_x\rangle = \int_t \int_f L_H(t,f) W_x(t,f)\, dt\, df \\ &\Downarrow \\ L_H(t,f) &\stackrel{\text{def}}{=} \int_\tau h\left(t+\frac{\tau}{2}, t-\frac{\tau}{2}\right) e^{-j2\pi f\tau} d\tau. \end{aligned}$$

The Wigner–Weyl framework originally appears in the area of phase space methods for quantum mechanics [369]. The Weyl symbol has also been applied in modern statistics [276] and, even more recently, the Weyl symbol was introduced in the signal processing literature [327, 205].

The Wigner distribution is somewhat more well–behaved as a time–frequency "energy distribution" compared to the Rihaczek distribution, a fact that carries over to the properties of the Weyl symbol in comparison to that of Zadeh's function. The Weyl symbol features two main advantages:

- The Weyl symbol of a self–adjoint operator is real–valued (consistent with the eigenvalue distribution interpretation) while Zadeh's function is generally complex valued. This is mathematically equivalent to the fact that one can compute the Weyl symbol of the adjoint operator by a simple conjugation:

  $$L_{H^*}(t,f) = L_H^*(t,f). \tag{4.24}$$

  This property was one of the key ingredients in our proof of the asymptotic equivalence of the classical time–varying power spectra for underspread processes (see Section 2.5.3).

- The celebrated performance of the Wigner distribution for chirp signals may be traced back to the mathematical issue of *unitary equivalence*: That is, shearing or rotating a Wigner distribution leads to a valid Wigner distribution of an (appropriately) modified signal, a fact that does not hold true for the Rihaczek distribution. For the associated system representations this carries over to the following facts: When the Weyl symbol of a system $\mathbf{H}$ gets subject to a symplectic

[10] To be historically correct we should mention what Bello [29] has called *frequency–dependent modulation function*:

$$L_H^{(-1/2)}(t,f) = \mathcal{F}_{\tau\to f}\, h(t+\tau, t)$$

(the superscript shall be explained in the following section). However, to the author's best knowledge this function never appeared elsewhere in the LTV system literature.

coordinate transform then one obtains a unitarily equivalent system $\widetilde{\mathbf{H}}$ (the singular/eigenvalues remain unchanged while the singular/eigensignals undergo a canonical unitary transform) [125]:

$$\begin{array}{rcll} L_{\widetilde{H}}(t,f) & = & L_H(at+bf, ct+df), & \text{with} \quad \left|\begin{array}{cc} a & b \\ c & d \end{array}\right| = 1. \\ & \Downarrow & & \\ \widetilde{\mathbf{H}} & = & \mathbf{UHU}^*, & \text{with} \quad \mathbf{UU}^* = \mathbf{U}^*\mathbf{U} = \mathbf{I}. \end{array}$$

This does not hold true for Zadeh's function. For example, rotating Zadeh's function leads to a modification of the system's singular values.

Note, that the eigenvalue distribution interpretation of a time–varying transfer function would indeed require the general type of unitary equivalence as satisfied by the Weyl symbol. Hence, loosely speaking, whenever an LTV system features "chirpy" eigensignals then one should prefer the Weyl symbol compared to Zadeh's function [205]. However, in this work we shall put the focus on underspread systems in the sense of (4.14) where Zadeh's function and the Weyl symbol are essentially equivalent.

We henceforth treat Zadeh's function and the Weyl symbol in parallel via the introduction of the generalized Weyl symbol [205].

4.3.4 Generalized Weyl Symbol

The *generalized Weyl symbol* (GWS) is a unitary representation of Hilbert–Schmidt (HS) operators[11] defined as follows:

$$L_H^{(\alpha)}(t,f) \stackrel{\text{def}}{=} \int_\tau h\left(t+\left(\frac{1}{2}-\alpha\right)\tau, t-\left(\frac{1}{2}+\alpha\right)\tau\right) e^{-j2\pi f\tau} d\tau,$$

it gives the Weyl symbol for $\alpha = 0$ [369, 125], Zadeh's time–varying transfer function (Kohn–Nirenberg symbol) for $\alpha = 1/2$ [385, 125], and Bello's "frequency–dependent modulation function" for $\alpha = -1/2$ [29]. In the choice of the α–parametrization we were led by the definition of the generalized Wigner distribution [176]

$$W_x^{(\alpha)}(t,f) \stackrel{\text{def}}{=} \int_\tau x\left(t+\left(\frac{1}{2}-\alpha\right)\tau\right) x^*\left(t-\left(\frac{1}{2}+\alpha\right)\tau\right) e^{-j2\pi f\tau} d\tau.$$

Using the generalized Wigner distribution we have a GWS–based time–frequency formulation of the quadratic form:

$$\langle \mathbf{H}x, x\rangle = \left\langle L_H^{(\alpha)}, W_x^{(\alpha)}\right\rangle. \tag{4.25}$$

We have already used the GWS in the foregoing chapter in our study of the generalized Wigner distribution of stochastic processes. One has the formal equivalence:

$$L_{R_x}^{(\alpha)}(t,f) = \mathrm{E}\left\{W_x^{(\alpha)}(t,f)\right\},$$

i.e., the GWS of a correlation operator is equivalent to the expected generalized Wigner distribution.

A lexical summary of the GWS properties can be found in the Appendix C, we here just mention only a some formulas relevant for the context of STFT–based system analysis and synthesis.

Interrelation with Spreading Function. The GWS is the (symplectic) Fourier transform of the generalized spreading function (as defined in (4.12))

$$L_H^{(\alpha)}(t,f) = \int_\tau \int_\nu S_H^{(\alpha)}(\tau,\nu) e^{j2\pi(\nu t-\tau f)} d\tau\, d\nu = \mathcal{F}^{-1}_{\nu\to t}\mathcal{F}_{\tau\to f}\left\{S_H^{(\alpha)}(\tau,\nu)\right\}. \tag{4.26}$$

[11] Hilbert–Schmidt operators correspond to LTV systems with square integrable kernel (impulse response.)

Hence, whenever the generalized spreading function is essentially concentrated about the origin, the GWS is a two–dimensional lowpass function.

LTI System. For a linear time–invariant system with convolution type kernel $h(t,s) = \bar{h}(t-s)$ the GWS gives the usual transfer function (Fourier transform of the operator's impulse response):

$$L_H^{(\alpha)}(t,f) = H(f) = \left(\mathcal{F}\bar{h}\right)(f).$$

LFI System. In the dual case of a linear, frequency–invariant system (e.g. a linear modulator) with distribution type impulse response $h(t,s) = \delta(t-s)m(t)$ the GWS is frequency independent:

$$L_H^{(\alpha)}(t,f) = m(t).$$

Normal HS Operator. The GWS of an LTV system which corresponds to a normal HS operator, i.e., $\mathbf{H}\mathbf{H}^* = \mathbf{H}^*\mathbf{H}$ is given by an eigenvalue–weighted sum of the generalized Wigner distribution of the eigensignals:

$$L_H^{(\alpha)}(t,f) = \sum_{k=1}^{\infty} \lambda_k W_{u_k}^{(\alpha)}(t,f). \tag{4.27}$$

This formula obviously shows how the intuitive interpretation of the GWS can be traced back to the interpretation of the generalized Wigner distribution. Without anticipation of later developments in this work, one can already expect troubles for those systems whose eigensignals show bad time–frequency localization, i.e., cover more than one "Heisenberg cell" (see the more detailed discussion in Section 2.3.2, p. 14).

Time–Frequency Shift–Covariance. In Section 2.3.1 we have defined time–frequency shifting of a system (linear operator) by[12]

$$\mathbf{H}^{(\tau,\nu)} = \mathbf{S}^{(\tau,\nu)}\mathbf{H}\mathbf{S}^{(\tau,\nu)*},$$

where $\mathbf{S}^{(\tau,\nu)}$ is an arbitrary version of a time–frequency shift operator as defined in (4.11). Such a time–frequency shift leaves the eigenvalues of $\mathbf{H}$ unchanged while the new eigensignals are just time–frequency shifted versions of the original ones, i.e., given

$$\left(\mathbf{H}x\right)(t) = \lambda_0 x(t)$$

it follows that

$$\left(\mathbf{H}^{(\tau,\nu)}x^{(\tau,\nu)}\right)(t) = \lambda_0 x^{(\tau,\nu)}(t),$$

with

$$x^{(\tau,\nu)}(t) = x(t-\tau)e^{j2\pi\nu t}.$$

Now, one of the fundamental properties of the GWS is its invariance w.r.t. a time–frequency shift of systems, or to put it the other way around: a translation of the GWS leads to a corresponding time–frequency shift of the eigensignals but leaves the eigenvalues of a normal HS operator unchanged:

$$L_{H^{(\tau,\nu)}}^{(\alpha)}(t,f) = L_H^{(\alpha)}(t-\tau, f-\nu). \tag{4.28}$$

This property is fundamental for the interpretation and theoretical manipulation of the GWS, specifically in the formulation of the stochastic WSSUS class. The time–frequency shift–covariance gives a

[12] Our notation is somewhat ambiguous, since the superscript (τ,ν) has a completely different meaning depending on the label of the operator. However, since we shall never need a time–frequency shift of the time–frequency shift operator this ambiguity does not cause further complications.

canonical example of unitary equivalent operators whose symbols are interrelated by a simple coordinate transform of the time–frequency plane (namely a translation). Along with the discussion of the Weyl symbol ($\alpha = 0$) we have already mentioned that other more general forms of unitary equivalence exist, which correspond to more general coordinate transforms of the symbol. However, the general GWS family satisfies only the shift–covariance (4.28). The following two properties are immediate consequences of this covariance.

Periodically Time–Varying Systems. A periodically time–varying linear system commutes with time–shifts that are multiples of its time–period T [62], i.e., given

$$x^{(lT)}(t) = x(t - lT), \qquad l \in \mathbb{Z},$$

one has

$$\left(\mathbf{H}x^{(lT)}\right)(t) = (\mathbf{H}x)(t + lT).$$

Using our notation, this definition can be compactly formulated as:

$$\mathbf{H}\mathbf{S}^{(lT,0)} = \mathbf{S}^{(lT,0)}\mathbf{H}.$$

This is equivalent to the following requirement for the kernel of $\mathbf{H}$:

$$h(t, s) = h(t + lT, s + lT).$$

The GWS of such an operator is periodic w.r.t. time:

$$L_H^{(\alpha)}(t, f) = L_H^{(\alpha)}(t + lT, f), \tag{4.29}$$

for $\alpha = 1/2$ this result is well–known [54].

Time–Frequency Periodic Systems. Time–frequency periodic systems arise in the context of the Gabor expansion as Weyl–Heisenberg frame operator and (near–perfect reconstruction) DFT filter banks. Such systems correspond to operators which satisfy a canonical commutation relation:

$$\mathbf{H}\mathbf{S}^{(lT,mF)} = \mathbf{S}^{(lT,mF)}\mathbf{H}, \qquad l, m \in \mathbb{Z},$$

where T is the time period and F is the frequency period of a time–frequency sampling grid. One can show that such operators establish an *algebra*[13].

The commutation relation is equivalent to the following kernel requirement:

$$h(t, s) = h(t + lT, s + lT)e^{j2\pi mF(t-s)}.$$

We call such systems *time–frequency periodic*, because the GWS is doubly periodic:

$$L_H^{(\alpha)}(t, f) = L_H^{(\alpha)}(t + lT, f + mF). \tag{4.30}$$

The generalized Weyl correspondence is matched to the algebra of time–frequency periodic operators in the sense of a Gelfand transform, because given two time–frequency periodic operators $\mathbf{H}$,$\mathbf{G}$ one has validity of the perfect symbol calculus for integer oversampling:

$$L_H^{(\alpha)}(t, f)L_G^{(\alpha)}(t, f) \quad = \quad L_{HG}^{(\alpha)}(t, f),$$

$$\text{for} \qquad |\alpha| = \frac{1}{2} \qquad \text{and} \qquad TF = \frac{1}{n}, \qquad n \in \mathbb{N}.$$

[13] Algebra implies in particular that this class of operators is closed w.r.t. composition or inversion. Time–frequency periodic operators are "big" in the sense of being non–Hilbert–Schmidt (thus non–compact), which means that they basically fall out of our usual domain of the Weyl correspondence. A rigorous theory without the intuitively appealing but slightly dangerous use of the generalized Weyl correspondence can be found in [181].

It is curious that in this specific respect the Weyl symbol ($\alpha = 0$) is inferior to the Kohn–Nirenberg symbol in the sense that the perfect symbol calculus holds only for $TF = \frac{1}{2n}$. (This means in particular that the Weyl symbol does not allow to study the specifically interesting case of critical sampling, $TF = 1$.)

The Weyl–Heisenberg frame operator is defined as ($\gamma(t)$ is the prototype):

$$\mathbf{M}_\gamma \stackrel{\text{def}}{=} \sum_m \sum_l \mathbf{S}^{(lT,mF)} \mathbf{P}_\gamma \mathbf{S}^{(lT,mF)*}.$$

where $\mathbf{P}_\gamma$ is the rank–one projection onto the prototype (Gabor analysis window):

$$(\mathbf{P}_\gamma)(t,s) = \gamma(t)\gamma^*(s).$$

Validity of the symbol calculus suggests that the GWS indeed establishes the spectral representation of the (non–compact) frame operator $\mathbf{M}_\gamma$. Indeed, for $TF = 1$ (critical sampling) the GWS with $|\alpha| = 1/2$ is essentially equivalent to the magnitude–squared Zak transform of the prototype γ:

$$L_{M_\gamma}^{(1/2)}(t,f) = L_{M_\gamma}^{(-1/2)}(t,f) = T\,|\mathcal{Z}_\gamma(t,f)|^2\,,$$

where the Zak transform is defined as [179, 37]

$$\mathcal{Z}_\gamma(t,f) = \sum_l \gamma(t+lT)e^{-j2\pi flT}.$$

For integer oversampling and multiwindowing one ends up with formal short–cut derivations of well–known results of Gabor theory [181, 393, 158] that support the fact that the GWS with $|\alpha| = 1/2$ establishes the generalized eigenvalue distribution of the (self–adjoint, non–compact) frame operator of integer–oversampled, generalized Weyl–Heisenberg frames. More details about this aspect of the generalized Weyl correspondence will appear elsewhere [113].

Continuous Weyl–Heisenberg Expansion. In the engineering applications of the GWS, namely as LTV system's transfer function or time–varying stochastic signal spectrum we tend to interpret the symbol as time–frequency–selective multiplier. Clearly, this interpretation is in conflict with Heisenberg uncertainty, nevertheless one can formally write any HS operator in the form of what we call a *continuous Weyl–Heisenberg expansion*:

$$\mathbf{H} = \int_t \int_f L_H^{(\alpha)}(t,f)\mathbf{P}^{(t,f)}(\alpha)\,dt\,df, \tag{4.31}$$

where $\mathbf{P}(\alpha)$ is an α–dependent infinitesimal prototype operator whose symbol is a 2D delta distribution:

$$L_{P(\alpha)}^{(\alpha)}(t,f) = \delta(t)\delta(f).$$

A more detailed discussion of $\mathbf{P}(\alpha)$ was the topic of Section 2.3.2. One may view $\mathbf{P}(\alpha)$ as "ideal time–frequency localizator". However, the severe α–dependence reflects how the different correspondence rules set up pragmatic definitions which, of course, can not overcome Heisenberg's uncertainty.

We can formally write the GWS in the form of an operator inner product:

$$L_H^{(\alpha)}(t,f) = \left\langle \mathbf{H}, \mathbf{P}^{(t,f)}(\alpha) \right\rangle.$$

Note that the generalized Wigner distribution can be written as a quadratic form based on just the same prototype operator:

$$W_x^{(\alpha)}(t,f) = \left\langle \mathbf{P}^{(t,f)}(\alpha)x, x \right\rangle,$$

studying the action of $\mathbf{P}(\alpha)$ is an easy way to explain the well–known, undesired behavior ("interference terms", [159]) of generalized Wigner distributions for natural signals.

The continuous Weyl–Heisenberg expansion (4.31) is the dual of the above discussed (τ,ν)–parametrized expansion as induced by the generalized spreading function (4.13) :

$$\mathbf{H} = \int_{\tau}\int_{\nu} S_H^{(\alpha)}(\tau,\nu)\mathbf{S}^{(\tau,\nu)}(\alpha)\,d\tau\,d\nu.$$

However, in contrast to the time–frequency shift operator $\mathbf{S}^{(\tau,\nu)}(\alpha)$, which admits a clear physical interpretation, the prototype operator $\mathbf{P}(\alpha)$ is not at all an ideally time–frequency–selective projection as our GWS interpretation would require. In the following section we will see that this lack of a useful point–wise interpretation can be removed in the case of underspread systems.

4.3.5 Properties of the GWS of Underspread Systems

The action of a general LTV system may be divided in time–frequency shifting and time–frequency selective multiplication, but it is only the latter effect that we take into account in our interpretation of the GWS. Hence, it is near at hand to restrict the discussion to systems which do not cause considerable time–frequency shifts. We have already specified this class via the underspread limitation (4.14). In fact, one can show that the desirable properties of the GWS get approximately valid with decreasing spreading product $\tau_0\nu_0$[14].

More mathematical details can be found in the following chapter, now we mention just the most important properties of the GWS of underspread systems.

Discrete Weyl–Heisenberg Expansion. The Fourier relationship between the generalized spreading function and the GWS (see (4.26)) allows to apply the sampling theorem for 2D functions to the GWS of an underspread system. Accordingly, the GWS is uniquely characterized by its samples on a rectangular grid with time–period $T = 1/2\nu_0$ and frequency–period $F = 1/2\tau_0$. The reconstruction equation of the GWS corresponds to what we call a *discrete Weyl–Heisenberg expansion* of the system:

$$\mathbf{H} = \sum_{l}\sum_{m} L_H^{(\alpha)}(lT,mF)\mathbf{P}^{(lT,mF)}(\alpha). \tag{4.32}$$

where $\mathbf{P}(\alpha)$ is an α–dependent prototype operator with sinc–type GWS (more details will be discussed in Chapter 5, Section 5.2). Note that for the special case $\alpha = 1/2$ this essential result was originally published by Kailath in [190]. Such a discrete Weyl–Heisenberg expansion leads immediately to a canonical realization of underspread LTV systems via multirate filter banks. We shall return to this point when we study the STFT–based system realization.

Approximate α–invariance. The GWS of an underspread system is approximately α–invariant (for a proof see the following chapter, Theorem 5.5):

$$L_H^{(\alpha_1)}(t,f) \approx L_H^{(\alpha_2)}(t,f). \tag{4.33}$$

Hence, when we treat underspread LTV systems, the choice of a particular α is a matter of convenience. In a discrete implementation the choice $\alpha = 1/2$ is preferable as it carries over to a discrete implementation.

Approximate Multiplicative Calculus. The GWS of the product of two jointly underspread operators is approximately equal to the product of the GWS of the operators:

[14] In this respect, the case $\alpha = 0$ is marked out by the fact that one could admit more general definitions of underspread systems, i.e., defined by a sheared or rotated version of the rectangular constraint in (4.14). However, to the author's knowledge the doubly symmetric underspread support (4.14) covers the practically predominant applications: Optimum filtering of natural signals and the slowly time—varying systems with limited memory. Moreover, in case of "sheared/rotated underspread" systems one can easily return to the usual underspread theory by putting a "down–chirp" and "up–chirp" multiplication/convolution before and after the system to be analyzed or synthesized.

$$\boxed{L^{(\alpha)}_{H_2H_1}(t,f) \approx L^{(\alpha)}_{H_1}(t,f) L^{(\alpha)}_{H_2}(t,f),} \tag{4.34}$$

where the approximation gets better with decreasing spreading product $\tau_0\nu_0$ (for details see Chapter 5, Theorem 5.1). Approximate validity of the symbol calculus (4.34) is of fundamental importance for the approximate equalization and inversion of underspread LTV systems and for the estimation of underspread processes. It shows that it is the class of underspread systems where one can use Zadeh's time–varying transfer function in a way just as the conventional transfer function of LTI systems. As an example consider a (zero–mean) underspread process $x(t)$ applied to an underspread LTV system $\mathbf{H}$. The correlation operators of the input and output process are related as

$$\mathbf{R}_{Hx} = \mathbf{H}\mathbf{R}_x\mathbf{H}^*. \tag{4.35}$$

Using the (4.33), (4.34) for $\alpha = 0$ and

$$\begin{aligned} L^{(\alpha)}_{R_x}(t,f) &= EW^{(\alpha)}_x(t,f), \\ L^{(0)}_{H^*}(t,f) &= L^{(0)*}_H(t,f), \end{aligned}$$

allows to reformulate the operator relation (4.35) via time–frequency representations:

$$EW^{(\alpha)}_{Hx}(t,f) \approx \left|L^{(\alpha)}_H(t,f)\right|^2 EW^{(\alpha)}_x(t,f), \tag{4.36}$$

where the L_2/L_∞–error of this approximation decreases with decreasing total spread of the process and system, respectively.

Note that (4.36) is the time–frequency–parametrized generalization of the well–known relation for stationary environments:

$$S_{Hx}(f) = |H(f)|^2 \, S_x(f),$$

where $S_x(f)$ is the power spectrum of a stationary input process, $H(f)$ is the transfer function of the LTI system, and $S_{Hx}(f)$ is the power spectrum of the stationary output process.

Another important example is approximate operator inversion. The identity operator is "totally underspread",

$$S^{(\alpha)}_I(\tau,\nu) = \delta(\tau)\delta(\nu) \qquad \Longleftrightarrow \qquad L^{(\alpha)}_I(t,f) = 1,$$

it is the neutral element of any class of jointly underspread operators. Hence, we can expect that the following approximation holds

$$L^{(\alpha)}_{H^{-1}}(t,f) \approx \frac{1}{L^{(\alpha)}_H(t,f)},$$

valid in a domain where the inverse symbol is bounded. Note, however, that the true inverse operator of an underspread operator may typically not be underspread.

Approximate Eigenpairs. Any time–frequency shifted version of an appropriately time–frequency localized prototype signal $\gamma(t)$ is an approximate eigensignal of an underspread system $\mathbf{H}$ and the generalized Weyl symbol $L^{(\alpha)}_H(t,f)$ determines the corresponding eigenvalue:

$$\mathbf{H}\mathbf{S}^{(t,f)}\gamma \approx L^{(\alpha)}_H(t,f)\mathbf{S}^{(t,f)}\gamma,$$

where $\mathbf{S}^{(t,f)}$ is an arbitrarily defined time–frequency shift operator (any member of the family $\mathbf{S}^{(t,f)}(\alpha)$, see (4.11)).

Appropriate localization of a prototype signal $\gamma(t)$ can be specified, e.g., by our standard window matching rule:

$$\frac{T_\gamma}{F_\gamma} = \frac{\tau_0}{\nu_0},$$

where T_γ is the temporal duration (square root of the second order temporal moment), F_γ denotes the bandwidth (precisely the square–root of the second order spectral moment), and τ_0, ν_0 defined the system's spreading constraint as defined by (4.14). (The mathematical details can be found in Chapter 5, Theorem 5.6.)

Supremum Norm. We have mentioned that the GWS of a general LTV system does not faithfully reflect the maximum/minimum eigenvalues/singular values of a system, i.e., a positive operator may correspond to a negative symbol and vice versa[15]. However, the GWS of an underspread system gives a reliable information on the maximum singular value, one has

$$\left| \sup |L_H^{(\alpha)}(t,f)|^2 - \|\mathbf{H}\|_\infty^2 \right| < \rho \left(\sum_{k=0}^{\infty} \sigma_k \right)^2,$$

where ρ is a constant that gets smaller with decreasing total spread of the system and σ_k are the operator's singular values (for the proof of this statement we refer to Section 5.4.6).

4.4 STFT Based System Analysis

In the previous chapter, Section 3.3 we have seen that the short–time Fourier transform (STFT)

$$STFT_x^{(\gamma)}(t,f) \stackrel{\text{def}}{=} \int_{t'} x(t')\gamma^*(t'-t)e^{-j2\pi f t'}dt' = \left\langle x, \gamma^{(t,f)} \right\rangle$$

leads to a satisfactory second order representation of underspread processes (in terms of the spectrogram) provided that one adapts the window length to the spreading constraint of the process. In view of the mathematical equivalence of second–order process and LTV system representation, we expect that these results carry over to STFT–based system analysis and design. Recall that we always assume a normalized analysis window:

$$\|\gamma\|^2 = \int_t |\gamma(t)|^2 dt = 1.$$

We start with the formal introduction of an STFT kernel $k_{STFT}^{(H,\gamma)}(t,f,t',f')$ which defines a linear integral operator that maps the STFT of the input signal onto the STFT of the output signal:

$$STFT_{Hx}^{(\gamma)}(t,f) = \int_{t'} \int_{f'} k_{STFT}^{(H,\gamma)}(t,f,t',f') STFT_x^{(\gamma)}(t',f')\, dt'\, df',$$

[15] The operator norm $\|\mathbf{H}\|_\infty$ is defined as

$$\|\mathbf{H}\|_\infty \stackrel{\text{def}}{=} \sup \left\{ \frac{\|\mathbf{H}x\|}{\|x\|} : x \neq 0 \right\} \qquad \text{with} \quad \|x\|^2 = \langle x, x \rangle.$$

In a system theoretic context such operators are called L_2–stable (finite energy input leads always to finite energy output) whenever $\|\mathbf{H}\|_\infty$ is bounded. It is important to note that we deviate from the usual notation in the mathematics literature, where "our" $\|\mathbf{H}\|_\infty$ is the standard operator norm, while the Hilbert–Schmidt norm is often denoted by $\|\mathbf{H}\|_{HS}$ or by $|||\mathbf{H}|||$. Both norms are compatible with a signal's L_2–norm in the sense that

$$\|\mathbf{H}x\| \leq \|\mathbf{H}\|_\infty \|x\| \leq \|\mathbf{H}\|_{HS} \|x\|.$$

the STFT kernel is thus given by:

$$k_{STFT}^{(H,\gamma)}(t,f,t',f') \stackrel{\text{def}}{=} \left\langle \mathbf{H}\gamma^{(t',f')}, \gamma^{(t,f)} \right\rangle = \int_{t_1}\int_{t_2} h(t_1,t_2)\gamma(t_2-t')\gamma^*(t_1-t)e^{-j2\pi(t_1 f - t_2 f')}dt_1\,dt_2,$$

which is equivalent to computing the STFT of the output signal when $\gamma^{(t',f')}$ is the input signal:

$$k_{STFT}^{(H,\gamma)}(t,f,t',f') = STFT_{H\gamma^{(t',f')}}^{(\gamma)}(t,f).$$

By setting $\mathbf{H} = \mathbf{R}_x$ we have the formal equivalence with the correlation of the STFT of a zero–mean nonstationary process as discussed in the foregoing chapter (see (2.44)):

$$k_{STFT}^{(R_x,\gamma)} = \left\langle \mathbf{R}_x\gamma^{(t',f')}, \gamma^{(t,f)} \right\rangle = R_{STFT}^{(x,\gamma)}(t,f,t',f').$$

This formal equivalence allows to utilize (2.46) by replacing $EA_x(\tau,\nu)$ by $S_H^{(1/2)}(\tau,\nu)$ such that we have a useful expression for the STFT kernel in terms of the system's spreading function and the ambiguity function of the window:

$$k_{STFT}^{(H,\gamma)}(t,f,t',f') = \int_{\tau}\int_{\nu} S_H^{(1/2)}(\tau,\nu)A_\gamma^{(1/2)}(t-t'-\tau,f-f'-\nu)e^{j2\pi[(\nu-f)(t'+\tau)+f't']}d\tau\,d\nu. \tag{4.37}$$

4.4.1 Short–Time Transfer Function

The STFT is a highly redundant, unique 2D linear signal representation, hence the STFT kernel establishes a redundant, unique 4D linear representation of LTV systems. In practice this means that the representation is too detailed to be useful. In fact, when we study time–frequency–selective multiplication it is only the *diagonal* of the STFT–kernel we are basically interested in. (The off–diagonal contributions essentially characterize the time–frequency displacement effects.) We call this diagonal *short–time transfer function*, it is defined as

$$T_H^{(\gamma)}(t,f) = k_{STFT}^{(H,\gamma)}(t,f,t,f) = \left\langle \mathbf{H}\gamma^{(t,f)}, \gamma^{(t,f)} \right\rangle = \left\langle \mathbf{H}, \mathbf{P}_\gamma^{(t,f)} \right\rangle, \tag{4.38}$$

i.e., in practice, this means applying a time–frequency shifted version of a prototype signal at the input and matched filtering of the output signal ($\mathbf{P}_\gamma$ is the rank–one projection onto the window function)[16].

Of course, the short–time transfer function does not give a necessarily unique representation of a general LTV system. We can easily trace back this issue to the theory of underspread processes using the fact that the short–time transfer function of a correlation operator is just the expected spectrogram of the underlying process $x(t)$ (see (3.9)):

$$T_{R_x}^{(\gamma)}(t,f) = ESPEC_x^{(\gamma)}(t,f). \tag{4.39}$$

Before we reformulate the window matching criterion we mention a few other properties of the short–time transfer function (by the way of replacing $\mathbf{R}_x$ by $\mathbf{H}$ we can exploit all STFT–relevant results of the foregoing chapters).

Relation to the Generalized Weyl Symbol. The short–time transfer function of a system can be written as smoothed version of the system's generalized Weyl symbol (GWS) where the smoothing

[16] In the mathematical physics literature an operator quantization of the form

$$\sigma(\rho) = \left\langle \mathbf{H}\gamma^{(\rho)}, \gamma^{(\rho)} \right\rangle$$

is called *Berezin symbol* [274] or also *lower symbol* [305], where $\gamma^{(\rho)}$ denotes the action of a certain group (or projective representation of a group) of unitary operators acting on γ. Hence, from this point of view one may call $T_H^{(\gamma)}(t,f)$ "Weyl–Heisenberg–Berezin" symbol of $\mathbf{H}$.

kernel is the generalized Wigner distribution of the window (analogously to the relation between the physical spectrum and the Wigner–Ville spectrum (2.33)):

$$T_H^{(\gamma)}(t,f) = L_H^{(\alpha)}(t,f) * * W_\gamma^{(\alpha)*}(-t,-f). \tag{4.40}$$

Hence, the short–time transfer function is a natural, simple estimator of an LTV system's time–varying transfer function Moreover, (4.40) immediately shows that $T_H^{(\gamma)}(t,f)$ is an alternative, nonunitary Weyl–Heisenberg symbol of a linear system as it satisfies the shift–covariance property: A shift of the symbol corresponds to a unitarily equivalent system:

$$T_{H^{(\tau,\nu)}}^{(\gamma)}(t,f) = T_H^{(\gamma)}(t-\tau, f-\nu).$$

Normal HS operator. For a normal Hilbert–Schmidt operator the short–time transfer function leads to an eigenvalue–weighted sum of the spectrograms of the eigensignals (see (2.24)):

$$T_H^{(\gamma)}(t,f) = \sum_{k=1}^{\infty} \lambda_k SPEC_{u_k}^{(\gamma)}(t,f). \tag{4.41}$$

Self–Adjoint Operator. It is a well–known fact of the theory of self–adjoint operators that the range interval of the quadratic form for a normalized input signal is equal to the eigenvalue range interval [252]:

$$\lambda_{\min} \leq \langle \mathbf{H}x, x \rangle \leq \lambda_{\max} \qquad \text{for} \qquad \|x\| = 1. \tag{4.42}$$

Hence, from the definition of the short–time transfer function (4.38) we can immediately conclude that for a self–adjoint operator, the range of the short–time transfer function is bounded by the maximum/minimum eigenvalues:

$$\lambda_{\min} \leq T_H^{(\gamma)}(t,f) \leq \lambda_{\max}. \tag{4.43}$$

It depends on the operator and the window whether equality actually takes place for some (t,f). More generally, one can show that (for bounded operators on $L_2(\mathbb{R})$):

$$\left\| T_H^{(\gamma)} \right\|_\infty \leq \|\mathbf{H}\|_\infty ,$$

which motivates to call $T_H^{(\gamma)}(t,f)$ *lower symbol* of $\mathbf{H}$[305].

LTI System. The short–time transfer function of an LTI system leads to a smoothed version of the LTI systems's true transfer function and the smoothing kernel is the spectrum of the window (see (2.17)):

$$T_{H_{LTI}}^{(\gamma)}(t,f) = H(f) * |\Gamma(-f)|^2,$$

Note that i) this relation is well–known and underlies most practical LTI transfer function measurement systems[17], ii) in contrast to its time–varying generalization (4.40) this estimate can be made arbitrarily precise by increasing the window length.

[17] With regard to correlative methods (based on pseudorandom excitation) it should be noted that, in essence, computing the usual cross–correlation estimates is equivalent to evaluating a quadratic form as underlies the definition of the short–time transfer function. In fact, the quadratic form is the fundamental mathematical eigenvalue estimator such that any sound (non–parametric) transfer function estimator must implicitly or explicitly evaluate such a quadratic form.

4.4.2 Window Matching for STFT–based System Analysis

What is a good window for the above discussed STFT–based system analysis? Recall that the short–time transfer function does not uniquely characterize a general LTV system. Hence, we define a good window by the requirement of getting as much information about the system as possible via the short–time transfer function. However, since the short–time transfer function is the diagonal of the STFT kernel this requirement is essentially equal to *minimizing an off–diagonal norm of the STFT kernel.* Analog to the statistical STFT–window optimization of the foregoing chapter we define the optimum window by

$$\gamma_{opt} = \arg\min_{\gamma} \int_t \int_f \int_\tau \int_\nu \left| K_{STFT}^{(H,\gamma)}(t,f,t-\tau,f-\nu) \right|^2 (1 - W(\tau,\nu))\, dt\, df\, d\tau\, d\nu,$$

where $W(\tau,\nu)$ is a radially non–increasing smoothing kernel (see (3.3)). Utilizing the results of the foregoing chapter, a compact form is obtained by replacing the expected ambiguity function $EA_x^{(\alpha)}(\tau,\nu)$ in (3.6) by the spreading function of the operator, $S_H^{(\alpha)}(\tau,\nu)$ (we suppress the α superscript since all terms are α–invariant)

$$\gamma_{opt} = \arg\max_{\gamma} \left\langle |S_H|^2 * *W, |A_\gamma|^2 \right\rangle, \qquad \|\gamma\| = 1. \tag{4.44}$$

However, it is often the nature of system analysis that we do not a priori know the kernel. But what we usually have is some a priori knowledge on the support of the spreading function. In this case one can apply our matching rule (3.20)

$$\frac{T_\gamma}{F_\gamma} = \frac{\tau_0}{\nu_0},$$

where T_γ^2, F_γ^2 are the temporal and spectral moments of a good time–frequency localized window function, and τ_0, ν_0, characterize the a priori knowledge about the rectangular spreading constraint of the system. The choice of such a matched window function assures the following advantages:

- The short–time transfer function leads to a unique representation of an underspread LTV system. This fact is best explained by switching from the (t,f)–domain convolution relation (4.40) to the corresponding (τ,ν)–domain multiplication:

$$\begin{aligned} T_H^{(\gamma)}(t,f) &= L_H^{(\alpha)}(t,f) * *W_\gamma^{(\alpha)*}(-t,-f) \\ &\Updownarrow \\ \left(\mathcal{F}_{t\to\nu}\mathcal{F}_{f\to\tau}^{-1} T_H^{(\gamma)}\right)(\tau,\nu) &= S_H^{(\alpha)}(\tau,\nu) A_\gamma^{(\alpha)*}(\tau,\nu). \end{aligned} \tag{4.45}$$

  Since $S_H^{(\alpha)}(\tau,\nu)$ is a unique system representation, it suffices to study whether one can recover $S_H^{(\alpha)}(\tau,\nu)$ from given $T_H^{(\gamma)}(t,f)$. In fact, whenever the essential support of $A_\gamma^{(\alpha)}(\tau,\nu)$ covers the support of $S_H^{(\alpha)}(\tau,\nu)$, then $S_H^{(\alpha)}(\tau,\nu)$ is well–defined via a minimum–norm deconvolution (equivalent to the inversion of the map from the expected spectrogram to the Wigner–Ville spectrum as discussed in Section 3.3):

$$S_H^{(\alpha)}(\tau,\nu) = \begin{cases} \dfrac{\left(\mathcal{F}_{t\to\nu}\mathcal{F}_{f\to\tau}^{-1} T_H^{(\gamma)}\right)(\tau,\nu)}{A_\gamma^{(\alpha)*}(\tau,\nu)}, & |A_\gamma(\tau,\nu)| > \epsilon, \\ 0, & |A_\gamma(\tau,\nu)| \le \epsilon, \end{cases} \tag{4.46}$$

  where ϵ has to be selected such that the ϵ–support $A_\gamma(\tau,\nu)$ covers the (a priori known) support of $S_H(\tau,\nu)$.

- For underspread systems, the bias of the short–time transfer function as an estimator of the GWS is comparatively small, since one has $A_\gamma^{(\alpha)}(\tau,\nu) \approx 1$ within the support of $S_H^{(\alpha)}(\tau,\nu)$ in (4.45). (For a more detailed discussion see Section 3.3.7, where the analogous problem of the bias of spectrogram based Wigner–Ville spectrum estimation is considered.)

  For strongly underspread systems one may well replace the GWS by the short–time transfer functions in the approximate symbol calculus:

$$T_{H_1H_2}^{(\gamma)}(t,f) \approx T_{H_1}^{(\gamma)}(t,f) T_{H_2}^{(\gamma)}(t,f).$$

  However, we conjecture that the approximation error will be generally larger than in the GWS based symbol calculus. For example, in case of LTI systems where one has a GWS–based perfect symbol calculus, it is easy to see that the symbol calculus based on the short–time transfer function leads to a (window–dependent) error.

- Used as a prototype signal, the matched window and any time–frequency shifted version thereof is an approximate eigensignal of the system, we shall follow this idea more specifically in our study of WSSUS channels.

In this section we have seen that the STFT is an appropriate tool for the analysis of underspread LTV systems. It may be expected that the STFT is also useful for the realization of underspread systems. This is the topic of the following section.

4.5 STFT–based Synthesis of LTV Systems

Multiplicative modification of the short–time Fourier transform (STFT) is a well–known concept for the design of linear time–varying filters [283]. The basic idea is to insert a modification step between STFT analysis and synthesis. The overall system consists of three parts:

1. STFT analysis with analysis window $\gamma(t)$,

$$STFT_x^{(\gamma)}(t,f) = \int_s x(s)\gamma^*(s-t)e^{-j2\pi fs}ds,$$

2. Multiplicative modification of the STFT outcome $STFT_x^{(\gamma)}(t,f)$ by a time–frequency–parametrized multiplier function $M(t,f)$,

$$\widetilde{STFT}_x^{(\gamma)}(t,f) = M(t,f)STFT_x^{(\gamma)}(t,f),$$

3. STFT synthesis applied to the modified STFT outcome $\widetilde{STFT}_x^{(\gamma)}(t,f)$ using a so–called synthesis window $g(t)$; this yields the output signal $(\mathbf{H}_{STFT}x)(t)$

$$(\mathbf{H}_{STFT}x)(t) = \int_{t'}\int_{f'} \widetilde{STFT}_x^{(\gamma)}(t',f')g(t-t')e^{j2\pi f't}dt'\,df'.$$

This concept is illustrated in Figure 4.2.

For simplicity of the discussion we shall concentrate on the choice $g(t) = \gamma(t)$ (equal analysis and synthesis window).

Our usual terminology for the time–frequency shifting of operators provides a compact notation: recall that STFT synthesis may be viewed as a continuous Weyl–Heisenberg resolution of the identity

$$\mathbf{I} = \int_t\int_f \mathbf{P}_\gamma^{(t,f)}\,dt\,df,$$

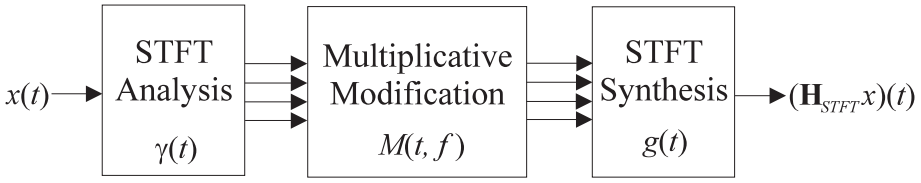

Figure 4.2: On the synthesis of linear time–varying systems via multiplicative modification of the short–time Fourier transform.

the STFT analysis–modification–synthesis scheme is a perturbation of this resolution of the identity:

$$\mathbf{H}_{STFT} = \int_t \int_f M(t,f)\ \mathbf{P}_{\gamma}^{(t,f)}\, dt\, df. \tag{4.47}$$

In view of the structural parallelism with the GWS–based Weyl–Heisenberg expansion (4.31):

$$\mathbf{H} = \int_t \int_f L_H^{(\alpha)}(t,f)\mathbf{P}^{(t,f)}(\alpha) dt\, df,$$

we interpret $M(t,f)$ as another Weyl–Heisenberg symbol of a linear operator $\mathbf{H}_{STFT}$ (see for the parallelism to the generalized Weyl symbol). We shall call this symbol "*STFT multiplier*". Clearly, the map $M(t,f) \mapsto \mathbf{H}_{STFT}$ is not unitary just as the map underlying the "short–time transfer function". We have a simple convolution relation with the GWS:

$$L^{(\alpha)}_{H_{STFT}}(t,f) = M(t,f) * * W_{\gamma}^{(\alpha)}(t,f), \tag{4.48}$$

which shows how the spectral and temporal width of the window blur the spectral and temporal selectivity of the multiplier function $M(t,f)$. Recall that this is analogous to the case of STFT–based system analysis via the short–time transfer function:

$$T_H^{(\gamma)}(t,f) = L_H^{(\alpha)}(t,f) * * W_{\gamma}^{(\alpha)*}(-t,-f).$$

The map $M(t,f) \mapsto \mathbf{H}_{STFT}$ is an ongoing area of research in mathematics [261, 305], where operators of the form $\mathbf{H}_{STFT}$ are called *Toeplitz* and $M(t,f)$ is usually referred to as *upper symbol*. We here just discuss the issue of minimum/maximum eigenvalues (L_2–stability and L_2–invertibility) as this is of obvious interest in the design of LTV filters (and it clarifies the mathematical terminology). In this respect, the disadvantage of the non–unitarity of $M(t,f)$ is somewhat compensated by the fact that $M(t,f)$ reliably reflects the minimum/maximum eigenvalue. Based on the spectrogram of an input signal,

$$SPEC_x^{(\gamma)}(t,f) = \left| STFT_x^{(\gamma)}(t,f) \right|^2 = \left\langle \mathbf{P}_{\gamma}^{(t,f)} x, x \right\rangle$$

and (4.47), one can easily derive a time–frequency parametrized formulation of the quadratic form:

$$\langle \mathbf{H}_{STFT} x, x \rangle = \left\langle M, SPEC_x^{(\gamma)} \right\rangle. \tag{4.49}$$

Real–Valued Multiplier. Eq. (4.49) immediately shows that real–valued multipliers always yield self–adjoint operators. For normalized $x(t)$ the range interval of the quadratic form is equal to the eigenvalue range interval (see (4.43)). Recognizing the positivity and boundedness of the spectrogram of a normalized window,

$$0 \le SPEC_x^{(\gamma)}(t,f) \le 1 \qquad \text{for} \qquad \|x\| = 1,$$

one has

$$\inf M(t,f) \leq \lambda_k \leq \sup M(t,f), \tag{4.50}$$

hence, the term *upper* symbol in the mathematics literature. We furthermore mention without proof that this property can be essentially generalized to complex–valued multipliers in so far as bounded multipliers correspond to bounded symbols. Hence, the map $M(t,f) \mapsto \mathbf{H}_{STFT}$ preserves boundedness and positivity. This is in striking contrast to the GWS where a bounded symbol may correspond to an unbounded operator and positive symbols can lead to indefinite operators.

Spreading Function of STFT–based Systems. The convolution relation of the Weyl–Heisenberg symbols, $L_H^{(\alpha)}(t,f)$ and $M(t,f)$, (4.48) correspond to multiplicative relations in the (τ,ν)–domain

$$S_{H_{STFT}}^{(\alpha)}(\tau,\nu) = m(\tau,\nu) A_\gamma^{(\alpha)}(\tau,\nu), \tag{4.51}$$

where $m(\tau,\nu)$ denotes the symplectic Fourier transform of $M(t,f)$,

$$m(\tau,\nu) \stackrel{\text{def}}{=} \mathcal{F}_{t\to\nu}\mathcal{F}_{f\to\tau}^{-1}\{M(t,f)\}. \tag{4.52}$$

and $A_\gamma^{(\alpha)}(\tau,\nu)$ is the generalized ambiguity function of the window $\gamma(t)$. As expected, for the practically reasonable assumptions of

1. a usual (good time–frequency localized) window function,
2. a well–behaved (non–oscillating) multiplier $M(t,f)$,

multiplicative modification of the STFT leads always to systems with restricted time–frequency shifting, i.e., *STFT based systems are always approximately underspread.*

In the following section we shall see that the converse statement is also true: any system with well–restricted time–frequency shifting is realizable by multiplicative STFT–modification based on an appropriate window function $\gamma(t)$ and bounded multiplier function $M(t,f)$.

4.5.1 Window Matching for STFT–based Systems

For a given spreading constraint (τ_0,ν_0) we have the matching requirement that any operator satisfying this constraint be *realizable via STFT analysis–modification–synthesis* using the matched window. Realizability means that one can find a bounded multiplier function by a regular deconvolution of (4.48) corresponding to a regular division in the (τ,ν)–domain:

$$m(\tau,\nu) = \begin{cases} \dfrac{S_{H_{STFT}}^{(\alpha)}(\tau,\nu)}{A_\gamma^{(\alpha)}(\tau,\nu)}, & \left|A_\gamma^{(\alpha)}(\tau,\nu)\right| \geq \epsilon, \\ 0, & \left|A_\gamma^{(\alpha)}(\tau,\nu)\right| < \epsilon, \end{cases} \tag{4.53}$$

where ϵ is a small constant that would depend on the numerical implementation. Just as for the STFT–based system analysis and nonstationary process representation the matching is obtained by adapting the essential support of the window's ambiguity function such that it covers the (a priori knowledge of) the essential support of the system's spreading function (see Fig. 3.1, Section 3.3). Such an adaption can be obtained by our usual matching rule:

$$\frac{T_\gamma}{F_\gamma} = \frac{\tau_0}{\nu_0},$$

where now τ_0, ν_0 characterize the support of $S_H(\tau,\nu)$ and T_γ, F_γ are the duration and bandwidth of the window.

4.5.2 Weyl–Heisenberg Frames as STFT–Multipliers

We now show how the above discussed STFT filter theory turns out useful for the context of the Gabor expansion (which in a discrete setting corresponds to DFT–type filterbanks [38]).

Recall that the frame operator of a Weyl–Heisenberg frame can be written as a sum of the rank–one projection operator onto the prototype, shifted on the grid:

$$\mathbf{M}_\gamma = \sum_m \sum_l \mathbf{P}_\gamma^{(lT,mF)}.$$

When we compare this formula with the continuous Weyl–Heisenberg expansion of STFT–multiplication systems (4.47), it is easy to see that $\mathbf{M}_\gamma$ can be interpreted as an STFT–multiplication system with the following distribution–valued multiplier:

$$M^{(T,F)}(t,f) = \sum_l \sum_m \delta(t - lT)\delta(f - mF). \tag{4.54}$$

Moreover, we have shown that the spreading function of STFT–multiplication systems is given by

$$S^{(\alpha)}_{H_{STFT}}(\tau,\nu) = m(\tau,\nu) A^{(\alpha)}_\gamma(\tau,\nu),$$

where $m(\tau,\nu)$ is the symplectic 2D Fourier transform of the multiplier. In the specific case of the "delta grid multiplier" (4.54) we obtain a 2D delta pulse train on a "dual" grid:

$$m^{(T,F)}(\tau,\nu) = \frac{1}{TF} \sum_l \sum_m \delta\left(\tau - \frac{l}{F}\right) \delta\left(\nu - \frac{m}{T}\right), \tag{4.55}$$

which allows to conclude immediately that the spreading function of the Weyl–Heisenberg frame operator is given by:

$$S^{(\alpha)}_{M_\gamma}(\tau,\nu) = \frac{1}{TF} \sum_l \sum_m \delta\left(\tau - \frac{l}{F}\right) \delta\left(\nu - \frac{m}{T}\right) A^{(\alpha)}_\gamma\left(\frac{l}{F}, \frac{m}{T}\right),$$

an essentially well–known result ("Janssen's representation" of the frame operateror [181]), revisited in an alternative, time–frequency symmetric way.

Studying the perfect reconstruction problem for the Gabor expansion means that we start with defining ($\gamma(t)$ is the analysis, $g(t)$ is the synthesis prototype)

$$\mathbf{M}_{g,\gamma} \stackrel{\text{def}}{=} \sum_l \sum_m \mathbf{P}_{g,\gamma}^{(lT,mF)}, \quad \text{with} \quad (\mathbf{P}_{g,\gamma})(t,s) = g(t)\gamma^*(s), \tag{4.56}$$

which should be equal to the identity operator on $L_2(\mathbb{R})$. Now, based on the above discussed STFT–multiplier trick we get a compact proof of the famous Wexler–Raz result:

$$\begin{aligned}
\mathbf{M}_{g,\gamma} &= \mathbf{I} \\
&\Updownarrow \\
L^{(\alpha)}_{M_{g,\gamma}}(t,f) &\equiv 1 \\
&\Updownarrow \\
S^{(\alpha)}_{M_{g,\gamma}}(\tau,\nu) &= \delta(\tau)\delta(\nu) \\
&\Updownarrow \\
A^{(\alpha)}_{g,\gamma}\left(\frac{l}{F}, \frac{m}{T}\right) &= TF\delta_{l0}\delta_{m0},
\end{aligned}$$

i.e., perfect reconstruction requires to solve for a biorthogonality condition on a dual grid. This method of studying Weyl–Heisenberg frames allows more than just a formal rederivation of existing results, namely it leads to straightforward important structural generalizations:

- One can generalize the rectangular sampling grid to arbitrary *non–separable sampling grid* (for example Quincunx–type) in (4.54) and by exploiting classical results from 2D sampling theory [88] we obtain a dual grid and a generalized version of the Wexler–Raz result. This provides a simple proof of a conjecture in a recent (so far unpublished work) of Zibulski and Zeevi [393] concerning the structure of the dual frame of Quincunx–type generalized Weyl–Heisenberg frames.

  It should be emphasized, that regarding the underspread operator diagonalization problems considered in this work, the Quincunx–type time–frequency sampling pattern is particularly promising whenever the actual support of the operator's spreading function tends to an elliptical shape rather than a rectangular.

- One can replace the rank–one prototype operator $\mathbf{P}_{g,\gamma}$ in (4.56) by a general rank–N operator and generalize the Wexler–Raz result to the case of multiwindowing.

- Our method of derivation uses only time–frequency–separable Fourier transforms which means that it allows for a trivial generalization to arbitrary signal dimensions (which is not possible for the usual Zak transform based derivations), when we try to cover the case of *non–separable prototypes*.

4.5.3 Discussion

By way of summarizing the discussion on STFT–based system analysis and design we state the following facts:

- The use of the STFT leads to non–unitary Weyl–Heisenberg symbols of linear systems: From an analysis point of view to the "short–time transfer function" $T_H^{(\gamma)}(t,f)$, from a synthesis point of view to the "STFT–multiplier" $M(t,f)$. These representations have the same basic interpretation as the GWS, i.e., viewing the system in the sense of a time–frequency–parametrized multiplier (time–varying transfer function).

- Both the "short–time transfer function" and the "STFT–multiplier" are interrelated to the GWS via a convolution, where the generalized Wigner distribution of the window acts as smoothing kernel:

$$\begin{aligned} L_{H_{STFT}}^{(\alpha)}(t,f) &= M(t,f) ** W_\gamma^{(\alpha)}(t,f), \\ T_H^{(\gamma)}(t,f) &= L_H^{(\alpha)}(t,f) ** W_\gamma^{(\alpha)*}(-t,-f). \end{aligned}$$

  This convolution relation shows (i) the shift–covariance and (ii) the non–unitarity of both symbols. The non–unitarity has the following practical consequences: The "short–time transfer function" $T_H^{(\gamma)}(t,f)$ does not uniquely characterize a general LTV system; it is not possible to assign an "STFT–multiplier" $M(t,f)$ to a general LTV system such that this system can be realized by multiplicative modification of the STFT using $M(t,f)$ (because the deconvolution is ill–conditioned whenever the window's ambiguity function vanishes within the essential support of the system's spreading function). However, in contrast to the unitary Weyl correspondence, the maps $\mathbf{H} \mapsto T_H^{(\gamma)}(t,f)$ and $M(t,f) \mapsto \mathbf{H}$ preserve the eigenvalue range of self–adjoint operators in the sense that:

$$\begin{aligned} \lambda_{\min} \leq \quad T_H^{(\gamma)}(t,f) \quad &\leq \lambda_{\max}, \\ \inf M(t,f) \leq \qquad \lambda_k \qquad &\leq \sup M(t,f), \end{aligned}$$

  i.e., $T_H^{(\gamma)}(t,f)$ and $M(t,f)$ are "lower" and "upper" Weyl–Heisenberg symbols of $\mathbf{H}$. These properties are of fundamental relevance for the numerical stability of STFT–based system analysis and design.

- When we restrict ourselves to underspread LTV systems and assume a matched window both the "short–time transfer function" $T_H^{(\gamma)}(t,f)$ and the "STFT–multiplier" $m(t,f)$ provide a unique representation of underspread LTV systems (since restricting the maps $\mathbf{H} \mapsto T_H^{(\gamma)}(t,f)$ and $M(t,f) \mapsto \mathbf{H}$ onto the subspace of underspread HS operators and to the subspace of accordingly bandlimited multipliers lead to invertible one–to–one mappings corresponding to minimum norm deconvolutions as considered in (4.46) and (4.53)).

- STFT–based linear filtering is the classical "mother concept" for any DFT–based perfect reconstruction systems (modulated filter banks, FFT–based transmultiplexers). The discussion in Section (4.5.2) has shown that the presented STFT–filter theory establish a conceptual basis for a more flexible design of DFT–based perfect reconstruction systems.

4.6 Signal Design for WSSUS Channels

Stochastic LTV systems provide an adequate model for various time–varying communication channels. Such a stochastic model is the so–called wide–sense stationary uncorrelated scattering (WSSUS) channel, originally defined by Bello [29], widely accepted for important multipath wave propagation channels such as the troposcatter channel, underwater acoustic channel, and (with some amount of controversy) the mobile radio channel [197, 188, 162, 116]. The WSSUS concept is also adopted in radar theory, namely, as a stochastic method for target identification [355].

Both historically and in the modern literature one can distinguish two alternative motivations for the WSSUS setup.

A *physical* way of reasoning which, in essence, assumes a large number of statistically independent point scatterers that cause (i) a (narrowband) Doppler effect, (ii) a time–delay, (iii) a path loss (scalar multiplication by a factor smaller than one), and (iv) an unimodular phase factor (which, as a random variable, is assumed to be uniformly distributed). For more details about this line of argumentation see [197]. The physical motivation is intuitively appealing for the troposcatter channel and the underwater acoustic channel but questionable for the mobile radio scenario.

A *pragmatic* formulation of the incomplete a priori knowledge about an LTV system, with the usual motivation of most statistical concepts: If we are to match a system to unknown physical situations we assume that this situations are random and we match our system to the statistics. Now, in case of nonparametric LTV system theory we are confronted with 2D random processes which, in principle, lead to 4D second order statistics. However, it is not realistic to work with 4D functions. Hence, it is necessary to assume some additional stationarity constraint such that these 4D statistics essentially degenerate to 2D functions. The only nonparametric system representation where 2D stationarity is not totally unphysical is the transfer function, but stationarity of the (time–varying) transfer function is exactly the key feature of the WSSUS concept.

In the original WSSUS model the system representations are considered as 2D stochastic processes with specific second order statistic. In the (most natural) spreading domain the WSSUS channel is characterized by i) being zero–mean, ii) showing uncorrelated amplitudes of the time–frequency shifts[18]:

$$\mathrm{E}\left\{S_H^{(1/2)}(\tau,\nu)\right\} = 0, \tag{4.57}$$

$$\mathrm{E}\left\{S_H^{(1/2)}(\tau,\nu)S_H^{(1/2)*}(\tau',\nu')\right\} = C_H(\tau,\nu)\delta(\tau-\tau')\delta(\nu-\nu'), \tag{4.58}$$

[18] Bello's original denomination is based on the statistics of the $h(t,t-\tau)$ when $h(t,s)$ denotes the system's kernel. The function $h(t,t-\tau)$ is *wide–sense stationary* with respect to t and *uncorrelated* with respect to τ. This is the way how Bello set up the classical denomination of wide–sense stationary uncorrelated scattering (WSSUS). Notwithstanding the undeniable relevance of Bello's classical work the reader may agree that the terminology is somewhat of a historical misnomer as it does not take into account the symmetric behaviour of the WSSUS channel w.r.t. time and frequency, i.e., WSSUS can be much more concisely characterized by *uncorrelated spreading*.

here, the function $C_H(\tau,\nu)$ is the so–called *scattering function*. A WSSUS channel is uniquely characterized by its scattering function.

Using the fact that

$$S_H^{(\alpha)}(\tau,\nu) = S_H^{(1/2)}(\tau,\nu)e^{-j2\pi\nu\tau(\alpha-1/2)}.$$

it is easy to show the statistical α–invariance of the WSSUS concept [205]:

$$\begin{aligned}
\mathrm{E}\left\{S_H^{(\alpha)}(\tau,\nu)S_H^{(\alpha)*}(\tau',\nu')\right\} &= \mathrm{E}\left\{S_H^{(1/2)}(\tau,\nu)S_H^{(1/2)*}(\tau',\nu')\right\} e^{-j2\pi\{\nu\tau-\nu'\tau'\}(\alpha-1/2)} \\
&= C_H(\tau,\nu)\delta(\tau-\tau')\delta(\nu-\nu')e^{-j2\pi\{\nu\tau-\nu'\tau'\}(\alpha-1/2)} \\
&= C_H(\tau,\nu)\delta(\tau-\tau')\delta(\nu-\nu'). \qquad (4.59)
\end{aligned}$$

The statistics of the generalized Weyl symbol are accordingly given by

$$\mathrm{E}\{L_H^{(\alpha)}(t,f)\} = 0, \qquad (4.60)$$

$$\mathrm{E}\{L_H^{(\alpha)}(t,f)L_H^{(\alpha)^*}(t',f')\} = R_L(t-t',f-f'). \qquad (4.61)$$

Here, $R_L(t,f)$ denotes the so–called *time–frequency correlation function* of a WSSUS system which is in Fourier correspondence to the scattering function. Due to its physical interpretation the scattering function is the most natural way to present WSSUS systems [197]. But also in our particular context of matching WH families the scattering function leads to a compact representation.

4.6.1 Optimum Single Pulse

In general, a stochastic operator has no eigensignals. Nevertheless, it makes sense to ask about that particular signal which is the best approximation to an eigensignal in an average sense over the ensemble of LTV systems. This can be made precise by evaluating the expectation of the orthogonal distortion. The output signal can be split up as follows:

$$(\mathbf{H}\gamma)(t) = \langle \mathbf{H}\gamma,\gamma\rangle\,\gamma(t) + \epsilon_{\min}(t),$$

where

$$\langle \gamma, \epsilon_{\min}\rangle = 0$$

such that using Pythagoras:

$$\|\epsilon_{min}\|^2 = \|\mathbf{H}\gamma\|^2 - |\langle \mathbf{H}\gamma,\gamma\rangle|^2.$$

We are interested in the expectation over the WSSUS ensemble

$$\mathrm{E}\left\{\|\epsilon_{min}\|^2\right\} = \mathrm{E}\left\{\|\mathbf{H}\gamma\|^2\right\} - \mathrm{E}\left\{|\langle \mathbf{H}\gamma,\gamma\rangle|^2\right\}. \qquad (4.62)$$

This expectation can be expressed in terms of the channel's scattering function $C_H(\tau,\nu)$:

$$\mathrm{E}\{\|\epsilon_{min}\|^2\} = \int_\tau \int_\nu C_H(\tau,\nu)d\tau\,d\nu - \left\langle C_H,\, |A_\gamma|^2\right\rangle, \qquad (4.63)$$

where we have used properties (B.17), (B.24) and (B.33) of the generalized spreading function and the α–invariant definition of WSSUS (4.59) in the following way

$$\begin{aligned}
\mathrm{E}\left\{\|\mathbf{H}\gamma\|^2\right\} &= \mathrm{E}\left\{\langle \mathbf{H}^*\mathbf{H}\gamma,\gamma\rangle\right\} \\
&= \mathrm{E}\left\{\left\langle S_{H^*H}^{(0)}, A_\gamma\right\rangle\right\} \\
&= \int_\tau\int_\nu\int_{\tau'}\int_{\nu'} \mathrm{E}\left\{S_H^{(0)*}(-\tau',-\nu')S_H^{(0)}(\tau-\tau',\nu-\nu')\right\} e^{-j\pi(\tau'\nu-\nu'\tau)} \\
&\quad \cdot A_\gamma^{(0)*}(\tau,\nu)d\tau\, d\nu\, d\tau'\, d\nu' \\
&= \int_\tau\int_\nu\int_{\tau'}\int_{\nu'} C_H(\tau-\tau',\nu-\nu')\delta(\tau)\delta(\nu)e^{-j\pi(\tau'\nu-\nu'\tau)}A_\gamma^{(0)*}(\tau,\nu)d\tau\, d\nu\, d\tau'\, d\nu' \\
&= \int_{\tau'}\int_{\nu'} C_H(\tau',\nu')A_\gamma^{(0)}(0,0)d\tau'\, d\nu',
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{E}\left\{|\langle\mathbf{H}\gamma,\gamma\rangle|^2\right\} &= \mathrm{E}\left\{\left|\left\langle S_H^{(\alpha)}, A_\gamma^{(\alpha)}\right\rangle\right|^2\right\} \\
&= \int_\tau\int_\nu\int_{\tau'}\int_{\nu'} \mathrm{E}\left\{S_H^{(\alpha)}(\tau,\nu)S_H^{(\alpha)*}(\tau',\nu')\right\} A_\gamma^{(\alpha)}(\tau,\nu)A_\gamma^{(\alpha)*}(\tau',\nu')d\tau\, d\nu\, d\tau'\, d\nu' \\
&= \int_\tau\int_\nu C_H(\tau,\nu)\left|A_\gamma^{(\alpha)}(\tau,\nu)\right|^2 d\tau\, d\nu.
\end{aligned}$$

Minimizing the expected orthogonal distortion leads to a compact optimization problem

$$\gamma_{opt} = \arg\max_\gamma \left\langle C_H,\, |A_\gamma|^2\right\rangle, \qquad \|\gamma\| = 1. \tag{4.64}$$

which is seen to be structurally equivalent to the STFT window optimization criterion (4.44) when one replaces the magnitude squared spreading function $|S_H(\tau,\nu)|^2$ by the channel's scattering function $C_H(\tau,\nu)$.

4.6.2 Optimum Weyl–Heisenberg Transmission Set

The above discussed optimum single pulse may be of interest as a test signal for measuring channel characteristics (channel sounding). However, for the digital communication over a time–varying channel we have to select a whole set of transmission signals. The time–frequency plane is a natural domain to study the basic signal design principles (modulation schemes) of multiple access systems. Existing digital mobile communication systems can be coarsely split up into two categories [269, 336, 379]:

- Time–division/frequency division multiple access systems where the effective transmission pulse covers approximately one Heisenberg cell and orthogonality is obtained by time–frequency–disjointness.
- Spread spectrum systems where the effective transmission pulse (the signal component corresponding to one information bearing symbol) covers a region much larger than one Heisenberg cell, orthogonality is obtained by digital modulation via specific binary sequences (the effective pulses of different users are time–frequency–overlapping).

Within the existing standards (such as the European GSM, or the U.S. IS–54) there is no design freedom for an explicit matching of the transmission signal[19]. However, for the development of future systems the need for bandwidth efficiency and the increasing digital signal processing capabilities make it reasonable to return to the roots and ask about how to adapt a (reasonably structured) signal set to the a priori knowledge about a WSSUS channel. Clearly, the result of such an optimization will always have only philosophical impact in so far as there are many other practical side constraints that have to be met in the design of a practical digital communication system.

We here consider the following setup: The input signal is a weighted linear combination of a WH set based on a transmission pulse $\gamma(t)$:

$$x(t) = \sum_m \sum_n p(m,n)\gamma^{(mT,nF)}(t), \tag{4.65}$$

where $p(m,n)$ are considered as information bearing pulse amplitudes and T may e.g. be considered as the symbol rate and F as the channel separation (see Figure 4.3). However, we treat inter– and cochannel interference in a common manner such that our results are independent of the actual rule for channel access. Note, that depending on the definition of the prototype (which need not be a transmission pulse in the classical sense) (4.65) allows to cover any practical modulation scheme (it is the mapping from the user's bit stream onto the pulse amplitudes $p(m,n)$ that allows distinguish the variety of digital communication systems).

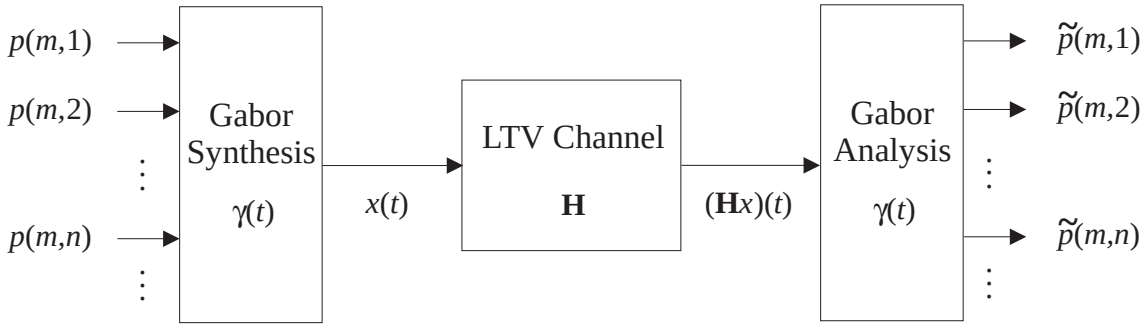

Figure 4.3: The considered digital communication setup.

We furthermore assume uncorrelated pulse amplitudes with normalized power:

$$\mathrm{E}_p\left\{p(m,n)p^*(m',n')\right\} = \delta_{mm'}\delta_{nn'}, \tag{4.66}$$

where $\delta_{m,m'}$ is the Kronecker symbol ($\delta_{m,m'} = 1$ for $m = m'$ else zero). The condition (4.66) is, in principle, realistic for modern digital communication systems but it certainly leads to a considerable narrowing of the modulation schemes that can be covered in connection with (4.65). In particular, the so–called Gaussian minimum–shift keying (GMSK) principle that underlies the European GSM standard would require statistical dependence between temporally adjacent pulse amplitudes.

A matched filter receiver for the symbol associated to $\gamma^{(mT,nF)}(t)$ evaluates an inner product of $\gamma^{(mT,nF)}(t)$ and the channel's output signal $(\mathbf{H}x)(t)$ (we assume noise free observation) :

$$\hat{p}(m,n) = \left\langle \mathbf{H}x, \gamma^{(mT,nF)} \right\rangle = \sum_{m',n'} p(m',n')H_\gamma(m,n,m',n'),$$

where we have a introduced a discretized version of the STFT kernel defined as

$$H_\gamma(m,n,m',n') \stackrel{\text{def}}{=} \left\langle \mathbf{H}\gamma^{(m'T,n'F)}, \gamma^{(mT,nF)} \right\rangle = k_{STFT}^{(H,\gamma)}(m'T,n'F,mT,nF).$$

[19] It should be noted that channel coding can be interpreted as implicitly matching the effective transmission signal by linear combination of transmission pulses. The weights of such a combination are, of course, restricted to binary numbers which leads to problems beyond the scope of this work. Yet, the combined consideration of channel coding and modulation is a promising perspective for future systems (see [379] for an excellent recent contribution in this spirit).

Due to the distortion caused by the channel the matched filter output contains undesired contributions from other pulses. The optimum transmission pulse minimizes the total expected energy of all interfering distorted pulses:

$$\gamma_{opt} = \arg\min_{\gamma} \mathrm{E}_{H,p} \left\{ \left| \sum_{m',n'} p(m',n') H_\gamma(m,n,m',n') W(m-m',n-n') \right|^2 \right\}, \tag{4.67}$$

where the expectation is over the WSSUS channel and the amplitude ensemble and the weight function selects contributions from all pulses with different TF localization:

$$W(m-m',n-n') \stackrel{\text{def}}{=} 1 - \delta_{nn'}\delta_{mm'}.$$

In order to compute this expectation we note that based on (4.66) one has the fact that

$$\mathrm{E}_p \left\{ \left| \sum_{m',n'} p(m',n') H_\gamma(m,n,m',n') \right|^2 \right\} = \sum_{m',n'} \left| H_\gamma(m,n,m',n') \right|^2,$$

and we introduce (4.37) in order to compute the expectation over the WSSUS ensemble:

$$\begin{aligned}
&\mathrm{E}_H \left\{ \left| H_\gamma(m,n,m',n') \right|^2 \right\} \\
&= \mathrm{E}_H \left\{ \left| \int_\tau \int_\nu S_H^{(1/2)}(\tau,\nu) A_\gamma^{(1/2)}((m-m')T-\tau,(n-n')F-\nu) e^{j2\pi\{n'Fm'T+(\nu-nF)mT\}} d\tau\, d\nu, \right|^2 \right\} \\
&= \int_{\tau_1}\int_{\nu_1}\int_{\tau_2}\int_{\nu_2} \mathrm{E}_H \left\{ S_H^{(1/2)}(\tau_1,\nu_1) S_H^{(1/2)*}(\tau_2,\nu_2) \right\} A_\gamma^{(1/2)}((m-m')T-\tau_1,(n-n')F-\nu_1) \\
&\quad \cdot A_\gamma^{(1/2)*}((m-m')T-\tau_2,(n-n')F-\nu_2) e^{j2\pi(\nu_1-\nu_2)mT} d\tau_1\, d\nu_1\, d\tau_2\, d\nu_2 \\
&= \int_{\tau_1}\int_{\nu_1}\int_{\tau_2}\int_{\nu_2} C_H(\tau_1,\nu_1)\delta(\tau_1-\tau_2)\delta(\nu_1-\nu_2) A_\gamma^{(1/2)}((m-m')T-\tau_1,(n-n')F-\nu_1) \\
&\quad \cdot A_\gamma^{(1/2)*}((m-m')T-\tau_2,(n-n')F-\nu_2) e^{j2\pi(\nu_1-\nu_2)mT} d\tau_1\, d\nu_1\, d\tau_2\, d\nu_2 \\
&= \int_\tau \int_\nu C_H(\tau,\nu) \left| A_\gamma(\tau-(m-m')T,\nu-(n-n')F) \right|^2 d\tau\, d\nu.
\end{aligned}$$

Combining these results we have

$$\begin{aligned}
&\mathrm{E}_{H,p} \left\{ \left| \sum_{m',n'} p(m',n') H_\gamma(m,n,m',n') W(m-m',n-n') \right|^2 \right\} \\
&= \left\langle \sum_m \sum_n W(m,n) C_H(\tau-mT,\nu-nF) \ , \ \left| A_\gamma(\tau,\nu) \right|^2 \right\rangle.
\end{aligned}$$

Hence, the optimization of the transmission pulse according to (4.67) is given by

$$\boxed{\begin{aligned} &\gamma_{opt,G} = \arg\min_{\gamma} \left\langle \textstyle\sum_m \sum_n W(m,n) C_H(\tau-mT,\nu-nF) \ , \ \left| A_\gamma(\tau,\nu) \right|^2 \right\rangle, \\ &\qquad\qquad \text{subject to} \qquad \|\gamma\|^2 = 1. \end{aligned}} \tag{4.68}$$

which is structurally equivalent to the matching criterion of the Gabor expansion of a nonstationary process (3.40) (by replacing $|EA_x(\tau,\nu)|^2$ by $C_H(\tau,\nu)$). From the general viewpoint of multirate signal processing this equivalence is not unexpected: In a discrete–time setting, the Gabor expansion corresponds to DFT filter banks, a subclass of uniform filter banks, whose theory is well–known for its mathematical equivalence to transmultiplexer theory [54, 62, 357]. In view of the parallelism with the Gabor matching theory we refer to Section 3.4 for a detailed discussion of (4.68). However, recall that the matching of a Weyl–Heisenberg set includes matching the sampling grid. In this respect, we emphasize that all of the arguments in Section 3.4 that led to the matched grid:

$$\frac{T}{F} = \frac{\tau_0}{\nu_0},$$

essentially apply to the WSSUS case. That is, the presented theory essentially allows to design an efficient (DFT based), perfect reconstruction transmultiplexer whose prototypes are approximate eigensignals of an underspread WSSUS channel. This means that any user "sees" an almost nondispersive channel with random gain (regardless of the multiple access rule).

4.7 Extension to Matched Multiwindow Expansions

In the previous sections we have shown that underspread LTV systems can be uniquely analyzed and realized via STFT–based methods. In both cases we have used a minimum–norm deconvolution in order to link the STFT–based non–unitary system representations (the short–time transfer function $T_H^{(\gamma)}(t,f)$ and the multiplier symbol $M(t,f)$) to the unitary Weyl correspondence. However, while such a minimum–norm deconvolution was apt to proof invertibility on a subspace of underspread operators, it is of limited practical relevance due to its huge numerical expense. In order to realize an underspread operator with given GWS we now discuss multiwindow methods both for the theoretically simple continuous and the practically important discrete setting.

4.7.1 Continuous Expansion

Any operator with restricted spreading characterized by

$$S_H^{(\alpha)}(\tau,\nu) = S_H^{(\alpha)}(\tau,\nu)\,\chi_H(\tau,\nu),$$

can be written as a *continuous Weyl–Heisenberg expansion*

$$\mathbf{H} = \int_t \int_f L_H^{(\alpha)}(t,f)\,\mathbf{P}^{(t,f)}(\alpha)dt\,df, \tag{4.69}$$

where the prototype operator is determined by the spreading constraint:

$$S_{P(\alpha)}^{(\alpha)}(\tau,\nu) = \chi_H(\tau,\nu).$$

In Section 4.5 we have shown how STFT analysis–modification–synthesis allows to realize underspread operators in the form of a Weyl–Heisenberg expansion based upon the rank–one projection operator $\mathbf{P}_\gamma$ (see Section 4.5):

$$\mathbf{H}_{STFT} = \int_t \int_f M(t,f)\mathbf{P}_\gamma^{(t,f)}\,dt\,df,$$

where $M(t,f)$ is the multiplier and $\mathbf{P}_\gamma$ is the rank–one projection onto the window. The prototype operator in (4.69) is not rank–one. However, by using the singular value decomposition of the prototype operator,

$$\mathbf{P}(\alpha) = \sum_k \sigma_k \mathbf{P}_{u_k,v_k},$$

one can set up a multiwindow expansion in the following form:

$$\mathbf{H} = \sum_k \sigma_k \mathbf{H}_{STFT,k} \tag{4.70}$$

with

$$\mathbf{H}_{STFT,k} = \int_t \int_f L_H^{(\alpha)}(t,f) \mathbf{P}_{u_k,v_k}^{(t,f)} dt\, df,$$

i.e., we have a weighted parallel combination of STFT–based systems $\mathbf{H}_{STFT,k}$ with (i) equal multiplier function given by the GWS $L_H^{(\alpha)}(t,f)$, (ii) the analysis/synthesis windows given by the (α–dependent) singular signals of the prototype operator, $u_k(t), v_k(t)$, and (iii) the weights determined by the (α–dependent) singular values σ_k of the prototype operator [205].

The number of windows required is theoretically infinite. In practice, a small number of windows will lead to sufficient precision. The Hilbert–Schmidt norm of the approximation error depends on the singular value distribution of the prototype operator (we presuppose that the singular values are indexed in the order of non–increasing magnitude):

$$\left\| \mathbf{H} - \sum_{k=1}^{N} \sigma_k \mathbf{H}_{STFT,k} \right\| \leq \sum_{k=N+1}^{\infty} \sigma_k \left\| \mathbf{H}_{STFT,k} \right\| < \|\mathbf{H}\| \sum_{k=N+1}^{\infty} \sigma_k. \tag{4.71}$$

A further simplification may be based on the fact that the prototype operator is approximately normal such that one can use identical analysis and synthesis windows $u_k(t) = v_k(t)$, then, the "weights" are the generally complex eigenvalues. Note, furthermore, that the problem of finding the optimum approximation by a *single* window STFT system leads back to a window optimization criterion of the previous chapter:

$$\gamma_{opt,2}(t) = \arg\min_{\gamma} \|\mathbf{P}(\alpha) - \mathbf{P}_{\gamma}\| = \arg\min_{\gamma} \left\langle \chi_H, A_{\gamma}^{(\alpha)} \right\rangle \qquad \text{subject to} \qquad \|\gamma\| = 1.$$

4.7.2 Discrete Expansion

The GWS of an underspread system is a 2D–lowpass function, in Section 4.3.5 we have already pointed out that this leads to a discrete Weyl–Heisenberg expansion in the following form:

$$\mathbf{H} = \sum_l \sum_m L_H^{(\alpha)}(lT, mF)\, \mathbf{P}^{(lT,mF)}(\alpha).$$

From the just discussed continuous case we know that this can be realized as a multiwindow, but now also multirate STFT filter, i.e., consisting of STFT analysis, sampling on a rectangular grid, multiplicative modification and STFT synthesis. That is, we have a structurally simple, directly parallelizable multirate realization of underspread operators. This is what makes the approximate symbol calculus,

$$L_{H_1 H_2}^{(\alpha)}(lT, mF) \approx L_{H_1}^{(\alpha)}(lT, mF) L_{H_2}^{(\alpha)}(lT, mF),$$

to a feasible, efficient alternative to existing methods for the equalization, inversion or more general treatment of underspread LTV systems.

4.7.3 Illustrative Example: Practical Nonstationary Wiener Filter

Again, the nonstationary Wiener filtering (minimum mean–squared error filtering) provides a good example for the relevance of the presented theory. The nonstationary Wiener filter is given by

$$\mathbf{H}_{MMSE} = \mathbf{R}_x (\mathbf{R}_x + \mathbf{R}_n)^{-1},$$

where $\mathbf{R}_x, \mathbf{R}_n$ are the correlation operators of the signal and noise process, respectively. More details have been discussed in Section 2.1.1, appropriate references are [268, 326, 355] and in particular [199]. We furthermore assume incomplete a priori knowledge in the form of a given spreading constraint τ_0, ν_0 of an underspread process and contamination with white noise with correlation:

$$\mathbf{R}_n = \sigma_n^2 \mathbf{I}.$$

The Wigner–Ville spectrum is the Weyl symbol of the correlation operator, for the white noise we have:

$$EW_n(t,f) = \sigma_n^2.$$

If we would have given the Wigner–Ville spectrum of the signal process then the Weyl symbol of the nonstationary Wiener filter were in good approximation given by[20]:

$$L_{H_{MMSE}}(t,f) \approx \frac{EW_x(t,f)}{EW_x(t,f) + \sigma_n^2}.$$

However, since we do not know the exact second order statistics of the signal process we have to replace the expected Wigner distribution by its estimate based on the observation. Moreover, we can switch to the rectangular sampling grid $T = \frac{1}{2\nu_0}$ and $F = \frac{1}{2\tau_0}$ since we presuppose that all involved operators are (at least approximately) jointly underspread with the spreading constraint determined by the a priori knowledge about the signal process:

$$L_{H_{MMSE}}(lT,mF) \approx \frac{\widehat{E}W_x(lT,mF)}{\widehat{E}W_x(lT,mF) + \sigma_n^2}.$$

For the derivation of an optimum estimator we refer the reader to the Appendix D. There it is shown that given a noisy observation

$$y(t) = x(t) + n(t),$$

where $x(t)$ is a circular complex, zero–mean, nonstationary Gaussian process with known spreading constraint $\chi_x(\tau,\nu)$ (with regard to $EA_x(\tau,\nu)$) and $n(t)$ is circular complex, zero–mean stationary white Gaussian noise with variance σ_n^2, one can show that the MVUB estimator of the process' Wigner–Ville spectrum is given by

$$\widehat{E}W_x(t,f) = \left\langle \widehat{\mathbf{P}}^{(t,f)} y, y \right\rangle,$$

where $\widehat{\mathbf{P}}$ is the prototype operator of the minimum–norm Weyl–Heisenberg expansion of the process' correlation operator $\mathbf{R}_x$ according to the spreading constraint $\chi_x(\tau,\nu)$. The practical realization of this estimator is given by the optimum finite–rank approximation as a weighted sum of spectrograms:

$$\widehat{E}W_x(t,f) = \sum_{k=1}^{N} \lambda_k SPEC_y^{(u_k)}(t,f),$$

where the window functions are the eigensignals of the prototype operator corresponding to the spreading constraint. The MMSE filter can then be realized as a finite–rank multiwindow STFT–based system in the form

$$\mathbf{H}_{MMSE} = \sum_{k=1}^{N} \lambda_k \sum_l \sum_m L_{H_{MMSE}}(lT,mF) \mathbf{P}_{u_k}^{(lT,mF)},$$

where the window functions $u_k(t)$ again are the eigensignals of the prototype operator corresponding to the spreading constraint of the signal process.

This leads to a highly efficient realization as we can use *one and the same set of orthogonal windows for the optimum spectral estimation and the optimum filtering of an underspread process.* Figure 4.4

[20] For simplicity we suppress the α–parametrization of the GWS, all of the results are basically α–invariant.

shows a schematic illustration of the overall signal processing scheme. It should be emphasized that this concept is not as "revolutionary" as it seems on the first sight. Existing concepts for speech enhancement based on a filter bank which is used both for the estimation of the power spectrum and the filtering turn out to be conceptually similar. One has to recognize that averaging over the instantaneous channel power of succesive blocks is equivalent to the use of a multiwindow estimator based on time–disjoint (thereby orthogonal) windows. We have already pointed out this equivalence at the end of the previous chapter. Hence our Weyl–Heisenberg theory may also be seen as an nonstationary reinterpretation of existing "quasistationary methods". We feel that the presented theory provides a basis for a more systematic design of nonstationary enhancement filters in a way such that the choice of less parameters is subject to a trial and error procedure.

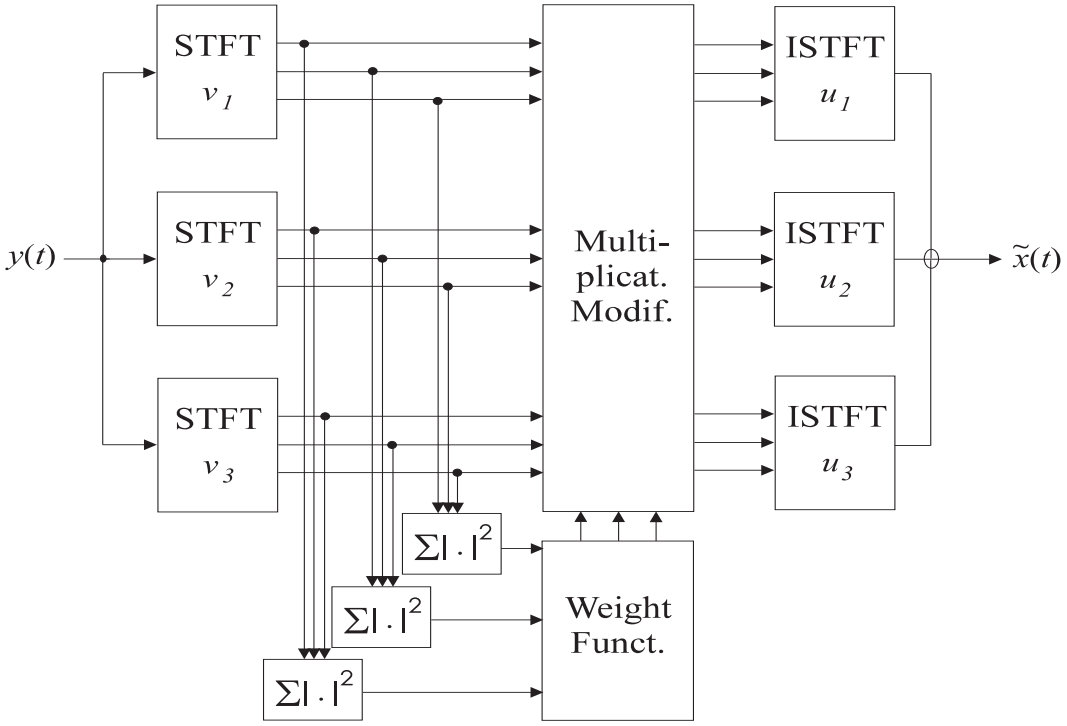

Figure 4.4: Realization of nonstationary Wiener filter via multiple STFT multiplication systems.

4.8 Numerical Experiments

The multiwindow realization of the Wiener filter (as discussed in the previous section) is the topic of this numerical experiment. We consider a synthetical zero–mean, complex–valued nonstationary signal process with time–varying bandpass–like innovations system. The observation of this process is subject to additive, white Gaussian noise. We study the performance of the following exact and approximate Wiener filter realizations for different noise levels:

1. The exact Wiener filter based on complete knowledge of $\mathbf{R}_x$ and the noise level, computed via matrix inversion.

2. An approximate Wiener filter also based on complete second–order knowledge but computed via the symbolic (transfer function) calculus using Zadeh's function.

3. Another approximate realization based on complete knowledge now realized via the multiwindow–STFT scheme.

	SNR of observation				
Filter method	–3	–2	–1	0	1
Exact Wiener filter compl. 2nd–ord.–knowl.	4.0	3.3	2.5	2.1	1.2
Multi–STFT realization compl. 2nd–ord.–knowl.	3.9	3.1	2.3	1.6	0.9
Multi–STFT realization incompl. 2nd–ord.–knowl.	3.1	2.1	1.2	0.2	−0.7

Table 4.1: Average SNR–improvement in dB of the various filter methods for different noise levels

4. A "practical" enhancement scheme based on the incomplete knowledge of (i) the support of $EA_x(\tau,\nu)$, and (ii) the noise level using a multiwindow STFT both for the estimation and the filtering (see Figure 4.4). It should be emphasized that this concept uses no *a priori* knowledge about the (absolute) time–frequency localization of the signal process.

Table 4.1 lists the average SNR improvement of the filter schemes for different levels of input SNR measured in dB. We have not included method 2. in this table as it shows no significant difference to the exact Wiener filter. (A specific experiment about the symbolic calculus for underspread and overspread operators can be found at the end of the following chapter.) The practical enhancement scheme shows good performance for negative SNR levels while it breaks down at positive SNR levels, see the last line of Table 4.1. Figure 4.5 shows the underlying window functions, the Rihaczek spectrum of the process, and the observation–dependent features of the "practical" filtering scheme for typical noise and signal realizations at input–SNR=0dB.

4.9 Summary

In this chapter we have studied linear time–varying systems via time–frequency parametrized representations. The generalized spreading function (GSF) $S_H^{(\alpha)}(\tau,\nu)$ characterizes the time–frequency shifts caused by a linear system, while the generalized Weyl symbol (GWS) $L_H^{(\alpha)}(t,f)$ reflects time–frequency selective multiplication in the sense of a time–varying transfer function (within the limits of Heisenberg's uncertainty principle). The generalized Weyl symbol combines and unifies various classical definitions of a time–frequency parametrized linear operator representation: Zadeh's time–varying transfer function, the Kohn–Nirenberg symbol, Bello's "frequency–dependent modulation function", and the Weyl symbol. The mutual interrelation between the Weyl symbol and the spreading function is essentially a 2D Fourier transform (symplectic Fourier transform). By defining a coordinate transform acting on the impulse response $h(t,t')$ and the bifrequency function $B_H(f,f')$ as:

$$\begin{aligned}
\left(\mathcal{A}_\alpha h\right)(t,\tau) &= h^{(\alpha)}(t,\tau) \stackrel{\text{def}}{=} h\left(t+\left(\frac{1}{2}-\alpha\right)\tau, t-\left(\frac{1}{2}+\alpha\right)\tau\right), \\
\left(\mathcal{A}_\alpha B_H\right)(f,\nu) &= B_H^{(\alpha)}(f,\nu) \stackrel{\text{def}}{=} B_H\left(f+\left(\frac{1}{2}-\alpha\right)\nu, f-\left(\frac{1}{2}+\alpha\right)\nu\right),
\end{aligned}$$

we have the following commutative map of relations:

$$\begin{array}{ccccc} B_H(f,f') = \sum_{k=1}^{\infty} \sigma_k U_k(f) V_k^*(f') & \xrightarrow{\mathcal{A}_\alpha} & B_H^{(\alpha)}(f,\nu) & \xrightarrow{\mathcal{F}} & L_H^{(\alpha)}(t,f) = \sum_{k=1}^{\infty} \sigma_k W_{u_k,v_k}^{(\alpha)}(t,f) \\ \downarrow \mathcal{F}\mathcal{F}^{-1} & & \downarrow \mathcal{F}\mathcal{F}^{-1} & & \downarrow \mathcal{F}\mathcal{F}^{-1} \\ h(t,t') = \sum_{k=1}^{\infty} \sigma_k u_k(t) v_k^*(t') & \xrightarrow{\mathcal{A}_\alpha} & h^{(\alpha)}(t,\tau) & \xrightarrow{\mathcal{F}} & S_H^{(\alpha)}(\tau,\nu) = \sum_{k=1}^{\infty} \sigma_k A_{u_k,v_k}^{(\alpha)}(\tau,\nu), \end{array}$$

where $W_{u_k,v_k}^{(\alpha)}(t,f)$ and $A_{u_k,v_k}^{(\alpha)}(\tau,\nu)$ are the cross–Wigner distribution and cross–ambiguity function of the singular signals, and $\mathcal{F}$ denotes the Fourier transform.

Via the product of an LTV system's maximum time shift and maximum frequency shift ("total spread") we have defined a precised version of the classical *underspread/overspread* classification: Underspread systems feature a total spread much smaller than one and overspread systems larger than one. The generalized Weyl symbol of an underspread system satisfies a number of asymptotic properties that establish a time–frequency–parametrized generalization of any of the well–known properties of an LTI system's transfer function.

We have defined a "short–time transfer function" (STTF) as a practical point estimator of the system's generalized Weyl symbol. This STTF is based on time–frequency shifted versions of a prototype signal applied to the system's input followed by matched filtering applied to the system's output. We have shown that this STTF establishes a unique representation of an underspread LTV system, provided that the prototype is matched to the underspread support by our standard matching rule

$$\frac{T_\gamma}{F_\gamma} = \frac{\tau_0}{\nu_0},$$

where T_γ/F_γ are the duration/bandwidth of the prototype and τ_0/ν_0 are the maximum time/frequency shift of the system. Analogously to the system analysis issue, the short–time Fourier transform allows to *synthesize* underspread systems by putting a multiplicative modification between the analysis and synthesis stage. Moreover, we have pointed out how the presented STFT–filter theory can be used for the design of generalized perfect reconstruction systems.

We have extended multiwindow methods from their classical use in time–invariant spectral estimation to a combined time–varying filtering and spectral estimation. Specifically, we have shown that one and the same set of windows is matched to the spectral estimation and optimum filtering of an underspread process. These theoretical results lead to a structurally simple and highly parallel implementation of a practical nonstationary Wiener filter.

We have reviewed and extended the theory of the stochastic WSSUS class. In particular, we have shown: (i) The WSSUS definition is invariant within the family of the generalized Weyl symbol, (ii) Asking about approximate eigensignals of a WSSUS channel leads to a prototype matching criterion that parallels our STFT–window optimization theory. (iii) An optimum time–frequency tiling for the digital transmission over WSSUS channels can be found analog to matched Gabor grid of the previous chapter.

We emphasize that most of the results of this chapter carry over to the discrete time case when we restrict ourselves to $|\alpha| = 1/2$. In particular, all of the presented STFT results carry over to nonsubsampled DFT filterbank theory and the WSSUS matching theory is essentially applicable for the optimization of DFT–based transmultiplexers. With regard to $\alpha = 0$, it appears to be impossible to reformulate the continuous time theory in a way that maintains the unitarity of the Weyl correspondence. Nevertheless, the specific mathematical properties of the Weyl correspondence make it to the fundamental theoretical tool for time–frequency–parametrized operator representation.

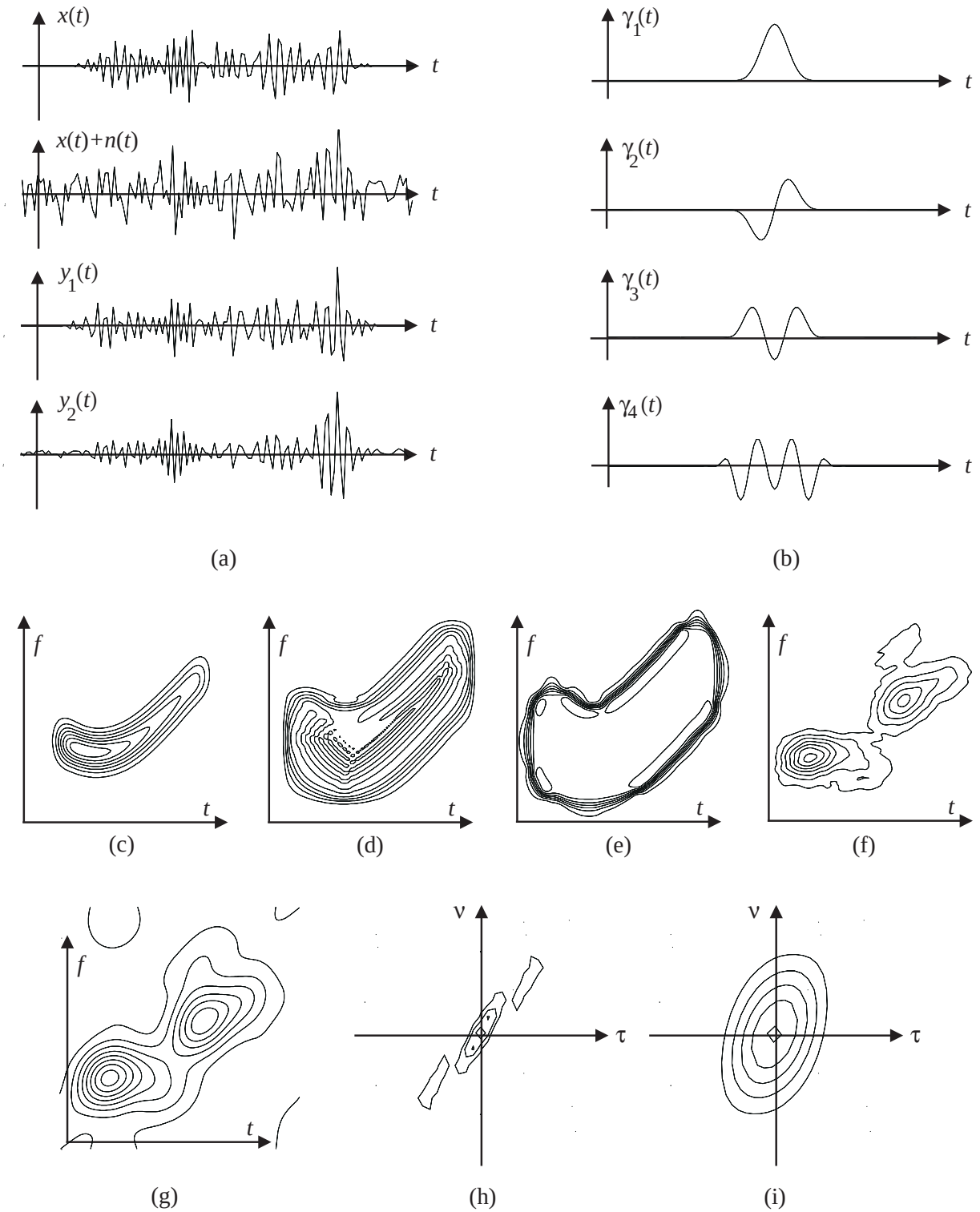

Figure 4.5: Nonstationary Wiener filter via multiplicative modification of multiwindow STFT: (a) $x(t)$—realization of signal process, $x(t)+n(t)$—noisy version, $y_1(t)$–output of theoretical, exact Wiener filter, $y_2(t)$ output of practical, approximate Wiener filter, (b) the matched window set, (c) time–varying spectrum of signal process, (d) time–varying transfer function of exact Wiener filter, (e) time–varying transfer function of theoretical approximation via KN–symbol calculus, (f) multiwindow spectral estimate based on the realization (as shown in (a)), (g) time–varying transfer function of practical, approximate Wiener filter, (h) expected ambiguity function of signal process (i) effective spreading function of multiwindow kernel.

Chapter 5

On Underspread Operators

This chapter is devoted to a theoretical study of Hilbert–Schmidt operators with restricted spreading function. We start reviewing the idea of a symbolic calculus, which in turn leads to the spreading constraint as a prerequisite for the applicability of the generalized Weyl symbol (GWS). It is shown that underspread operators form an approximate commutative algebra. In particular we prove that the product of the generalized GWSs of two operators is approximately equal to the symbol of the product operator (approximate homomorphism). Moreover, we show that (i) underspread operators are approximately normal, *(ii) the GWS of an underspread operator is asymptotically α–invariant, (iii) the GWS and a properly time–frequency localized signal form an approximate eigenpair, and (iv) any underspread operator can be realized by multiplicative modification of the short–time Fourier transform using an appropriately localized analysis/synthesis window.*

Along with the abstract formulation of quantitative results we point out their relevance in signal and system theory.

5.1 On the Concept of a Symbolic Calculus

As already mentioned in the previous chapter the basic idea of a symbolic calculus belongs to common engineers' knowledge as far as linear time–invariant (LTI) operators are concerned. Here, it is a well–known and often used fact that the operator product corresponds to a product of the symbols (transfer functions). Mathematically, one says that LTI operators form a commutative algebra. In the dual case of linear frequency–invariant (LFI) operators (i.e., time–domain multiplication operators), the mathematical structure is identical. Thus, whenever $\mathbf{H}$ and $\mathbf{G}$ are both LTI or both LFI one has a perfect symbol calculus,

$$L_G^{(\alpha)}(t,f)L_H^{(\alpha)}(t,f) = L_{GH}^{(\alpha)}(t,f) = L_{HG}^{(\alpha)}(t,f), \tag{5.1}$$

where $L_H^{(\alpha)}(t,f)$ is the generalized Weyl symbol (GWS).

Clearly, the identity operator is both LTI and LFI, its symbol is time–frequency–invariant equal to one (the neutral element of the field of complex numbers):

$$L_I^{(\alpha)}(t,f) = 1,$$

a fact that together with (5.1) allows to extend the symbolic calculus to the computation of the inverse operator (existence provided, exactly valid for both LTI or LFI operators):

$$L_{H^{-1}}^{(\alpha)}(t,f) = \frac{1}{L_H^{(\alpha)}(t,f)}.$$

Operator composition or inversion appears in widespread applications particularly in statistical signal processing. An important example is minimum mean–squared error (MMSE) filtering [199] as

already discussed in more detail in previous chapters. Given statistically independent signal and noise processes with correlation $\mathbf{R}_x$ and $\mathbf{R}_n$, respectively, the general MMSE filter

$$\mathbf{H}_{MMSE} = \mathbf{R}_x(\mathbf{R}_x + \mathbf{R}_n)^{-1}$$

can be formulated by scalar operations:

$$L^{(\alpha)}_{H_{MMSE}}(t,f) = \frac{L^{(\alpha)}_{R_x}(t,f)}{L^{(\alpha)}_{R_x}(t,f) + L^{(\alpha)}_{R_n}(t,f)}, \tag{5.2}$$

valid whenever the involved operators are either all LTI or all LFI (i.e., the signal and noise processes are either both stationary or both nonstationary white) and the inverse is well–defined[1].

Recall that the LTI and LFI operators are characterized by a severe spreading constraint, i.e., the spreading function of the operator is ideally concentrated on either the τ–axis

$$S^{(\alpha)}_{H_{LTI}}(\tau,\nu) = h(\tau)\delta(\nu),$$

or the ν–axis

$$S^{(\alpha)}_{H_{LFI}}(\tau,\nu) = \delta(\tau)M(\nu),$$

as discussed in Section 4.2.3.

Intuitively, we expect that the symbolic calculus remains approximately valid for operators which are slightly perturbed versions of ideal LTI or LFI operators, i.e., "slowly" time–varying or "slowly" frequency–varying operators. Although this idea is quite old (at least in its implicit consequences), the present work seems to present the first detailed analysis of the conceptual limitation of time–frequency parametrized operator symbols that form the GWS class.

5.1.1 Orthogonal Resolutions of the Identity and Symbolic Calculus

LTI and LFI operators are never Hilbert–Schmidt (HS). However, in case of LTV systems or nonstationary processes HS operators are practically predominant. We now review the well–known discrete symbolic calculus of commuting normal HS operators [252, 311].

As mentioned in the discussion of the Karhunen–Loeve transform (Chapter 3), a normal HS operator $\mathbf{H}$ can be expanded into a weighted sum of orthogonal projection operators, i.e., a perturbed orthogonal resolution of the identity [252]:

$$\mathbf{H} = \sum_{k=1}^{\infty} \lambda_k \mathbf{P}_{u_k} \qquad \text{and} \qquad \mathbf{I} = \sum_{k=1}^{\infty} \mathbf{P}_{u_k}, \tag{5.3}$$

where the eigenvalues can be formally written as a HS operator inner product

$$\lambda_k = \langle \mathbf{H}, \mathbf{P}_{u_k} \rangle . \tag{5.4}$$

Two normal HS operators $\mathbf{H}, \mathbf{G}$ commute if and only if they have a common eigenbasis $\{u_k\}$. One has (using the fact that $\mathbf{P}_{u_k}\mathbf{P}_{u_l} = \mathbf{P}_{u_k}\delta_{kl}$)

$$\mathbf{GH} = \mathbf{HG} = \sum_{k=1}^{\infty}\sum_{l=1}^{\infty} \lambda_{k,G}\lambda_{l,H}\mathbf{P}_{u_k}\mathbf{P}_{u_l} = \sum_{k=1}^{\infty} \lambda_{k,G}\lambda_{k,H}\mathbf{P}_{u_k},$$

where $\lambda_{k,H}, \lambda_{k,G}$ are the eigenvalues of the operators $\mathbf{H}, \mathbf{G}$. Once we consider the eigenbasis $\{u_k\}$ as a priori knowledge, the product operator is uniquely characterized by the multiplication of the eigenvalue distributions [252]:

$$\lambda_{k,GH} = \lambda_{k,HG} = \lambda_{k,G}\lambda_{k,H}. \tag{5.5}$$

[1] Within the framework of a symbolic calculus, the generalized inverse (in the Moore–Penrose sense) can be defined in a straightforward way, by setting $L^{(\alpha)}_{H_{MMSE}}(t,f)$ to zero where the magnitude of the denominator in the left side of (5.2) is below some threshold.

In our context it is appropriate to call these eigenvalue distributions discrete *symbols* of the operator and (5.5) is then referred to as *perfect symbol calculus*. Clearly, the identity is the neutral element of any such commutative operator algebra, one has

$$\lambda_{k,I} = 1,$$

and the generalized operator inversion can be essentially written as an algebraic inversion in the field of complex numbers:

$$\lambda_{k,H^{-1}} = \begin{cases} \frac{1}{\lambda_{k,H}}, & \text{for} \quad \lambda_{k,H} \geq \epsilon, \\ \\ 0, & \text{for} \quad \lambda_{k,H} < \epsilon. \end{cases}$$

5.1.2 Nonorthogonal Weyl–Heisenberg Resolutions of the Identity

STFT–based Resolutions of the Identity. Recall that the reproducing formula underlying STFT synthesis can be written as (for details see Section 4.5)

$$\mathbf{I} = \int_t \int_f \mathbf{P}_\gamma^{(t,f)} dt\, df, \tag{5.6}$$

where $\mathbf{P}_\gamma$ is the rank–one projection onto the window and the superscript denotes time–frequency shifting as has been introduced in Section 2.3.1:

$$\mathbf{P}_\gamma^{(t,f)} = \mathbf{S}^{(t,f)} \mathbf{P}_\gamma \mathbf{S}^{(t,f)*},$$

where $\mathbf{S}^{(t,f)}$ is a time–frequency shift operator. When the underlying window is well time–frequency localized, then $\mathbf{P}_\gamma^{(t,f)}$ is time–frequency selective about (t, f).

STFT based system design is based on a multiplicative perturbation of (5.6):

$$\mathbf{H} = \int_t \int_f M(t,f)\, \mathbf{P}_\gamma^{(t,f)} dt\, df.$$

Idealized Interpretation. The transfer function of an LTI system and the power spectrum of a stationary process are *continuous eigenvalue distributions* of linear time–invariant operators (mathematically precise one should speak about an approximate point spectrum, because the point spectrum is always discrete). The common idealized interpretation of Weyl–Heisenberg operator representations (as e.g. the generalized Weyl symbol) may be interpreted as a *time–frequency parametrized eigenvalue distribution* $\Lambda(t, f)$ of a linear time–varying operator in the sense of

$$\mathbf{H} \stackrel{?}{=} \int_t \int_f \Lambda(t,f)\, \mathbf{P}^{(t,f)} dt\, df.$$

This interpretation is in obvious conflict with Heisenberg's uncertainty since $\mathbf{P}^{(t,f)}$ should be an ideally time–frequency selective localization operator. In particular one would have to require orthogonality of the time–frequency shifted versions of the prototype operator:

$$\mathbf{P}^{(t,f)} \mathbf{P}^{(t',f')} \stackrel{?}{=} \begin{cases} \mathbf{P}^{(t,f)}, & (t = t') \wedge (f = f'), \\ \\ 0, & (t \neq t') \vee (f \neq f'). \end{cases} \tag{5.7}$$

However, such orthogonality is generally impossible. It is particularly unrealistic when $\mathbf{P}$ is rank–one, i.e., $\mathbf{P} = \mathbf{P}_\gamma$ since the STFT expansion set $\gamma^{(t,f)}$ is highly linear dependent. But a rank–one

projection onto a well time–frequency localized signal is obviously the optimally concentrated time–frequency localization operator.

Discrete Expansion. In this work we usually consider LTV operators which are of Hilbert–Schmidt (HS) type. HS operators always have a *discrete* eigenvalue spectrum (see Section 5.1.1). This is another fundamental contradiction with the idea of a *continuous* time–frequency parametrized eigenvalue spectrum. From this point of view one may already expect that time–frequency discretization should come up naturally whenever a time—frequency parametrized eigenvalue spectrum makes sense.

A discrete Weyl–Heisenberg resolution of the identity is given by

$$\mathbf{I} = \sum_l \sum_m \mathbf{P}^{(lT,mF)}.$$

Mutual orthogonality of the (l,m)–indexed projection operators $\mathbf{P}^{(lT,mF)}$,

$$\mathbf{P}^{(lT,mF)}\mathbf{P}^{(l'T,m'F)} = \mathbf{P}^{(lT,mF)}\delta_{l,l'}\delta_{m,m'}$$

is now theoretically possible such that there exist operators which are exactly diagonalized by a discrete Weyl–Heisenberg basis:

$$\mathbf{H} = \sum_l \sum_m \Lambda(l,m)\mathbf{P}^{(lT,mF)},$$

where $\Lambda(l,m)$ is the hypothetical eigenvalue distribution (point spectrum) of $\mathbf{H}$.

However, for the rank–one case $\mathbf{P} = \mathbf{P}_\gamma$ it is well–known that orthonormal Weyl–Heisenberg bases are difficult to obtain and lead always to bad time–frequency localized prototype signals γ [184, 28]. Hence, the existence of a diagonalizing Weyl–Heisenberg basis $\gamma^{(lT,mF)}$ is a far too restrictive condition for the classification of operators that allow an approximate Weyl–Heisenberg symbol calculus.

5.2 Operators with Restricted Spreading

A general LTV operator can be formally written as a continuous Weyl–Heisenberg expansion in terms of the generalized Weyl symbol $L_H^{(\alpha)}(t,f)$,

$$\mathbf{H} = \int_t \int_f L_H^{(\alpha)}(t,f)\,\mathbf{P}^{(t,f)}(\alpha)\,dt\,df, \tag{5.8}$$

where the infinitesimal prototype operator $\mathbf{P}(\alpha)$ is defined by

$$S_{P(\alpha)}^{(\alpha)}(\tau,\nu) \equiv 1 \qquad \Leftrightarrow \qquad L_{P(\alpha)}^{(\alpha)}(t,f) = \delta(t)\delta(f),$$

so that, formally :

$$L_H^{(\alpha)}(t,f) = \left\langle \mathbf{H}, \mathbf{P}^{(t,f)}(\alpha) \right\rangle. \tag{5.9}$$

The prototype operator characterizes the properties and interpretation of the GWS. Recall that the eigenvalues of a normal HS operator can be written as (see (5.4))

$$\lambda_k = \langle \mathbf{H}, \mathbf{P}_{u_k} \rangle,$$

where $\mathbf{P}_{u_k}$ is the rank–one projection onto the window. Hence, the interpretation of the GWS as a time–frequency parametrized eigenvalue distribution would require that $\mathbf{P}(\alpha)$ be a time–frequency selective projection operator. But actually $\mathbf{P}(\alpha)$ is far from an orthogonal projection operator. Specifically for $\alpha = 0$, it is even unitary up to a constant factor 2 (for more details, see the discussion in Section 2.3.2). Hence, we have the fact that for a general LTV operator $\mathbf{H}$ the prototype operator

of the continuous Weyl–Heisenberg expansion does not at all support the idea of a time–frequency parametrized eigenvalue distribution. However, under the assumption of restricted spreading one can set up a more adequate version of the continuous Weyl–Heisenberg expansion as follows.

Many practically important LTV operators satisfy a certain spreading constraint, in so far as the spreading function $S_H^{(\alpha)}(\tau,\nu)$ is restricted to a domain centered about the origin. We characterize such a spreading constraint via an (0/1–valued) indicator function $\chi_H(\tau,\nu)$ with minimum area such that:

$$S_H^{(\alpha)}(\tau,\nu) = S_H^{(\alpha)}(\tau,\nu)\chi_H(\tau,\nu). \tag{5.10}$$

Note that this equation is α–invariant since the magnitude of the generalized spreading function does not depend on α.

For such operators, the multiplication of (5.10) carries over to an idempotent convolution of the generalized Weyl symbol which we write in terms of the GWS of an α–dependent prototype operator $\mathbf{P}(\alpha)$:

$$L_H^{(\alpha)}(t,f) = L_H^{(\alpha)}(t,f) * *L_{P(\alpha)}^{(\alpha)}(t,f) = \int_{t'}\int_{f'} L_H^{(\alpha)}(t',f')L_{P(\alpha)}^{(\alpha)}(t-t',f-f')dt'\,df'. \tag{5.11}$$

This means that the spreading function of $\mathbf{P}(\alpha)$ is just the indicator function

$$S_{P(\alpha)}^{(\alpha)}(\tau,\nu) = \chi_H(\tau,\nu). \tag{5.12}$$

Using the time–frequency shift–covariance of the GWS (C.15) we have

$$L_{P(\alpha)}^{(\alpha)}(t-t',f-f') = L_{P^{(t',f')}(\alpha)}^{(\alpha)}(t,f).$$

The idempotent convolution (5.11) thus leads to an alternative continuous Weyl–Heisenberg operator decomposition:

$$\mathbf{H} = \int_t\int_f L_H^{(\alpha)}(t,f)\,\mathbf{P}^{(t,f)}(\alpha)\,dt\,df, \tag{5.13}$$

structurally equivalent to (5.8) but now based on a Hilbert–Schmidt prototype operator. Such a prototype operator is still no projection operator but — in view of (5.12) — it cannot introduce arbitrarily large time–frequency shifts. Hence, it is basically time–frequency selective and the dominant singular signals are typically well time–frequency concentrated about the origin of the time–frequency plane. With increasing spread of $\chi_H(\tau,\nu)$ the prototype operator $\mathbf{P}(\alpha)$ gets increasingly α–invariant, positive and its essential range gets larger.

According to the alternative Weyl–Heisenberg expansion (5.13) we can redefine the generalized Weyl symbol for any operator satisfying the spreading constraint in terms of the associated prototype operator:

$$L_H^{(\alpha)}(t,f) = \left\langle \mathbf{H}, \mathbf{P}^{(t,f)}(\alpha) \right\rangle.$$

Clearly, the continuous Weyl–Heisenberg decomposition (5.13) is highly ambiguous since the choice of any alternative indicator function $\tilde{\chi}(\tau,\nu)$ that contains the domain of $\chi_H(\tau,\nu)$,

$$\chi_H(\tau,\nu) = \tilde{\chi}(\tau,\nu)\chi_H(\tau,\nu), \tag{5.14}$$

leads to a different prototype operator $\widetilde{\mathbf{P}}(\alpha)$ according to

$$S_{\tilde{P}(\alpha)}^{(\alpha)}(\tau,\nu) = \tilde{\chi}(\tau,\nu).$$

It is intuitively appealing to take the prototype operator with optimally concentrated spreading function but we now show that there is a more concrete reasoning for adopting this operator.

Recall that the integral over the magnitude squared spreading function equals the Hilbert–Schmidt norm of the operator:

$$\|\widetilde{\mathbf{P}}(\alpha)\|^2 = \int_\tau \int_\nu \tilde{\chi}(\tau,\nu) d\tau \, d\nu.$$

This shows that among all indicator functions satisfying (5.14) the indicator function $\chi_H(\tau,\nu)$ has minimum area and corresponds to the characteristic operator with minimum Hilbert–Schmidt norm. Hence, one may define the prototype operator with spreading function $\chi_H(\tau,\nu)$ by the following optimization principle:

$$\mathbf{P} = \arg\min_{\widetilde{\mathbf{P}}} \|\widetilde{\mathbf{P}}\|^2 \qquad \text{subject to} \qquad \mathbf{H} = \int_t \int_f L_H^{(\alpha)}(t,f) \widetilde{\mathbf{P}}^{(t,f)} dt \, df. \tag{5.15}$$

The formal setup of such a *minimum–norm Weyl–Heisenberg expansion* provides an abstract "short–cut" derivation for minimum–variance unbiased (MVUB) time–varying auto– and cross spectrum estimators (the details are discussed in the Appendices D and E):

- Given a noisy observation of a circular complex, zero–mean, nonstationary Gaussian process $x(t)$ with known spreading constraint $\chi_x(\tau,\nu)$ (with regard to $EA_x(\tau,\nu)$):

  $$y(t) = x(t) + n(t),$$

  where $n(t)$ is circular complex, zero–mean, stationary white Gaussian noise with variance σ_n^2, one can show that the MVUB estimator of the process' Wigner–Ville spectrum is given by

  $$\widehat{EW}_x(t,f) = \left\langle \widehat{\mathbf{P}}^{(t,f)} y, y \right\rangle,$$

  where $\widehat{\mathbf{P}}$ is the prototype operator of the minimum–norm Weyl–Heisenberg expansion of the process' correlation operator $\mathbf{R}_x$ according to the spreading constraint $\chi_x(\tau,\nu)$.

- Given an observation of a noise–free, circular complex, stationary white, zero–mean, unit–variance Gaussian input process and noisy output process,

  $$y(t) = (\mathbf{H}x)(t) + n(t),$$

  where $\mathbf{H}$ is a linear time–varying system with known spreading constraint $\chi_H(\tau,\nu)$ and $n(t)$ is zero–mean, circular complex, stationary white Gaussian noise, one can show that the MVUB estimator of the system's Weyl symbol is given by

  $$\widehat{L}_H(t,f) = \left\langle y, \widehat{\mathbf{P}}^{(t,f)} x \right\rangle,$$

  where $\widehat{\mathbf{P}}$ is the prototype operator of the minimum–norm Weyl–Heisenberg expansion of the system $\mathbf{H}$ according to the spreading constraint $\chi_H(\tau,\nu)$.

5.3 Rectangular Constraint

Now and for the rest of this chapter we consider a rectangular spreading constraint defined as[2]:

$$S_H^{(\alpha)}(\tau,\nu) = S_H^{(\alpha)}(\tau,\nu) \, \chi_{[-\tau_0,\tau_0]}(\tau) \, \chi_{[-\nu_0,\nu_0]}(\nu). \tag{5.16}$$

[2]Such a spreading constraint basically corresponds to a non–causal LTV system. Causal systems can, however, easily be incorporated into this framework by extracting an ideal time–delay system such that the spreading function of the remaining subsystem lies inside the centered rectangle.

Note that apart from the analytical feasibility of quantitative studies such a spreading constraint corresponds to the practical setup of minimum a priori knowledge about a nonstationary environment. Take for example the mobile radio channel [269] where ν_0 is determined by the maximum admissible velocity of the radio stations while τ_0 comes out of the pragmatic modeling necessity of using Viterbi equalizers with severely constrained tap number.

According to (5.16), the generalized Weyl symbol is a two–dimensional low–pass function which is uniquely determined by its samples on a rectangular time–frequency grid with constants $T \leq \frac{1}{2\nu_0}$ and $F \leq \frac{1}{2\tau_0}$, respectively. We have the following reconstruction equation:

$$L_H^{(\alpha)}(t,f) = TF \sum_l \sum_m L_H^{(\alpha)}(lT,mF)\, L_{P(\alpha)}^{(\alpha)}(t-lT, f-mF), \tag{5.17}$$

where the generalized Weyl symbol of the prototype operator is given by:

$$L_{P(\alpha)}^{(\alpha)}(t,f) = \frac{\sin(2\pi\nu_0 t)\sin(2\pi\tau_0 f)}{\pi^2 tf}. \tag{5.18}$$

Equation (5.17) corresponds to a *discrete Weyl–Heisenberg expansion* of the operator $\mathbf{H}$:

$$\mathbf{H} = TF \sum_l \sum_m L_H^{(\alpha)}(lT,mF)\, \mathbf{P}^{(lT,mF)}(\alpha). \tag{5.19}$$

As pointed out in Section 5.1.2 such a discrete Weyl–Heisenberg expansion comes conceptually closest to the idea of a time–frequency parametrized perturbed resolution of the identity. It must be emphasized that the eigenvalue interpretation of the GWS samples implies the assumption of an approximate multiplicity TF. This becomes obvious in the discrete trace formula, for a normal HS operator one has:

$$\sum_l \sum_m L_H^{(\alpha)}(lT,mF) = \frac{1}{TF} \sum_{k=1}^{\infty} \lambda_k.$$

However, such an eigenvalue counting argument starts to make sense for

$$TF \geq 1,$$

or, equivalently,

$$\nu_0 \tau_0 \leq 1/4.$$

This threshold corresponds exactly to what we have introduced as *critical spread*. Hence, asymptotic eigenvalue counting (in the spirit of [367, 106]) gives both an operator theoretic reasoning for the underspread/overspread classification but also an intuitive indication of the critical Gabor density (critical phase space density) as discussed in [28, 43, 173, 292].

Oversampling, i.e. , $T > \frac{1}{2\nu_0}$ and $F > \frac{1}{2\tau_0}$ leaves a certain freedom in the choice of $\mathbf{P}(\alpha)$. The spreading function of the prototype operator has to satisfy the anti–aliasing condition:

$$S_{P(\alpha)}^{(\alpha)}(\tau,\nu) \sum_k \sum_m S_H^{(\alpha)}\left(\tau - \frac{k}{F}, \nu - \frac{m}{T}\right) = S_H^{(\alpha)}(\tau,\nu).$$

This requirement is more severe than its counterpart in the continuous Weyl–Heisenberg expansion (5.14). However, the freedom in the choice of the prototype operator may be helpful when it comes to the realization of underspread operators via multiwindow methods as discussed in the previous chapters.

5.4 Properties of Underspread Operators

So far our discussion of underspread operators was a manycoloured collection of intuitive arguments with the main conclusion that for operators with constrained spreading the idea of a time–frequency parametrized symbolic calculus *may* make sense. Following is a list of concrete quantitative results that prove this conjecture. Moreover, we discuss further interesting properties of underspread operators.

5.4.1 Jointly Underspread Operators

In order to formulate a theorem on the approximate symbol calculus we need to formalize the subspace of underspread operators that satisfy the same spreading constraint.

Definition. *Two operators* $\mathbf{H}$ *and* $\mathbf{G}$ *are jointly underspread with spreading constants* τ_0 *and* ν_0 *when their spreading functions satisfy:*

$$S_H^{(\alpha)}(\tau,\nu) = S_H^{(\alpha)}(\tau,\nu)\,\chi_{[-\tau_0,\tau_0]}(\tau)\,\chi_{[-\nu_0,\nu_0]}(\nu). \tag{5.20}$$

Equivalently, one can say that two underspread operators $\mathbf{G}$ and $\mathbf{H}$ are jointly underspread whenever any linear combination $\alpha\mathbf{G}+\beta\mathbf{H}$ is underspread. Jointly underspread HS operators form a subspace of the Hilbert space of Hilbert–Schmidt operators with the reproducing formula:

$$\mathbf{H} = \int_t \int_f \left\langle \mathbf{H}, \mathbf{P}^{(t,f)}(\alpha) \right\rangle \mathbf{P}^{(t,f)}(\alpha)\,dt\,df, \tag{5.21}$$

where $\mathbf{P}(\alpha)$ is the jointly underspread prototype operator given by (5.18). The orthogonal projection of a general HS operator $\mathbf{H}$ onto such a *reproducing kernel Hilbert space* can be formulated as a simple multiplication in the spreading domain:

$$S_{H_P}^{(\alpha)}(\tau,\nu) = S_H^{(\alpha)}(\tau,\nu)\,\chi_{[-\tau_0,\tau_0]}(\tau)\,\chi_{[-\nu_0,\nu_0]}(\nu),$$

where $\mathbf{H}_P$ is the projected version of $\mathbf{H}$.

In this chapter, we identify underspread operators as operators with restricted spreading (with no general requirement on τ_0,ν_0). In any of the following theorems we will explicitly define the maximum total spread $\tau_0\nu_0$ for which the result is valid. However, all of these bounds start to make practical sense for underspread operators in the sense of the previous chapters, defined by the asymptotic requirement $\tau_0\nu_0 \ll 1$.

5.4.2 Approximate Multiplicative Symbol Calculus

In order to investigate the possibility of replacing the operator product by the product of the symbols we need the spreading function of a product of two HS operators:

Proposition 1.1 *The generalized spreading function of the composition of two Hilbert–Schmidt operators with sufficient decay of their spreading function is given by:*

$$S_{GH}^{(\alpha)}(\tau,\nu) = \int_{\tau'}\int_{\nu'} S_G^{(\alpha)}(\tau',\nu')\,S_H^{(\alpha)}(\tau-\tau',\nu-\nu')\,e^{-j2\pi\{\tau'\nu(\alpha+1/2)+\tau\nu'(\alpha-1/2)-2\tau'\nu'\alpha\}}d\tau'\,d\nu'. \tag{5.22}$$

We shall refer to this expression as *twisted convolution* which is the mathematical terminology for the case $\alpha=0$ [125].

Proof: Based on the inversion formula for the generalized spreading function,

$$h(t,s) = \int_\nu S_H^{(\alpha)}(t-s,\nu)e^{j2\pi\nu[(1/2+\alpha)t+(1/2-\alpha)s]}d\nu,$$

one can write the kernel of the composite operator $\mathbf{GH}$ in terms of the spreading functions of the operators:

$$\begin{aligned}(\mathbf{GH})(t,s) &= \int_r (\mathbf{G})(t,r)\,(\mathbf{H})(r,s)dr \\ &= \int_r\int_{\nu_1}\int_{\nu_2} S_G^{(\alpha)}(t-r,\nu_1)S_H^{(\alpha)}(r-s,\nu_2) \\ &\quad \cdot e^{j2\pi\{\nu_1[(1/2+\alpha)t+(1/2-\alpha)r]+\nu_2[(1/2+\alpha)r+(1/2-\alpha)s]\}}dr\,d\nu_1\,d\nu_2.\end{aligned}$$

Then, using the definition of the generalized spreading function,

$$S_H^{(\alpha)}(\tau,\nu) \stackrel{\text{def}}{=} \int_t h\left(t+\left(\frac{1}{2}-\alpha\right)\tau, t-\left(\frac{1}{2}+\alpha\right)\tau\right) e^{-j2\pi\nu t}dt,$$

the spreading function of **HG** is given by a four–dimensional integral

$$S_{GH}^{(\alpha)}(\tau,\nu) = \int_t \int_r \int_{\nu_1} \int_{\nu_2} S_G^{(\alpha)}(t+(1/2-\alpha)\tau - r, \nu_1)\, S_H^{(\alpha)}(r-t+(1/2+\alpha)\tau, \nu_2)$$

$$\cdot e^{j2\pi\{\nu_1[(1/2+\alpha)(t+(1/2-\alpha)\tau)+(1/2-\alpha)r]+\nu_2[(1/2+\alpha)r+(1/2-\alpha)(t-(1/2+\alpha)\tau)]\}}\, e^{-j2\pi\nu t}dt\, dr\, d\nu_1\, d\nu_2.$$

Since we have required sufficient decay, the integrand is in L_1 and we can interchange the order of integration. Using the substitution $t \to t_1 + t_2/2, r \to t_1 - t_2/2$, two integrals collapse by inverse Fourier transform and by setting $\tau' = t_2 + (1/2-\alpha)\tau, \nu' = \nu_1$ one arrives at the final result (5.22). ∎

The twisted convolution (5.22) is not far from a conventional convolution which would correspond to the perfect symbol calculus by virtue of the Fourier duality between $L_H(t,f)$ and $S_H(\tau,\nu)$ (see (B.14)):

$$L_{GH} \stackrel{?}{=} L_G L_H \qquad \Leftrightarrow \qquad S_{GH} \stackrel{?}{=} S_G * * S_H.$$

Closer inspection of the disturbing complex exponential factor in (5.22) shows that for well–concentrated spreading functions the twisted convolution approaches a conventional convolution. This thought can be made precise by the following theorem [208]:

Theorem 5.1 *The generalized Weyl symbol of two jointly underspread Hilbert–Schmidt operators* **H** *and* **G** *with* $\tau_0\nu_0 < 1/2$ *satisfies:*

$$\left|L_{GH}^{(\alpha)} - L_G^{(\alpha)} L_H^{(\alpha)}\right| \;<\; 2\sin\left(\pi\tau_0\nu_0(1+2|\alpha|)\right) \|S_G\|_1 \|S_H\|_1,$$

$$\left\|L_{GH}^{(\alpha)} - L_H^{(\alpha)} L_G^{(\alpha)}\right\|^2 \;<\; 64\,\tau_0\nu_0 \sin^2\left(\pi\tau_0\nu_0(1+2|\alpha|)\right) \|\mathbf{G}\|^2 \|\mathbf{H}\|^2,$$

where $\|S_H\|_1$ *denotes the* L_1*–norm of the spreading function*

$$\|S_H\|_1 \stackrel{\text{def}}{=} \int_\tau \int_\nu |S_H(\tau,\nu)|\, d\tau\, d\nu,$$

and $\|\mathbf{H}\|^2$ *denotes the Hilbert–Schmidt norm of the operator* **H**.

Proof: Starting from (5.22) we use the integral triangle inequality and estimate the integrand as follows

$$\begin{aligned}
&\left|L_{GH}^{(\alpha)} - L_G^{(\alpha)} L_H^{(\alpha)}\right| \\
&= \left| \int_\tau \int_\nu \left\{ S_{GH}^{(\alpha)}(\tau,\nu) - \left(S_G^{(\alpha)} * * S_H^{(\alpha)}\right)(\tau,\nu) \right\} e^{j2\pi(t\nu - f\tau)} d\tau\, d\nu \right| \\
&= \left| \int_\tau \int_\nu \int_{\tau'} \int_{\nu'} S_G^{(\alpha)}(\tau',\nu') S_H^{(\alpha)}(\tau-\tau',\nu-\nu') \right. \\
&\quad \left. \cdot \left(e^{j2\pi[\tau\nu'(1/2-\alpha) - \tau'\nu(1/2+\alpha) + 2\tau'\nu'\alpha]} - 1\right) e^{j2\pi(t\nu - f\tau)} d\tau\, d\nu d\tau'\, d\nu' \right| \\
&\leq 2 \int_{-\tau_0}^{\tau_0} \int_{-\nu_0}^{\nu_0} \int_{-\tau_0}^{\tau_0} \int_{-\nu_0}^{\nu_0} \left|S_G^{(\alpha)}(\tau_1,\nu_1)\right| \left|S_H^{(\alpha)}(\tau_2,\nu_2)\right| \\
&\quad \cdot \sin\left(\pi[\tau_2\nu_1(1/2-\alpha) - \tau_1\nu_2(1/2+\alpha) + 2\tau_1\nu_1\alpha]\right) d\tau_1\, d\nu_1\, d\tau_2\, d\nu_2 \\
&< \max_{|\tau_i|<\tau_0, |\nu_i|<\nu_0} \left\{\sin\left(\pi[\tau_2\nu_1(1/2-\alpha) - \tau_1\nu_2(1/2+\alpha) + 2\tau_1\nu_1\alpha]\right)\right\} \\
&\quad \cdot \int_{-\tau_0}^{\tau_0} \int_{-\nu_0}^{\nu_0} \int_{-\tau_0}^{\tau_0} \int_{-\nu_0}^{\nu_0} \left|S_G^{(\alpha)}(\tau_1,\nu_1)\right| \left|S_H^{(\alpha)}(\tau_2,\nu_2)\right| d\tau_1\, d\nu_1\, d\tau_2\, d\nu_2 \\
&= 2\sin\left(\pi\tau_0\nu_0(1+2|\alpha|)\right) \|S_H\|_1 \|S_G\|_1.
\end{aligned}$$

In the last step we have used the fact that

$$\max\left\{\sin\pi[\tau\nu'(1/2-\alpha) - \tau'\nu(1/2+\alpha) + 2\tau'\nu'\alpha)]\right\} = \sin(\pi\tau_0\nu_0(1+2|\alpha|))$$
$$\text{for} \qquad |\tau| \leq \tau_0, \quad |\nu| \leq \nu_0, \quad |\alpha| \leq 1/2, \quad \tau_0\nu_0 \leq 1/2.$$

which is easy to check.

For asymptotic analysis it is appropriate to make this bound slightly coarser. For trace–class underspread operators one has the fact that

$$\|S_H\|_1 < 4\tau_0\nu_0 \mathrm{tr}\sqrt{\mathbf{H}^*\mathbf{H}},$$

since

$$\left|S_H^{(\alpha)}(\tau,\nu)\right| \leq \sum_{k=1}^{\infty} \sigma_k = \mathrm{tr}\sqrt{\mathbf{H}^*\mathbf{H}},$$

as discussed in Appendix B, (B.22). Furthermore using $\sin x < x$, one has

$$\left|L_{GH}^{(\alpha)} - L_H^{(\alpha)} L_G^{(\alpha)}\right| < 32\pi\tau_0^3\nu_0^3(1+2|\alpha|)\mathrm{tr}\sqrt{\mathbf{G}^*\mathbf{G}}\mathrm{tr}\sqrt{\mathbf{H}^*\mathbf{H}}.$$

Again starting with (5.22) we use the Schwarz inequality and based on (B.19):

$$\int_\tau \int_\nu |S_H(\tau,\nu)|^2 \, d\tau\, d\nu = \|\mathbf{H}\|^2,$$

we obtain the L_2–bound as follows

$$
\begin{aligned}
&\left\| L_{GH}^{(\alpha)} - L_G^{(\alpha)} L_H^{(\alpha)} \right\|^2 \\
&= \left\| S_{GH}^{(\alpha)} - S_G^{(\alpha)} * * S_H^{(\alpha)} \right\|^2 \\
&= \int_{-2\tau_0}^{2\tau_0} \int_{-2\nu_0}^{2\nu_0} \left| \int_{\tau'} \int_{\nu'} S_G^{(\alpha)}(\tau',\nu') S_H^{(\alpha)}(\tau-\tau',\nu-\nu') \right. \\
&\quad \left. \cdot \left(e^{-j2\pi[\tau\nu'(1/2-\alpha)-\tau'\nu(1/2+\alpha)+2\tau'\nu'\alpha)]} - 1 \right) d\tau'\, d\nu' \right|^2 d\tau\, d\nu \\
&\leq \int_{-2\tau_0}^{2\tau_0} \int_{-2\nu_0}^{2\nu_0} \left(\int_{\tau_1} \int_{\nu_1} |S_G^{(\alpha)}(\tau_1,\nu_1)|^2 d\tau_1\, d\nu_1 \right) \\
&\quad \cdot \left(4 \int_{\tau_2} \int_{\nu_2} |S_H^{(\alpha)}(\tau-\tau_2,\nu-\nu_2)|^2 \sin^2\left(\pi[\tau\nu_2(1/2-\alpha) - \tau_2\nu(1/2+\alpha) + 2\tau_2\nu_2\alpha]\right) d\tau_2\, d\nu_2 \right) d\tau\, d\nu \\
&\leq \int_{-2\tau_0}^{2\tau_0} \int_{-2\nu_0}^{2\nu_0} \left(\int_{\tau_1} \int_{\nu_1} |S_G^{(\alpha)}(\tau_1,\nu_1)|^2 d\tau_1\, d\nu_1 \right) \\
&\quad \cdot \left(4 \int_{\tau_2} \int_{\nu_2} |S_H^{(\alpha)}(\tau_2,\nu_2)|^2 \sin^2\left(\pi[\tau_2\nu(1/2-\alpha) - \tau\nu_2(1/2+\alpha) + 2\tau_2\nu_2\alpha]\right) d\tau_2\, d\nu_2 \right) d\tau\, d\nu \\
&< 64\,\tau_0\nu_0 \|\mathbf{G}\|^2 \max_{|\tau_i|<\tau_0, |\nu_i|<\nu_0,\ i\in\{1,2\}} \left\{ \sin^2\left(\pi[\tau_2\nu_1(1/2-\alpha) - \tau_1\nu_2(1/2+\alpha) + 2\tau_2\nu_2\alpha]\right) \right\} \\
&\quad \cdot \left(\int_{\tau_2} \int_{\nu_2} |S_H^{(\alpha)}(\tau_2,\nu_2)|^2 d\tau_2\, d\nu_2 \right) \\
&= 64\,\tau_0\nu_0 \sin^2\left(\pi\tau_0\nu_0(1+2|\alpha|)\right) \|\mathbf{G}\|^2 \|\mathbf{H}\|^2.
\end{aligned}
$$

∎

5.4.3 Approximate Commutativity

Theorem (5.1) suggests that two underspread operators commute in an approximate sense. A bound can be immediately obtained by

$$
\begin{aligned}
\|\mathbf{GH} - \mathbf{HG}\| &= \left\| L_{GH}^{(0)} - L_{HG}^{(0)} \right\| \\
&\leq \left\| L_{GH}^{(0)} - L_H^{(0)} L_G^{(0)} \right\| + \left\| L_H^{(0)} L_G^{(0)} - L_{HG}^{(0)} \right\| \\
&\leq 16\, \|\mathbf{G}\| \|\mathbf{H}\| \sqrt{\tau_0\nu_0} \sin \pi\tau_0\nu_0.
\end{aligned}
$$

However, starting directly with (5.22) allows to obtain a slightly sharper bound, expressed by the following theorem:

Theorem 5.2 *The Hilbert–Schmidt norm of the commutator of two jointly underspread operators* $\mathbf{H}$ *and* $\mathbf{G}$ *with* $\tau_0\nu_0 < 1/4$ *is bounded according to*

$$
\|\mathbf{GH} - \mathbf{HG}\|^2 < 64\tau_0\nu_0 \sin^2\left(2\pi\tau_0\nu_0\right) \|\mathbf{G}\|^2 \|\mathbf{H}\|^2.
$$

Proof: For the case $\alpha = 0$ (5.22) gives:

$$
S_{GH}^{(0)}(\tau,\nu) = \int_{\tau'} \int_{\nu'} S_G^{(0)}(\tau',\nu') S_H^{(0)}(\tau-\tau',\nu-\nu') e^{-j\pi\{\tau'\nu+\tau\nu'\}} d\tau'\, d\nu',
$$

on the other hand one has:

$$S^{(0)}_{HG}(\tau,\nu) = \int_{\tau'}\int_{\nu'} S^{(0)}_{G}(\tau',\nu') S^{(0)}_{H}(\tau-\tau',\nu-\nu') e^{j\pi\{\tau'\nu+\tau\nu'\}} d\tau'\, d\nu'.$$

We use these two formulas to write the HS–norm of the commutator in terms of the spreading functions of the two operators which in turn gives the desired bound:

$$\begin{aligned}
&\|\mathbf{GH}-\mathbf{HG}\|^2 \\
&= \left\| S^{(0)}_{GH} - S^{(0)}_{HG} \right\|^2 \\
&\leq \int_{\tau}\int_{\nu} \left| \int_{\tau'}\int_{\nu'} S^{(0)}_{G}(\tau',\nu') S^{(0)}_{H}(\tau-\tau',\nu-\nu') \left(e^{-j\pi(\tau\nu'-\tau'\nu)} - e^{j\pi(\tau\nu'-\tau'\nu)} \right) \right|^2 d\tau\, d\nu\, d\tau'\, d\nu' \\
&\leq \int_{-2\tau_0}^{2\tau_0}\int_{-2\nu_0}^{2\nu_0} \left(\int_{\tau_1}\int_{\nu_1} |S^{(\alpha)}_{G}(\tau_1,\nu_1)|^2 d\tau_1\, d\nu_1 \right) \\
&\quad \cdot \left(4 \int_{\tau_2}\int_{\nu_2} |S^{(\alpha)}_{H}(\tau_2,\nu_2)|^2 \sin^2\left(\pi[\tau\nu_2-\tau_2\nu]\right) d\tau_2\, d\nu_2 \right) d\tau\, d\nu \\
&< 64\, \tau_0\nu_0 \sin^2(2\pi\tau_0\nu_0)\, \|\mathbf{G}\|^2 \|\mathbf{H}\|^2.
\end{aligned}$$

∎

5.4.4 Approximate Normality

Underspread operators are approximately normal, a fact that follows immediately from Theorem 5.2 considering the commutation of the underspread operator $\mathbf{H}$ and its adjoint $\mathbf{H}^*$. For an underspread operator with $\tau_0\nu_0 < 1/4$ one has the following bound:

$$\|\mathbf{H}\mathbf{H}^* - \mathbf{H}^*\mathbf{H}\|^2 \leq 64\, \tau_0\nu_0 \sin^2(2\pi\tau_0\nu_0)\, \|\mathbf{H}\|^4.$$

In practice, (approximate) normality means that one can use eigendecomposition based methods instead of singular–value decomposition based methods. Typically, this halves the numerical expense as the singular–value decomposition requires the solution of two eigenproblems.

5.4.5 Realization via Multiplicative STFT Modification

The realization of linear time–varying operators via multiplicative modification of the short–time Fourier transform (STFT) has been discussed in the foregoing chapter. There, we have seen that the GWS of a system based on multiplicative modification of the STFT is given by

$$L^{(\alpha)}_{H_{STFT}}(t,f) = M(t,f) **_{} W^{(\alpha)}_{\gamma}(t,f),$$

where $M(t,f)$ is the multiplier function and $W^{(\alpha)}_{\gamma}(t,f)$ is the generalized Wigner distribution of the analysis/synthesis window. The question whether a given operator can be realized by STFT modification can be studied by the generalized spreading function of the STFT–based system. It is given by

$$S^{(\alpha)}_{H_{STFT}}(\tau,\nu) = m(\tau,\nu)\, A^{(\alpha)}_{\gamma}(\tau,\nu), \tag{5.23}$$

where

$$m(\tau,\nu) \stackrel{\text{def}}{=} \mathcal{F}_{t\to\nu}\mathcal{F}^{-1}_{f\to\tau}\{M(t,f)\},$$

and $A^{(\alpha)}_\gamma(\tau,\nu)$ is the generalized ambiguity function of the window. Since $A^{(\alpha)}_\gamma(\tau,\nu)$ is usually well concentrated about the origin, (5.23) indicates that STFT based systems generically show small time–frequency displacement. On the other hand, one may conclude that underspread operators can be realized via STFT modification in so far as $M(t,f)$ can be found by a stable deconvolution. Precisely, one can say the following:

Theorem 5.3 *Any underspread operator* $\mathbf{H}$ *with* τ_0,ν_0 *can be realized via multiplicative STFT modification, i.e., in the form*

$$\mathbf{H} = \int_t \int_f M(t,f)\,\mathbf{P}^{(t,f)}_\gamma dt\,df,$$

where $\mathbf{P}_\gamma$ *is the rank–one projection onto an appropriately localized window, i.e., the ambiguity function satisfies*

$$1 - \left|A^{(\alpha)}_\gamma(\tau,\nu)\right| \le \epsilon_\gamma \qquad for \qquad |\tau|\le\tau_0, |\nu|\le\nu_0,$$

and $M(t,f)$ *is a bounded multiplier function.*

The maximum deviation between $M(t,f)$ *and the Weyl symbol* $L^{(0)}_H(t,f)$ *can be bounded by*

$$\left|M(t,f) - L^{(0)}_H(t,f)\right| < \frac{\epsilon_\gamma}{1-\epsilon_\gamma}\|S_H\|_1.$$

This theorem establishes both a coarse and a fine matching aspect for STFT–based system design:

- With coarse matching aspect we mean the fact that any underspread system can be realized via multiplicative STFT modification provided that the essential support of the window covers the support of the system's spreading function.
- The fine matching aspect regards the approximate realization of an underspread operator by using the Weyl symbol as multiplier function (thus avoiding the numerical expense of the deconvolution that leads to $M(t,f)$ in an exact realization). The goodness of approximation depends on the choice of the window, it can be measured in terms of ϵ_γ. Note that ϵ_γ can never vanish due to the radar uncertainty principle (see Appendix F, (F.29)).

Proof: The theorem is totally restricted to the case $\alpha = 0$, for the sake of notational simplicity we suppress the superscript.

Based on (5.23) the multiplier function is given by a well–defined deconvolution:

$$M(t,f) = \int_{-\tau_0}^{\tau_0}\int_{-\nu_0}^{\nu_0} \frac{S_H(\tau,\nu)}{A_\gamma(\tau,\nu)} e^{j2\pi(\nu t-\tau f)}\,d\tau\,d\nu. \tag{5.24}$$

Using the triangle inequality we can estimate the difference between $M(t,f)$ and $L_H(t,f)$ as follows:

$$\begin{aligned}
|M(t,f) - L_H(t,f)| &= \left|\int_{-\tau_0}^{\tau_0}\int_{-\nu_0}^{\nu_0}\left(\frac{S_H(\tau,\nu)}{A_\gamma(\tau,\nu)} - S_H(\tau,\nu)\right) e^{j2\pi(\nu t-\tau f)}\,d\tau\,d\nu\right| \\
&\le \int_{-\tau_0}^{\tau_0}\int_{-\nu_0}^{\nu_0}\left|\frac{S_H(\tau,\nu)}{A_\gamma(\tau,\nu)} - S_H(\tau,\nu)\right| d\tau\,d\nu \\
&< \left|\frac{1}{\min_{|\tau|<\tau_0,|\nu|<\nu_0} A_\gamma(\tau,\nu)} - 1\right| \int_{-\tau_0}^{\tau_0}\int_{-\nu_0}^{\nu_0} |S_H(\tau,\nu)|\,d\tau\,d\nu \\
&= \frac{\epsilon_\gamma}{1-\epsilon_\gamma}\|S_H\|_1.
\end{aligned}$$

Using the matched Gaussian window defined as

$$\gamma_0(t) = \sqrt[4]{\frac{2\nu_0}{\tau_0}} e^{-\pi \frac{\nu_0}{\tau_0} t^2},$$

the ambiguity function of the matched Gaussian window is given by

$$A_\gamma(\tau, \nu) = e^{-\frac{\pi}{2}\left(\frac{\nu_0}{\tau_0}\tau^2 + \frac{\tau_0}{\nu_0}\nu^2\right)},$$

and its minimum on the underspread support is

$$\min_{|\tau|<\tau_0, |\nu|<\nu_0} A_\gamma(\tau, \nu) = e^{-\frac{\pi}{2}\tau_0\nu_0}. \tag{5.25}$$

Among general (normalized) Gaussian windows characterized by their ratio of the temporal and spectral moment T_γ^2 and F_γ^2 as

$$\gamma(t) = \sqrt[4]{2F_\gamma/T_\gamma} e^{-\pi(F_\gamma/T_\gamma)t^2},$$

it can be easily shown that the minimum (5.25) is maximum for the matched Gaussian. This is consistent with our usual matching rule introduced in Section 3.3.5, Eq. (3.20):

$$\frac{T_\gamma}{F_\gamma} = \frac{\tau_0}{\nu_0}.$$

For the matched Gaussian window the deviation between the Weyl symbol and the STFT multiplier function is given by

$$|M(t,f) - L_H(t,f)| < \left(e^{\frac{\pi}{2}\tau_0\nu_0} - 1\right) \|S_H\|_1.$$

5.4.6 Local Stability of the Weyl Symbol

As already mentioned, for a general LTV operator the Weyl symbol is not a stable representation in the sense that boundedness of the operator is reflected in the symbol and vice versa. Specific examples can be found in [125]. However, in case of an underspread operator the Weyl symbol gives a locally reliable information about the eigenvalue interval of an operator. One has the following theorem:

Theorem 5.4 *The Weyl symbol of a self–adjoint underspread operator* $\mathbf{H}$ *with* τ_0, ν_0 *satisfies:*

$$\begin{aligned} \left|\sup L_H^{(0)}(t,f) - \lambda_{\max}\right| &< \rho\|S_H\|_1, \\ \left|\inf L_H^{(0)}(t,f) - \lambda_{\min}\right| &< \rho\|S_H\|_1, \end{aligned}$$

where ρ *is defined as*

$$\rho = \left(e^{\frac{\pi}{2}\tau_0\nu_0} - 1\right).$$

Proof: From Theorem 5.3 we know that an underspread operator with τ_0, ν_0 can be realized via multiplicative modification of the STFT using a matched Gaussian window and the deviation of the multiplier function $M(t,f)$ from the Weyl symbol can be bounded as

$$|L_H^{(0)}(t,f) - M(t,f)| \le \left(e^{\frac{\pi}{2}\tau_0\nu_0} - 1\right) \|S_H\|_1. \tag{5.26}$$

In order to couple the range of the Weyl symbol to the eigenvalue interval we also need the operator's short–time transfer function $T_H^{(\gamma)}(t,f)$ as introduced in the previous chapter,

$$T_H^{(\gamma)}(t,f) \stackrel{\text{def}}{=} \left\langle \mathbf{H}, \mathbf{P}_\gamma^{(t,f)} \right\rangle = \langle \mathbf{H}\mathbf{M}_f\mathbf{T}_t\gamma\,,\, \mathbf{M}_f\mathbf{T}_t\gamma \rangle\,.$$

Analog to the proof of Theorem 5.3 one can show that the bias of the short–time transfer function using a matched Gaussian is bounded as:

$$|L_H^{(0)}(t,f) - T_H^{(\gamma)}(t,f)| \leq \left(1 - e^{-\frac{\pi}{2}\tau_0\nu_0}\right) \|S_H\|_1. \tag{5.27}$$

Moreover, in the previous chapter we have seen that $T_H^{(\gamma)}(t,f)$, $M(t,f)$ are lower/upper Weyl–Heisenberg operator symbols in the sense that:

$$\begin{aligned} \lambda_{\min} \leq \quad & T_H^{(\gamma)}(t,f) \quad \leq \lambda_{\max}, \\ \inf M(t,f) \leq \quad & \lambda_k \quad \leq \sup M(t,f), \end{aligned}$$

which by combination with (5.26), (5.27) and by recognizing that

$$1 - e^{-x} \leq e^x - 1, \qquad x \in \mathbb{R},$$

conclude the proof. ∎

This result is in the spirit of various (much more sophisticated) mathematical work [238, 65, 180]. Moreover, the theorem shows that the Wigner–Ville spectrum of an underspread process is essentially positive, an issue that has received interest in the engineering literature [118].

General HS operator. Theorem 5.4 can be easily generalized to non–selfadjoint Hilbert–Schmidt operators. Let $\mathbf{H}$ be an underspread operator with spreading constants τ_0, ν_0, then the composite operator

$$\mathbf{R} = \mathbf{H}\mathbf{H}^*$$

is positive, self–adjoint and underspread with spreading constants $2\tau_0, 2\nu_0$. The maximum eigenvalue of $\mathbf{R}$ is the squared maximum singular value of $\mathbf{H}$,

$$\|\mathbf{R}\|_\infty = \lambda_{\max,R} = \|\mathbf{H}\mathbf{H}^*\|_\infty = \|\mathbf{H}\|_\infty^2 = \sigma_{\max,H}^2.$$

Theorem 5.4 says that

$$\left|\sup\left|L_R^{(0)}(t,f)\right| - \|\mathbf{R}\|_\infty\right| < \left(e^{2\pi\tau_0\nu_0} - 1\right) \|S_{HH^*}\|_1.$$

and according to Theorem 5.1 one has

$$\left|L_R^{(0)}(t,f) - \left|L_H^{(0)}\right|^2\right| < 2\sin(\pi\tau_0\nu_0) \|S_H\|_1^2.$$

Combining these inequalities allows to bound the difference between the magnitude squared symbol and the squared supremum norm of the operator:

$$\left|\sup\left|L_H^{(0)}(t,f)\right|^2 - \|\mathbf{H}\|_\infty^2\right| < \left(e^{2\pi\tau_0\nu_0} - 1\right) \|S_{HH^*}\|_1 + 2\sin(\pi\tau_0\nu_0) \|S_H\|_1^2. \tag{5.28}$$

The last two theorems are restricted to the case $\alpha = 0$. However, they essentially hold for arbitrary α. This α–invariance of the GWS for underspread operators is the topic of the next section.

5.4.7 Approximate α–Invariance

Starting with $\alpha = 0$ any member of the family of generalized spreading functions can be obtained via a unimodular multiplier function (B.15):

$$S_H^{(\alpha)}(\tau,\nu) = S_H^{(0)}(\tau,\nu)\, e^{-j2\pi\nu\tau\alpha}.$$

This relation immediately shows the α–invariance for LTI and LFI operators, because we have for LTI operators $\nu = 0$ for $S_H^{(\alpha)}(\tau,\nu) \neq 0$ and for LFI operators $\tau = 0$ for $S_H^{(\alpha)}(\tau,\nu) \neq 0$. Indeed, LTI and LFI operators are limit cases of underspread operators. For a general underspread operator one can formulate the following α–invariance theorem (we remind the reader that we always require $|\alpha| \leq 1/2$):

Theorem 5.5 *The generalized Weyl symbol of an underspread operator is approximately α–invariant. In particular, for underspread operators* $\mathbf{H}$ *with the specified spreading constants one has the following L_∞–bounds valid for $|\alpha| \leq 1/2$:*

$$\begin{aligned}
\left| L_H^{(\alpha_1)} - L_H^{(\alpha_2)} \right| &< 2\sin\left(\pi\tau_0\nu_0|\alpha_1-\alpha_2|\right)\|S_H\|_1, \qquad \tau_0\nu_0 < \frac{1}{2}, \\
\left| |L_H^{(\alpha_1)}|^2 - |L_H^{(\alpha_2)}|^2 \right| &< 2\sin\left(2\pi\tau_0\nu_0|\alpha_1-\alpha_2|\right)\|S_H\|_1^2, \qquad \tau_0\nu_0 < \frac{1}{4}.
\end{aligned}$$

For underspread Hilbert–Schmidt operators, the corresponding L_2–bounds are given by:

$$\begin{aligned}
\left\| L_H^{(\alpha_1)} - L_H^{(\alpha_2)} \right\|^2 &< 4\sin^2\left(\pi\tau_0\nu_0(\alpha_1-\alpha_2)\right)\|\mathbf{H}\|^2, \qquad \tau_0\nu_0 < \frac{1}{2}, \\
\left\| |L_H^{(\alpha_1)}|^2 - |L_H^{(\alpha_2)}|^2 \right\|^2 &< 64\tau_0\nu_0\sin^2\left(8\pi\tau_0\nu_0(\alpha_1-\alpha_2)\right)\|\mathbf{H}\|^4, \qquad \tau_0\nu_0 < \frac{1}{16}.
\end{aligned}$$

This theorem allows to show the approximate equivalence of various classical definitions of a time–varying spectrum for underspread processes (see Section 2.5.3) [212]. Just as discussed for Theorem 5.1, for trace class operators one has coarser but simpler L_∞–bounds:

$$\begin{aligned}
\left| L_H^{(\alpha_1)} - L_H^{(\alpha_2)} \right| &< 8\pi\tau_0^2\nu_0^2|\alpha_1-\alpha_2|\mathrm{tr}\sqrt{\mathbf{H}^*\mathbf{H}}, \qquad \tau_0\nu_0 < \frac{1}{2}, \\
\left| |L_H^{(\alpha_1)}|^2 - |L_H^{(\alpha_2)}|^2 \right| &< 64\pi\tau_0^3\nu_0^3|\alpha_1-\alpha_2|\mathrm{tr}^2\sqrt{\mathbf{H}^*\mathbf{H}}, \qquad \tau_0\nu_0 < \frac{1}{4}.
\end{aligned}$$

Similarly, we have the following bounds for the L_2–bounds for Hilbert–Schmidt underspread operators:

$$\begin{aligned}
\left\| L_H^{(\alpha_1)} - L_H^{(\alpha_2)} \right\|^2 &< 4\pi^2\tau_0^2\nu_0^2(\alpha_1-\alpha_2)^2\|\mathbf{H}\|^2, \qquad \tau_0\nu_0 < \frac{1}{2}, \\
\left\| |L_H^{(\alpha_1)}|^2 - |L_H^{(\alpha_2)}|^2 \right\|^2 &< 512\pi^2\tau_0^3\nu_0^3(\alpha_1-\alpha_2)^2\|\mathbf{H}\|^4, \qquad \tau_0\nu_0 < \frac{1}{16}.
\end{aligned}$$

Proof: For the proof of the first L_∞–bound we use the integral triangle inequality as follows:

$$\begin{aligned}
&\left| L_H^{(\alpha_1)} - L_H^{(\alpha_2)} \right| \\
&= \left| \int_\tau \int_\nu \left\{ S_H^{(\alpha_1)}(\tau,\nu) - S_H^{(\alpha_2)}(\tau,\nu) \right\} e^{j2\pi(\nu t - \tau f)} d\tau\, d\nu \right| \\
&\leq \int_\tau \int_\nu \left| S_H^{(0)}(\tau,\nu) \right| \left| e^{-j2\pi\tau\nu\alpha_1} - e^{-j2\pi\tau\nu\alpha_2} \right| d\tau\, d\nu \\
&= \int_{\tau_0}^{\nu_0} \int_{\tau_0}^{\nu_0} \left| S_H^{(0)}(\tau,\nu) \right| 2\sin\left(\pi\tau\nu|\alpha_1-\alpha_2|\right) d\tau\, d\nu \\
&< 2\sin\left(\pi\tau_0\nu_0|\alpha_1-\alpha_2|\right)\|S_H\|_1,
\end{aligned}$$

where in the last step we have made use of the requirements $\tau_0\nu_0 < 1/2$ in the following sense

$$\max_{|\tau|<\tau_0, |\nu|<\nu_0} \left\{ \sin\left(\pi\tau\nu|\alpha_1-\alpha_2|\right) \right\} = \sin\left(\pi\tau_0\nu_0|\alpha_1-\alpha_2|\right),$$

which is always valid since:

$$\pi\tau_0\nu_0|\alpha_1-\alpha_2| < \pi/2 \qquad |\alpha_1-\alpha_2| < 1.$$

The proof of the second L_∞–bound goes along the same lines:

$$\begin{aligned}
&\left| |L_H^{(\alpha_1)}|^2 - |L_H^{(\alpha_2)}|^2 \right| \\
&= \left| \int_\tau \int_\nu \int_{\tau'} \int_{\nu'} \left\{ S_H^{(\alpha_1)}(\tau',\nu') S_H^{(\alpha_1)*}(\tau'-\tau,\nu'-\nu) \right. \right. \\
&\quad \left. \left. - S_H^{(\alpha_2)}(\tau',\nu') S_H^{(\alpha_2)*}(\tau'-\tau,\nu'-\nu) \right\} e^{j2\pi(\nu t - \tau f)} d\tau\, d\nu\, d\tau'\, d\nu' \right| \\
&\leq 2 \int_\tau \int_\nu \int_{\tau'} \int_{\nu'} \left| S_H^{(0)}(\tau',\nu') \right| \left| S_H^{(0)}(\tau'-\tau,\nu'-\nu) \right| \\
&\quad \cdot \left| e^{-j2\pi[\tau'\nu+\nu'\tau-\tau\nu]\alpha_1} - e^{-j2\pi[\tau'\nu+\nu'\tau-\tau\nu]\alpha_2} \right| d\tau\, d\nu\, d\tau'\, d\nu' \\
&\leq 2 \int_{\tau_1} \int_{\nu_1} \int_{\tau_2} \int_{\nu_2} \left| S_H^{(0)}(\tau_1,\nu_1) \right| \left| S_H^{(0)}(\tau_2,\nu_2) \right| 2\sin\left(\pi(\tau_1\nu_1 - \tau_2\nu_2)|\alpha_1-\alpha_2|\right) d\tau_1\, d\nu_1\, d\tau_2\, d\nu_2 \\
&< 2\sin\left(2\pi\tau_0\nu_0|\alpha_1-\alpha_2|\right) \|S_H\|_1^2,
\end{aligned}$$

where we have used

$$\max_{|\tau_i|<\tau_0, |\nu_i|<\nu_0} \{\tau_1\nu_1 - \tau_2\nu_2\} = 2\tau_0\nu_0,$$

and the requirements $\tau_0\nu_0 < 1/4$ and $|\alpha_1 - \alpha_2| < 1$ such that

$$2\pi\tau_0\nu_0\,|\alpha_1-\alpha_2| < \pi/2.$$

For the proof of the L_2–bounds we need the unitarity of the spreading function:

$$\int_\tau \int_\nu |S_H(\tau,\nu)|^2 \, d\tau\, d\nu = \|\mathbf{H}\|^2 .$$

The proof of the first L_2–bound is straightforward:

$$\begin{aligned}
&\left\| L_H^{(\alpha_1)} - L_H^{(\alpha_2)} \right\|^2 \\
&= \left\| S_H^{(\alpha_1)} - S_H^{(\alpha_2)} \right\|^2 \\
&= \int_\tau \int_\nu \left| S_H^{(0)}(\tau,\nu) \right|^2 \left| e^{-j2\pi\tau\nu\alpha_1} - e^{-j2\pi\tau\nu\alpha_2} \right|^2 d\tau\, d\nu \\
&= \int_\tau \int_\nu \left| S_H^{(0)}(\tau,\nu) \right|^2 \sin^2\left(\pi\tau\nu(\alpha_1-\alpha_2)\right) d\tau\, d\nu \\
&< 4\sin^2\left(\pi\tau_0\nu_0(\alpha_1-\alpha_2)\right) \|\mathbf{H}\|^2,
\end{aligned}$$

where in the last step we have used the fact that $\alpha_1 - \alpha_2 < 1$ and $\tau_0\nu_0 < 1/2$ with the reasoning analogous to the proof of the L_∞–bound.

In the proof of the second L_2–bound we use the Schwarz inequality and proceed similarly to the foregoing proof:

$$\begin{aligned}
&\left\| |L_H^{(\alpha_1)}|^2 - |L_H^{(\alpha_2)}|^2 \right\|^2 \\
&= \left\| S_H^{(\alpha_1)}(\tau,\nu) * * S_H^{(\alpha_1)*}(-\tau,-\nu) - S_H^{(\alpha_2)}(\tau,\nu) * * S_H^{(\alpha_2)*}(-\tau,-\nu) \right\|^2 \\
&= \int_{-2\tau_0}^{2\tau_0} \int_{-2\nu_0}^{2\nu_0} \left| \int_{\tau'} \int_{\nu'} S_H^{(0)}(\tau',\nu') S_H^{(0)*}(\tau'-\tau,\nu'-\nu) \right. \\
&\quad \left. \cdot \left\{ e^{-j2\pi(\tau\nu'+\tau'\nu-\tau\nu)\alpha_1} - e^{-j2\pi(\tau\nu'+\tau'\nu-\tau\nu)\alpha_2} \right\} d\tau'\, d\nu' \right|^2 d\tau\, d\nu \\
&\leq \int_{-2\tau_0}^{2\tau_0} \int_{-2\nu_0}^{2\nu_0} \left(\int_{\tau_1} \int_{\nu_1} \left| S_H^{(0)}(\tau_1,\nu_1) \right|^2 d\tau_1\, d\nu_1 \right) \\
&\quad \cdot \left(\int_{\tau_2} \int_{\nu_2} \left| S_H^{(0)}(\tau_2-\tau,\nu_2-\nu) \right|^2 \left| e^{-j2\pi(\tau\nu_2+\tau_2\nu-\tau\nu)\alpha_1} - e^{-j2\pi(\tau\nu_2+\tau_2\nu-\tau\nu)\alpha_2} \right|^2 d\tau_2\, d\nu_2 \right) d\tau\, d\nu \\
&= \int_{-2\tau_0}^{2\tau_0} \int_{-2\nu_0}^{2\nu_0} \|\mathbf{H}\|^2 \int_{\tau'} \int_{\nu'} \left| S_H^{(0)}(\tau',\nu') \right|^2 4\sin^2\left(\pi(\tau\nu'+\tau'\nu+\tau\nu)(\alpha_1-\alpha_2)\right) d\tau\, d\nu\, d\tau'\, d\nu' \\
&< 64\tau_0\nu_0 \sin^2\left(8\pi\tau_0\nu_0(\alpha_1-\alpha_2)\right) \|\mathbf{H}\|^4,
\end{aligned}$$

where in the last step we have used the spreading requirement $\tau_0\nu_0 < 1/16$, which together with

$$\begin{aligned}
&\max\left\{\tau\nu' + \tau'\nu + \tau\nu\right\} = 8\tau_0\nu_0, \\
\text{subject to} \qquad &|\tau| < 2\tau_0, |\tau'| < \tau_0, |\nu| < 2\nu_0, |\nu'| < \nu_0,
\end{aligned}$$

and $|\alpha_1 - \alpha_2| < 1$, implies

$$8\pi\tau_0\nu_0|\alpha_1-\alpha_2| < \pi/2.$$

∎

These α–invariance results show that many of the classical time–frequency parametrized operator representations are essentially equal for underspread operators and it is a matter of convenience (ease of a numerical implementation) which version of the generalized Weyl correspondence is used.

5.4.8 Approximate Eigenpairs

The optimum design of structured sets of approximate eigensignals was one of the main topics of the previous chapters. More than once, we have observed that given the mere knowledge of the spreading support of an operator, one has a time–frequency shift–covariant characterization of approximate eigensignals. By covariance we mean that one can produce a manifold of approximate eigensignals by time–frequency–shifting of an appropriate prototype. However, while we have presented both optimum and low–cost design criteria for approximate eigensignals, we still have to investigate the goodness of the approximation. We stress that it does not make sense to measure the deviation from a true eigensignal (in e.g. L_2–sense) because these true eigensignals are in general complicated, unstructured signals. Moreover, determination of eigenfunctions is always tied to the determination of the corresponding eigenvalue. For underspread operators we expect that when we shift an appropriately localized prototype function to (t_0, f_0) it is just the value of the Weyl symbol at (t_0, f_0) which determines the (approximate) eigenvalue:

$$\mathbf{H}\gamma^{(t_0,f_0)} \stackrel{?}{=} L_H^{(\alpha)}(t_0,f_0)\gamma^{(t_0,f_0)}$$

Hence, we define the L_2–deviation from this idealized property to quantify the concept of approximate eigensignals.

Theorem 5.6 *Given an underspread operator* $\mathbf{H}$ *with* τ_0, ν_0 *and a normalized prototype function* γ *which achieves appropriate localization in the sense that*

$$|A_\gamma(\tau,\nu) - 1|\chi_{[-2\tau_0,2\tau_0]}(\tau)\,\chi_{[-2\nu_0,2\nu_0]}(\nu) \leq \epsilon_\gamma,$$

then, the time–frequency shifted versions $\gamma^{(t,f)}$ *and the corresponding values of the Weyl symbol* $L_H(t,f)$ *form an approximate eigenpair in the sense that*

$$\|\mathbf{H}\gamma^{(t,f)} - L_H(t,f)\gamma^{(t,f)}\|^2 < 2\sin(\pi\tau_0\nu_0)\|S_H\|_1 + \epsilon_\gamma\left(\|S_{H^*H}\|_1 + 2\|S_H\|_1^2\right),$$

with $\gamma^{(\tau,\nu)}(t) = \gamma(t-\tau)e^{j2\pi\nu t}$.

Proof: Normalized prototype implies:

$$\|\gamma\| = 1 \qquad \Longleftrightarrow \qquad A_\gamma(0,0) = 1.$$

The assumption of "appropriate localization" allows to split up the ambiguity function of the prototype as

$$A_\gamma(\tau,\nu) - 1 \stackrel{\text{def}}{=} \rho_\gamma(\tau,\nu),$$

where, within the doubled underspread support, the magnitude of $\rho_\gamma(\tau,\nu)$ is bounded by ϵ_γ:

$$|\rho_\gamma(\tau,\nu)|\,\chi_{[-2\tau_0,2\tau_0]}(\tau)\,\chi_{[-2\nu_0,2\nu_0]}(\nu) \leq \epsilon_\gamma.$$

The ambiguity function of a time–frequency shifted version of the prototype function is then given by (see (F.26)):

$$A_{\gamma^{(t,f)}}(\tau,\nu) = A_\gamma(\tau,\nu)e^{-j2\pi(\nu t-\tau f)} = (1+\rho_\gamma(\tau,\nu))\,e^{-j2\pi(\nu t-\tau f)}.$$

First, we estimate the difference between the quadratic form and the Weyl symbol:

$$\begin{aligned}
&\left|\left\langle \mathbf{H}\gamma^{(t,f)}, \gamma^{(t,f)}\right\rangle - L_H(t,f)\right| \\
&= \left|\left\langle S_H, A_{\gamma^{(t,f)}}\right\rangle - L_H(t,f)\right| \\
&= \left|\int_{-\tau_0}^{\tau_0}\int_{-\nu_0}^{\nu_0} S_H(\tau,\nu)A_\gamma^*(\tau,\nu)e^{j2\pi(\nu t-\tau f)}\,d\tau\,d\nu - L_H(t,f)\right| \\
&= \left|\int_{-\tau_0}^{\tau_0}\int_{-\nu_0}^{\nu_0} S_H(\tau,\nu)\rho_\gamma^*(\tau,\nu)e^{j2\pi(\nu t-\tau f)}\,d\tau\,d\nu\right| \\
&\leq \int_{-\tau_0}^{\tau_0}\int_{-\nu_0}^{\nu_0} |S_H(\tau,\nu)\rho_\gamma(\tau,\nu)|\,d\tau\,d\nu \\
&< \epsilon_\gamma\,\|S_H\|_1\,.
\end{aligned}$$

Second, we estimate the difference between the output energy and the magnitude–squared Weyl symbol, to this end we employ Theorem 5.1 and (C.11)

$$\begin{aligned}
&\left|\left\|\mathbf{H}\gamma^{(t,f)}\right\|^2 - |L_H(t,f)|^2\right| \\
&= \left|\left\langle S_{H^*H}, A_{\gamma^{(t,f)}}\right\rangle - |L_H(t,f)|^2\right| \\
&= \left|\int_{-\tau_0}^{\tau_0}\int_{-\nu_0}^{\nu_0} S_{H^*H}(\tau,\nu)A^*_\gamma(\tau,\nu)e^{j2\pi(\nu t-\tau f)}\,d\tau\,d\nu - |L_H(t,f)|^2\right| \\
&= \left|L_{H^*H}(t,f) + \int_{-2\tau_0}^{2\tau_0}\int_{-2\nu_0}^{2\nu_0} S_{H^*H}(\tau,\nu)\rho^*_\gamma(\tau,\nu)e^{j2\pi(\nu t-\tau f)}\,d\tau\,d\nu - |L_H(t,f)|^2\right| \\
&\le \left|L_{H^*H}(t,f) - |L_H(t,f)|^2\right| + \int_{-2\tau_0}^{2\tau_0}\int_{-2\nu_0}^{2\nu_0} |S_{H^*H}(\tau,\nu)\rho_\gamma(\tau,\nu)|\;d\tau\,d\nu \\
&< 2\sin\pi\tau_0\nu_0\,\|S_H\|_1^2 + \epsilon_\gamma\,\|S_{H^*H}(\tau,\nu)\|_1\,.
\end{aligned}$$

For the sake of a compact notation we introduce:

$$\begin{aligned}
\epsilon_{H,1}(t,f) &\stackrel{\text{def}}{=} \left\langle \mathbf{H}\gamma^{(t,f)}, \gamma^{(t,f)}\right\rangle - L_H(t,f), \\
\epsilon_{H,2}(t,f) &\stackrel{\text{def}}{=} \left\|\mathbf{H}\gamma^{(t,f)}\right\|^2 - |L_H(t,f)|^2\,.
\end{aligned}$$

We split up the L_2–norm of the theorem via,

$$\|x-y\|^2 = \|x\|^2 - 2\mathrm{Re}\left\{\langle x,y\rangle\right\} + \|y\|^2$$

and get:

$$\begin{aligned}
&\|\mathbf{H}\gamma^{(t,f)} - L_H(t,f)\gamma^{(t,f)}\|^2 \\
&= \left\|\mathbf{H}\gamma^{(t,f)}\right\|^2 - 2\mathrm{Re}\left\{L^*_H(t,f)\left\langle \mathbf{H}\gamma^{(t,f)}, \gamma^{(t,f)}\right\rangle\right\} + |L_H(t,f)|^2 \\
&= |L_H(t,f)|^2 + \epsilon_{H,2}(t,f) - 2\mathrm{Re}\left\{L^*_H(t,f)\left(L_H(t,f) + \epsilon_{H,1}(t,f)\right)\right\} + |L_H(t,f)|^2 \\
&= \epsilon_{H,2}(t,f) - 2\mathrm{Re}\left\{L^*_H(t,f)\epsilon_{H,1}(t,f)\right\} \\
&< 2\sin(\pi\tau_0\nu_0)\|S_H\|_1 + \epsilon_\gamma\left(\|S_{H^*H}\|_1^2 + 2\left|\mathrm{Re}\left\{L_H(t,f)\right\}\right|\right).
\end{aligned}$$

Finally, we estimate the Weyl symbol by the L_1–norm of the spreading function as follows:

$$\begin{aligned}
\left|\mathrm{Re}\left\{L_H(t,f)\right\}\right| &< |L_H(t,f)| \\
&= \left|\int_\tau\int_\nu S_H(\tau,\nu)e^{-j2\pi(\nu t-\tau f)}\,d\tau\,d\nu\right| \\
&\le \|S_H\|_1,
\end{aligned}$$

which concludes the proof. ∎

An appropriately localized prototype function can be found by our usual matching rule:

$$\frac{T_\gamma}{F_\gamma} = \frac{\tau_0}{\nu_0}.$$

Error Norm	Overspread (Fig.5.1)	Critical spread (Fig. 5.2)	Underspread (Fig. 5.3)
$\frac{\|L_{GH} - L_G L_H\|}{\|\mathbf{G}\|\|\mathbf{H}\|}$	196	52	9.5
$\frac{\|\mathbf{GH} - \mathbf{G} \otimes \mathbf{H}\|_\infty}{\|\mathbf{G}\|_\infty \|\mathbf{H}\|_\infty}$	1.39	0.35	0.035

Table 5.1: Error norms corresponding to the experiments of Figs.5.1–5.3

As already pointed in Section 5.4.5, for a Gaussian prototype the matching rule obtains the optimum duration/bandwidth in the sense of minimum ϵ_γ. That is, given the matched Gaussian pulse:

$$\gamma_0(t) = \sqrt[4]{\frac{2\nu_0}{\tau_0}} e^{-\pi \frac{\nu_0}{\tau_0} t^2},$$

we have (see (5.25)):

$$\epsilon_\gamma = 1 - e^{-\frac{\pi}{2}\tau_0\nu_0}.$$

Such that the eigenpair deviation reduces to a bound approximately proportional to the square of the operator's total spread ($\sigma_H = 4\tau_0\nu_0$):

$$\begin{aligned} &\|\mathbf{H}\gamma^{(t,f)} - L_H(t,f)\gamma^{(t,f)}\|^2 \\ &< \quad 2\pi\tau_0\nu_0\|S_H\|_1 + \left(1 - e^{-\frac{\pi}{2}\tau_0\nu_0}\right)\left(\|S_{H^*H}\|_1 + 2\|S_H\|_1^2\right) \\ &\approx \quad \frac{\pi}{2}\sigma_H^2 \mathrm{tr}\sqrt{\mathbf{H}^*\mathbf{H}}\left(1 + \mathrm{tr}\sqrt{\mathbf{H}^*\mathbf{H}}\right). \end{aligned}$$

5.5 Numerical Experiments

The applicability of the symbolic calculus of underspread operators has already been demonstrated in the previous chapter by the Wiener filter example. We now consider the simple multiplicative symbol calculus ("twisted product") for simply designed underspread, critical and overspread operators with "down–chirp"/"up–chirp" structure, see the Figures 5.1–5.3. The *twisted product* of two operators is defined by an obvious, three–step procedure consisting of: (i) computing the generalized Weyl symbols of both operators, (ii) pointwise multiplication of the symbols, (iii) applying the inverse generalized Weyl correspondence to this product. Based on the continuous Weyl–Heisenberg expansion one can define the twisted product formally by:

$$\mathbf{G} \otimes \mathbf{H} \stackrel{\mathrm{def}}{=} \int_t \int_f L_G(t,f) L_H(t,f)\, \mathbf{P}(t,f) dt\, df,$$

where $\mathbf{P}(t,f)$ is the infinitesimal prototype or its discrete–time counterpart.

In Table 5.1 we have listed the Hilbert–Schmidt norm and the standard operator norm (maximum singular value) for the experiments corresponding to Figures 5.1–5.3.

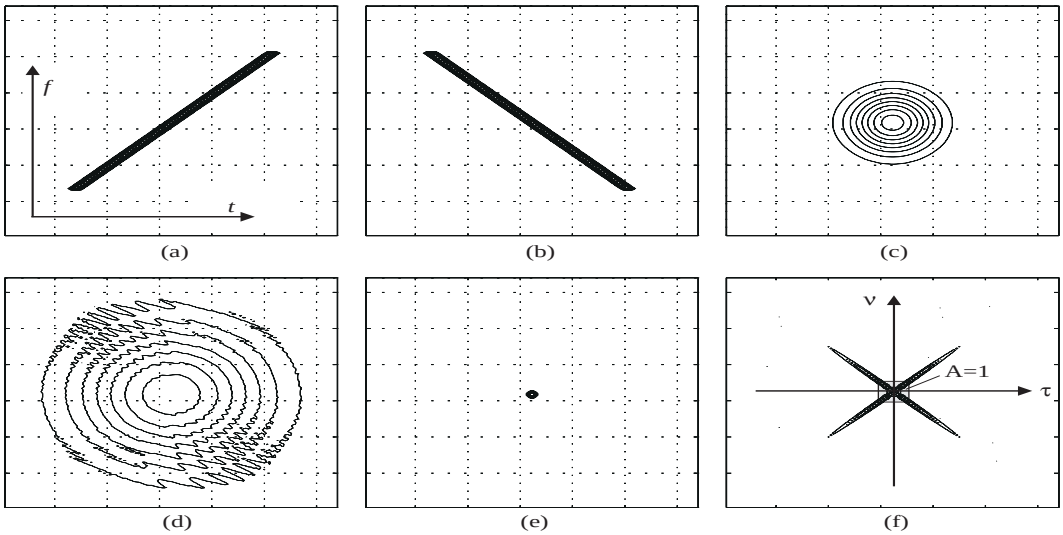

Figure 5.1: Symbolic calculus for overspread operators: (a) $L_G(t,f)$, (b) $L_H(t,f)$ (c) $|L_{GH}(t,f)|$, (d) $|L_{H^*G^*}(t,f)|$, (e) $L_G(t,f)L_H(t,f)$, (f) $|S_{G+H}(\tau,\nu)|$.

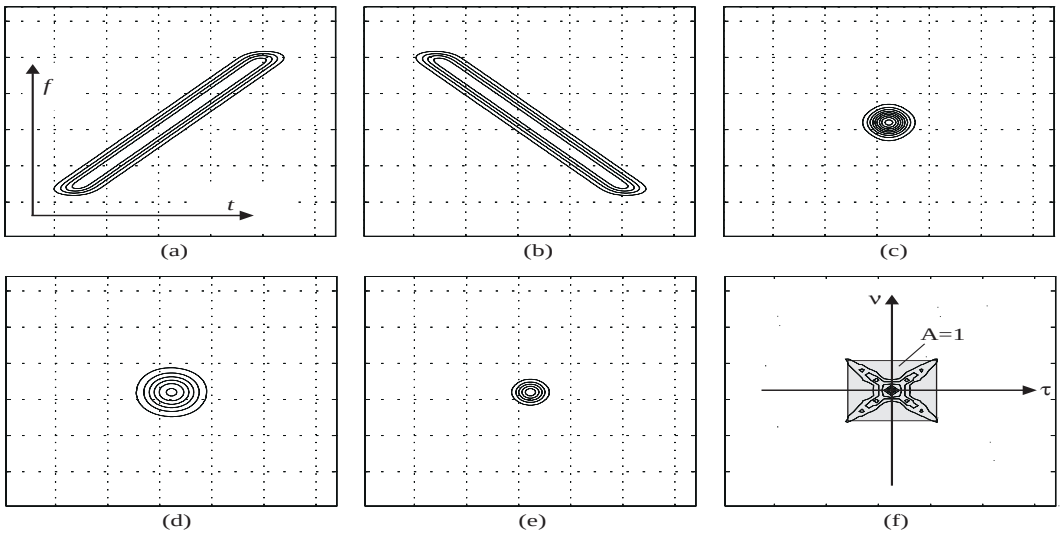

Figure 5.2: Symbolic calculus for operators with critical spread.

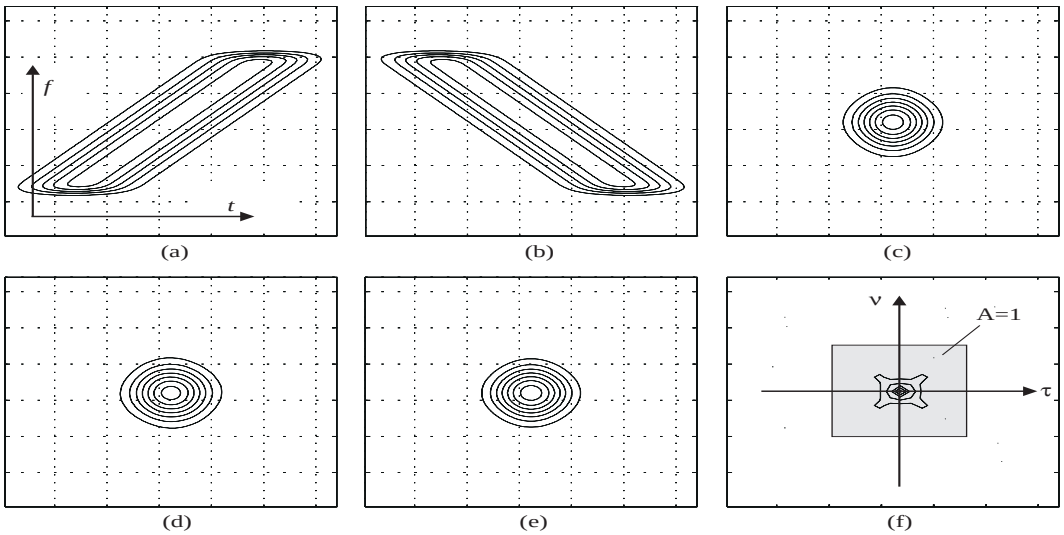

Figure 5.3: Symbolic calculus for underspread operators.

5.6 Summary

We have studied a subspace of the Hilbert space of Hilbert–Schmidt operators defined by a compactly supported spreading function:

$$S_H(\tau,\nu) = S_H(\tau,\nu)\,\chi_{[-\tau_0,\tau_0]}(\tau)\,\chi_{[-\nu_0,\nu_0]}(\nu).$$

Such operators correspond to slowly time–varying systems with limited memory or to the covariance kernel of a nonstationary process with limited temporal/spectral correlation width. The total spread

$$\sigma_H \stackrel{\text{def}}{=} 4\tau_0\nu_0,$$

is a fundamental parameter for the identification (estimation), realization and interpretation of underspread operators and their Weyl–Heisenberg symbols.

For bounded $S_H(\tau,\nu)$, we have shown that underspread operators and their generalized Weyl symbols are asymptotically consistent with LTI operator theory. The total spread σ_H is the critical parameter of all asymptotic results:

- Asymptotic validity of symbol calculus for jointly underspread operators:

$$\frac{\|L_{GH} - L_G L_H\|^2}{\|\mathbf{G}\|^2\,\|\mathbf{H}\|^2} = \mathcal{O}(\sigma_H^3).$$

- Asymptotic commutativity of jointly underspread operators:

$$\frac{\|\mathbf{HG} - \mathbf{GH}\|^2}{\|\mathbf{G}\|^2\,\|\mathbf{H}\|^2} = \mathcal{O}(\sigma_H^3).$$

- Asymptotic normality:

$$\frac{\|\mathbf{HH}^* - \mathbf{H}^*\mathbf{H}\|^2}{\|\mathbf{H}\|^4} = \mathcal{O}(\sigma_H^3).$$

- Asymptotic local stability:

$$\frac{\left|\sup|L_H(t,f)|^2 - \|\mathbf{H}\|_\infty^2\right|}{\mathrm{tr}^2\sqrt{\mathbf{H}^*\mathbf{H}}} = \mathcal{O}(\sigma_H^2).$$

- Asymptotic α–invariance:

$$\frac{\left|L_H^{(\alpha_1)}(t,f) - L_H^{(\alpha_2)}(t,f)\right|}{\mathrm{tr}\sqrt{\mathbf{H}^*\mathbf{H}}} = \mathcal{O}(\sigma_H^2).$$

- Approximate equivalence with (matched window) STFT–multiplier symbol:

$$\frac{|M(t,f) - L_H(t,f)|}{\mathrm{tr}\sqrt{\mathbf{H}^*\mathbf{H}}} = \mathcal{O}(\sigma_H^2).$$

- Approximate eigenpairs:

$$\frac{\left\|\,\mathbf{H}\,\gamma^{(t,f)} - L_H(t,f)\,\gamma^{(t,f)}\,\right\|^2}{\mathrm{tr}\sqrt{\mathbf{H}^*\mathbf{H}}\left(1 + \mathrm{tr}\sqrt{\mathbf{H}^*\mathbf{H}}\right)} = \mathcal{O}(\sigma_H^2).$$

Chapter 6

Conclusions

The main results are summarized and open problems for future research are pointed out.

6.1 Summary

The time–varying transfer function of an LTV system and the time–varying power spectrum of a nonstationary process can be reduced to the same mathematical concept: One maps a linear operator onto a function of time and frequency, a *Weyl–Heisenberg symbol* of the operator.

We have studied various definitions of such Weyl–Heisenberg (WH) symbols, in particular the *generalized Weyl symbol* $L_H^{(\alpha)}(t,f)$ which comprises *Zadeh's time–varying transfer function (Kohn–Nirenberg symbol)* and the *Weyl symbol*. The (generalized) Weyl symbol of a correlation operator is equivalent to the (generalized) Wigner–Ville spectrum of a nonstationary process and Zadeh's time–varying transfer function is closely related to Priestley's evolutionary spectrum. The "naive" interpretation of the WH symbols is given by the concept of a time–frequency–parametrized eigenvalue distribution (in the sense of generalizing the frequency–parametrized eigenvalue distributions of translation invariant operators).

We have introduced a formal interpretation of the generalized Weyl symbol (GWS) in the form of a *continuous Weyl–Heisenberg expansion* of a Hilbert–Schmidt (HS) operator (we suppress the α–parametrization):

$$\mathbf{H} = \int_t \int_f L_H(t,f)\mathbf{P}^{(t,f)}\,dt\,df \qquad \text{with} \qquad L_H(t,f) = \left\langle \mathbf{H}, \mathbf{P}^{(t,f)} \right\rangle, \tag{6.1}$$

where $\mathbf{P}$ is an infinitesimal prototype operator which does not admit a helpful physical interpretation as it introduces arbitarily large time–frequency shifts (while our "naive" interpretation would require a perfect "time–frequency localizator"). This has led to the conclusion that one has to restrict the class of HS operators where the eigenvalue interpretation of the GWS works in an asymptotic sense.

The generalized spreading function $S_H(\tau,\nu)$ is the Fourier dual of the generalized Weyl symbol leading to a decomposition similar to (6.1) based on the time–frequency shift operator

$$\mathbf{H} = \int_\tau \int_\nu S_H(\tau,\nu)\mathbf{S}^{(\tau,\nu)}\,d\tau\,d\nu \qquad \text{with} \qquad S_H(\tau,\nu) = \left\langle \mathbf{H}, \mathbf{S}^{(\tau,\nu)} \right\rangle. \tag{6.2}$$

The time–frequency shift operator $\mathbf{S}^{(\tau,\nu)}$ admits a meaningful physical interpretation: It corresponds to the time–delay and Doppler–shift as it appears, e.g., in any time–varying multipath environment. Irrespective of the actual physical scenario, the spreading representation (6.2) holds for any HS operator and it can be extended to certain important non–HS operators. The magnitude of $S_H(\tau,\nu)$ is definition–independent and we have seen that particularly the support of $S_H(\tau,\nu)$ is helpful for classifying "natural" and "synthetic" linear operators.

We have studied operators with restricted spreading characterized by a limited support of the spreading function $S_H(\tau,\nu)$:

$$S_H(\tau,\nu) = S_H(\tau,\nu)\chi_H(\tau,\nu),$$

where $\chi_H(\tau,\nu)$ is a 0/1–valued indicator function. We have shown that based upon this indicator function the generalized Weyl symbol can be redefined via a different prototype operator $\widehat{\mathbf{P}}$:

$$L_H(t,f) = \left\langle \mathbf{H}, \widehat{\mathbf{P}}^{(t,f)} \right\rangle, \qquad \text{with} \qquad S_{\widehat{P}}(\tau,\nu) \stackrel{\text{def}}{=} \chi_H(\tau,\nu). \tag{6.3}$$

Considering $\chi_H(\tau,\nu)$ as incomplete, realistic *a priori* knowledge about a nonstationary environment, the prototype operator $\widehat{\mathbf{P}}$ is matched to the nonstationary environment both in the abstract way of minimum HS norm and in the concrete sense of conveying optimum estimators:

- In contrast to the infinitesimal prototype $\mathbf{P}$, the operator $\widehat{\mathbf{P}}$ is (i) Hilbert–Schmidt and (ii) it does not introduce arbitrarily large time–frequency shifts. Among all admissible prototype operators (for which (6.1) holds true) $\widehat{\mathbf{P}}$ is marked out by (i) minimum HS norm and (ii) minimum amount of potential time–frequency shifts.

- In a usual time–varying spectrum estimation setup the optimum estimate can be formulated as a quadratic form of the observation based on $\widehat{\mathbf{P}}$.

- In a usual system identification setup the optimum unbiased estimator is given by a bilinear form of the input/output observation in terms of $\widehat{\mathbf{P}}$.

In practice one has to consider finite–rank approximations of the prototype operator. An optimum rank–one reduction of $\widehat{\mathbf{P}}$ leads to a, typically pulse–like, prototype signal $\gamma(t)$ which is matched to nonstationary environments for various different setups:

- Optimum STFT/Gabor window for the representation of nonstationary processes.

- Matched STFT window for the linear filtering of nonstationary processes.

- Optimum distortion free pulse for the transmission over a WSSUS channel (here $\chi_H(\tau,\nu)$ characterizes the support of the channel's scattering function).

Operators with restricted spreading have been specialized to the case of a rectangular support:

$$\chi_H(\tau,\nu) = \chi_{[-\tau_0,\tau_0]}(\tau)\chi_{[-\nu_0,\nu_0]}(\nu),$$

where τ_0 is the maximum time shift and ν_0 is the maximum frequency shift. Based upon this definition we have introduced as precised version of the classical underspread/overspread classification of LTV systems which has been originally introduced for the case of nonstationary processes: We call a process/system underspread when $\tau_0\nu_0 \ll 1$ and overspread in the converse case.

For underspread operators we have shown practically relevant results as follows:

- For jointly underspread operators the generalized Weyl symbol of the product operator is approximately equal to the product of the symbols:

  $$L_{H_1H_2}(t,f) \approx L_{H_1}(t,f)L_{H_2}(t,f),$$

  where the approximation gets better (both in L_2 and L_∞ sense) with decreasing product $\tau_0\nu_0$. This result shows in particular that it is the class of underspread environments where one can use a time–varying power spectrum and a time–varying transfer function for applications involving operator algebra (Wiener filter, equalization, inverse filtering, etc.).

- For underspread operators the generalized Weyl symbol gets increasingly α–invariant with decreasing spreading product $\tau_0\nu_0$:

$$L_H^{(\alpha_1)}(t,f) \approx L_H^{(\alpha_2)}(t,f).$$

  This result can be related to the longstanding discussion about the "optimum" definition of a stochastic time–varying power spectrum: For the class of underspread processes (where the notion of a time–varying power spectrum makes sense) all of the classical definitions (in particular Priestley's spectrum and the Wigner–Ville spectrum) are essentially equivalent.

- The generalized Weyl symbol of an underspread operators is uniquely characterized by its samples on a rectangular grid ("matched grid") with constants characterized by

$$\frac{T}{F} = \frac{\tau_0}{\nu_0} \qquad \text{and} \qquad TF = \frac{1}{4\tau_0\nu_0}.$$

  Sampling the symbol leads to *discrete Weyl–Heisenberg expansion* as follows:

$$\mathbf{H} = TF \sum_k \sum_l L_H(lT, mF)\widehat{\mathbf{P}}^{(lT,mF)},$$

  the discrete counterpart of (6.1). We have pointed out the consequences of this result for (i) the estimation and optimum filtering of underspread processes, (ii) the representation, identification and realization of underspread LTV systems. Moreover, we have observed the intuitively appealing fact that our "naive" eigenvalue interpretation starts to make sense for

$$TF = 1,$$

  i.e., the critical density of WH–frames.

- For underspread environments the matching of a prototype signal (window, transmission pulse) has been approximately reduced to the simple rule:

$$\frac{T_\gamma}{F_\gamma} = \frac{\tau_0}{\nu_0},$$

  where T_γ and F_γ are the duration/bandwidth (defined via second order moments) of the prototype signal.

  Any time–frequency shifted version of such a matched prototype signal is an approximate eigensignal of the underspread operator and the generalized Weyl symbol determines the corresponding eigenvalue:

$$\mathbf{H}\gamma^{(t,f)} \approx L_H(t,f)\gamma^{(t,f)}, \qquad \text{with} \quad \gamma^{(t,f)} \stackrel{\text{def}}{=} \mathbf{S}^{(t,f)}\gamma.$$

6.2 Future Research

In what follows we list various open problems and potentially fruitful connections for future research:

- The asymptotic theory of underspread operators is conceptually equivalent to the asymptotic relation between *quantum* mechanics and *classical* mechanics [359] or between *wave* optics and *geometric* optics. These asymptotics are utilized in mathematical physics by the so–called Wentzel–Kramers–Brillouin (WKB) method (see [241, p.208] for the quantum mechanical context or [396, p.307] for the wave propagation context). In order to illustrate this parallelism recall that the basic idea of the underspread asymptotics can be formulated by the commutator of time and frequency shift operator:

$$\|\mathbf{M}_\nu\mathbf{T}_\tau - \mathbf{T}_\tau\mathbf{M}_\nu\|_\infty = \|\mathbf{S}^{(\tau,\nu)}(1/2) - \mathbf{S}^{(\tau,\nu)}(-1/2)\|_\infty = 2\sin\pi\tau\nu.$$

The analog result for the Schrödinger representation of the Heisenberg group [125] is

$$\left\| \mathbf{R}(-hp, q, 1/2\, h^2 pq) - \mathbf{R}(-hp, q, -1/2h^2 pq) \right\|_\infty = 2 \sin \pi pq h^2,$$

where h is Planck's constant. Hence, $\tau\nu \to 0$ is mathematically equivalent to $h \to 0$ in so far as the commutator of the fundamental building blocks (time and frequency shifts in our case, position and momentum in the quantum context) approaches zero. But $h \to 0$ characterizes exactly the asymptotic correspondence between classical and quantum mechanics. However, while the mathematical background is identical, the details are quite different [359]. The WKB method is used for the approximate, theoretical solution of eigenvalue problems involving (noncompact) partial differential operators while our underspread theory is based on a subspace of Hilbert–Schmidt operators in a way that is easily extended to a discrete, numerical setting. Notwithstanding these fundamental differences, we feel that the parallelism with the WKB method is certainly a promising direction for further investigations.

- One of the basic problems associated with our underspread definition (sharply restricted spreading function) is the fact that jointly underspread operators do not establish an algebra, because the twisted convolution enlarges the support of the spreading function just as a regular convolution. One possible way to overcome this problem is by defining a truncated, twisted product as:

$$\mathbf{G} \otimes \mathbf{H} \stackrel{\text{def}}{=} \sum_l \sum_m L_G(lT, mF) L_H(lT, mF) \widehat{\mathbf{P}}^{(lT, mF)},$$

  where $\widehat{\mathbf{P}}$ is the matched prototype operator of the underspread environment. It is easy to see that $\mathbf{G} \otimes \mathbf{H}$ defines a commutative algebra of jointly underspread operators. It is very likely that this property comes at the cost of an increased approximation error (w.r.t. the natural operator product $\mathbf{GH}$) compared to the continuously defined twisted product.

- For the derivation of key results about underspread operators we have often used a short–time Fourier transform (STFT) based on a (matched) Gaussian window. Such an STFT is closely related to the *Bargmann transform* [175, 125]:

$$(\mathcal{B}x)(\mu) \stackrel{\text{def}}{=} e^{(\pi/2)|\mu|^2} e^{-j2\pi tf/2} STFT_x^{(\gamma)}(t, f),$$

  with

$$\gamma(t) \stackrel{\text{def}}{=} 2^{1/4} e^{-\pi t^2} \qquad \text{and} \qquad \mu \stackrel{\text{def}}{=} t + jf.$$

  For $x \in \mathrm{L}_2(\mathbb{R})$, $\mathcal{B}x$ is an entire analytic function. Hence, we conjecture that via the Bargmann transform one can learn more about the "inner symmetry" of time–frequency representations by analytic function theory.

- This thesis was exclusively devoted to a continuous–time, $L_2(\mathbb{R})$ setup. However, we know that all basic results essentially carry over to the *finite (periodic), discrete, and finitely discrete case.* Instead of performing such a generalization in a piecemeal fashion (tedious computations leading to almost predictable results), we suggest to define underspread operators on *locally compact abelian (LCA) groups* that enable a combined, abstract treatment of the standard engineering setups [390, 155, 284].

- The use of *generalized time–frequency sampling lattices* is a natural extension of the Gabor window matching theory (where the discussion was restricted to rectangular sampling grids). We expect that given an underspread operator with elliptical spreading constraint, a hexagonal time–frequency sampling lattice leads to better performance in the sense of approximate diagonalization. Just as stated for the previous point, the theory of LCA groups allows a straightforward treatment of generalized (nonseparable) time–frequency sampling lattices.

APPENDIX A: Linear Operator Theory

This appendix summarizes some facts of linear operator theory matched to the scope of this thesis. With this summary we hope to give a reader with usual engineer's mathematical level the necessary mathematical definitions to follow the main ideas. A comprehensive monograph about standard linear operator theory (on engineer's level) is [252].

A.0.1 Representation of Linear Operators

A linear operator $\mathbf{H}$ maps an input signal $x(t)$ onto an output signal $(\mathbf{H}x)(t)$ in a way such that a linear superposition law holds:

$$(\mathbf{H}[\mu_1 x_1 + \mu_2 x_2])(t) = \mu_1 (\mathbf{H}x_1)(t) + \mu_2 (\mathbf{H}x_2)(t), \tag{A.1}$$

where μ_1 and μ_2 are two arbitrary complex numbers. Unique representation of $\mathbf{H}$ means implicit or explicit knowledge of any input–output mapping $x \mapsto \mathbf{H}x$. With the fundamental property (A.1) in mind it is clear that the explicit setup of all input–output mappings is highly redundant (this would be necessary in the case of a general nonlinear operator). Once we know the output $(\mathbf{H}x_0)(t)$ for a certain $x_0(t)$ we know the output for any scalar multiple $\mu x_0(t)$. In this work the linear operators usually act on $L_2(\mathbb{R})$, where we can write any signal in terms of an orthonormal basis:

$$x(t) = \sum_{k=1}^{\infty} \langle x, u_k \rangle u_k(t) \qquad \text{with} \qquad \langle x, u_k \rangle = \int_t x(t) u_k^*(t) dt. \tag{A.2}$$

The action of the operator on $x(t)$ can be traced back to its action on the basis elements $u_k(t)$, where we assume that the sum actually converges (the next section will provide the condition on $\mathbf{H}$ for the convergence of (A.2)):

$$(\mathbf{H}x)(t) = \sum_{k=1}^{\infty} \langle x, u_k \rangle (\mathbf{H}u_k)(t).$$

We can obtain an entirely discrete representation of $\mathbf{H}$ by expanding the output signal in terms of the basis:

$$\langle \mathbf{H}x, u_k \rangle = \sum_{k'=1}^{\infty} \langle x, u_{k'} \rangle \langle \mathbf{H}u_{k'}, u_k \rangle. \tag{A.3}$$

Hence, we have a discrete representation of the operator $\mathbf{H}$ by the matrix $H(k,k') = \langle \mathbf{H}u_{k'}, u_k \rangle$, provided that (A.2) converges.

However, the usual operator representation is by a continuous kernel $h(t,s)$ in the form of an integral operator:

$$(\mathbf{H}x)(t) = \int_s h(t,s) x(s) ds. \tag{A.4}$$

The kernel is just the continuous counterpart of the above defined matrix $H(k,k')$, with the main difference that now the representation is formally based on the continuously parametrized "basis" of delta pulses:

$$h(t,t_0) = (\mathbf{H}\delta_{t_0})(t), \qquad \text{with} \qquad \int_t x(t)\delta_{t_0}(t) dt = x(t_0).$$

A.0.2 Hilbert–Schmidt Operators

In the bulk of this work we restrict the discussion on operators with square–integrable kernel $h(t,s)$:

$$\int_t \int_s |h(t,s)|^2 \, dt \, ds = \|\mathbf{H}\|^2 < \infty. \tag{A.5}$$

An operator that meets this requirement is called Hilbert–Schmidt (HS). The norm $\|\mathbf{H}\|^2$ is called *Hilbert–Schmidt norm*[1] and may be defined — without explicit use of an operator representation — just in terms of the action of $\mathbf{H}$ onto an (arbitrary) orthonormal basis $\{u_k\}$:

$$\int_t \int_s |h(t,s)|^2 \, dtds = \|\mathbf{H}\|^2 = \sum_{k=1}^{\infty} \|\mathbf{H}u_k\|^2. \tag{A.6}$$

Hilbert–Schmidt operators always possess discrete representations in the form of a (potentially infinite dimensional) matrix $H(k,k') = \langle \mathbf{H}u_{k'}, u_k \rangle$. Once this is granted, the Hilbert–Schmidt norm of $\mathbf{H}$ is just the Frobenius norm of $H(k,k')$. It is furthermore of practical importance that one can always find finite–rank approximations $\mathbf{H}_N$ of a Hilbert–Schmidt operator $\mathbf{H}$ in a way such that

$$\|\mathbf{H} - \widetilde{\mathbf{H}}_N\|^2 \leq \epsilon \qquad \text{for some } N < \infty. \tag{A.7}$$

Thus by an appropriate choice of the basis $\{u_k\}$ in (A.3) we can *represent a Hilbert–Schmidt operator by a finite–dimensional matrix with arbitrary good precision.* This is of fundamental relevance for the numerical simulation of such operators. Unfortunately, there exist technically important operators that are not of HS–type, most prominently the linear time–invariant operators. The differences and parallels in the treatment of HS–type and non–HS operators will be discussed in the following sections.

HS Operators as Elements of a Hilbert Space. In various applications of linear operators it is useful to consider the operators as elements of a Hilbert space. This is particularly simple for Hilbert–Schmidt operators, where one defines an inner product according to

$$\langle \mathbf{H}, \mathbf{G} \rangle \stackrel{\text{def}}{=} \sum_{k=1}^{\infty} \langle \mathbf{H}u_k, \mathbf{G}u_k \rangle = \int_t \int_f (\mathbf{H})(t,s) \, (\mathbf{G})^*(t,s) dt \, ds. \tag{A.8}$$

Here, $\{u_k(t)\}$ is an arbitrary orthonormal basis of $L_2(\mathbb{R})$. The norm induced by the inner product (A.8) is indeed the Hilbert–Schmidt norm (see (A.5) for the definition). The linear map from the kernel $(\mathbf{H})(t,s)$ onto its operator $\mathbf{H}$ is thus an isometric isomorphism from $L_2(\mathbb{R}^2)$ onto the Hilbert space of Hilbert–Schmidt operators.

A.0.3 Eigenvalue and Singular Value Decompositions

One of the key ideas in the study of linear operators is to split the rather complicated general input–output relation into simpler "subsystems" in order to understand the behavior of the operator. One way of obtaining such decompositions is to consider the special case of an *eigensignal.* An input signal $x(t)$ is called eigensignal of the operator $\mathbf{H}$ when it meets the following equation

$$(\mathbf{H}x)(t) = \lambda x(t) \tag{A.9}$$

i. e., the output signal $(\mathbf{H}x)(t)$ is a weighted version of the input signal $x(t)$; the (generally complex) factor λ is the corresponding *eigenvalue.*

[1] In this respect, we deviate from the usual notation in the mathematical literature, where the operator norm is defined as $\|\mathbf{H}\|_\infty = \sup\{\|\mathbf{H}x\| : \|x\| = 1\}$. One has of course $\|\mathbf{H}\|_\infty \leq \|\mathbf{H}\|$.

Solution of (A.9) yields a manifold of eigensignals $u_k(t)$ and corresponding eigenvalues λ_k. The properties of these eigenvalues allow to distinguish subclasses of linear operators.

Normal Operator. If the eigensignals $u_k(t)$ set up a complete orthonormal basis of $L_2(\mathbb{R})$ then the operator is called *normal*. A usual, equivalent definition of a normal operator is given by the property

$$\mathbf{H}\mathbf{H}^* = \mathbf{H}^*\mathbf{H}, \tag{A.10}$$

i. e., a normal operator commutes with its adjoint. (The adjoint of an operator with kernel $h(t,s)$ has the kernel $h^*(s,t)$, where the asterisk denotes complex conjugation.)

Complete solution of the eigenvalue problem yields a useful representation of a normal operator, namely the spectral decomposition

$$h(t,s) = \sum_{k=1}^{\infty} \lambda_k u_k(t) u_k^*(s). \tag{A.11}$$

Self–Adjoint Operator. If, additionally, the eigenvalues are real–valued, then the operator $\mathbf{H}$ is self–adjoint. A self–adjoint operator is defined by the requirement:

$$\mathbf{H} = \mathbf{H}^*. \tag{A.12}$$

The self–adjoint operators represent a subclass of the normal operators.

Projection Operator. An important special case of a self–adjoint operator is the orthogonal projection operator which is marked out by being idempotent:

$$\mathbf{P}^2 = \mathbf{P}, \qquad \text{and} \qquad \mathbf{P} = \mathbf{P}^*. \tag{A.13}$$

The projection operator is in one–to–one correspondence to a linear signal space $\mathcal{S}$. It provides the formal solution of the fundamental problem of finding that signal $x_{\mathcal{S}} \in \mathcal{S}$ which has the minimal $L_2(\mathbb{R})$–distance to a given signal $x \notin \mathcal{S}$:

$$\left(\mathbf{P}_{\mathcal{S}} x\right)(t) = \arg\min_{x_{\mathcal{S}}} \|x - x_{\mathcal{S}}\|^2. \tag{A.14}$$

Note that the requirement for idempotency alone admits skew projections too. The eigenvalues of a projection operator are either 0 or 1, the spectral decomposition is given by

$$\left(\mathbf{P}\right)(t,s) = \sum_{k=1}^{N} u_k(t) u_k^*(s) = \sum_{k=1}^{N} \left(\mathbf{P}_k\right)(t,s), \tag{A.15}$$

where N is the dimension of the corresponding signal space and $\{u_k\}$ is any orthonormal basis spanning the corresponding signal space.

Having set up the notion of a projection operator we can reformulate the spectral decomposition of a normal operator (A.11) as a weighted sum of projections:

$$h(t,s) = \sum_{k=1}^{\infty} \lambda_k \mathbf{P}_k, \tag{A.16}$$

where the projection operators establish a so–called resolution of the identity,

$$\mathbf{I} = \sum_{k=1}^{\infty} \mathbf{P}_k, \qquad \mathbf{P}_k \mathbf{P}_l = \mathbf{P}_k \delta_{kl}. \tag{A.17}$$

Important examples of (infinite–rank) projections are the (ideal) band–limitation operator $\mathbf{B}_F$ and the time–limitation operator $\mathbf{D}_T$ with kernels:

$$\left(\mathbf{B}_F\right)(t,s) = \frac{\sin\left(2\pi f(t-s)\right)}{\pi(t-s)}; \qquad \left(\mathbf{D}_T\right)(t,s) = \delta(t-s)\chi_{[-T,T]}(t). \tag{A.18}$$

Unitary Operator. A unitary operator preserves the norm (energy) of a signal, i.e.

$$\|\mathbf{U}x\|^2 = \|x\|^2. \tag{A.19}$$

This property is equivalent to requiring

$$\mathbf{U}\mathbf{U}^* = \mathbf{U}^*\mathbf{U} = \mathbf{I}; \tag{A.20}$$

a unitary operator is thus normal and all eigenvalues are unimodular $|\lambda| = 1$. An, in our context, important example of a unitary operator is the time–frequency shift operator defined as

$$\left(\mathbf{S}^{(\tau,\nu)}x\right)(t) = x(t-\tau)e^{j2\pi\nu t}, \tag{A.21}$$

where the superscripts τ and ν denote the amount of time shift and frequency shift, respectively. The (generalized) eigenfunctions of the time–frequency shift operators are frequency–shifted versions of so–called chirp signals with chirp rate equal to $\frac{\nu}{\tau}$,

$$u_\mu(t) = e^{j2\pi(\frac{\nu}{2\tau}t^2+\mu t)}, \qquad \lambda_\mu = e^{j2\pi(\frac{\nu}{2}-\mu)\tau}, \qquad \mu \in \mathbb{R} \tag{A.22}$$

where λ_μ are the corresponding eigenvalues and $\mu \in \mathbb{R}$.

Singular Value Decomposition. For a general HS operator the eigenvalue problem may not yield a complete orthonormal basis of eigensignals. In this case, one has to study the eigenvalue problems of the composite operators $\mathbf{H}\mathbf{H}^*$ and $\mathbf{H}^*\mathbf{H}$ [252]; these operators are HS and nonnegative self–adjoint, and thus yield two complete orthonormal bases of so–called *singular signals*. The right singular signals $v_k(t)$ are the eigensignals of $\mathbf{H}\mathbf{H}^*$ and the left singular signals $u_k(t)$ are the eigensignals of $\mathbf{H}^*\mathbf{H}$. The two composite operators have the same (positive, real–valued) eigenvalues whose square root are the so–called *singular values* σ_k. This leads to a decomposition valid for any compact operator $\mathbf{H}$,

$$h(t,s) = \sum_{k=1}^{\infty} \sigma_k u_k(t) v_k^*(s), \tag{A.23}$$

the *singular value decomposition*, well–known in matrix theory. The Hilbert–Schmidt norm can be expressed by the singular values:

$$\sum_{k=1}^{\infty} \sigma_k^2 = \|\mathbf{H}\|^2. \tag{A.24}$$

Note that the singular value decomposition of a normal operator is closely related to the corresponding spectral decomposition, one has

$$\sigma_k = |\lambda_k| \qquad\qquad v_k(t) = e^{-j\arg\{\lambda_k\}}u_k(t),$$

where $u_k(t)$ are the eigensignals and λ_k are the eigenvalues of the normal operator. For a nonnegative self–adjoint operator the left and right singular signals are identical and the singular value and spectral decomposition are equivalent.

A specific example of a nonnormal HS operator is given by the cascade composition of the band–limitation and time–limitation operator:

$$\mathbf{H} = \mathbf{D}_T\mathbf{B}_F \tag{A.25}$$

the left singular signals of $\mathbf{H}$ are the prolate spheroidal wave functions and the right singular signals are just the Fourier dual, i.e., their Fourier transform is given by the prolate spheroidal wave functions [267].

Operator Square Root. To a given positive–semidefinite, self–adjoint operator $\mathbf{R}$ one can define a square–root operator formally by:

$$\mathbf{H} = \sqrt{\mathbf{R}} \quad \Longleftrightarrow \mathbf{R} = \mathbf{H}^2. \tag{A.26}$$

However, $\mathbf{H}$ is not uniquely defined because it is easy to see that given one specific $\mathbf{H}$ with $\mathbf{R} = \mathbf{H}^2$, any other operator defined as $\mathbf{H}' = \mathbf{HU}$ with unitary $\mathbf{U}$ also satisfies (A.26). One usually assumes that $\sqrt{\mathbf{R}}$ is self–adjoint and positive–semidefinite, which leads to a well–defined operator.

Trace Class Operators. The *trace* of a linear integral operator is defined as

$$\mathrm{tr}\mathbf{H} \stackrel{\mathrm{def}}{=} \int_t (\mathbf{H})(t,t)dt. \tag{A.27}$$

Operators with summable singular values, i.e.,

$$\sum_{k=1}^{\infty} \sigma_k = \mathrm{tr}\sqrt{\mathbf{HH}^*} = M < \infty. \tag{A.28}$$

are called *trace class*. Trace class operators form a subset of Hilbert–Schmidt operators.

Inner Product of Self–Adjoint HS Operators. For two Hilbert–Schmidt operators $\mathbf{H}$ and $\mathbf{G}$ the inner product is equal to the trace of the product of $\mathbf{H}$ and $\mathbf{G}^*$:

$$\langle \mathbf{H}, \mathbf{G} \rangle = \mathrm{tr}\left\{\mathbf{HG}^*\right\}. \tag{A.29}$$

This gives the following representation for the HS norm:

$$\|\mathbf{H}\|^2 = \mathrm{tr}\left\{\mathbf{HH}^*\right\} = \mathrm{tr}\left\{\mathbf{H}^*\mathbf{H}\right\} \tag{A.30}$$

A.0.4 Spectral Decomposition of Normal Non–HS Operator

Just as a formal completion we discuss the practically important case of normal operators whose kernel is not square–integrable. In the context of this work there are two important examples:

- **LTI operator.** Linear time–invariant (LTI) operators are predominant in the engineering literature appearing either as LTI systems or as the correlation operator of a wide–sense stationary process. The kernel of an LTI operator reduces to a one–dimensional function and its action is a time–domain convolution:

$$h(t,s) = \hat{h}(t-s), \qquad \Longleftrightarrow \qquad (\mathbf{H}x) = \int_\tau \hat{h}(\tau)x(t-\tau)d\tau. \tag{A.31}$$

- **LFI operator.** The Fourier dual of an LTI operator is the linear frequency–invariant operator that appears as a linear modulation, time–domain windowing operator or corresponding to the correlation kernel of nonstationary white noise; the kernel is given by

$$h(t,s) = m(t)\delta(t-s), \qquad \Longleftrightarrow \qquad (\mathbf{H}x) = m(t)x(t). \tag{A.32}$$

The generalized eigenvalues of an LTI or LFI operator form a *continuous spectrum*, and the summation of the spectral decomposition of HS operators (A.11) formally carries over to an integration,

$$h(t,s) = \int_\mu \Lambda(\mu)u_\mu(t)u_\mu^*(s)d\mu, \tag{A.33}$$

where Λ_μ is the continuous eigenvalue spectrum and $u_\mu(t)$ is continuously parametrized "orthogonal basis" of eigensignals (these are not in $L_2(\mathbb{R})$). In the case of an LTI operator one can write,

$$h(t,s) = \int_f H(f)e^{j2\pi ft}e^{-j2\pi fs}df = \mathcal{F}^{-1}_{f\to(t-s)}\left\{H(f)\right\}. \tag{A.34}$$

A.0.5 Analysis–Modification–Synthesis Interpretation

The key advantage of the eigen/singular value decompositions (A.11) and (A.23) lies in the fact that they allow to replace the rather nontransparent I–O–relation (4.3) by the more convenient analysis–modification–synthesis scheme. That means in particular, the action of a compact operator on the input signal $x(t)$ can be split up in three operations:

1. *Analysis* — expansion of the input signal $x(t)$ into a weighted sum of the basis $v_k(t)$, the coefficients c_k are given by

$$c_k = \langle x, v_k \rangle = \int_t x(t) v_k^*(t) dt. \tag{A.35}$$

2. *Multiplicative Modification* — i. e., weighting of the coefficients $c(k)$ by the singular values σ_k of the operator, according to

$$\tilde{c}_k = c_k \sigma_k. \tag{A.36}$$

3. *Synthesis* — summation of the modified expansion yields the output signal

$$(\mathbf{H}x)(t) = \sum_{k=1}^{\infty} \tilde{c}_k u_k(t). \tag{A.37}$$

If the operator is regarded as a filter, then the actual filtering is performed in the second step; the analysis and the synthesis step are based on unitary (energy–preserving) transforms. With the advent of powerful digital signal processing facilities this concept has been realized in various applications. In this work modifications of this concept are considered in the sense that one admits (theoretically suboptimal but numerically efficient) nonorthogonal Weyl–Heisenberg–structured bases in the analysis and synthesis parts.

A.0.6 Unitary Equivalence

Two linear operator $\mathbf{H}$ and $\mathbf{G}$ are called *unitarily equivalent* whenever there exists a unitary operator $\mathbf{U}$ such that:

$$\mathbf{G} = \mathbf{U}\mathbf{H}\mathbf{U}^* \qquad \Leftrightarrow \qquad \mathbf{H} = \mathbf{U}^*\mathbf{G}\mathbf{U}. \tag{A.38}$$

Unitarily equivalent operators have an identical eigen/singular value distribution, while the eigen/-singular signals are interrelated via the unitary transform $\mathbf{U}$.

In the context of this thesis the following case of unitary equivalence is particularly important:

$$\widetilde{\mathbf{H}} = \mathbf{S}^{(\tau,\nu)}\mathbf{H}\mathbf{S}^{(\tau,\nu)*}, \tag{A.39}$$

where $\mathbf{S}^{(\tau,\nu)}$ is the time–frequency shift operator as defined in (A.21). The spectral decomposition of $\widetilde{\mathbf{H}}$ is characterized by identical singular values as the original operator $\mathbf{H}$, while the singular signals are time–frequency shifted versions of the singular signals of $\mathbf{H}$:

$$(\mathbf{H})(t,s) = \sum_{k=1}^{\infty} \sigma_k u_k(t) v_k^*(s) \qquad \Rightarrow \qquad \left(\widetilde{\mathbf{H}}\right)(t,s) = \sum_{k=1}^{\infty} \sigma_k \tilde{u}_k(t) \tilde{v}_k^*(s) \tag{A.40}$$

with

$$\tilde{u}_k(t) = \left(\mathbf{S}^{(\tau,\nu)} u_k\right)(t) = u_k(t-\tau) e^{j2\pi\nu t} \qquad \text{and} \qquad \tilde{v}_k(t) = \left(\mathbf{S}^{(\tau,\nu)} v_k\right)(t) = v_k(t-\tau) e^{j2\pi\nu t}.$$

APPENDIX B: Generalized Spreading Function

Time–delay and (narrowband) Doppler shift are the predominant energy preserving (unitary) effects of a time–varying multipath wave propagation channel. This is the context where the name spreading function (short for delay–Doppler spread function) was introduced. In quantum mechanics, time–shift and frequency–shift operators play an important role in so far as they establish representations of the Heisenberg group. Before we discuss the definition of the generalized spreading function, we point out the basic mathematical problem that leads to the parametrization of the generalized spreading function.

The time—shift (translation) operator, defined by

$$\left(\mathbf{T}_\tau x\right)(t) = x(t-\tau)$$

may be seen as the fundamental, unitary building block of linear time–invariant operators in the sense that

$$\mathbf{H}x = \int_\tau \bar{h}(\tau)\mathbf{T}_\tau x \, d\tau,$$

where $\bar{h}(\tau)$ is the impulse response (convolution kernel) of $\mathbf{H}$.

Time–varying operators may also cause frequency shifts, defined by:

$$\left(\mathbf{M}_\nu x\right)(t) = x(t)e^{-j2\pi\nu t},$$

where, in engineering contexts, $\mathbf{M}_\nu$ is often called modulation operator.

Time shift operators and frequency shift operators, respectively, form a commutative group corresponding to the additive group on the real line:

$$\mathbf{T}_{\tau_1}\mathbf{T}_{\tau_2} = \mathbf{T}_{\tau_2}\mathbf{T}_{\tau_1} = \mathbf{T}_{\tau_1+\tau_2}, \tag{B.1}$$

$$\mathbf{M}_{\nu_1}\mathbf{M}_{\nu_2} = \mathbf{M}_{\nu_2}\mathbf{M}_{\nu_1} = \mathbf{M}_{\nu_1+\nu_2}. \tag{B.2}$$

However, the time–shift operator and the frequency–shift operator do not mutually commute:

$$\mathbf{T}_\tau\mathbf{M}_\nu = \mathbf{M}_\nu\mathbf{T}_\tau e^{-j2\pi\tau\nu},$$

i.e., they commute only up to an unimodular factor. This causes fundamental complications, (i) time–frequency shift operators do not represent the additive group on $\mathbb{R}\times\mathbb{R}$:

$$\mathbf{T}_{\tau_1+\tau_2}\mathbf{M}_{\nu_1+\nu_2} \neq \mathbf{T}_{\tau_1}\mathbf{M}_{\nu_1}\mathbf{T}_{\tau_2}\mathbf{M}_{\nu_2}, \tag{B.3}$$

and (ii) in the definition of a unitary, time–frequency shift operator one has to admit at least two natural definitions:

$$\mathbf{S}^{(\tau,\nu)}(1/2) \stackrel{\text{def}}{=} \mathbf{M}_\nu\mathbf{T}_\tau,$$

$$\mathbf{S}^{(\tau,\nu)}(-1/2) \stackrel{\text{def}}{=} \mathbf{T}_\tau\mathbf{M}_\nu.$$

However, by splitting up the time shift and the frequency shift one can define more general versions of a time–frequency shift operator, these various versions differ only by an unimodular factor. We define a general α–parametrized time–frequency shift operator by:

$$\left(\mathbf{S}^{(\tau,\nu)}(\alpha)x\right)(t) \stackrel{\text{def}}{=} x(t-\tau)e^{j2\pi\nu t}e^{j2\pi\nu\tau(\alpha-1/2)}. \tag{B.4}$$

Alternatively, one may define the general shift operator by using the following split–up:

$$\begin{aligned}\mathbf{S}^{(\tau,\nu)}(\alpha) &= \mathbf{M}_{\nu(1/2+\alpha)}\mathbf{T}_\tau\mathbf{M}_{\nu(1/2-\alpha)}\\ &= \mathbf{T}_{\tau(1/2-\alpha)}\mathbf{M}_\nu\mathbf{T}_{\tau(1/2+\alpha)}.\end{aligned}$$

For arbitrary, fixed α, the time–frequency shift operators do not form a 2D group as indicated in (B.3). However, the time–frequency shift operators are closely related to the 3D *Weyl–Heisenberg group* which is defined by the following group law:

$$(\tau,\nu,\mu)\odot(\tau',\nu',\mu') \stackrel{\text{def}}{=} \left(\tau+\tau',\nu+\nu',\mu+\mu'+\frac{1}{2}(\tau\nu'-\nu\tau')\right).$$

The unitary *Schrödinger* representation of the *Weyl–Heisenberg group* is defined as [125]

$$\left(\mathbf{R}(\tau,\nu,\mu)x\right)(t) \stackrel{\text{def}}{=} x(t+\tau)e^{j2\pi(\mu+\nu t+\frac{1}{2}\tau\nu)}. \tag{B.5}$$

And, in fact, one can view our α–parametrized time–frequency shift operator as a proper subset of the Schrödinger representation where the parameter α appears in the third coordinate as follows:

$$\mathbf{S}^{(\tau,\nu)}(\alpha) = \mathbf{R}(-\tau,\nu,\tau\nu\alpha).$$

This shows that the time–frequency shift operators carry much of the beautiful symmetry of the Weyl–Heisenberg group. In representation theory one calls such a family of unitary operators with "hidden group structure" *projective representation* [50] whenever one has a "quasi–homomorphism" in the form

$$\mathbf{H}(x)\mathbf{H}(y) = c(x,y)\mathbf{H}(x+y), \tag{B.6}$$

where $\mathbf{H}(x)$ is a parametrized family of unitary operators and $c(x,y)$ is an unimodular factor. In the context of this work there are two cases where the hidden group structure of the time–frequency shift operator comes out with striking consequences:

- For time–frequency periodic operators with $TF = 1/n$ (see (B.31)) the time–frequency shift operators form a commutative subgroup of the Weyl–Heisenberg group.

- Time–frequency shifting of operators in the sense of their symbols (see (C.15)) can be considered as the action of a tensor product of time–frequency shift operators. It is easy to see that such a tensor product is isomorphic to the additive group on $\mathbb{R}\times\mathbb{R}$.

The case $\alpha = 0$ corresponds to the Weyl correspondence and it is indeed historically correct to put our (τ,ν)–domain formula before the (t,f)–domain formulas (nowadays it is more natural to define the Weyl correspondence in the (t,f)–domain based on the Wigner distribution, see Appendix C). Among the general versions of a time–frequency shift operator Weyl's symmetrical time–frequency shift operator is marked out by various symmetry properties:

- Taking the adjoint leads to an inverse shift with identical unimodular factor:

  $$\mathbf{S}^{(\tau,\nu)*}(0) = \mathbf{S}^{(-\tau,-\nu)}(0).$$

  This is not true for any other definition of the time–frequency shift operator. For $\alpha = 1/2$ one has:

  $$\mathbf{S}^{(\tau,\nu)*}(1/2) = \mathbf{S}^{(-\tau,-\nu)}(-1/2).$$

- One may obtain Weyl's operator by an infinitesimal split up of the time and frequency shifts (in the sense of a skew shift that follows the line between $(0,0)$ and (τ,ν)):

  $$\mathbf{S}^{(\tau,\nu)*}(0) \stackrel{\text{def}}{=} \lim_{N\to\infty}\Pi_{n=1}^{N}\mathbf{T}\left(\frac{\tau}{N}\right)\mathbf{M}\left(\frac{\nu}{N}\right) = \lim_{N\to\infty}\Pi_{n=1}^{N}\mathbf{M}\left(\frac{\nu}{N}\right)\mathbf{T}\left(\frac{\tau}{N}\right).$$

- Weyl's original derivation was based on an intuitive argument based on exponential operators [125, p.79], i.e., one can formally define

$$\mathbf{S}^{(\tau,\nu)*}(0) \stackrel{\text{def}}{=} e^{j2\pi(\tau\mathbf{F}-\nu\mathbf{T})} \tag{B.7}$$

where $\mathbf{F}$ is essentially a differential operator,

$$(\mathbf{F}x)(t) = \frac{1}{2j\pi}\frac{dx}{dt},$$

and $\mathbf{T}$ is a specific multiplication operator,

$$(\mathbf{T}x)(t) = tx(t).$$

Weyl's essential line of reasoning was to replace the elementary functions of the symplectic Fourier transform (as a building block for 2D functions),

$$c(\tau,\nu) = e^{j2\pi(\tau f-\nu t)}$$

by an operator as a building block for linear operators. Weyl suggested to do this in a canonical way in so far as time t can be interpreted as the spectrum of the operator $\mathbf{T}$ and frequency f can be interpreted as the spectrum of the operator $\mathbf{F}$, i.e.,

$$\begin{array}{rcl} f & \mapsto & \mathbf{F} \\ t & \mapsto & \mathbf{T} \\ & \Downarrow & \\ c(\tau,\nu) & \mapsto & \mathbf{S}^{(\tau,\nu;0)} \end{array}$$

(Recall that the essential idea of a symbolic calculus is to replace the calculus of operators by the calculus of their spectra.)

After this mathematical introduction we return to the basic definition of a spreading function. The amount of potential time and frequency shifts caused by a linear operator $\mathbf{H}$ can be specified by the *asymmetrical spreading function* (also known as delay–Doppler–spread function [334]

$$S_H^{(1/2)}(\tau,\nu) \stackrel{\text{def}}{=} \int_t h(t,t-\tau)e^{-j2\pi\nu t}dt\,. \tag{B.8}$$

The asymmetrical spreading function establishes an infinitesimal decomposition of $\mathbf{H}$ into $\alpha = 1/2$–type time–frequency shift operators:

$$\mathbf{H} = \int_\tau \int_\nu S_H^{(1/2)}(\tau,\nu)\mathbf{S}^{(\tau,\nu)}(1/2)d\tau d\nu. \tag{B.9}$$

The classical asymmetrical spreading function corresponds to the Kohn–Nirenberg calculus. However, as pointed out, there is no mathematical (nor physical) reason to prefer the $\alpha = 1/2$ version of the time–frequency shift operator. Hence, we define a generalized spreading function (GSF) as:

$$S_H^{(\alpha)}(t,f) \stackrel{\text{def}}{=} \int_t h\left(t+\left(\frac{1}{2}-\alpha\right)\tau, t-\left(\frac{1}{2}+\alpha\right)\tau\right)e^{-j2\pi\nu t}dt, \tag{B.10}$$

which, for $\alpha = 0$ gives the recently introduced *symmetrical spreading function* [205].

Following is a list of relevant properties and relations:

1. **Inner Product Definition.** With the inner product defined as (A.8) we can write the GSF formally as:

$$S_H^{(\alpha)}(\tau,\nu) = \left\langle \mathbf{H}, \mathbf{S}^{(\tau,\nu)}(\alpha) \right\rangle. \tag{B.11}$$

The spreading function of the time–frequency shift operator is formally given by a two–dimensional delta distribution:

$$S^{(\alpha)}_{S^{(\tau_0,\nu_0)}(\alpha)}(\tau,\nu) = \delta(\tau-\tau_0)\delta(\nu-\nu_0). \tag{B.12}$$

2. **Inversion.** The inversion formula is given by:

$$h(t,s) = \int_\nu S_H^{(\alpha)}(t-s,\nu) e^{j2\pi\nu[(1/2+\alpha)t+(1/2-\alpha)s]} d\nu. \tag{B.13}$$

3. **Interrelation with GWS.** The GSF is in Fourier correspondence to the generalized Weyl symbol:

$$S_H^{(\alpha)}(\tau,\nu) = \int_t \int_f L_H^{(\alpha)}(t,f) e^{-j2\pi(\nu t-\tau f)} dt\, df = \mathcal{F}_{t\to\nu}\mathcal{F}^{-1}_{f\to\tau}\left\{Z_H(t,f)\right\}. \tag{B.14}$$

4. **Bifrequency Domain Definition.** The bifrequency function $B_H(f,s)$ is defined as the kernel of an integral operator corresponding to the action of $\mathbf{H}$ on the frequency domain representations of the input and the output signal:

$$\left(\mathcal{F}\mathbf{H}x\right)(f) = \int_s B_H(f,s)X(s)ds.$$

By the nature of the Fourier dualism, the bifrequency–domain definition of the GSF is structurally similar to the time–domain definition:

$$S_H^{(\alpha)}(\tau,\nu) = \int_f B_H\left(f+\left(\frac{1}{2}-\alpha\right)\nu, f-\left(\frac{1}{2}+\alpha\right)\nu\right) e^{j2\pi f\tau}\, df.$$

5. **Mutual Interrelation.** The relation between the various members of the GSF is given by the multiplication by an unimodular function:

$$S_H^{(\alpha_1)}(\tau,\nu) = S_H^{(\alpha_2)}(\tau,\nu) e^{-j2\pi\nu\tau(\alpha_1-\alpha_2)}. \tag{B.15}$$

Hence, the magnitude of $S_H^{(\alpha)}(\tau,\nu)$ is α–invariant

$$\left|S_H^{(\alpha_1)}(t,f)\right|^2 = \left|S_H^{(\alpha_2)}(t,f)\right|^2. \tag{B.16}$$

6. **Bilinear Form.** The bilinear form can be written as

$$\langle \mathbf{H}x, y\rangle = \left\langle S_H^{(\alpha)}, A_{y,x}^{(\alpha)} \right\rangle. \tag{B.17}$$

where $A_{y,x}^{(\alpha)}(\tau,\nu)$ is the generalized cross–ambiguity function (see (F.19)).

7. **Unitarity.** The GSF preserves the inner product of two HS operators $\mathbf{H}$,$\mathbf{G}$:

$$\langle \mathbf{H}, \mathbf{G}\rangle = \left\langle S_H^{(\alpha)}, S_G^{(\alpha)} \right\rangle, \tag{B.18}$$

as a consequence, we have for the HS norm:

$$\int_\tau \int_\nu \left|S_H^{(\alpha)}(\tau,\nu)\right|^2 d\tau\, d\nu = \|\mathbf{H}\|^2, \tag{B.19}$$

8. **Spectral Decomposition.** The GSF of a HS operator can be written in terms of the generalized cross–ambiguity functions of the singular signals

$$S_H^{(\alpha)}(\tau,\nu) = \sum_{k=1}^{\infty} \sigma_k A_{u_k,v_k}^{(\alpha)}(\tau,\nu). \tag{B.20}$$

In the origin, the GSF equals the operator's trace:

$$S_H^{(\alpha)}(0,0) = \mathrm{tr}\mathbf{H}, \tag{B.21}$$

which is the Fourier–dual property of the GWS's "trace formula" (C.14). Sometimes useful is the following upper bound

$$\left|S_H^{(\alpha)}(\tau,\nu)\right| \leq \sum_{k=1}^{\infty} \sigma_k = \mathrm{tr}\sqrt{\mathbf{H}^*\mathbf{H}}. \tag{B.22}$$

9. **Adjoint Operator.** The generalized spreading function of the adjoint operator is given by:

$$S_{H^*}^{(\alpha)}(\tau,\nu) = S_H^{(\alpha)*}(-\tau,-\nu)e^{-j4\pi\tau\nu\alpha}, \tag{B.23}$$

the symmetrical spreading function is marked out by the fact that

$$S_{H^*}^{(0)}(\tau,\nu) = S_H^{(0)*}(-\tau,-\nu). \tag{B.24}$$

Note, that (B.23) implies:

$$\left|S_{H^*}^{(\alpha)}(\tau,\nu)\right| = \left|S_H^{(\alpha)}(-\tau,-\nu)\right|. \tag{B.25}$$

10. **Positive Operator.** The GSF of a self–adjoint, positive operator takes on its maximum in the origin:

$$\left|S_H^{(\alpha)}(0,0)\right|^2 \geq \left|S_H^{(\alpha)}(\tau,\nu)\right|^2. \tag{B.26}$$

11. **LTI Operator.** A linear time–invariant (LTI) operator does not introduce any frequency–shift of the input signal; this is reflected in the GSF which is ideally concentrated on the τ–axis:

$$S_H^{(\alpha)}(\tau,\nu) = \delta(\nu)h(\tau). \tag{B.27}$$

12. **LFI Operator.** Linear frequency–invariant operators are dual to the LTI operators and do not introduce any time–shift, accordingly, the GSF does not show any spread into the direction of τ

$$S_H^{(\alpha)}(\tau,\nu) = M(\nu)\delta(\tau). \tag{B.28}$$

13. **Time–Frequency Shifting.** We define time–frequency shifting of operators by

$$\mathbf{H}^{(t,f)} = \mathbf{S}^{(t,f)}\mathbf{H}\mathbf{S}^{(t,f)*},$$

where $\mathbf{S}^{(t,f)}$ is an arbitrary version of a time–frequency shift operator. The generalized spreading function of such a time–frequency shifted operator is a modulated version of the original operator:

$$S_{H^{(t,f)}}^{(\alpha)}(\tau,\nu) = S_H^{(\alpha)}(\tau,\nu)e^{-j2\pi(\nu t-\tau f)}. \tag{B.29}$$

14. **Time–Varying Tapped Delay Line Operator.** The time–varying tapped delay line is a popular, relatively simple model for time–varying multipath propagation channel, it is defined by

$$h(t,s) = \sum_{k=1}^{N} m_k(t)\delta(t-s-kT).$$

Its spreading function is distribution valued, it is ideally concentrated on T–spaced lines parallel to the ν–axis :

$$S_H^{(\alpha)}(\tau,\nu) = \sum_{k=1}^{N} \delta(\tau - kT) M_k(\nu) e^{-j2\pi\nu kT(\alpha-1/2)},$$

with $M_k(\nu) = (\mathcal{F}m_k)(\nu)$.

15. **Periodically Time–Varying Operator.** The generalized spreading function of a periodically time–varying operator with the kernel property $h(t,t') = h(t+lT, t'+lT)$, $l \in \mathbb{Z}$ is separable into delta–pulses in frequency direction at the multiples of the fundamental frequency $1/T$ and into continuous functions of τ:

$$S_H^{(\alpha)}(\tau,\nu) = \sum_n \rho_n^{(\alpha)}(\tau)\delta(\nu - \frac{n}{T}) \tag{B.30}$$

with

$$\rho_n^{(\alpha)}(\tau) = \frac{1}{T}\int_0^T h(t+(1/2-\alpha)\tau, t-(1/2+\alpha)\tau)e^{-j2\pi\frac{n}{T}t}dt.$$

16. **Time–Frequency Periodic Operator.** We define time–frequency periodic operators by

$$\mathbf{H}\mathbf{S}^{(lT,mF)} = \mathbf{S}^{(lT,mF)}\mathbf{H}, \qquad l,m \in \mathbb{Z} \tag{B.31}$$

The spreading function of such a time–frequency periodic operator is ideally concentrated on a dual grid:

$$S_H^{(\alpha)}(\tau,\nu) = \sum_l \sum_m \rho_{l,m}^{(\alpha)} \delta\left(\tau - \frac{l}{F}\right)\delta\left(\nu - \frac{m}{T}\right), \tag{B.32}$$

where the coefficients can always be defined as samples of the spreading function of a specific prototype operator $\mathbf{P}$

$$\rho_{m,l}^{(\alpha)} = \frac{1}{TF} S_P^{(\alpha)}\left(\frac{l}{F}, \frac{m}{T}\right),$$

which is defined by (provided that this sum makes sense)

$$\mathbf{H} = \sum_l \sum_m \mathbf{S}^{(lT,mF)}\mathbf{P}\mathbf{S}^{(lT,mF)*}.$$

In the context of the Gabor expansion and Weyl–Heisenberg frames, (B.32) is known as *Janssen representation* of the frame operator.

17. **Underspread Operators.** Given an operator whose GSF support lies inside a centered rectangle:

$$S_H^{(\alpha)}(\tau,\nu) = S_H^{(\alpha)}(\tau,\nu)\,\chi_{[-\tau_0,\tau_0]}(\tau)\,\chi_{[-\nu_0,\nu_0]}(\nu),$$

we call the operator underspread when $4\tau_0\nu_0 \ll 1$ and overspread in the converse case.

18. **Operator Composition.** Of key interest in the study of underspread operators is the expression for the spreading function of the composite operators:

$$S_{GH}^{(\alpha)}(\tau,\nu) = \int_{\tau'}\int_{\nu'} S_G^{(\alpha)}(\tau',\nu') S_H^{(\alpha)}(\tau-\tau',\nu-\nu') e^{-j2\pi\{\tau'\nu(\alpha+1/2)+\tau\nu'(\alpha-1/2)-2\tau'\nu'\alpha\}} d\tau'\, d\nu', \tag{B.33}$$

this relation is known as *twisted convolution* [125].

APPENDIX C: Generalized Weyl Correspondence

The generalized Weyl correspondence has been introduced in [204] as the formal unification of the classical Weyl correspondence [369, 125] and the more recent Kohn–Nirenberg correspondence [203, 125].

The *Weyl symbol* of an operator $\mathbf{H}$ with kernel $h(t,s)$ is defined as

$$L_H(t,f) \stackrel{\text{def}}{=} \int_\tau h\left(t+\frac{\tau}{2}, t-\frac{\tau}{2}\right) e^{-j2\pi f\tau} d\tau,$$

where the operator acts as $(\mathbf{H}x)(t) = \int_s h(t,s)x(s)ds$. The map $\mathbf{H} \mapsto L_H(t,f)$ is called *Weyl correspondence* [10, 11, 12].

The *Kohn–Nirenberg symbol* [203, 125] or, equivalently, *Zadeh's time–varying transfer function* [386] is defined as

$$Z_H(t,f) \stackrel{\text{def}}{=} \int_\tau h(t,t-\tau)e^{-j2\pi f\tau} d\tau.$$

Both the Weyl symbol $L_H(t,f)$ and the Kohn–Nirenberg symbol $Z_H(t,f)$ establish a unitary representation of Hilbert–Schmidt (HS) operators[2], they share a number of common properties and provide a largely parallel concept for Weyl–Heisenberg operator representation. The *generalized Weyl symbol* (GWS) is an α–parametrized family of unitary operator representations defined as

$$L_H^{(\alpha)}(t,f) \stackrel{\text{def}}{=} \int_\tau h\left(t+\left(\frac{1}{2}-\alpha\right)\tau, t-\left(\frac{1}{2}+\alpha\right)\tau\right) e^{-j2\pi f\tau} d\tau, \tag{C.1}$$

it gives the Weyl symbol in particular for $\alpha = 0$ and the Kohn–Nirenberg symbol for $\alpha = 1/2$. The parameter α comes up more naturally in the spreading domain along with the nonuniqueness of the time–frequency shift operator (see Appendix B).

The basic mathematical interpretation of the generalized Weyl symbol may be phrased as time–frequency parametrized "eigenvalue distribution". When $\mathbf{H}$ is a linear time–varying system then, in engineering terms, we may interpret the symbol as *time–varying transfer function*. When $\mathbf{H}$ is equal to a (positive semidefinite) correlation operator $\mathbf{R}_x$ then the symbol is equivalent to the *generalized Wigner–Ville spectrum* [121],

$$L_{R_x}^{(\alpha)}(t,f) = EW_x^{(\alpha)}(t,f),$$

which is a classical definition of a *time–varying power spectrum*.

Following is a list of important properties and relations:

1. **Inversion.** The inversion formula is given by:

$$h(t,s) = \int_f L_H^{(\alpha)}\left(\left(\frac{1}{2}+\alpha\right)t + \left(\frac{1}{2}-\alpha\right)s, f\right) e^{j2\pi f(t-s)} df. \tag{C.2}$$

2. **Interrelation with Spreading Function.** The GWS is the (symplectic) Fourier transform of the generalized spreading function (for the definition of the spreading function see Appendix B)

$$L_H^{(\alpha)}(t,f) = \int_\tau \int_\nu S_H^{(\alpha)}(\tau,\nu) e^{j2\pi(\nu t-\tau f)} d\tau\, d\nu = \mathcal{F}^{-1}_{\nu\to t}\mathcal{F}_{\tau\to f}\left\{S_H^{(\alpha)}(\tau,\nu)\right\}.$$

[2] See Appendix A for the basic definitions of linear operator theory.

3. **Bifrequency Domain Definition.** In terms of the bifrequency function $B_H(f,s)$ (see (4.4)), the GWS is given by

$$L_H^{(\alpha)}(t,f) = \int_\nu B_H\left(f + \left(\frac{1}{2} - \alpha\right)\nu, f - \left(\frac{1}{2} + \alpha\right)\nu\right) e^{j2\pi t\nu} d\nu.$$

4. **Unitarity.** The GWS preserves the inner product of the kernels $g(t,s)$ and $h(t,s)$ of two HS operators $\mathbf{G}$ and $\mathbf{H}$,

$$\langle g, h\rangle = \left\langle L_G^{(\alpha)}, L_H^{(\alpha)}\right\rangle = \langle \mathbf{G}, \mathbf{H}\rangle . \tag{C.3}$$

As a consequence, the L_2–norm of the GWS is just the HS norm (see (A.5)) of the corresponding operator:

$$\int_t \int_f \left|L_H^{(\alpha)}(t,f)\right|^2 dt\, df = \left\|L_H^{(\alpha)}(t,f)\right\|^2 = \|\mathbf{H}\|^2. \tag{C.4}$$

5. **LTI Operator.** For a linear time–invariant operator with convolution type kernel $h(t,s) = \bar{h}(t-s)$ the GWS gives the usual transfer function (Fourier transform of the operator's impulse response):

$$L_H^{(\alpha)}(t,f) = H(f) = \mathcal{F}_{(t-s)\mapsto f}\bar{h}(t-s). \tag{C.5}$$

6. **LFI Operator.** In the dual case of a linear, frequency–invariant operator (i.e., multiplication operator) with distribution type kernel $h(t,s) = \delta(t-s)m(t)$ the GWS is frequency independent:

$$L_H^{(\alpha)}(t,f) = m(t). \tag{C.6}$$

7. **Periodically Time–Varying Operator.** A periodically time–varying linear operator commutes with time–shifts that are multiples of a time–period T:

$$\mathbf{H}\mathbf{S}^{(lT,0)} = \mathbf{S}^{(lT,0)}\mathbf{H}, \qquad l \in \mathbb{Z}.$$

This is equivalent with the following requirement for the kernel of $\mathbf{H}$:

$$h(t,s) = h(t+lT, s+lT).$$

The GWS of such an operator is periodic w.r.t. time:

$$L_H^{(\alpha)}(t,f) = L_H^{(\alpha)}(t+lT,f). \tag{C.7}$$

8. **Time–Frequency Periodic Operator.** We define time–frequency periodic operators by

$$\mathbf{H}\mathbf{S}^{(lT,mF)} = \mathbf{S}^{(lT,mF)}\mathbf{H},$$

where T is the time period, F is the frequency period and $l, m \in \mathbb{Z}$. The GWS of such a time–frequency–periodic operator is doubly periodic

$$L_H^{(\alpha)}(t,f) = L_H^{(\alpha)}(t+lT, f+mF). \tag{C.8}$$

Moreover, given two time–frequency periodic operators $\mathbf{H}$,$\mathbf{G}$ and either one of the following assumptions:

- $TF = \frac{1}{n}$, $n \in \mathbb{N}$, and $|\alpha| = 1/2$

- $TF = \frac{1}{2n}$, $n \in \mathbb{N}$, and $\alpha = 0$

then, one has validity of the perfect symbol calculus:

$$L_H^{(\alpha)}(t,f) L_G^{(\alpha)}(t,f) = L_{HG}^{(\alpha)}(t,f).$$

Such time–frequency periodic operator arise as (linear combinations of) Weyl–Heisenberg frame operators defined as ($\gamma(t)$ is the prototype):

$$\mathbf{M}_\gamma \stackrel{\text{def}}{=} \sum_m \sum_l \mathbf{S}^{(lT,mF)} \mathbf{P}_\gamma \mathbf{S}^{(lT,mF)*} \qquad \text{with} \qquad (\mathbf{P}_\gamma)(t,s) = \gamma(t)\gamma^*(s).$$

For the $TF = 1$ (critical sampling) the Kohn–Nirenberg symbol is essentially equivalent to the magnitude–squared Zak transform of the prototype γ:

$$L_{M_\gamma}^{(1/2)}(t,f) = T \left| \mathcal{Z}_\gamma(t,f) \right|^2,$$

where the Zak transform [388, 175, 179, 393, 158] is defined as

$$\mathcal{Z}_\gamma(t,f) = \sum_l \gamma(t+lT) e^{-j2\pi flT}.$$

The Zak transform is also known as *Weil–Brezin* transform [125].

9. **Generalized Wigner Distribution.** The GWS of a rank–one projection operator $\mathbf{P}_x$ is just the generalized Wigner distribution of the eigensignal

$$L_{P_x}^{(\alpha)}(t,f) = W_x^{(\alpha)}(t,f), \qquad \text{with} \qquad (\mathbf{P}_x)(t,s) = x(t)x^*(s). \tag{C.9}$$

10. **Spectral Decomposition.** The GWS of a HS operator can be expanded into a singular–value weighted sum of cross Wigner distributions

$$L_H^{(\alpha)}(t,f) = \sum_{k=1}^{\infty} \sigma_k W_{u_k,v_k}^{(\alpha)}(t,f), \tag{C.10}$$

where the weights $\sigma_k \geq 0$ are the singular values, and the orthonormal bases $\{u_k(t)\}$ and $\{v_k(t)\}$ are the left and right singular signals of the operator.

11. **Adjoint Operator.** The GWS of the adjoint operator can be obtained as

$$L_{H^*}^{(\alpha)}(t,f) = L_H^{(-\alpha)*}(t,f), \tag{C.11}$$

for the Weyl symbol in particular we thus have

$$L_{H^*}^{(0)}(t,f) = L_H^{(0)*}(t,f), \tag{C.12}$$

which shows that, for self–adjoint operators, $L_H^{(0)}(t,f)$ is real–valued.

12. **Marginals.** The time and frequency marginals are α–invariant,

$$\int_t L_H^{(\alpha)}(t,f)dt = B_H(f,f), \quad \int_f L_H^{(\alpha)}(t,f)df = h(t,t), \tag{C.13}$$

where $B_H(f,f')$ is the bifrequency function. Given a HS operator, the total integral of the GWS equals the operator's trace, the so–called "trace formula":

$$\int_t \int_f L_H^{(\alpha)}(t,f)dtdf = \sum_{k=1}^{\infty} \sigma_k \langle u_k, v_k \rangle = \text{tr}\mathbf{H}. \tag{C.14}$$

13. **Finite Support Symbol.** Let **D** be a time–limitation operator defined by

$$L_D^{(\alpha)}(t,f) = \chi_{[-t_0,t_0]}(t)$$

and **B** its dual, a band–limitation operator with symbol

$$L_B^{(\alpha)}(t,f) = \chi_{[-f_0,f_0]}(f)$$

then, given an arbitrary HS operator **H** one has the following finite support properties

$$\begin{aligned} L_{DHD}^{(\alpha)}(t,f) &= \chi_{[-t_0,t_0]}(t) L_{DHD}^{(\alpha)}(t,f) \\ L_{BHB}^{(\alpha)}(t,f) &= \chi_{[-f_0,f_0]}(f) L_{BHB}^{(\alpha)}(t,f) \end{aligned}$$

It should be emphasized that, for this to hold, our usual requirement $|\alpha| \leq 1/2$ is necessary (it is not necessary for most other properties of the GWS).

14. **Convolution–Product/Product–Convolution Operators.** Given an LTI operator **M** and an LFI operator **H** defined by

$$\begin{aligned} L_M^{(\alpha)}(t,f) &= m(t), \\ L_H^{(\alpha)}(t,f) &= H(f), \end{aligned}$$

the combinations are known as convolution–product operator **MH** and product–convolution operator **HM** [261]. The GWS of these specific class of operators does not generally lead to a separable symbol. However, for specific choices of α one has:

$$\begin{aligned} L_{HM}^{(1/2)}(t,f) &= m(t)H(f), \\ L_{MH}^{(-1/2)}(t,f) &= m(t)H(f), \end{aligned}$$

which, in case of band–limitation and time–limitation operators can be specialized to (also see the previous discussion on support properties):

$$\begin{aligned} L_{BD}^{(1/2)}(t,f) &= \chi_{[-t_0,t_0]}(t)\chi_{[-f_0,f_0]}(f), \\ L_{DB}^{(-1/2)}(t,f) &= \chi_{[-t_0,t_0]}(t)\chi_{[-f_0,f_0]}(f). \end{aligned}$$

15. **Time–Frequency Shift–Covariance.** A translation of the GWS corresponds to a respective time–frequency shift of the singular signals, while the singular values remain constant:

$$L_{H^{(\tau,\nu)}}^{(\alpha)}(t,f) = L_H^{(\alpha)}(t-\tau, f-\nu) \iff \mathbf{H}^{(\tau,\nu)} = \mathbf{S}^{(\tau,\nu)}\mathbf{H}\mathbf{S}^{(\tau,\nu)*}, \tag{C.15}$$

where $\mathbf{S}^{(\tau,\nu)}$ is an arbitrary combination of a time–shift with lag τ and frequency shift with lag ν.

16. **Continuous Weyl–Heisenberg Expansions.** The GWS leads to a formal expansion of a HS operator into a continuous superposition of time–frequency shifted versions of an infinitesimal prototype operator:

$$\mathbf{H} = \int_t \int_f L_H^{(\alpha)}(t,f)\mathbf{P}^{(t,f)}(\alpha)\,dt\,df \tag{C.16}$$

generally valid for the α–dependent non–HS prototype operator defined by

$$L_P^{(\alpha)}(t,f) = \delta(t)\delta(f). \tag{C.17}$$

For the most prominent choices of α, the infinitesimal prototypes act as follows

- For $\alpha = 1/2$, $\mathbf{P}^{(t_0,f_0)}$ is ideally frequency selective and the output is ideally temporally localized:
$$\left(\mathbf{P}^{(t_0,f_0)}x\right)(t) = \delta(t-t_0)X(f_0)e^{j2\pi f_0 t_0}.$$
- For $\alpha = -1/2$, one has ideal frequency concentration on the range and ideal temporal concentration on the domain:
$$\left(\mathbf{P}^{(t_0,f_0)}x\right)(t) = e^{j2\pi f_0 t}x(t_0)e^{-j2\pi f_0 t_0}.$$
- For the specific case $\alpha = 0$ the prototype operator acts in a highly symmetric way as a "time–frequency point mirror":
$$\begin{aligned}\left(\mathbf{P}^{(t_0,f_0)}x\right)(t) &= 2x(2t_0-t)e^{j2\pi 2f_0(t-t_0)},\\ \left(\mathcal{F}\mathbf{P}^{(t_0,f_0)}x\right)(f) &= 2X(2f_0-f)e^{j2\pi 2t_0(f_0-f)}.\end{aligned}$$

Whenever the operator at hand satisfies a sharp spreading constraint of the form

$$S_H^{(\alpha)}(\tau,\nu) = S_H^{(\alpha)}(\tau,\nu)\,\chi_H(\tau,\nu), \tag{C.18}$$

where $\chi_H(\tau,\nu)$ is an indicator function, then the prototype operator is HS and given by:

$$S_{P_\alpha}^{(\alpha)}(\tau,\nu) = \chi_H(\tau,\nu).$$

17. **Discrete Weyl–Heisenberg Expansions.** When we specialize (C.18) to a rectangular shape defined by
$$\chi_H(\tau,\nu) = \chi_{[-\tau_0,\tau_0]}(\tau)\,\chi_{[-\nu_0,\nu_0]}(\nu),$$
then the sampling theorem leads to a discrete Weyl–Heisenberg expansion of the form:
$$\mathbf{H} = \sum_l \sum_m L_H^{(\alpha)}(lT,mF)\,\mathbf{P}^{(lT,mF)}(\alpha), \tag{C.19}$$
where the symbol of the prototype operator $\mathbf{P}(\alpha)$ is given by a 2D sinc–type kernel:
$$L_P^{(\alpha)}(t,f) = \frac{\sin(2\pi\nu_0 t)\sin(2\pi\tau_0 f)}{4\pi^2\tau_0\nu_0 tf},$$
and we have a "discrete trace formula":
$$\sum_l \sum_m L_H^{(\alpha)}(lT,mF) = \frac{1}{TF}\mathrm{tr}\mathbf{H}.$$

18. **Input–Output Relations.** For $|\alpha| = 1/2$, the GWS can be essentially interpreted as kernel of an integral operators which maps from the time/frequency domain of the input signal to frequency/time domain of the output signal ($x(t), y(t)$ are the input/output signal and $X(f), Y(f)$ their Fourier transforms) :
$$\begin{aligned}y(t) &= \int_f L_H^{(1/2)}(t,f)X(f)e^{j2\pi ft}\,df,\\ Y(f) &= \int_t L_H^{(-1/2)}(t,f)x(t)e^{-j2\pi ft}\,dt.\end{aligned}$$

APPENDIX D: Time–Varying Spectral Estimation

In this appendix we derive closed form expressions for the bias and variance regarding the estimation of the expectation of a general real–valued, quadratic time–frequency shift–invariant signal representation. We assume a noisy observation of a circular complex, nonstationary Gaussian process with statistically independent circular complex, stationary white Gaussian noise. We specialize the results to the estimation of the Wigner–Ville spectrum of a process with known support of its expected ambiguity function. It is shown that in this setup the minimum variance unbiased (MVUB) estimator is given by a quadratic form with prototype operator whose spreading function equals the given indicator function of the support of the process' expected ambiguity function.

Estimation Target. We consider the estimation of a *real–valued, time–frequency shift invariant, time–varying spectrum*, which can be written as the expectation of a quadratic form:

$$P_x(t,f) = \mathrm{E}\left\{\left\langle \mathbf{P}^{(t,f)}x, x\right\rangle\right\},$$

where $\mathbf{P}^{(t,f)}$ is a time–frequency shifted version of the self–adjoint prototype operator $\mathbf{P}$. Recall that we define time–frequency shifting of operators by

$$\mathbf{P}^{(t,f)} = \mathbf{S}^{(t,f)}\mathbf{P}\mathbf{S}^{(t,f)*},$$

where $\mathbf{S}^{(\tau,\nu)}$ is an arbitrary version of the α–parametrized family of unitary time–frequency shift operators (see (B.4)).

Observation. We assume that $x(t)$ is a circular complex, zero–mean, nonstationary Gaussian process with trace–class correlation kernel:

$$\left(\mathbf{R}_x\right)(t,s) = \mathrm{E}\left\{x(t)x^*(s)\right\}, \qquad \text{with} \qquad \mathrm{tr}\mathbf{R}_x < \infty.$$

We furthermore consider a noisy observation of $x(t)$,

$$y(t) = x(t) + n(t) \qquad \text{with} \qquad \mathrm{E}\left\{n(t)n^*(s)\right\} = \sigma_n^2\delta(t-s),$$

where $n(t)$ is statistically independent, circular complex Gaussian white noise.

Estimator. Similar to the estimation target, the estimator is a quadratic form now in terms of the observation and based on a different prototype operator:

$$\widehat{P}_y(t,f) = \left\langle \widehat{\mathbf{P}}^{(t,f)}y, y\right\rangle,$$

In the bias/variance computation it is important to have a brief formal notation in order to avoid lengthy expressions. We thus perform the computation in terms of the correlation operators of the processes and the prototype operators of the estimator and the estimation target. By way of preparation we first note the following identity:

$$P_x(t,f) = \mathrm{E}\left\{\left\langle \mathbf{P}^{(t,f)}x, x\right\rangle\right\} = \mathrm{tr}\left\{\mathbf{R}_x\mathbf{P}^{(t,f)}\right\} = \left\langle \mathbf{R}_x, \mathbf{P}^{(t,f)}\right\rangle, \tag{D.1}$$

where tr denotes the trace operator (see (A.8)) and the inner product is a Hilbert–Schmidt compatible operator inner product (see (A.8)). We mention the inner product equivalence (D.1) since by using *unitary operator representations* such as the spreading function of $\mathbf{P}$ or spreading function of $\mathbf{R}_x$ (which is equivalent to the expected ambiguity function of $x(t)$), we shall replace the discrete trace operation by integral representations. Furthermore note the trace invariance

$$\mathrm{tr}\mathbf{P}^{(t,f)} = \mathrm{tr}\mathbf{P};$$

we shall henceforth suppress the (t, f) superscript wherever possible due to this trace invariance.

Trace Normalization. As a simple side–constraint we require that the original time–frequency representation (estimation target) "distributes" the Karhunen–Loeve eigenvalues of the process over the time–frequency plane:

$$\int_t \int_f P_x(t,f)dt\,df = \mathrm{tr}\mathbf{R}_x, \tag{D.2}$$

which corresponds to the deterministic requirement of an energy distribution:

$$\int_t \int_f \left\langle \mathbf{P}^{(t,f)}x, x \right\rangle dt\,df = \|x\|^2. \tag{D.3}$$

These equivalent requirements (D.2) and (D.3) can be reduced to a trace normalization of the prototype operator since

$$\int_t \int_f \mathrm{tr}\left\{\mathbf{P}^{(t,f)}\mathbf{R}_x\right\} dt\,df = \mathrm{tr}\left\{\left(\int_t \int_f \mathbf{P}^{(t,f)}dt\,df\right)\mathbf{R}_x\right\} = \mathrm{tr}\mathbf{P}\,\mathrm{tr}\mathbf{R}_x, \tag{D.4}$$

which follows immediately from the fact that

$$\int_t \int_f \mathbf{P}^{(t,f)}dt\,df = \mathrm{tr}\{\mathbf{P}\}\mathbf{I}, \tag{D.5}$$

which is easy to check via the trace formula of the Weyl correspondence (C.14). We henceforth restrict our attention to prototype operators that are Hilbert–Schmidt (HS) and satisfy:

$$\mathrm{tr}\mathbf{P} = \mathrm{tr}\widehat{\mathbf{P}} = 1.$$

At this point it should be already mentioned that the prototype operator of a time–frequency distribution may not be HS as e.g. in case of the Wigner distribution. We shall however prove in this appendix that finite variance estimators always have prototype operators that are HS. Moreover, we also show that the Wigner–Ville spectrum of a process with both limited nonstationarity and finite temporal correlation can always be written as a quadratic form of a Hilbert–Schmidt operator.

In the derivation of the unbiased estimators we shall represent the prototype operators via their spreading functions (for definition and properties see Appendix B). The HS requirement carries over to

$$\int_\tau \int_\nu |S_P(\tau,\nu)|^2\,d\tau\,d\nu \le M < \infty,$$

and the trace normalization is reflected in

$$S_P(0,0) = 1.$$

Expectation of the Estimate. With the statistical independence of signal and noise we have

$$\mathbf{R}_y = \mathbf{R}_x + \sigma_n^2\mathbf{I}. \tag{D.6}$$

(The correlation operator of stationary white noise is a scalar multiple of the identity, which is denoted by $\mathbf{I}$.) Based on (D.6) and (D.1) we immediately obtain the expectation of our estimate:

$$\mathrm{E}\left\{\widehat{P}_y(t,f)\right\} = \mathrm{tr}\left\{\widehat{\mathbf{P}}^{(t,f)}(\mathbf{R}_x + \sigma_n^2\mathbf{I})\right\} = \mathrm{E}\left\{\widehat{P}_x(t,f)\right\} + \sigma_n^2\mathrm{tr}\widehat{\mathbf{P}}. \tag{D.7}$$

Bias. According to (D.7), the bias of the estimator is given by

$$B(t,f) \stackrel{\text{def}}{=} \mathrm{E}\left\{\widehat{P}_y(t,f)\right\} - P_x(t,f) = \mathrm{tr}\left\{\widetilde{\mathbf{P}}^{(t,f)}\mathbf{R}_x\right\} + \sigma_n^2 \mathrm{tr}\widehat{\mathbf{P}}, \tag{D.8}$$

where we have introduced the bias operator $\widetilde{\mathbf{P}}$ as follows:

$$\widetilde{\mathbf{P}} \stackrel{\text{def}}{=} \widehat{\mathbf{P}} - \mathbf{P}.$$

Using the Schwarz inequality for operator inner products and triangle inequality we immediately get a tight bound for the maximum bias:

$$\|B\|_\infty \leq \|\widetilde{\mathbf{P}}\|\|\mathbf{R}_x\| + \sigma_n^2 \left|\mathrm{tr}\widehat{\mathbf{P}}\right|.$$

However, in the choice of an estimator the maximum bias alone is no sufficient criterion, since an estimator that achieves small $\|B\|_\infty$ may still lead to a comparatively large average bias. We thus define an *integrated squared bias* as the total integral of the time–frequency dependent part of the squared bias:

$$B_0^2 \stackrel{\text{def}}{=} \int_t \int_f \left\{B^2(t,f) - \sigma_n^4 \mathrm{tr}^2\widehat{\mathbf{P}}\right\} dt\, df.$$

In the computation of B_0^2 we can use (D.4) but we have to evaluate another nontrivial time–frequency integral, namely over $\mathrm{tr}^2\{\widetilde{\mathbf{P}}^{(t,f)}\mathbf{R}_x\}$. In order to obtain a useful expression we introduce the spreading functions of both $\mathbf{P}$ and $\mathbf{R}_x$ (where the latter is just the expected ambiguity function of the process, $EA_x(\tau,\nu)$):

$$\begin{aligned}
\int_t \int_f \mathrm{tr}^2\left\{\widetilde{\mathbf{P}}^{(t,f)}\mathbf{R}_x\right\} dt\, df &= \int_t \int_f \left|\left\langle S_{\widetilde{P}^{(t,f)}}, EA_x\right\rangle\right|^2 dt\, df \\
&= \int_t \int_f \int_{\tau_1} \int_{\nu_1} \int_{\tau_2} \int_{\nu_2} S_{\widetilde{P}}(\tau_1,\nu_1) EA_x(\tau_1,\nu_1) \\
&\quad \cdot S^*_{\widetilde{P}}(\tau_2,\nu_2) EA^*_x(\tau_2,\nu_2) e^{-j2\pi[(\nu_1-\nu_2)t-(\tau_1-\tau_2)f]} dt\, df\, d\tau_1\, d\nu_1\, d\tau_2\, d\nu_2 \\
&= \int_\tau \int_\nu \left|S_{\widetilde{P}}(\tau,\nu)\right|^2 |EA_x(\tau,\nu)|^2 d\tau\, d\nu \\
&= \left\langle \left|S_{\widetilde{P}}\right|^2, |EA_x|^2 \right\rangle.
\end{aligned} \tag{D.9}$$

With (D.4) and (D.9) we finally obtain the following result for the integrated bias:

$$B_0^2 = \left\langle \left|S_{\widetilde{P}}\right|^2, |EA_x|^2 \right\rangle + 2\sigma_n^2 \mathrm{tr}\widehat{\mathbf{P}}\,\mathrm{tr}\widetilde{\mathbf{P}}\,\mathrm{tr}\mathbf{R}_x.$$

Due to (D.8) unbiased estimation requires at first exact knowledge of the noise level σ_n^2. Presupposing this *a priori* knowledge we can distinguish two basic classes of unbiased estimators:

1. Estimators based on an "unbiased" prototype operator, such that $\widetilde{\mathbf{P}} = 0$. This is a trivial solution that will be usually inadmissible as it leads to infinite variance (in case e.g. of the Wigner distribution).

2. Estimators with "biased" prototype operator where the bias operator satisfies

$$S_{\widetilde{P}}(\tau,\nu) EA_x(\tau,\nu) \equiv 0. \tag{D.10}$$

Variance. The computation of the variance of the estimator, which is defined as

$$V^2(t,f) \stackrel{\text{def}}{=} \mathrm{E}\left\{\widehat{P}_y^2(t,f)\right\} - \left(\mathrm{E}\left\{\widehat{P}_y(t,f)\right\}\right)^2, \tag{D.11}$$

is less straightforward.

In the determination of $\mathrm{E}\left\{\widehat{P}_y^2(t,f)\right\}$ one has to evaluate the following sum of the expectation of various combinations of quadratic forms in terms of the self–adjoint operator $\widehat{\mathbf{P}}^{(t,f)}$

$$\begin{aligned}\mathrm{E}\left\{\widehat{P}_y^2(t,f)\right\} = \mathrm{E}\Big\{&\left\langle\widehat{\mathbf{P}}^{(t,f)}x,x\right\rangle^2 + \left\langle\widehat{\mathbf{P}}^{(t,f)}n,n\right\rangle^2 + 2\left\langle\widehat{\mathbf{P}}^{(t,f)}x,x\right\rangle\left\langle\widehat{\mathbf{P}}^{(t,f)}n,n\right\rangle \\ &+\left\langle\widehat{\mathbf{P}}^{(t,f)}x,n\right\rangle^2 + \left\langle\widehat{\mathbf{P}}^{(t,f)}n,x\right\rangle^2 + 2\left\langle\widehat{\mathbf{P}}^{(t,f)}x,n\right\rangle\left\langle\widehat{\mathbf{P}}^{(t,f)}n,x\right\rangle\Big\}.\end{aligned} \tag{D.12}$$

The first term requires to compute the following expectation:

$$\mathrm{E}\left\{\left|\left\langle\widehat{\mathbf{P}}^{(t,f)}x,x\right\rangle\right|^2\right\} = \mathrm{E}\left\{\int_{t_1}\int_{t_2}\int_{t_3}\int_{t_4} h(t_1,t_2)x(t_2)x^*(t_1)h(t_3,t_4)x(t_4)x^*(t_3)dt_1\,dt_2\,dt_3\,dt_4\right\}, \tag{D.13}$$

where $h(t_1,t_2)$ is the kernel of $\mathbf{P}^{(t,f)}$ (for notational simplicity we have intermediately suppressed the superscript). The here appearing fourth–order moment of a Gaussian process can be written in terms of the correlation functions by using Isserlis' formula for the special case of circular complex processes:

$$\mathrm{E}\left\{x(t_2)x^*(t_1)x(t_3)x^*(t_4)\right\} = \left(\mathbf{R}_x\right)(t_2,t_1)\left(\mathbf{R}_x\right)(t_3,t_4) + \left(\mathbf{R}_x\right)(t_2,t_4)\left(\mathbf{R}_x\right)(t_3,t_1).$$

The expectation of (D.13) is thus split into a sum of two terms:

$$\begin{aligned}\mathrm{E}\left\{\left|\left\langle\widehat{\mathbf{P}}^{(t,f)}x,x\right\rangle\right|^2\right\} &= \left\{\int_{t_1}\int_{t_2} h(t_1,t_2)\left(\mathbf{R}_x\right)(t_2,t_1)dt_1\,dt_2\right\}\left\{\int_{t_3}\int_{t_4} h(t_3,t_4)\left(\mathbf{R}_x\right)(t_4,t_3)dt_3\,dt_4\right\} \\ &\quad+\int_{t_1}\int_{t_2}\int_{t_3}\int_{t_4} h(t_1,t_2)\left(\mathbf{R}_x\right)(t_2,t_3)h(t_3,t_4)\left(\mathbf{R}_x\right)(t_4,t_1)dt_1\,dt_2\,dt_3\,dt_4 \\ &= \mathrm{tr}^2\left\{\widehat{\mathbf{P}}^{(t,f)}\mathbf{R}_x\right\} + \mathrm{tr}\left\{\left(\widehat{\mathbf{P}}^{(t,f)}\mathbf{R}_x\right)^2\right\}.\end{aligned} \tag{D.14}$$

Using this result we can immediately evaluate the second term of (D.12) (recall $\mathbf{R}_n = \sigma_n^2\mathbf{I}$):

$$\mathrm{E}\left\{\left|\left\langle\widehat{\mathbf{P}}^{(t,f)}n,n\right\rangle\right|^2\right\} = \sigma_n^4\left(\mathrm{tr}^2\widehat{\mathbf{P}} + \mathrm{tr}\widehat{\mathbf{P}}^2\right), \tag{D.15}$$

it is time–frequency invariant.

The third term can be written as

$$\mathrm{E}\left\{\left\langle\widehat{\mathbf{P}}^{(t,f)}x,x\right\rangle\left\langle\widehat{\mathbf{P}}^{(t,f)}n,n\right\rangle\right\} = \sigma_n^2\mathrm{tr}\widehat{\mathbf{P}}\,\mathrm{tr}\left\{\widehat{\mathbf{P}}^{(t,f)}\mathbf{R}_x\right\}. \tag{D.16}$$

For the fourth term,

$$\mathrm{E}\left\{\left|\left\langle\widehat{\mathbf{P}}^{(t,f)}x,n\right\rangle\right|^2\right\} = \mathrm{E}\left\{\int_{t_1}\int_{t_2}\int_{t_3}\int_{t_4} h(t_1,t_2)x(t_2)n^*(t_1)h(t_3,t_4)x(t_3)n^*(t_3)dt_1\,dt_2\,dt_3\,dt_4\right\},$$

and the fifth term,

$$\mathrm{E}\left\{\left|\left\langle\widehat{\mathbf{P}}^{(t,f)}n,x\right\rangle\right|^2\right\} = \mathrm{E}\left\{\int_{t_1}\int_{t_2}\int_{t_3}\int_{t_4} h(t_1,t_2)n(t_2)x^*(t_1)h(t_3,t_4)n(t_4)x^*(t_3)dt_1\,dt_2\,dt_3\,dt_4\right\},$$

we note that for circular complex variables:

$$\mathrm{E}\left\{x(t_1)x(t_2)\right\} = 0$$

thus,

$$\mathrm{E}\left\{\left|\left\langle \widehat{\mathbf{P}}^{(t,f)}x, n\right\rangle\right|^2\right\} = \mathrm{E}\left\{\left|\left\langle \widehat{\mathbf{P}}^{(t,f)}n, x\right\rangle\right|^2\right\} = 0. \tag{D.17}$$

The last term of (D.12) can be evaluated as follows:

$$\begin{aligned}
&\mathrm{E}\left\{\left\langle \widehat{\mathbf{P}}^{(t,f)}x, n\right\rangle\left\langle \widehat{\mathbf{P}}^{(t,f)}n, x\right\rangle\right\} \\
&= \mathrm{E}\left\{\int_{t_1}\int_{t_2}\int_{t_3}\int_{t_4} h(t_1,t_2)x(t_2)n^*(t_1)h(t_3,t_4)n(t_4)x^*(t_3)dt_1\,dt_2\,dt_3\,dt_4\right\} \\
&= \int_{t_1}\int_{t_2}\int_{t_3} h(t_1,t_2)\left(\mathbf{R}_x\right)(t_2,t_3)h(t_3,t_1)dt_1\,dt_2\,dt_3 \\
&= \sigma_n^2\mathrm{tr}\left\{\left(\widehat{\mathbf{P}}^{(t,f)}\right)^2\mathbf{R}_x\right\}.
\end{aligned} \tag{D.18}$$

Inserting (D.14), (D.15), (D.16), (D.17), (D.18) in (D.12) we finally obtain the following expression for the variance of the estimator

$$V^2(t,f) = \mathrm{tr}\left\{\left(\widehat{\mathbf{P}}^{(t,f)}\mathbf{R}_x\right)^2\right\} + 2\sigma_n^2\mathrm{tr}\left\{\left(\widehat{\mathbf{P}}^{(t,f)}\right)^2\mathbf{R}_x\right\} + \sigma_n^4\mathrm{tr}\widehat{\mathbf{P}}^2. \tag{D.19}$$

The Schwarz inequality for the operator inner product leads to the following tight bound on the maximum variance:

$$\left\|V^2\right\|_\infty \leq \left\|\widehat{\mathbf{P}}\right\|^2\left(\left\|\mathbf{R}_x\right\| + \sigma_n^2\right)^2.$$

Just as discussed for the bias, we now obtain an integrated version of the variance. To this end, we have to evaluate the total time–frequency integral over the two time–frequency–dependent terms of (D.19). In order to compute the first integral we again use the spreading function of both $\widehat{\mathbf{P}}^{(t,f)}$ and $\mathbf{R}_x$. Based on the formula for the spreading function of the composite operators (B.33) we have

$$S_{\widehat{P}^{(t,f)}R_x}(\tau,\nu) = \int_{\tau_1}\int_{\nu_1} S_{\widehat{P}}(\tau_1,\nu_1)e^{-j2\pi(\nu_1 t-\tau_1 f)}EA_x(\tau-\tau_1,\nu-\nu_1)e^{j\pi(\tau_1\nu-\nu_1\tau)}d\tau_1\,d\nu_1;$$

which we now use for the first variance term integral,

$$\begin{aligned}
&\int_t\int_f \mathrm{tr}\left\{\left(\widehat{\mathbf{P}}^{(t,f)}\mathbf{R}_x\right)^2\right\}dt\,df \\
&= \int_t\int_f\int_\tau\int_\nu \left|S_{\widehat{P}^{(t,f)}R_x}(\tau,\nu)\right|^2 dt\,df\,d\tau\,d\nu \\
&= \int_t\int_f\int_\tau\int_\nu\int_{\tau_1}\int_{\nu_1}\int_{\tau_2}\int_{\nu_2} S_{\widehat{P}}(\tau_1,\nu_1)EA_x(\tau-\tau_1,\nu-\nu_1)S^*_{\widehat{P}}(\tau_2,\nu_2)EA^*_x(\tau-\tau_2,\nu-\nu_2)\cdot \\
&\quad \cdot e^{-j2\pi[(\nu_1-\nu_2)t-(\tau_1-\tau_2)f+1/2(\tau_1\nu-\nu_1\tau-\tau_2\nu+\nu_2\tau)]}dt\,df\,d\tau\,d\nu\,d\tau_1\,d\nu_1\,d\tau_2\,d\nu_2 \\
&= \int_\tau\int_\nu\int_{\tau_1}\int_{\nu_1}\left|S_{\widehat{P}}(\tau_1,\nu_1)\right|^2|EA_x(\tau-\tau_1,\nu-\nu_1)|^2 d\tau\,d\nu\,d\tau_1\,d\nu_1 \\
&= \mathrm{tr}\widehat{\mathbf{P}}^2\mathrm{tr}\mathbf{R}_x^2.
\end{aligned} \tag{D.20, D.21}$$

For the second integral it suffices to note that

$$\left(\mathbf{P}^{(t,f)}\right)^2 = \mathbf{S}^{(t,f)}\mathbf{P}\mathbf{S}^{(t,f)*}\mathbf{S}^{(t,f)}\mathbf{P}\mathbf{S}^{(t,f)*} = \mathbf{S}^{(t,f)}\mathbf{P}^2\mathbf{S}^{(t,f)*} = \left(\mathbf{P}^2\right)^{(t,f)},$$

whence we have with (D.5):

$$\int_t \int_f \mathrm{tr}\left\{\left(\widehat{\mathbf{P}}^{(t,f)}\right)^2 \mathbf{R}_x\right\} dt\, df = \mathrm{tr}\widehat{\mathbf{P}}^2 \mathrm{tr}\mathbf{R}_x. \tag{D.22}$$

For the integrated variance we thus obtain a compact result:

$$V_0^2 \stackrel{\mathrm{def}}{=} \int_t \int_f \left\{V^2(t,f) - \sigma_n^4 \mathrm{tr}\widehat{\mathbf{P}}^2\right\} dt\, df = \|\widehat{\mathbf{P}}\|^2 \left(\mathrm{tr}\mathbf{R}_x^2 + 2\sigma_n^2 \mathrm{tr}\mathbf{R}_x\right).$$

Observe that *the maximum variance, the time–frequency invariant variance term and the integrated variance are proportional to the HS–norm of the prototype operator of the estimator*, $\widehat{\mathbf{P}}$. (Note that the trace of a squared self–adjoint HS–operator equals its HS–norm, i.e., $\mathrm{tr}\mathbf{P}^2 = \|\mathbf{P}\|^2$.) Thus we can *a priori* exclude a $\widehat{\mathbf{P}}$ that is not Hilbert–Schmidt as it leads to infinite variance. However, as stated above, the prototype operator of the prominent Wigner–Ville spectrum is not Hilbert–Schmidt. Thus, at this point we can already exclude *finite–variance unbiased estimation* of the Wigner–Ville spectrum of a *general* nonstationary process. This is basically a well–known fact for the noiseless observation [121] and similar to the estimation problem of Priestley's evolutionary spectrum [296].

MVUB Estimator for the WVS. The intuitive idea of smoothing frequency parametrized estimates in order to reduce the variance is a classical principle of time–invariant spectrum estimation. Of course, time–invariant spectrum estimation relies on the fundamental implicit *a priori* knowledge (or *a priori* assumption) of stationarity which conceptually allows "total smoothing" in time direction. It is thus not astonishing that we have to require a smoothness condition on the target of our time–varying spectral estimator in order to obtain unbiased estimators with finite–variance. The prototype operator of the Wigner–Ville spectrum (WVS) can be characterized by

$$S_P(\tau,\nu) \equiv 1, \tag{D.23}$$

as has been already discussed it is not Hilbert–Schmidt and the estimation problem only starts to make sense when we have some additional *a priori* knowledge on the process at hand.

For the existence of an unbiased, finite–variance estimator of the WVS we have two requirements:

1. For an unbiased estimate we need to know the noise level σ_n^2.
2. For a finite–variance unbiased estimate we need to know a spreading constraint with finite support area of the process' expected ambiguity function, or with other words, a smoothness condition on the time–varying spectrum to be estimated.

We henceforth tacitly assume knowledge of the noise level such that the bias term due to noise, i.e. $\sigma_n^2 \mathrm{tr}\widehat{\mathbf{H}}$, can be simply corrected. According to (D.10) the requirement for the bias operator of any unbiased estimator is:

$$S_{\widetilde{P}}(\tau,\nu) EA_x(\tau,\nu) \equiv 0. \tag{D.24}$$

Since, $S_{\widetilde{P}}(\tau,\nu) = S_{\widehat{P}}(\tau,\nu) - S_P(\tau,\nu) = S_{\widehat{P}}(\tau,\nu) - 1$, we have the following requirement for the prototype operator:

$$S_{\widehat{P}_{UB}}(\tau,\nu) = \begin{cases} 1, & \text{where} \quad EA_x(\tau,\nu) \neq 0, \\ \text{arbitrary}, & \text{where} \quad EA_x(\tau,\nu) = 0. \end{cases} \tag{D.25}$$

Now, we understand minimum variance in the sense of the global variance measure as defined by the integrated variance V_0^2. We thus have to select the unbiased estimator whose prototype operator achieves minimum Hilbert–Schmidt norm. This optimization turns out to be trivial since the Hilbert–Schmidt norm is equal to the total integral of the magnitude squared spreading function:

$$\|\widehat{\mathbf{P}}\|^2 = \int_\tau \int_\nu |S_{\widehat{P}}(\tau,\nu)|^2 d\tau\, d\nu. \tag{D.26}$$

The minimum–variance unbiased (MVUB) quadratic estimator is thus obtained by setting the spreading function of the prototype operator zero wherever possible:

$$S_{\widehat{P}_{MVUB}}(\tau,\nu) = \begin{cases} 1, & \text{where} \quad EA_x(\tau,\nu) \neq 0, \\ 0, & \text{where} \quad EA_x(\tau,\nu) = 0. \end{cases} \tag{D.27}$$

This result is in obvious accordance with intuition, we may loosely characterize the MVUB estimator by "smooth as much as possible without introducing bias".

Minimum Norm Weyl–Heisenberg Expansion. In the context of this thesis we have a nice abstract interpretation of the MVUB estimator in terms of Weyl–Heisenberg operator decomposition as follows.

The correlation operator of a nonstationary process with support–restricted $EA_x(\tau,\nu)$ admits a decomposition in terms of a weighted integral of time–frequency shifted versions of a prototype operator where the weight function is the WVS:

$$\mathbf{R}_x = \int_t \int_f EW_x(t,f)\, \mathbf{P}^{(t,f)} dt\, df \qquad \text{with} \qquad S_P(\tau,\nu) EA_x(\tau,\nu) = EA_x(\tau,\nu), \tag{D.28}$$

this decomposition is highly ambiguous w.r.t. the choice of the prototype operator. Now, there exist a *minimum norm Weyl–Heisenberg decomposition* that is marked out by the minimum Hilbert–Schmidt norm of its prototype operator:

$$\widehat{\mathbf{P}} = \arg\min_{\mathbf{P}} \|\mathbf{P}\|^2 \qquad \text{subject to} \qquad \mathbf{R}_x = \int_t \int_f EW_x(t,f) \mathbf{P}^{(t,f)} dt\, df,$$

which leads just to the prototype operator of the MVUB estimator.

Extension to Real–Part of Generalized WVS. The MVUB estimator for the WVS can be easily extended to the expected real–part of the generalized WVS. Since one has

$$\mathrm{Re}\left\{ EW_x^{(\alpha)}(t,f) \right\} = \mathcal{F}_{\tau\to f} \mathcal{F}^{-1}_{\nu\to t} \left\{ EA_x(\tau,\nu) \cos(2\pi\nu\tau\alpha) \right\}, \tag{D.29}$$

the prototype operator of the MVUB estimator for $\mathrm{Re}\{EW_x^{(\alpha)}(t,f)\}$ is a slightly modified version of the MVUB estimator for the WVS:

$$S_{\widehat{P}_{MVUB}}(\tau,\nu) = \chi_x(\tau,\nu) \cos(2\pi\nu\tau\alpha) \qquad \text{with} \qquad \chi_x(\tau,\nu) \stackrel{\text{def}}{=} \begin{cases} 1, & \text{where} \quad EA_x(\tau,\nu) \neq 0, \\ 0, & \text{where} \quad EA_x(\tau,\nu) = 0. \end{cases}$$

It should be emphasized that for a small spread of $EA_x(\tau,\nu)$ the cosine factor will be negligible.

APPENDIX E: LTV System Identification

In this appendix we derive bias and variance expressions regarding the estimation of a Weyl–Heisenberg LTV system representation under the assumption of the observation of (i) a circular complex, stationary white Gaussian input process and (ii) an output process contaminated by statistically independent, circular complex, stationary white Gaussian noise. We furthermore discuss the resulting theoretical MVUB estimator for the generalized Weyl symbol (includes the usual time–varying transfer function of Zadeh and the Weyl symbol) of a system with known support of the spreading function. The MVUB estimator is a quadratic form in the observation with a prototype operator whose generalized spreading function is given by the known support of the system's spreading function.

While the main results are analog to the case of auto–spectral analysis as discussed in Appendix D, the details are different.

Estimation Target. The estimation target is an arbitrary Weyl–Heisenberg system representation $T_H(t,f)$ characterized by a prototype operator $\mathbf{P}$ in the form:

$$T_H(t,f) = \left\langle \mathbf{H}, \mathbf{P}^{(t,f)} \right\rangle.$$

The generalized Weyl symbol $L_H^{(\alpha)}(t,f)$ is a subclass of such system representations.

Based on a zero–mean, circular complex, Gaussian, stationary white input process with variance

$$\mathrm{E}\{x(t)x^*(s)\} = \sigma_x^2 \delta(t-s),$$

the output signal

$$y(t) = (\mathbf{H}x)(t)$$

is a zero–mean, circular complex nonstationary process and the estimation problem is equivalent to the time–varying cross spectral estimation of a time–frequency shift–invariant signal representation:

$$\mathbf{R}_{yx} = \sigma_x^2 \mathbf{H} \qquad \Longrightarrow \qquad T_H(t,f) = \mathrm{E}\left\{P_{yx}(t,f)\right\} = \left\langle \mathbf{R}_{yx}, \mathbf{P}^{(t,f)} \right\rangle,$$

where $\mathbf{R}_{yx}$ is the cross–correlation operator of the processes $y(t), x(t)$ and $P_{yx}(t,f)$ is a quadratic time–frequency shift invariant cross signal representation with prototype operator $\mathbf{P}$.

Observation. We assume a noiseless observation of the input process $x(t)$ and a noisy observation of the output process:

$$y(t) = (\mathbf{H}x)(t) + n(t),$$

where $n(t)$ is circular complex, zero–mean, stationary white Gaussian noise with correlation function:

$$\mathrm{E}\{n(t)n^*(s)\} = \sigma_n^2 \delta(t-s).$$

Estimator. The estimator is based on a generally different prototype operator $\widehat{P}$:

$$\widehat{T}_H(t,f) = \frac{1}{\sigma_x^2} \left\langle y, \widehat{\mathbf{P}}^{(t,f)} x \right\rangle.$$

Bias. The expectation of the estimate can be compactly written as

$$\mathrm{E}\left\{\widehat{T}_H(t,f)\right\} = \frac{1}{\sigma_x^2}\mathrm{E}\left\{\left\langle \mathbf{H}x + n, \widehat{\mathbf{P}}^{(t,f)} x \right\rangle\right\} = \frac{1}{\sigma_x^2}\mathrm{E}\left\{\left\langle \mathbf{H}x, \widehat{\mathbf{P}}^{(t,f)} x \right\rangle\right\} = \left\langle \mathbf{H}, \widehat{\mathbf{P}}^{(t,f)} \right\rangle.$$

The bias is accordingly given by:

$$B(t,f) \stackrel{\mathrm{def}}{=} \mathrm{E}\left\{\widehat{T}_H(t,f)\right\} - T_H(t,f) = \left\langle \mathbf{H}, \widetilde{\mathbf{P}}^{(t,f)} \right\rangle, \tag{E.1}$$

where we have introduced a "biased" prototype operator defined as:

$$\widetilde{\mathbf{P}} \stackrel{\text{def}}{=} \widehat{\mathbf{P}} - \mathbf{P}.$$

The maximum bias is proportional to the HS norm of the bias operator:

$$\|B\|_\infty \leq \|\widetilde{\mathbf{P}}\|\|\mathbf{H}\|,$$

note that this is in general no tight bound. Whenever one considers *a priori* knowledge on the support of the spreading function of the system $\mathbf{H}$ then it is possible to define unbiased estimators with $\|\widetilde{\mathbf{P}}\| \neq 0$.

We furthermore compute the integrated squared bias in terms of the spreading functions of the involved operators:

$$\begin{aligned}
B_0^2 &\stackrel{\text{def}}{=} \int_t \int_f |B(t,f)|^2 \, dt \, df \\
&= \int_t \int_f \left| \left\langle \mathbf{H}, \widetilde{\mathbf{P}}^{(t,f)} \right\rangle \right| dt \, df \\
&= \int_t \int_f \left| \left\langle S_H, S_{\widetilde{P}^{(t,f)}} \right\rangle \right|^2 dt \, df \\
&= \int_t \int_f \int_{\tau_1} \int_{\nu_1} \int_{\tau_2} \int_{\nu_2} S_H(\tau_1,\nu_1) S^*_{\widetilde{P}}(\tau_1,\nu_1) \\
&\quad \cdot S^*_H(\tau_2,\nu_2) S_{\widetilde{P}}(\tau_2,\nu_2) e^{j2\pi[(\nu_1-\nu_2)t-(\tau_1-\tau_2)f]} dt \, df \, d\tau_1 \, d\nu_1 \, d\tau_2 \, d\nu_2 \\
&= \int_\tau \int_\nu |S_H(\tau,\nu)|^2 \left| S_{\widetilde{P}}(\tau,\nu) \right|^2 d\tau \, d\nu \\
&= \left\langle |S_H|^2, \left| S_{\widetilde{P}} \right|^2 \right\rangle.
\end{aligned}$$

Just as in the auto–case of spectral estimation we have two essentially different ways for unbiased estimation:

- Unbiased estimators with $\widehat{\mathbf{P}} = \mathbf{P}$. With regard to the estimation of the generalized Weyl symbol, this trivial version of unbiased estimation does not lead to finite variance estimators as will be discussed subsequently.
- Unbiased estimators with $\widehat{\mathbf{P}} \neq \mathbf{P}$ but nonoverlapping spreading functions of the bias operator and the system

  $$|S_H(\tau,\nu)|^2 \left| S_{\widetilde{P}}(\tau,\nu) \right|^2 \equiv 0.$$

  This is the practically interesting case for the estimation of a system's generalized Weyl symbol with known support of the spreading function.

Variance. The variance of the estimator is defined as

$$V^2(t,f) \stackrel{\text{def}}{=} \mathrm{E}\left\{ \left| \widehat{T}_H(t,f) \right|^2 \right\} - \left| \mathrm{E}\left\{ \widehat{T}_H(t,f) \right\} \right|^2.$$

Due to our assumption of a zero–mean and statistically independent noise there remain two terms in the expectation of the magnitude squared estimate:

$$\mathrm{E}\left\{ \left| \widehat{T}_H(t,f) \right|^2 \right\} = \frac{1}{\sigma_x^4} \left(\mathrm{E}\left\{ \left| \left\langle \mathbf{H}x, \mathbf{P}^{(t,f)} x \right\rangle \right|^2 \right\} + \mathrm{E}\left\{ \left| \left\langle n, \mathbf{P}^{(t,f)} x \right\rangle \right|^2 \right\} \right). \tag{E.2}$$

For the first term we need to evaluate the following expectation:

$$\mathrm{E}\left\{\left|\left\langle \mathbf{H}x, \mathbf{P}^{(t,f)}x\right\rangle\right|^2\right\} = \mathrm{E}\left\{\int_{t_1}\int_{t_2}\int_{t_3} h(t_1,t_2)x(t_2)p^*(t_1,t_3)x^*(t_3)dt_1\,dt_2\,dt_3 \cdot \int_{t_4}\int_{t_5}\int_{t_6} h^*(t_4,t_5)x^*(t_5)p(t_4,t_6)x(t_6)dt_4\,dt_5\,dt_6\right\}.$$

Based on Isserli's fourth order moment formula for complex Gaussian variables we have

$$\mathrm{E}\left\{x(t_2)x^*(t_3)x^*(t_5)x(t_6)\right\} = \sigma_x^4\left(\delta(t_2-t_3)\delta(t_5-t_6)+\delta(t_2-t_5)\delta(t_3-t_6)\right)$$

and thus

$$\begin{aligned}\mathrm{E}\left\{\left|\left\langle \mathbf{H}x, \mathbf{P}^{(t,f)}x\right\rangle\right|^2\right\} &= \sigma_x^4\int_{t_1}\int_{t_2}\int_{t_4}\int_{t_5} h(t_1,t_2)p^*(t_1,t_3)h^*(t_4,t_5)p(t_4,t_5)dt_1\,dt_2\,dt_4\,dt_5 \\ &\quad +\sigma_x^4\int_{t_1}\int_{t_2}\int_{t_3}\int_{t_4} h(t_1,t_2)p^*(t_1,t_3)h^*(t_4,t_2)p(t_4,t_3)dt_1\,dt_2\,dt_3\,dt_4 \\ &= \left|\left\langle \mathbf{H}, \mathbf{P}^{(t,f)}\right\rangle\right|^2 + \left\langle \mathbf{H}\mathbf{H}^*, \mathbf{P}^{(t,f)}\mathbf{P}^{(t,f)*}\right\rangle \quad \text{(E.3)} \\ &= \left|\left\langle \mathbf{H}, \mathbf{P}^{(t,f)}\right\rangle\right|^2 + \left\langle \mathbf{H}\mathbf{H}^*, (\mathbf{P}\mathbf{P}^*)^{(t,f)}\right\rangle, \quad \text{(E.4)}\end{aligned}$$

where in the last step we have used the fact that

$$\mathbf{P}^{(t,f)}\mathbf{P}^{(t,f)*} = \mathbf{S}^{(t,f)}\mathbf{P}\mathbf{S}^{(t,f)*}\mathbf{S}^{(t,f)}\mathbf{P}^*\mathbf{S}^{(t,f)*} = \mathbf{S}^{(t,f)}\mathbf{P}\mathbf{P}^*\mathbf{S}^{(t,f)*} = (\mathbf{P}\mathbf{P}^*)^{(t,f)}.$$

The computation of the second term of (E.2) is straightforward:

$$\begin{aligned}\mathrm{E}\left\{\left|\left\langle n, \mathbf{P}^{(t,f)}x\right\rangle\right|^2\right\} &= \int_{t_1}\int_{t_2}\int_{t_3}\int_{t_4} n(t_1)p^*(t_1,t_2)x^*(t_2)n^*(t_3)p(t_3,t_4)x(t_4)dt_1\,dt_2\,dt_3\,dt_4 \\ &= \sigma_n^2\sigma_x^2\int_{t_1}\int_{t_3} p^*(t_1,t_3)p(t_1,t_3)dt_1\,dt_3 \quad \text{(E.5)} \\ &= \sigma_n^2\sigma_x^2\|\widehat{\mathbf{P}}\|^2. \quad \text{(E.6)}\end{aligned}$$

Such that finally we get a compact result for the variance of the estimator based on (E.4) and (E.6):

$$V^2(t,f) = \left\langle \mathbf{H}\mathbf{H}^*, \left(\widehat{\mathbf{P}}\widehat{\mathbf{P}}^*\right)^{(t,f)}\right\rangle + \frac{\sigma_n^2}{\sigma_x^2}\|\widehat{\mathbf{P}}\|^2. \quad \text{(E.7)}$$

The maximum variance is proportional to the HS norm of the prototype operator:

$$\|V^2\|_\infty \le \|\widehat{\mathbf{P}}\|^2\left(\|\mathbf{H}\|^2 + \frac{\sigma_n^2}{\sigma_x^2}\right).$$

We also compute an integrated variance by integrating over the time–frequency–dependent part of the

variance:

$$\begin{aligned} V_0^2 &\stackrel{\text{def}}{=} \int_t \int_f \left(V^2(t,f) - \frac{\sigma_n^2}{\sigma_x^2} \|\widehat{\mathbf{P}}\|^2 \right) dt\, df \\ &= \int_t \int_f \left\langle \mathbf{H}\mathbf{H}^*, \left(\widehat{\mathbf{P}}\widehat{\mathbf{P}}^*\right)^{(t,f)} \right\rangle dt\, df \\ &= \left\langle \mathbf{H}\mathbf{H}^*, \int_t \int_f \left(\widehat{\mathbf{P}}\widehat{\mathbf{P}}^*\right)^{(t,f)} dt\, df \right\rangle \\ &= \|\widehat{\mathbf{P}}\|^2 \langle \mathbf{H}\mathbf{H}^*, \mathbf{I} \rangle \\ &= \|\widehat{\mathbf{P}}\|^2 \|\mathbf{H}\|^2, \end{aligned}$$

where we have used the generally valid Weyl–Heisenberg resolution of the identity in the form of:

$$\int_t \int_f (\mathbf{P}\mathbf{P}^*)^{(t,f)} dt df = \operatorname{tr}\{\mathbf{P}\mathbf{P}^*\}\, \mathbf{I} = \|\mathbf{P}\|^2 \mathbf{I}.$$

It is seen that just as in the auto–case of spectral estimation all reasonable variance measures, the constant variance term, the integrated variance and the maximum variance, are proportional to the Hilbert–Schmidt norm of the estimator's prototype operator. Thus, given a class of unbiased estimators one has to select the estimator with minimum HS norm of its prototype operator in order to minimize the variance both in a local and global sense.

MVUB Estimator for $L_H^{(\alpha)}(t,f)$. It is well–known that identification of a general, i.e., unconstrained LTV system is no reasonably defined statistical task [191]. In this context, we view LTV system identification as the estimation of a unitary Weyl–Heisenberg representation of the system as is given by the generalized Weyl symbol. The prototype operator can be defined by

$$S_P^{(\alpha)}(\tau,\nu) \equiv 1,$$

since

$$\|\mathbf{P}\|^2 = \int_\tau \int_\nu \left| S_P^{(\alpha)}(\tau,\nu) \right|^2 d\tau\, d\nu$$

it is seen that $\mathbf{P}$ is never Hilbert–Schmidt irrespective of the choice of α. Hence, as expected, we can exclude finite–variance estimation of the GWS of a general LTV system. However, based upon *a priori* knowledge about a spreading constraint of the system $\mathbf{H}$ one can obtain an optimum unbiased estimator in the sense of minimum global and local variance. Analog to the case of auto–spectrum estimation the prototype operator of the optimum estimator is given by:

$$S^{(\alpha)}_{\widehat{P}_{MVUB}(\alpha)}(\tau,\nu) = \begin{cases} 1, & \text{where} \quad S_H(\tau,\nu) \neq 0, \\ 0, & \text{where} \quad S_H(\tau,\nu) = 0, \end{cases} \tag{E.8}$$

and it can be interpreted as a minimum norm Weyl–Heisenberg expansion of the system $\mathbf{H}$

$$\widehat{\mathbf{P}}_{MVUB}(\alpha) = \arg\min_{\mathbf{P}} \|\mathbf{P}\|^2, \qquad \text{subject to} \qquad \mathbf{H} = \int_t \int_f L_H^{(\alpha)}(t,f) \mathbf{P}^{(t,f)}(\alpha) dt\, df.$$

APPENDIX F: Time–Frequency Signal Representations

F.0.1 Short–Time Fourier Transform/Spectrogram

In the analysis of many natural signals such as speech or audio, one is interested in obtaining a temporally localized information about the spectral content of the signal. To this end, it is near at hand to apply a sliding window prior to the Fourier transform. This yields the *short–time Fourier transform*, a classical tool for time–varying signal processing [135, 7, 283, 251, 27, 250, 264, 14, 145, 249, 207, 115]:

$$STFT_x^{(\gamma)}(t,f) \stackrel{\text{def}}{=} \int_{t'} x(t')\gamma^*(t'-t)e^{-j2\pi ft'}dt', \tag{F.1}$$

where $\gamma(t)$ is the underlying analysis window. The STFT $STFT^{(\gamma)}(t,f)$ can be interpreted in various different ways:

- As the Fourier transform of the τ–dependent windowed signals $x(t)\gamma(t-\tau)$.
- As the output of a filter bank with impulse responses $h^{(f)}(\tau) = \gamma(\tau)e^{j2\pi\tau f}$, i.e., modulated versions of the analysis window $\gamma(t)$.
- As the coefficients of an expansion of the signal $x(t)$ time–frequency shifted versions of the analysis window $\gamma(t)$,

$$x(t) = \int_\tau \int_\nu \left\langle x, \mathbf{S}^{(\tau,\nu)}\gamma \right\rangle \left(\mathbf{S}^{(\tau,\nu)}\gamma\right)(t)d\tau d\nu = \int_\tau \int_\nu STFT_x^{(\gamma)}(\tau,\nu)\left(\mathbf{S}^{(\tau,\nu)}\gamma\right)(t)d\tau d\nu, \tag{F.2}$$

  where $\mathbf{S}^{(\tau,\nu)}$ is the time–frequency shift operator (cf. (A.21)). This shows that the signal can be recovered from its STFT. Here and in all of the following results concerning the STFT, $\|\gamma\|^2 = 1$ is presupposed.

The STFT preserves inner signal products in the following sense:

$$\left\langle STFT_x^{(\gamma)},\ STFT_y^{(\gamma)} \right\rangle = \langle x, y\rangle, \qquad \|\gamma\|^2 = 1. \tag{F.3}$$

Spectrogram. The STFT as a complex–valued signal representation is inappropriate for subjective signal analysis. The squared magnitude of the STFT, the spectrogram [202, 2, 8, 293]

$$SPEC_x^{(\gamma)}(t,f) = \left|STFT_x^{(\gamma)}(t,f)\right|^2 \tag{F.4}$$

may be interpreted as a time–frequency–parametrized energy distribution since

$$\int_t \int_f SPEC_x^{(\gamma)}(t,f)dtdf = \|x\|^2. \tag{F.5}$$

Note furthermore that the spectrogram is time–frequency shift–invariant in the following sense:

$$SPEC^{(\gamma)}_{S^{(\tau,\nu)}x}(t,f) = SPEC_x^{(\gamma)}(t-\tau, f-\nu).$$

Representation via Time–Frequency–Shifted Projection Operator. In the context of this thesis it is furthermore illuminating to write the spectrogram formally as a quadratic form

$$SPEC_x^{(\gamma)}(t,f) = \left\langle \mathbf{P}_\gamma^{(t,f)}x, x\right\rangle,$$

where $\mathbf{P}_\gamma^{(t,f)}$ is the rank–one projection operator onto the time–frequency shifted versions of the analysis window.

F.0.2 Gabor Expansion

The short–time Fourier transform entails considerable redundancy due to the fact that time–frequency shifted versions of the analysis window are highly linearly dependent. This redundancy can be removed (or at least largely reduced) by sampling the STFT on a rectangular grid,

$$G_x^{(\gamma)}(m,k) \stackrel{\text{def}}{=} \int_{t'} x(t')\gamma^*(t'-mT)e^{-j2\pi kFt'}dt', \tag{F.6}$$

where T and F denote the time and frequency sampling periods, respectively. This is a signal analysis point of view. Note, however, that Gabor's original idea [134] starts from a signal synthesis point of view, postulating a signal expansion in the form of

$$x(t) = \sum_m \sum_k G_x^{(\gamma)}(m,k)\, g(t-mT)e^{j2\pi kFt}, \tag{F.7}$$

where $g(t)$ is the so–called Gabor synthesis window or, alternatively, Gabor logon [24, 153, 174, 177, 281, 282, 128, 366, 263, 392, 206, 131, 391, 129, 288]. The Gabor expansion theory is mathematically equivalent to the concept of "coherent states" is it appears in mathematical physics [175, 274, 66, 6, 292, 69]. It should be emphasized that (i) validity of (F.7) imposes certain constraints on the sampling density and the synthesis window $g(t)$ [18, 28, 31] and (ii) the analysis window $\gamma(t)$ (and thus the coefficients $G_x^{(\gamma)}(m,k)$) may not be uniquely defined [291, 290, 307, 72, 181, 309, 308]. An in–depth discussion of these questions is beyond the scope of this appendix. However, for the context of this work it is interesting to emphasize the relevance of the fundamental threshold

$$TF = 1, \tag{F.8}$$

which is usually called *critical density* of the Gabor expansion. This choice was suggested in the original work of Gabor, who was inspired by heuristic, information theoretic considerations. Nowadays, it is well–known that his idea makes theoretical sense in so far as one can indeed span the whole $L_2(\mathbb{R})$ via a critical Gabor family $\{g^{(mT,kF)}\}$ for $TF = 1$. However, the Balian–Low theorem [68] essentially shows the nonexistence of a pair of "nice" (finite spectral and temporal second order moments) analysis and synthesis windows such that (F.7) works in the critical case $TF = 1$. Moreover, for the critical density, the numerical stability of the representation is poor. Hence, in practice one has to consider an *oversampled* Gabor expansion with density,

$$TF < 1.$$

F.0.3 Generalized Wigner Distribution

The spectrogram as a time–varying signal spectrum has the fundamental deficiency to depend on the arbitrary choice of a window function. One way to avoid the arbitrary choice of a window is to start with an ad hoc definition of a time–varying signal spectrum. By far the most prominent definition of a time–varying signal spectrum is the Wigner distribution [370, 76, 23, 51, 52, 55, 26, 41, 34, 229, 35, 13]

$$W_x(t,f) \stackrel{\text{def}}{=} \int_\tau x\left(t+\frac{\tau}{2}\right) x^*\left(t-\frac{\tau}{2}\right) e^{-j2\pi f\tau}d\tau. \tag{F.9}$$

Another classical definition is the Rihaczek distribution [302]

$$R_x(t,f) \stackrel{\text{def}}{=} \int_\tau x(t)x^*(t-\tau)e^{-j2\pi f\tau}d\tau = x(t)X^*(f)e^{-j2\pi ft}. \tag{F.10}$$

The Wigner distribution and the Rihaczek distribution share a number of attractive properties [159]. These representations can written in a unified way by defining an α–parametrized generalized (cross) Wigner distribution (GWD) [53, 161, 176]

$$W_{x,y}^{(\alpha)}(t,f) \stackrel{def}{=} \int_{\tau} x\left(t+\left(\frac{1}{2}-\alpha\right)\tau\right) y^*\left(t-\left(\frac{1}{2}+\alpha\right)\tau\right) e^{-j2\pi f\tau} d\tau. \tag{F.11}$$

The GWD $W_x^{(\alpha)}(t,f)$ reduces to the Wigner distribution for $\alpha = 0$ and to the Rihaczek distribution for $\alpha = 1/2$. In the present work we always presuppose

$$|\alpha| \leq 1/2.$$

Note that the spectrogram can be written as a smoothed GWD,

$$S_x^{(\gamma)}(t,f) = W_x^{(\alpha)}(t,f) * * W_\gamma^{(\alpha)*}(-t,-f). \tag{F.12}$$

We mention three properties of the GWD that are specifically relevant for this work (more properties can be found in the classical paper [51]):

- **Sesquilinear Form.** The GWD preserves inner signal products in the sense of a unitary, sesquilinear map ("Moyal's formula")

$$\left\langle W_{x_1,y_1}^{(\alpha)}, W_{x_2,y_2}^{(\alpha)} \right\rangle = \langle x_1, x_2 \rangle \langle y_1, y_2 \rangle^*. \tag{F.13}$$

- **Time–Frequency Shift–Covariance.** A time–frequency shift of the signal results in a corresponding shift of the GWD,

$$W_{S^{(\tau,\nu)}x}^{(\alpha)}(t,f) = W_x^{(\alpha)}(t-\tau, f-\nu). \tag{F.14}$$

- **Integral.** The total integral yields the signal's energy:

$$\int_t \int_f W_x^{(\alpha)}(t,f)\, dt\, df = \int_t |x(t)|^2 dt = \|x\|^2. \tag{F.15}$$

F.0.4 Generalized Ambiguity Function

The implicit or explicit use of time–frequency shifted versions of a prototype signal appears in different applications such as signal analysis, filtering, sonar/radar, and digital communication. In any of these applications the following question is important: How does the inner product of a signal $x(t)$ and its time–frequency shifted version

$$x^{(\tau,\nu)} \stackrel{\text{def}}{=} (\mathbf{S}^{(\tau,\nu)}x)(t) = x(t-\tau)e^{j2\pi\nu t}$$

depend on the time and frequency shifts τ and ν, respectively. This dependence can be formally represented by the *asymmetrical ambiguity function*, also called *time–frequency autocorrelation function* (the meaning of the superscript will be clarified soon) [371, 339, 338, 285, 220, 17, 320, 321, 350, 333, 15, 222, 16, 349] which is defined as:

$$A_x^{(1/2)}(\tau,\nu) \stackrel{def}{=} \left\langle x, x^{(\tau,\nu)} \right\rangle = \int_t x(t)x^*(t-\tau)e^{-j2\pi\nu t} dt. \tag{F.16}$$

In the context of this work we shall need the fact that the ambiguity function can be seen as an invertible one–to–one mapping of the signal's rank–one product $x(t)x^*(t')$ onto the correlative time–frequency plane [157]. The inversion formula is then given by[3]:

$$x(t)x^*(t') = \int_\nu A_x^{(1/2)}(t-t',\nu)e^{j2\pi\nu t} d\nu. \tag{F.17}$$

[3] The signal $x(t)$ itself is uniquely determined by the rank–one product $x(t)x^*(t')$ up to a constant phase factor.

The freedom in the definition of a time–frequency shift operator leads to a corresponding freedom in the definition of an ambiguity function. We have most prominently the just mentioned asymmetrical ambiguity function (F.16) and the symmetrical ambiguity function[4],

$$A_x^{(0)}(\tau,\nu) \stackrel{def}{=} \int_t x\left(t+\frac{\tau}{2}\right) x^*\left(t-\frac{\tau}{2}\right) e^{-j2\pi\nu t} dt. \tag{F.18}$$

Analog to the introduction of the generalized Wigner distribution, one can treat these definitions in parallel by using the family of the generalized (cross–)ambiguity function

$$A_{x,y}^{(\alpha)}(\tau,\nu) \stackrel{def}{=} \int_t x\left(t+\left(\frac{1}{2}-\alpha\right)\tau\right) y^*\left(t-\left(\frac{1}{2}+\alpha\right)\tau\right) e^{-j2\pi\nu t} dt. \tag{F.19}$$

In most applications one is merely interested in the (squared) magnitude of the ambiguity functions which is fortunately invariant w. r. t. α,

$$\left|A_x^{(\alpha_1)}(\tau,\nu)\right|^2 = \left|A_x^{(\alpha_2)}(\tau,\nu)\right|^2. \tag{F.20}$$

The ambiguity functions satisfy a number of useful properties [159]. The ambiguity function is the fundamental time–frequency working tool in this thesis, it leads to a compact formulation of the main results. We here mention just those properties which are of specific interest in our context:

- **Mutual Relation.** Any member of the family $A_x^{(\alpha)}(\tau,\nu)$ can be expressed in terms of the symmetrical ambiguity function multiplied by an α–dependent unimodular factor:

$$A_x^{(\alpha)}(\tau,\nu) = A_x^{(0)}(\tau,\nu) e^{-j2\pi\nu\tau\alpha}. \tag{F.21}$$

- **Sesquilinear Form.** The generalized ambiguity function preserves inner signal products in the sense of a unitary, sesquilinear map:

$$\left\langle A_{x_1,y_1}^{(\alpha)}, A_{x_2,y_2}^{(\alpha)} \right\rangle = \langle x_1, x_2\rangle \langle y_1, y_2\rangle^*. \tag{F.22}$$

- **Fourier Correspondence to the GWD.** The generalized ambiguity function is the symplectic Fourier transform to the generalized Wigner distribution,

$$A_{x,y}^{(\alpha)}(\tau,\nu) = \mathcal{F}_{t\to\nu}\mathcal{F}_{f\to\tau}^{-1} W_{x,y}^{(\alpha)}(t,f). \tag{F.23}$$

- **Temporal/Spectral Correlation.** As a time–frequency correlation function, $A_x^{(\alpha)}(\tau,\nu)$ is consistent with the conventional (α–invariant) temporal and spectral correlation functions, respectively:

$$A_x^{(\alpha)}(\tau,0) = \int_t x(t)x^*(t-\tau)dt, \tag{F.24}$$

$$A_x^{(\alpha)}(0,\nu) = \int_f X(f)X^*(f-\nu)df. \tag{F.25}$$

- **Time–frequency Shifting.** Time–frequency shifting of a signal corresponds to a modulation of its ambiguity function:

$$A_{S^{(t,f)}x}^{(\alpha)}(\tau,\nu) = A_x^{(\alpha)}(\tau,\nu) e^{-j2\pi(\nu t-\tau f)}. \tag{F.26}$$

  Hence, a time–frequency shift of the signal leaves the magnitude of the generalized ambiguity function unchanged:

$$\left|A_{Sx}^{(\alpha)}(\tau,\nu)\right|^2 = \left|A_x^{(\alpha)}(\tau,\nu)\right|^2. \tag{F.27}$$

[4]The underlying definition of the symmetrical time–frequency shift operator is discussed in the Appendix B.

- **Maximum.** The generalized ambiguity function takes on its maximum in the origin (as expected for a two–dimensional *autocorrelation function*)

$$\left|A_x^{(\alpha)}(0,0)\right|^2 = \|x\|^4 \geq \left|A_x^{(\alpha)}(\tau,\nu)\right|^2. \tag{F.28}$$

- **Radar Uncertainty Principle.** The volume under the magnitude–squared ambiguity function equals the squared energy of the signal:

$$\int_\tau \int_\nu \left|A_x^{(\alpha)}(\tau,\nu)\right|^2 d\tau\, d\nu = \|x\|^4. \tag{F.29}$$

  By the combined consideration of (F.28) and (F.29),

$$\int_\tau \int_\nu \left|A_x^{(\alpha)}(\tau,\nu)\right|^2 d\tau\, d\nu = \left|A_x^{(\alpha)}(0,0)\right|^2$$

  it gets obvious that the ambiguity function of any signal cannot be arbitrarily concentrated about the origin. This property is usually referred to as "radar uncertainty principle" because it forbids the existence of a radar pulse with ideally concentrated $\left|A_x^{(\alpha)}(\tau,\nu)\right|^2$. (Such a signal would be desirable for the classical radar problem.)

- **Temporal/Spectral Moments.** In the origin, the second derivatives of the ambiguity function equal the temporal and spectral moments:

$$-\frac{1}{\pi}\frac{\partial^2 A_x^{(\alpha)}}{\partial\nu^2}(0,0) = 4\pi \int_t t^2 \left|x(t)\right|^2 dt = T_x^2, \tag{F.30}$$

$$-\frac{1}{\pi}\frac{\partial^2 A_x^{(\alpha)}}{\partial\tau^2}(0,0) = 4\pi \int_f f^2 \left|X(f)\right|^2 df = F_x^2. \tag{F.31}$$

  One can use these properties to approximate $A_x^{(\alpha)}(\tau,\nu)$ via Taylor's formula (valid for small τ,ν and real–valued $x(t)$):

$$A_x^{(\alpha)}(\tau,\nu) \approx 1 - \frac{1}{2\pi}\left(F_x^2\tau^2 + T_x^2\nu^2\right), \tag{F.32}$$

$$\left|A_x^{(\alpha)}(\tau,\nu)\right|^2 \approx 1 - \frac{1}{\pi}\left(F_x^2\tau^2 + T_x^2\nu^2\right). \tag{F.33}$$

Bibliography

[1] M. Abramowitz and I. Stegun. *Handbook of Mathematical Functions.* Dover, New York, 1965.

[2] M.H. Ackroyd. Short–time spectra and time–frequency energy distributions. *J. Acoust. Soc. Amer.*, 50:1229–1231, 1970.

[3] D. Adler and S. Raz. Synthesis of legal Wigner filter. In *Proc. ICASSP 91*, pages 2057–2060, 1991.

[4] V.R. Algazi and D.J. Sakrison. On the optimality of Karhunen-Loeve expansion. *IEEE T-IT*, pages 319–321, March 1969.

[5] S.T. Ali, J.-P. Antoine, and J.-P. Gazeau. Continuous frames in Hilbert space. *Annals of Physics*, pages 1–37, February 1993.

[6] S.T. Ali, J.-P. Antoine, J.-P. Gazeau, and U.A. Müller. Coherent states and their generalizations. *Reviews in Math. Physics*, 7/7:1013–1104, 1995.

[7] J.B. Allen and L.R. Rabiner. A unified approach to short-time Fourier analysis and synthesis. *IEEE Proceedings*, 65(11), 1977.

[8] R.A. Altes. Detection, estimation and classification with spectrograms. *J. Acoust. Soc. Am.*, 67(4):1232–1246, Apr. 1980.

[9] R.A. Altes. Wide–band, proportional–bandwidth Wigner–Ville analysis. *IEEE T-ASSP*, 38(6):1005–1012, 1990.

[10] R.F.V. Anderson. The Weyl functional calculus. *J. Funct. Anal.*, 4:240–267, 1969.

[11] R.F.V. Anderson. On the Weyl functional calculus. *J. Funct. Anal.*, 6:110–115, 1970.

[12] R.F.V. Anderson. The multiplicative Weyl functional calculus. *J. Funct. Anal.*, 9:423–440, 1972.

[13] T. Arai. Some extensions of semiclassical limit $\hbar \to 0$ for Wigner functions on phase space. *J. Math. Phys.*, 36(2):622–630, 1995.

[14] L. Auslander, C. Buffalano, R. Orr, and R. Tolimieri. A comparison of the Gabor and short-time Fourier transforms for signal detection and feature extraction in noise environments. In *SPIE Proc. 1348*, 1990.

[15] L. Auslander and I. Gertner. Wide–band ambiguity function and the $a.x + b$ group. In L. Auslander, T. Kailath, and S. Mitter, editors, *Signal Processing Part I*, volume IMA 22, pages 1–12. Springer Verlag, New York, 1990.

[16] L. Auslander, F. Geshwind, and F. Warner. Radar waveform design and the Heisenberg group. *J. Appl. Comp. Harm. Anal.*, 2(4):350–362, 1995.

[17] L. Auslander and R. Tolimieri. Characterizing the radar ambiguity functions. *IEEE T-IT*, 30(6):832–836, 1984.

[18] R. Balian. Un principe d'incertitude fort en théorie du signal on en mécanique quantique. *C. R. Acad. Sci. Paris*, 292:1357–1362, 1981.

[19] R. Bamler. Linear shift–variant processing of 2–D signals via sequence convolution: Special cases and experimental results. In *Proc. IEEE ICASSP84*, pages 151–154, 1984.

[20] R. Bamler. *Mehrdimensionale lineare Systeme: Fourier–Transformation und Delta–Funktionen.* Springer Verlag, Berlin, 1989.

[21] R. Bamler and J. Hofer-Alfeis. 2D linear space–variant processing by coherent optics: a sequence convolution approach. *Optics Communications*, 43(2):97–102, Sept. 1982.

[22] M. Basseville. Detecting changes in signals and systems—a survey. *Automatica*, 24(3):309–326, 1988.

[23] M.J. Bastiaans. Transport equations for the Wigner distribution function. *Optica Acta*, 26(10):1265–1272, 1979.

[24] M.J. Bastiaans. Gabor's expansion of a signal into Gaussian elementary signals. *IEEE Proceedings*, 68:538–539, April 1980.

[25] M.J. Bastiaans. Uncertainty principle for spatially quasi-stationary, partially coherent light. *J. Opt. Soc. Am.*, 72(20):1441–1443, October 1982.

[26] M.J. Bastiaans. Use of the Wigner distribution function in optical problems. In *SPIE ECOOSA'84*, volume 492, pages 251–262, Amsterdam, October 1982.

[27] M.J. Bastiaans. On the sliding-window representation in digital signal processing. *IEEE Trans. Acoust. Speech Sig. Proc.*, 33:868–872, August 1985.

[28] G. Battle. Heisenberg proof of the Balian-Low theorem. *Lett. Math. Phys.*, 15:175–177, 1988.

[29] P.A. Bello. Characterization of randomly time–variant linear channels. *IEEE T-CS*, 11:360–393, 1963.

[30] J.J. Benedetto. Stationary frames and spectral estimation. In J. S. Byrnes and J. L., editors, *Probabilistic and Stochastic Advances in Analysis, with Applications*, volume NATO-ASI Series C. Kluwer Academic Publishers, to appear.

[31] J.J. Benedetto, C. Heil, and D.F. Walnut. Differentiation and the balian–low theorem. *J. Four. Anal. Appl.*, 1(4):355–403, 1995.

[32] C.A. Berger and L.A. Coburn. Berezin–Toeplitz estimates, 1995.

[33] J.O. Berger. *Statistical Decision Theory and Bayesian Analysis*. Springer Verlag, New York, 1985.

[34] J. Bertrand and P. Bertrand. A tomographic approach to Wigner's function. *Found. of Physics*, 17:397–405, 1987.

[35] M.U. Bikdash and K.-B. Yu. Analysis and filtering using the optimally smoothed Wigner distribution. *IEEE T-SP*, 41(4), 1993.

[36] B. Boashash. Time–frequency signal analysis. In S. Haykin, editor, *Advances in Spectrum Estimation*, volume 1, pages 418–517. Prentice Hall, Englewood Cliffs (NJ), 1990.

[37] H. Bölcskei. Gabor expansion and frame theory. Master's thesis, Techn.Univ.Vienna, 1994.

[38] H. Bölcskei, F. Hlawatsch, and H.G. Feichtinger. Equivalence of DFT filter banks and the Gabor expansions. In *SPIE Proc., Wavelet Applications in Signal and Image Processing III*, volume 2569, pages 128–139, San Diego (CA), July 1995.

[39] M. Boon, J. Zak, and J. Zucker. Rational von Neumann lattice. *J. Math. Phys.*, 24(2):316–323, 1983.

[40] R.K. Bose. Realizability theory of continuous linear operators on groups. *SIAM J. Math. Anal.*, 10(4):767–780, 1979.

[41] G.F. Boudreaux-Bartels and T.W. Parks. Time–varying filtering and signal estimation using Wigner distribution functions. *IEEE T-SP*, 34:442–451, 1986.

[42] W.R. Braun and U. Dersch. A physical mobile radio channel model. *IEEE T-VT*, 40(2):472–482, 1991.

[43] S. Brekke and K. Seip. Density theorems for sampling and interpolation in the Bargmann-Fock space, III. *Math. Scand.*, pages 112–126, 1993.

[44] J.B. Buckheit and D.L. Donoho. Time–frequency tilings which best expose the non–gaussian behavior of a stochastic process. In *Proc. IEEE-SP Int. Sympos. Time-Frequency Time-Scale Analysis*, pages 1–4, Paris, France, June 1996.

[45] S. Cambanis and E. Masry. Performance of discrete–time predictors of continuous–time stationary processes. *Philips J.Research*, 34(4):655–668, 1988.

[46] S.L. Campbell. Regularizations of linear time–varying singular systems. *Automatica*, 20(3):365–370, 1984.

[47] R.M. Canetti and M.D. Espana. Convergence analysis of the least–squares identification algorithm with a variable forgetting factor for time–varying linear systems. *Automatica*, 25(4):609–612, 1989.

[48] T. Chen and B.A. Francis. Linear time–varying $\langle_2$–optimal control of sampled–data systems. *Automatica*, 27(6):963–974, 1991.

[49] T. Chen and P.P. Vaidyanathan. Vector space framework for unification of one– and multidimensional filter bank theory. *IEEE T-SP*, 42(8):2006–2021, 1994.

[50] O. Christensen. *Frame decompositions in Hilbert Spaces.* PhD thesis, Aarhus Univ. and Univ. of Vienna, Aarhus (Dk) / Vienna (Austria), 1993.

[51] T.A.C.M. Claasen and W.F.G. Mecklenbräuker. The Wigner distribution — a tool for time–frequency signal analysis; part i: Continuous–time signals. *Philips J.Research*, 35(3):217–250, 1980.

[52] T.A.C.M. Claasen and W.F.G. Mecklenbräuker. The Wigner distribution — a tool for time–frequency signal analysis; part ii: Discrete–time signals. *Philips J.Research*, 35(4/5):276–300, 1980.

[53] T.A.C.M. Claasen and W.F.G. Mecklenbräuker. The Wigner distribution — a tool for time–frequency signal analysis; part iii: Relations with other time–frequency signal transformations. *Philips J.Research*, 35(6):372–389, 1980.

[54] T.A.C.M. Claasen and W.F.G. Mecklenbräuker. On stationary linear time–varying systems. *IEEE T-CS*, 29(3):169–184, 1982.

[55] T.A.C.M. Claasen and W.F.G. Mecklenbräuker. The aliasing problem in discrete time Wigner distributions. *IEEE Trans. Acoustics, Speech, and Signal Processing*, 31(5):1067–72, 1983.

[56] L. Cohen. Generalized phase–space distribution functions. *J. Math. Phys.*, 7:781–786, 1966.

[57] R. Coifman and M. Wickerhauser. Entropy–based algorithms for best basis selection. *IEEE T-IT*, 38(2):713–718, March 1992.

[58] C.E. Cook and M. Bernfeld. *Radar Signals—An Introduction to Theory and Applications.* Academic Press, New York, 1967.

[59] J.W. Cooley and J.W. Tukey. An algorithm for the machine computation of complex Fourier series. *Mathematics of Computation*, 19:297–301, April 1965.

[60] A. Córdoba and C. Fefferman. Wave packets and Fourier integral operators. *Comm. Partial Diff. Eq.*, 3:979–1005, 1978.

[61] G. Cristobal and R. Navarro. Space and frequency variant image enhancement based on a Gabor representation. *Pattern Recognition Letters*, 15(3):273–277, 1994.

[62] R.E. Crochiere and L.R. Rabiner. *Multirate Digital Signal Processing.* Prentice–Hall, New Jersey, 1983.

[63] Z. Cvetkovic and M. Vetterli. Overcomplete expansions and robustness. In *Proc. IEEE-SP Int. Sympos. Time-Frequency Time-Scale Analysis*, pages 325–328, Paris, France, June 1996.

[64] M. Dahleh and M.A. Dahleh. On slowly time–varying systems. *Automatica*, 27(1):201–205, 1991.

[65] I. Daubechies. On the distributions corresponding to bounded operators in the Weyl quantization. *Comm. Math. Phys.*, 75:229–238, 1980.

[66] I. Daubechies. Discrete sets of coherent states and their use in signal analysis. In I.W. Knowles and Y. Saito, editors, *Proceedings of the 1986 Int. Conf. Differential Equations and Mathematical Physics*, pages 73–82, Birmingham (Alabama), 1987. Springer.

[67] I. Daubechies. Time-frequency localization operators: a geometric phase space approach. *IEEE Trans. Inform. Theory*, 34/4, July 1988.

[68] I. Daubechies. The wavelet transform, time-frequency localization and signal analysis. *IEEE T-IT*, pages 961–1005, 1990.

[69] I. Daubechies. Orthonormal bases of coherent states: the canonical case and the $ax + b$-group. In D.S. Feng and J.R. Klauder, editors, *Proc. of the Oak Ridge Symposium on Coherent States*, pages 103–117, Singapore, June 1993 1995. World Scientific.

[70] I. Daubechies, S. Jaffard, and J.L. Journe. A simple Wilson orthonormal basis with exponential decay. *SIAM J. Math. Anal.*, 22:554–573, 1991.

[71] I. Daubechies and A.J.E.M. Janssen. Two theorems on lattice expansions. *IEEE Trans. Info. Theory*, 39(1):3–6, 1993.

[72] I. Daubechies, H. Landau, and Z. Landau. Gabor time-frequency lattices and the Wexler-Raz identity. *J. Four. Anal. Appl.*, 1(4):437–478, 1995.

[73] I. Daubechies and T. Paul. Time-frequency localization operators: a geometric phase space approach ii. the use of dilations. *Inverse Problems*, 4:661–680, 1988.

[74] M.J. Davis and E.J. Heller. Semiclassical Gaussian basis set method for molecular vibrational wave functions. *J.Chem.Phys.*, 71/8:3383–3395, 1979.

[75] N.G. De Bruijn. Uncertainty principles in Fourier analysis. In O. Shisha, editor, *Inequalities*. Academic Press, New York, 1967.

[76] N.G. De Bruijn. A theory of generalized functions with applications to Wigner distribution and Weyl correspondence. *Nieuw Archief voor Wiskunde*, 21(3):361–371, 1973.

[77] E. Degli, G. Mirko, and S. Isola. Classical limit of the quantized hyperbolic toral automorphisms. *Comm. Math. Phys.*, 167(3):471–507, 1995.

[78] M.H. DeGroot. *Optimal Statistical Decisions*. McGraw–Hill, New York, 1970.

[79] E. Del Re. Adaptive channel estimation for mobile radio, 1990.

[80] P.A. Delaney and D.O. Walsh. A bibliography of higher–order spectra and cumulants. *IEEE SP Magazine*, July 1994.

[81] C.A. Desoer. Slowly time–varying systems $\dot{x} = a(t)x$. *IEEE T-AC*, 14:780–781, 1969.

[82] E.J. Diethorn and D.C. Munson. A linear time–varying system framework for noniterative discrete–time band–limited signal extrapolation. *IEEE T-SP*, 39(1):–, 1991.

[83] G. Doblinger. An adaptive subspace filter for noise reduction. In *Proc.EUSIPCO–94*, 1994.

[84] D.L. Donoho, S. Mallat, and R. von Sachs. Estimating covariances of locally stationary processes: consistency of best basis methods. In *Proc. IEEE-SP Int. Sympos. Time-Frequency Time-Scale Analysis*, pages 337–340, Paris, France, June 1996.

[85] J.L. Doob. *Stochastic processes*. Wiley, New York, 1965.

[86] D. A. Dubin, M. A. Hennings, and T.B. Smith. Quantization in polar coordinates and the phase operator. *Publ. Res. Inst. Math. Sci.*, 30(3):479–532, 1994.

[87] E. Dubois. The sampling and reconstruction of time–varying imagery with application in video systems. *IEEE Proc.*, 73(4):502–522, April 1985.

[88] D.E. Dudgeon and R.M. Merserau. *Multidimensional Signal Processing*. Prentice Hall, Englewood Cliffs (NJ), 1984.

[89] D. Dunn, W. Higgins, and J. Wakeley. Texture segmentation using 2-D Gabor elementary functions. *IEEE Trans Patt. Anal. Machine Intell.*, 22/2:130–149, 1994.

[90] D.F. Dunn and W.E. Higgins. Optimal Gabor filters for texture segmentation. *IEEE Transactions on Image Processing*, 4(7):947–964, 1995.

[91] D.L. Duttweiler and T. Kailath. RKHS approach to detection and estimation problems— part iv: Non-gaussian detection. *IEEE T-IT*, 19(1):19–28, 1973.

[92] D.L. Duttweiler and T. Kailath. RKHS approach to detection and estimation problems— part v: Parameter estimation. *IEEE T-IT*, 19(1):29–37, 1973.

[93] T. Ebrahimi and M. Kunt. Image compression by Gabor expansion. *Opt. Eng.*, 30/7:873–880, 1991.

[94] B. Edler. Coding of audio signals with overlapping block transform and adaptive window functions. *Frequenz (in German)*, 43:252–256, 1989.

[95] P.D. Einziger. Gabor expansion of an aperture field in exponential elementary beams,. *IEE Electron. Lett.*, 24:665–666, 1988.

[96] P.D. Einziger, S. Raz, and M. Shapira. Gabor representation and aperture theory. *J.Opt.Soc.Am. (A)*, 3:508–522, 1986.

[97] Y. Ephraim and H.L. Van Trees. A signal subspace approach for speech enhancement. In *Proc. IEEE ICASSP–93*, pages 339–358, Minneapolis (MN), April 1993.

[98] E. Eweda and O. Macchi. Tracking error bounds of adaptive nonstationary filtering. *Automatica*, 21(3):249–254, 1985.

[99] L.D. Faddeev. Discrete Heisenberg–Weyl group and modular group. *Lett. Math. Phys.*, 34(3):249–254, 1995.

[100] Q. Fan. Phase space analysis of the identity. *J. Math. Phys.*, 34(8):3471–3477, 1994.

[101] A. Faridani. A generalized sampling theorem for locally compact abelian groups. *Comp. Math.*, 63:307–327, 1994.

[102] S. Farkash and S. Raz. An approach to Gabor-space filtering. Technical Report EE Publ. No.227, Dept.Elect.Eng., Technion, Israel, 1989.

[103] S. Farkash and S. Raz. The legality problem of linear systems in Gabor time–frequency space. *Sig. Proc.*, 34(3):283–295, 1993.

[104] S. Farkash and S. Raz. Linear systems in Gabor time–frequency space. *IEEE T-SP*, 42(3):611–617, 1994.

[105] M.V. Fedoryuk. The stationary phase method and pseudodifferential operators. *Russian Math. Surveys*, 26:65–115, 1971.

[106] C.L. Fefferman. The uncertainty principle. *Bull. Amer. Math. Soc. (N.S.)*, 9(2), 1983.

[107] H. G. Feichtinger, K. Gröchenig, and D. Walnut. Wilson bases and modulation spaces. *Math.Nachrichten*, 155:7–17, 1992.

[108] H.G. Feichtinger, O. Christensen, and T. Strohmer. A group-theoretical approach to Gabor analysis. *Optical Engineering*, 34(6):1697–1704, 1995.

[109] H.G. Feichtinger and K. Gröchenig. Banach spaces related to integrable group representations and their atomic decompositions, i. *J. Funct. Anal.*, 86:307–340, 1989.

[110] H.G. Feichtinger and K. Gröchenig. Banach spaces related to integrable group representations and their atomic decompositions, ii. *Monatsh. f. Mathematik*, 108:129–148, 1989.

[111] H.G. Feichtinger and K. Gröchenig. Gabor wavelets and the Heisenberg group: Gabor expansions and Short Time Fourier transform from the group theoretical point of view. In C.K. Chui, editor, *Wavelets – A Tutorial in Theory and Applications*, pages 359–397. Academic Press, Boston, 1992.

[112] H.G. Feichtinger and W. Hörmann. Harmonic analysis of generalized stochastic processes, 1991. unpublished manuscript.

[113] H.G. Feichtinger, W. Kozek, and P. Prinz. Gabor analysis and locally compact abelian groups. In *Gabor Analysis and Algorithms: Theory and Applications*. Birkhäuser, Boston, 1997. (in preparation).

[114] H.G. Feichtinger, P. Prinz, and W. Kozek. Gabor systems with good TF-localization and applications to image processing. In *Proc. ICIP-96*, volume 1, pages 249–252, Lausanne, 1996.

[115] B.G. Ferguson and B.G. Quinn. Application of the short-time Fourier transform and the Wigner-Ville distribution to the acoustic localization of aircraft. *J. Acoust. Soc. Am.*, 96/2:821–827, 1994.

[116] V. Filimon, W. Kozek, W. Kreuzer, and G. Kubin. LMS and RLS tracking analysis for WSSUS channels. In *Proc. IEEE ICASSP–93*, pages III/348–351, Minneapolis (MN), 1993.

[117] J.L. Flanagan. Speech technology and computing: A unique partnership. *IEEE Communications Magazine*, pages 84–89, May 1992.

[118] P. Flandrin. On the positivity of the Wigner–Ville spectrum. *Sig. Proc.*, 11(2):187–189, 1986.

[119] P. Flandrin. Maximal signal energy concentration in a time–frequency domain. In *Proc. IEEE ICASSP–88*, pages 2176–2179, 1988.

[120] P. Flandrin. A time–frequency formulation of optimum detection. *IEEE T-ASSP*, 36(9):1377–1384, 1988.

[121] P. Flandrin. Time–dependent spectra for nonstationary stochastic processes. In G. Longo and B.Picinbono, editors, *Time and Frequency Representation of Signals and Systems*, pages 69–124. Springer, 1989.

[122] P. Flandrin. Time–scale analyses and self–similar stochastic processes. In *NATO Advanced Study Institute on "Wavelets and their Applications"*, 1992.

[123] B. Fleury. *Charakterisierung von Mobil– und Richtfunkkanälen mit schwach stationären Fluktuationen und unkorrelierter Streuung (WSSUS).* PhD thesis, ETH Zürich, Switzerland, (in German), 1990.

[124] D.H. Foley and J.W. Sammon. An optimal set of discriminant vectors. *IEEE Trans. Comput.*, 24(3):130–138, 1975.

[125] G.B. Folland. *Harmonic Analysis in Phase Space.* Annals of Math. Studies. Princeton Univ. Press, Princeton (NJ), 1989.

[126] B. Francis. Snippets of $h^\infty-$ control theory. In F.A. Grünbaum, J.W. Helton, and P. Khargonekar, editors, *Signal Processing Part II*, volume IMA 23, pages 79–98. Springer Verlag, New York, 1990.

[127] A. Francos and M. Porat. Parametric estimation of multicomponent signals using minimum cross entropy time–frequency distributions. In *Proc. IEEE Int. Symp. Time–Frequency/Time–Scale Analysis*, 1996.

[128] B. Friedlander and B. Porat. Detection of transient signal by the Gabor representation. *IEEE Trans. ASSP*, 37(2):169–180, February 1989.

[129] B. Friedlander and A. Zeira. Oversampled Gabor representations for transient signals. *IEEE Trans. Signal Proc.*, 43(9):2088–2095, September 1995.

[130] A. Friedmann and M.L. Honig. On the spread of continuous–time linear systems. *SIAM J. Matrix Anal. Appl.*, 21(3):757–770, 1990.

[131] M. Friedmann and J. Shamir. Submicron feature analysis using optical Gabor transforms. In *The 9th Meeting on Optical Engineering in Isreal*, Tel–Aviv, Oct. 1994.

[132] K. Fukunaga and W.L.G. Koontz. Application of the Karhunen–Loeve expansion to feature selection and ordering. *IEEE Trans. Comput.*, 19:311–318, 1970.

[133] S.A. Gaal. *Linear Analysis and Representation Theory.* Springer–Verlag, New York, 1973.

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

[135] G. Gambardella. A contribution to the theory of short–time spectral analysis with nonuniform filters. *IEEE T-CT*, 18(4):455–460, 1971.

[136] W.A. Gardner. The spectral correlation theory of cyclostationary time–series. *Sig. Proc.*, 11:13–36, 1986.

[137] W.A. Gardner. Correlation estimation and time–series modeling for nonstationary processes. *Sig. Proc.*, 15:31–41, 1988.

[138] W.A. Gardner. *Statistical Spectral Analysis.* Prentice Hall, New Jersey, 1988.

[139] W.A. Gardner. Exploitation of spectral redundancy in cyclostationary signals. *IEEE Signal Processing Magazine*, pages 14–36, April 1991.

[140] W. Gersch and G. Kitagawa. Smoothness priors transfer function estimation. *Automatica*, 25(8):603–608, 1989.

[141] A. Gersho. Characterization of time–varying linear systems. *IEEE Proc.*, 51:238, January 1963.

[142] O.D. Grace. Instantaneous power spectra. *J. Acoust. Soc. Am.*, 69(1):191–198, 1981.

[143] O.M. Grasselli and S. Longhi. Disturbance localization by measurement feedback for linear perodic discrete–time systems. *Automatica*, 24(3):375–385, 1988.

[144] Y. Grenier. Time–dependent ARMA modeling of nonstationary signals. *IEEE Trans. Acoust. Speech and Signal Processing*, 31(4):899–911, 1984.

[145] K. Gröchenig. Irregular sampling of wavelet and short time Fourier transforms. *Constructive Approx. Theory.*, 9:283–297, 1993.

[146] F.A. Grünbaum. Some new explorations into the mystery of time and band limiting. *Advances in Applied Mathematics*, 13:328–349, 1992.

[147] S. Gunnarson and L. Ljung. Frequency domain tracking characteristics of adaptive algorithms. *IEEE T-ASSP*, 37(7):1072–1089, 1989.

[148] P.R. Halmos and V.S. Sunder. *Bounded Integral Operators on L^2 Spaces.* Springer–Verlag, New York, 1978.

[149] J.K. Hammond and R.F. Harrison. Wigner–Ville and evolutionary spectra for covariance equivalent nonstationary random processes. In *Proc. IEEE ICASSP–85*, pages 27.4.1–27.4.4, Tampa (FL), 1985.

[150] B. Hassibi, A.H. Sayed, and T. Kailath. h^∞ optimality of the LMS algorithm. *IEEE T-SP*, 44(2):267–280, 1996.

[151] S. Haykin. *Adaptive Filter Theory.* Prentice Hall, New Jersey, 1991.

[152] C. Heil, J. Ramanathan, and P. Topiwala. Asymptotic singular value decay of time–frequency localization operators. In *Proc. SPIE 94*, volume 2303, pages 15–24, 1994.

[153] C. Heil and D. Walnut. Gabor and wavelet expansions. In J. Byrnes, editor, *Recent Advances in Fourier Analysis and Its Applications*, pages 441–454. Kluwer Publ., Dordrecht, 1990.

[154] C.W. Helstrom. *Quantum Detection and Estimation Theory.* Academic Press, New York, 1976.

[155] M.A. Hennings. The Weyl transformation and quantisation for locally compact abelian groups. *Pubi. RIMS Kyoto University*, 21:1223–1235, 1985.

[156] S. Hermann, U. Martin, R. Reng, H.W. Schuessler, and K. Schwarz. High resolution channel measurement for mobile radio. In L. Torres, E. Masgrau, and M.A. Lagunas, editors, *Signal Processing V: Theories and Applications*, pages 1903–1906. Elsevier Science Publishers B.V., 1990.

[157] F. Hlawatsch. Regularity and unitarity of bilinear time–frequency signal representations. *IEEE T-IT*, 38(1):82–94, 1992.

[158] F. Hlawatsch and H. Bölcskei. Time–frequency analysis of frames. In *Proc.IEEE-SP Int. Sympos. Time-Frequency Time-Scale Analysis*, pages 52–55, Philadelphia (PA), Oct. 1995.

[159] F. Hlawatsch and G.F. Boudreaux-Bartels. Linear and quadratic time–frequency signal representations. *IEEE Signal Process.Magazine*, 9:21–67, April 1992.

[160] F. Hlawatsch and W. Kozek. Time–frequency projection filters and TF signal expansions. *IEEE T-SP*, 42(12):3321–3334, 1994.

[161] F. Hlawatsch and R. Urbanke. Bilinear time–frequency representations of signals: The shift–scale invariant class. *IEEE T-SP*, 42(2):357–366, 1994.

[162] P. Höher. A statistical discrete–time model for the WSSUS multipath channel. *IEEE T-VT*, 41(4):461–468, 1992.

[163] R.B. Holmes. Mathematical foundations of signal processing. *SIAM Rev.*, 21(3):361–388, 1979.

[164] M.L. Honig, K. Steiglitz, B. Gopinath, and S.P. Boyd. Bounds on the maximum throughput for digital communications with finite–precision and amplitude constraints. *IEEE T-IT*, 36(3):472–484, 1990.

[165] K. Horiuchi. Sampling principle for continuous signals with time–varying bands. *Information and Control*, 13:53–61, 1968.

[166] L. Hörmander. The Weyl calculus of pseudo–differential operators. *Comm. Pure Appl. Math.*, 32:359–443, 1979.

[167] H. Hotelling. Analysis of a complex of statistical variables into principle components. *J.Educ.Psychology*, 24:427–441 and 498–520, 1933.

[168] R. Howe. Quantum mechanics and partial differential equations. *J. Funct. Anal.*, 38:188–254, 1980.

[169] N.-C. Huang and J.K. Aggarwal. On linear shift–variant digital filters. *IEEE T-CS*, 27(8):672–679, 1980.

[170] R.L. Hudson. When is the Wigner quasi–probability density non–negative. *Rep. Math. Phys*, 6:249–252, 1974.

[171] M. Iwadare, A. Sugiyama, F. Hazu, A. Hirano, and T. Nishitani. A 128 kb/s Hi–Fi audio CODEC based on adaptive transform coding with adaptive block size. *IEEE J. on Selected Areas of Communications*, 10(1):138–144, 1992.

[172] A.K. Jain. A fast Karhunen–Loeve transform for a class of random processes. *IEEE T-C*, 24:1023–1029, 1976.

[173] A.J.E.M. Janssen. Gabor representation of generalized functions. *J. Math. Anal. Appl.*, 83:377–394, 1981.

[174] A.J.E.M. Janssen. Bargmann transform, Zak transform, and coherent states. *, J. Math. Phys.*, 23(5):720–731, May 1982.

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

[176] A.J.E.M. Janssen. Gabor representations and Wigner distributions for signals. In *Proc. IEEE Conf. ASSP , April 1983. 41B.2.1*, 1983.

[177] A.J.E.M. Janssen. A note on Hudson's theorem about functions with nonnegative Wigner distributions. *SIAM J. Math. Anal.*, 15/1:170–176, 1984.

[178] A.J.E.M. Janssen. The Zak transform: A signal transform for sampled time-continuous signals. *Philips J. Res.*, 43(1):23–69, 1988.

[179] A.J.E.M. Janssen. Wigner weight functions and Weyl symbols of non–negative definite linear operators. *Philips J. Res.*, 44:7–42, 1989.

[180] A.J.E.M. Janssen. Duality and biorthogonality for Weyl-Heisenberg frames. *J. Four. Anal. Appl.*, 1(4):403–437, 1995.

[181] A.J.E.M. Janssen. A density theorem for time–continuous filter banks. In J. Zeevi and R. Coifman, editors, *Signal and Image Representation in Combined Spaces*. Academic Press, New York, San Francisco, London, to appear, 1997. to appear.

[182] A.J.E.M. Janssen and S. Zelditch. Szegö limit theorems for the harmonic oscillator. *Trans. Amer. Math. Soc.*, 280(2):563–587, 1983.

[183] N.S. Jayant and P. Noll. *Digital Coding of Waveforms*. Prentice Hall, New Jersey, 1984.

[184] H.E. Jensen, T. Hoeholdt, and J. Justesen. Double series representation of bounded signals. *IEEE T-IT*, 34(4):613–624, 1988.

[185] C.R.Jr. Johnson. On the interaction of adaptive filtering, identification, and control. *IEEE Signal Processing Magazine*, pages 22–37, March 1995.

[186] D.L. Jones and R.G. Baraniuk. A simple scheme for adapting time–frequency representations. *IEEE T-SP*, 42(12):3530–3535, 1994.

[187] D.L. Jones and R.G. Baraniuk. An adaptive optimal–kernel time–frequency representation. *IEEE T-SP*, 43(10):2361–2371, 1995.

[188] G. Jourdain and G. Tziritas. Communication over fading dispersive channels. *Sig. Proc.*, 27:672–679, 1980.

[189] Y.T. Juang. Robust stability and robust pole assignment of linear systems with structured uncertainty. *IEEE T-AC*, 36(5):635–637, 1991.

[190] T. Kailath. Channel characterization: time–variant dispersive channels. In E. J. Baghdady, editor, *Lectures on Communication Theory*, pages 95–123. McGraw–Hill, 1961.

[191] T. Kailath. Measurements on time–variant communication channels. *IEEE T-IT*, 8(5):229–236, Sept. 1962.

[192] T. Kailath. RKHS approach to detection and estimation problems— part i: Deterministic signals in non-gaussian noise. *IEEE T-IT*, 17(5):530–549, 1971.

[193] R.E. Kalman. Mathematical description of linear dynamical systems. *SIAM J. Control*, 1(2):152–192, 1963.

[194] H. Karhunen. Über lineare Methoden in der Wahrscheinlichkeitsrechnung. *Ann.Acad.Science Fenn*, Ser. A.I.37, 1947.

[195] Y. Katznelson. *An Introduction to Harmonic Analysis.* Dover, New York, 1976.

[196] A.S. Kayhan, A. El-Jaroudi, and L.F. Chaparro. Data–adaptive evolutionary spectral estimation. *IEEE T-SP*, 43(1), 1995.

[197] R.S. Kennedy. *Fading Dispersive Communication Channels.* Wiley, New York, 1969.

[198] M. Kirby and L. Sirovich. Application of the Karhunen–Loeve procedure for the characterization of human faces. *IEEE T-PAMI*, 12, January 1990.

[199] H. Kirchauer, F. Hlawatsch, and W. Kozek. Time–frequency formulation and design of nonstationary Wiener filters. In *Proc. IEEE ICASSP–95*, pages 1089–1092, Detroit, 1995.

[200] J.R. Klauder. The design of radar signals having both high range resolution and high velocity resolution. *Bell Syst. Tech. J.*, pages 809–820, July 1960.

[201] K. Kodera, R. Gendrin, and C. Villedary. Analysis of time–varying signals with small BT values. *IEEE T-ASSP*, 26(1):64–76, 1978.

[202] R. Koenig, H.K. Dunn, and L.Y. Lacy. The sound spectrograph. *J. Acoust. Soc. Amer.*, 18:19–49, 1946.

[203] J.J. Kohn and L. Nirenberg. An algebra of pseudo–differential operators. *Comm. Pure Appl. Math.*, 18:269–305, 1965.

[204] W. Kozek. On the generalized Weyl correspondence and its application to time–frequency analysis of linear time–varying systems. In *Proc. IEEE Int.Symp.Time–Frequency/Time–Scale Analysis*, pages 167–170, Victoria (Canada), 1992.

[205] W. Kozek. Time–frequency signal processing based on the Wigner–Weyl framework. *Signal Processing*, 29:77–92, 1992.

[206] W. Kozek. Matched generalized Gabor expansion of nonstationary processes. In *Proc. IEEE Int.Conf. Signals, Systems and Computers*, pages 499–503, Pacific Grove (CA), 1993.

[207] W. Kozek. Optimally Karhunen–Loeve like STFT expansion of nonstationary processes. In *Proc. IEEE ICASSP–93*, pages 428–431, Minneapolis (MN), 1993.

[208] W. Kozek. On the transfer function calculus for underspread LTV channels. *to appear in IEEE Trans. Signal Processing, Special Issue on Advanced Communications*, 1996.

[209] W. Kozek. On the underspread/overspread classification of nonstationary random processes. In *Proc. ICIAM 95, Special Issue of ZAMM*, volume 3, pages 63–66, Hamburg, 1996.

[210] W. Kozek and F. Hlawatsch. A comparative study of linear and nonlinear time–frequency filters. In *Proc. IEEE Int.Symp.Time–Frequency/Time–Scale Analysis*, pages 163–166, Victoria (Canada), 1992.

[211] W. Kozek, F. Hlawatsch, H. Kirchauer, and U. Trautwein. Correlative time–frequency analysis and classification of nonstationary processes. In *Proc. IEEE Int.Symp.Time–Frequency/Time–Scale Analysis*, pages 417–420, Philadelphia (PA), 1994.

[212] W. Kozek and K. Riedel. Quadratic time–varying spectral estimation for underspread processes. In *Proc. IEEE Int.Symp.Time–Frequency/Time–Scale Analysis*, pages 460–463, Philadelphia (PA), 1994.

[213] B.V.K.V. Kumar and K. DeVos. Linear system description using Wigner distribution functions. In *Proc. SPIE Advanced Algorithms and Architectures for Signal Processing II*, volume 826, pages 115–124, 1987.

[214] R.R. Kurth. Distributed–parameter state–variable model for time–variant channels. *IEEE T-IT*, 17(5):558–565, 1971.

[215] H.K. Kwok and D.L. Jones. Improved FM demodulation in a fading environment. In *Proc. IEEE-SP Int. Sympos. Time-Frequency Time-Scale Analysis*, pages 9–12, Paris, France, June 1996.

[216] H.J. Landau and H.O. Pollak. Prolate spheroidal wave functions, Fourier analysis and uncertainty—ii. *Bell System Technical Journal*, 40(1):65–84, 1961.

[217] H.J. Landau and H.O. Pollak. Prolate spheroidal wave functions, Fourier analysis and uncertainty—iii: The dimension of the space of essentially time– and band–limited signals. *Bell System Technical Journal*, 41(4):1295–1336, 1962.

[218] M.G. Larimore, S.L. Wood, and J.R. Treichler. Tracking speed requirements for time–varying adaptive channel equalizers. In *Proc. IEEE ICASSP'87*, pages 2157–2160, Dallas (Texas), 1987.

[219] W.H. Lau, J. Austin, A. Hewitt, W. Vilar, and L. Martin. Analysis of the time–variant structure of microwave line–of–sight multipath phenomena. *IEEE T-C*, 39(6):847–855, 1991.

[220] M.J. Levin. Instantaneous spectra and ambiguity functions. *IEEE T-IT*, 13:95–97, 1967.

[221] Y. Li and H.-F. Chen. Robust adaptive pole placement for linear time–varying systems. *IEEE T-AC*, 41(5):714–719, 1996.

[222] E.H. Lieb. Integral bounds for radar ambiguity functions and Wigner distributions. *J. Math. Phys.*, 31(3):594–599, 1990.

[223] D.V. Lindley. *Making Decisions*. Wiley, London, 1971.

[224] L. Ljung. *System Identification: Theory for the User*. Prentice Hall, Englewood cliffs, 1987.

[225] M. Loeve. Functions aleatoires de seconde ordre. In P. Levy, editor, *Processus Stochastiques et Mouvement Brownien*. Hermann, Paris, France, 1948.

[226] D. Ludwig. The Radon transform on euclidean space. *Comm. Pure Appl. Math.*, 19:49–81, 1966.

[227] H.D. Lueke. *Signalübertragung*. Springer Verlag, Berlin, 1983.

[228] S. Mallat, G. Papanicolaou, and Z. Zhang. Adaptive covariance estimation of locally stationary processes. Technical report, New York University, 1996.

[229] N. Marcuvitz. The quasiparticle view of wave propagation. *IEEE Proc.*, pages 1350–1358, Oct. 1991.

[230] W.D. Mark. Spectral analysis of the convolution and filtering of nonstationary random processes. *J. Sound Vib.*, 11(1):19–63, 1970.

[231] R.J. Marks, J.F. Walkup, and M.O. Hagler. Sampling theorems for linear shift–variant systems. *IEEE T-CS*, 25(4):226–233, 1978.

[232] J.-B. Martens. The Hermite transform—theory. *IEEE T-ASSP*, 38(9):1595–1606, 1990.

[233] W. Martin and P. Flandrin. Wigner–Ville spectral analysis of nonstationary processes. *IEEE T-ASSP*, 33(6):1461–1470, 1985.

[234] R. Mathias. The Hadamard operator norm of a circulant and applications. *SIAM J. Matrix Anal. Appl.*, 14(4):1152–1167, Oct. 1993.

[235] G. Matz and F. Hlawatsch. Time–frequency formulation and design of optimal detectors. In *TFTS'96 - Symposium on Time-Frequency and Time-Scale Analyis*, Paris, June 1996.

[236] G. Matz, F. Hlawatsch, and W. Kozek. Weyl spectral analysis of nonstationary random processes. In *TFTS'95 - Symposium on Applications of Time–Frequency and Time–Scale Methods University of Warwick, Coventry, UK*, pages 120–127, Warwick, August 1995.

[237] G. Matz, F. Hlawatsch, and W. Kozek. The Weyl spectrum of nonstationary random processes. *submitted to IEEE Trans. Signal Processing*, 1995.

[238] G. Mauceri. The Weyl transform and bounded operators on $l^p(\mathbb{R}^n)$,. *J. Funct. Anal.*, 39:408–429, 1980.

[239] R.J. McAuly and T.F. Quatieri. Low–rate speech coding based on the sinusoidal model. In S. Furui and M.M. Sondhi, editors, *Advances in Speech Signal Processing*. Marcel Dekker, New York, 1992.

[240] A. Megretski. Frequency–domain criteria of robust stability for slowly time–varying systems. *IEEE T-AC*, 40:153–155, 1995.

[241] A. Messiah. *Quantenmechanik*. de Gruyter, Berlin, New York, 1976.

[242] Y. Meyer. Wavelets and operators. In I. Daubechies, editor, *AMS Proc. of Symposia in Appl. Math: Different Perspectives on Wavelets*, pages 35–58. AMS, Providence (RI), 1993.

[243] K.S. Miller. *Complex Stochastic Processes.* Addison Wesley, 1974.

[244] L.K. Montgomery. A generalization of the Gabor–Helstrom transform. *IEEE Trans. Information Theory*, pages 344–345, April 1967.

[245] C.T. Mullis and L.L. Sharf. Quadratic estimators of the power spectrum. In S. Haykin, editor, *Advances in Spectrum Analysis and Array Processing.* Prentice Hall, New Jersey, 1991.

[246] N.J. Munch. Noise reduction in tight Weyl-Heisenberg frames. *IEEE Trans. Info. Theory*, 38(2/II):608–616, 1992.

[247] D.C. Munson. Minimum sampling rates for linear shift–variant discrete–time systems. *IEEE T-ASSP*, 33(6):1556–1561, 1985.

[248] H. Naparst. Radar signal choice and processing for a dense target environment. In F.A. Grünbaum, J.W. Helton, and P. Khargonekar, editors, *Signal Processing Part II*, volume IMA 23, pages 294–319. Springer Verlag, New York, 1990.

[249] S.H. Nawab and E. Dorken. Efficient STFT approximation using a quantization and differencing method. In *Proc. IEEE ICASSP–93*, volume 3, pages 587–590, 1993.

[250] S.H. Nawab and T. Quatieri. Short–time Fourier transform. In J.S. Lim and A.V. Oppenheim, editors, *Advanced Topics in Signal Processing.* Prentice Hall, Englewood Cliffs (New Jersey), 1988.

[251] S.H. Nawab, T. Quatieri, and J.S. Lim. Signal reconstruction from short–time Fourier transform magnitude. *IEEE T-ASSP*, 31(4), 1983.

[252] A.W. Naylor and G.R. Sell. *Linear Operator Theory in Engineering and Science.* Springer, New York, 1982.

[253] M. Niedzwiecki. Steady–state and parameter tracking properties of self–tuning minimum variance regulators. *Automatica*, 25(4):597–602, 1989.

[254] M. Niedzwiecki. Identification of time–varying systems usings combined parameter estimation and filtering. *IEEE T-ASSP*, 38:679–686, 1990.

[255] M. Niedzwiecki. Recursive functional series modeling estimators for time–varying plants — more bad news than good. *IEEE T-AC*, 35(5):610–616, 1990.

[256] H. Niemi. Stochastic processes as Fourier transforms of stochastic measures. *Ann. Acad. Sci. Fenn. Ser. A. I. 591*, pages 1–47, 1975.

[257] H. Niemi. On the linear prediction problem of certain non–stationary stochastic processes. *Math. Scand.*, 39:146–160, 1976.

[258] C.L. Nikias and J.M. Mendel. Signal processing with higher–order spectra. *IEEE Signal Processing Magazine*, pages 10–37, July 1993.

[259] P. Noll. Wideband speech and audio coding. *IEEE Communications Magazine*, 31(11):34–44, 1993.

[260] K. Nowak. Some eigenvalue estimates for wavelet related Toeplitz operators. *Colloquium Mathematicum*, LXV, Fasc. 1:149–156, 1993.

[261] K. Nowak. On Calderon–Toeplitz operators. *Monatshefte f. Mathematik*, 116:49–72, 1994.

[262] Gaarder N.T. Scattering function estimation. *IEEE T-IT*, 14(5):684–693, 1968.

[263] R.S. Orr. The order of computation for finite discrete Gabor transform. *IEEE Trans. Signal Proc.*, 41(2):122–130, 1993.

[264] F.J. Owens and M.S. Murphy. A short–time Fourier transform. *Signal Process.*, 14:3–10, 1988.

[265] C.H. Page. Instantaneous power spectra. *J. Applied Physics*, 23(1):103–106, 1952.

[266] A. Papoulis. *Systems and Transforms with Applications in Optics.* McGraw Hill, New York, 1968.

[267] A. Papoulis. *Signal Analysis.* McGraw Hill, New York, 1984.

[268] A. Papoulis. *Probability, Random Variables, and Stochastic Processes.* McGraw Hill, New York, 1991.

[269] J.D. Parsons. *The Mobile Radio Propagation Channel.* Pentech Press, London, 1992.

[270] E. Parzen. Probability density functionals and reproducing kernel Hilbert spaces. In M. Rosenblatt, editor, *Time Series Analyis*, pages 155–169. John Wiley & Sons, New York, 1963.

[271] J. Peetre. The Weyl transform and Laguerre polynomials. Le Mathematiche, 27:301–323, 1972.

[272] S.C. Pei and T.Y. Wang. The Wigner distribution of linear time–invariant systems. *IEEE T-SP*, 36(10):1681–1684, 1988.

[273] S.C. Pei and T.Y. Wang. Modulation–based time–variant filter design using Wigner distribution. *Sig. Proc.*, 40(2-3):335–340, 1994.

[274] A. Perelomov. *Generalized Coherent States and Their Applicatons.* Texts and Monographs in Physics. Springer, Berlin-Heidelberg, 1986.

[275] N. Peterfreund and Y.Y. Zeevi. Nonuniform image representation in area-of-interest systems. *IEEE T-IP*, 4(10):1202–1212, 1995.

[276] P.C.B. Phillips. The exact distribution of the Stein–rule estimator. *J. Econ*, 25:123–131, 1984.

[277] J.C.T. Pool. Mathematical aspects of the Weyl correspondence. *J. Math. Phys.*, 7(1):66–76, 1966.

[278] H.V. Poor. *An Introduction to Signal Detection and Estimation.* Springer–Verlag, 1988.

[279] B. Porat. *Digital Processing of Random Signals: Theory and Methods.* Englewood Cliffs (NJ), 1994.

[280] B. Porat and B. Friedlander. A frequency domain algorithm for multiframe detection and estimation of dim targets. *IEEE T-PAMI*, 12(4):398–401, 1990.

[281] M. Porat and Y.Y. Zeevi. The generalized Gabor scheme of image representation in biological and machine vision. *IEEE Trans. PAMI*, 10(4):452–468, 1988.

[282] M. Porat and Y.Y. Zeevi. Localized texture processing in vision: Analysis and synthesis in the Gaborian space. *IEEE Trans. on Biomedical Eng.*, BME-36/1:115–129, 1989.

[283] M.R. Portnoff. Time–frequency representation of digital signals and systems based on short-time Fourier analysis. *IEEE Trans. Sign. Proc.*, 28:55–69, 1980.

[284] J.F. Price and A. Sitaram. Local uncertainty inequalities for locally compact groups. *Trans. Amer. Math. Soc.*, 308(1):105–114, 1988.

[285] R. Price and E.M. Hofstetter. Bounds on the volume and height distributions of the ambiguity function. *IEEE T-IT*, 11:207–214, 1965.

[286] M. Priestley. Evolutionary spectra and non–stationary processes. *J.Roy.Statist.Soc.Ser. B*, 27:204–237, 1965.

[287] M. Priestley. *Spectral Analysis and Time Series*, volume I–II. Academic Press, London, 1981.

[288] P. Prinz. Calculating the dual Gabor window for general sampling sets. *IEEE Trans. Signal Proc.*, 44(8):2078–2082, 1996.

[289] J.G. Proakis. Adaptive equalization for TDMA digital mobile radio. *IEEE T-VT*, 40(2):333–341, 1991.

[290] S. Qian and D. Chen. Discrete Gabor transform. *IEEE Trans. Signal Proc.*, 41(7):2429–2438, 1994.

[291] S. Qian, Chen K., and S. Li. Optimal biorthogonal functions for finite discrete–time Gabor expansion. *Signal Process.*, 27(2):177–185, May 1992.

[292] J. Ramanathan and T. Steger. Incompleteness of sparse coherent states. *Appl. Comp. Harm. Anal.*, 2:148–153, 1995.

[293] J. Ramanathan and P. Topiwala. Time-frequency localization via the Weyl correspondence. *SIAM J. Matrix Anal. Appl.*, 24(5):1378–1393, 1993.

[294] S. Ramprashad, T.W. Parks, and R. Shenoy. Signal modeling and detection using cone classes. *IEEE T-SP*, 44(2):329–338, 1996.

[295] M.M. Rao. Harmonizable processes: Structure theory. *L. Enseign. Math.*, 28:295–351, 1982.

[296] K. Riedel. Optimal data based kernel estimation of evolutionary spectra. *IEEE T-SP*, 41(7):2439–2447, 1993.

[297] K. Riedel. Optimal kernel estimation of the instantaneous frequency. *IEEE T-SP*, 42(10):2644–2649, 1994.

[298] K. Riedel and A. Sidorenko. Minimum bias multiple taper spectral estimation. *IEEE T-SP*, 43(1):188–195, 1995.

[299] K. Riedel, D.J. Thomson, and A. Sidorenko. Spectral estimation of plasma fluctuations i: Comparison of methods. *Physics of Plasmas*, 1:485–500, March 1994.

[300] F. Riesz and B.Sz. Nagy. *Functional Analysis.* Dover, New York, 1990.

[301] A.W. Rihaczek. Doppler–tolerant signal waveforms. *IEEE Proc.*, 54(6):849–857, 1966.

[302] A.W. Rihaczek. Signal energy distribution in time and frequency. *IEEE T-IT*, 14:369–374, 1968.

[303] A.W. Rihaczek. *Principles of High–Resolution Radar.* McGraw Hill, New York, 1969.

[304] M.D. Riley. *Speech Time–Frequency Representations.* Kluwer Academic Publisher, 1989.

[305] R. Rochberg. Toeplitz and Hankel operators, wavelets, nwo sequences, and almost diagonalization of operators. *Lecture Notes in Pure and Appl. Math.*, 51:425–444, 1990.

[306] R. Rochberg. The use of decomposition theorems in the study of operators. In Benedetto J. and Frazier M., editors, *Wavelets: Mathematics and Applications*, pages 547–570. CRC Press, Boca Raton, 1994.

[307] A. Ron and Z. Shen. Frames and stable bases for subspaces of $\mathrm{l}_2(\mathrm{r}^d)$: The duality principle of Weyl-Heisenberg sets. In *Proc. Lanczos Int.Cent.Conf.*, pages 422–425, Raleigh (NC), 1993.

[308] A. Ron and Z. Shen. Frames and stable bases for shift–invariant subspaces $\mathrm{l}_2(\mathrm{r}^d)$. *Canadian Journal of Mathematics*, 47(5):1051–1094, 1995.

[309] A. Ron and Z. Shen. Weyl-Heisenberg frames and Riesz bases in $l_2(r^d)$. Technical Report 95-03, University of Wisconsin, Madison (WI), 1995.

[310] W. Rudin. *Fourier Analysis on Groups.* Wiley, New York, 1967.

[311] W. Rudin. *Functional Analysis.* McGraw Hill, New York, 1973.

[312] L. Rutkowski. On–line identification of time–varying systems by nonparametric techniques. *IEEE T-AC*, 27(1), 1982.

[313] L. Rutkowski. On nonparametric identification with prediction of time–varying systems. *IEEE T-AC*, 29(1), 1984.

[314] J.A. Saghri, A.G. Tescher, and J.T. Reagan. Practical transform coding of multispectral imagery. *IEEE SP Magazine*, pages 32–43, January 1995.

[315] A.D. Sams and V.Z. Marmarelis. Identification of linear periodically time–varying systems using white–noise test inputs. *Automatica*, 25(4):563–567, 1995.

[316] A.H. Sayed and T. Kailath. A state–space approach to adaptive RLS filtering. *IEEE SP Magazine*, pages 18–60, July 1994.

[317] A.M. Sayeed and D.L. Jones. Optimal kernels for Wigner–Ville spectral estimation. In *Proc. ICASSP–94*, pages –, Sydney (Australia), April 1994.

[318] A.M. Sayeed and D.L. Jones. Optimal quadratic detection and estimation using generalized joint signal representations. *submitted to IEEE Signal Process.*, pages –, April 1994.

[319] A.M. Sayeed and D.L. Jones. Optimal detection using bilinear time–frequency and time–scale representations. *IEEE T-SP*, 43(12):2872–2883, 1996.

[320] W. Schempp. Radar ambiguity functions, nilpotent harmonic analysis, and holomorphic theta series. In R.A. Askey and et al, editors, *Special Functions: Group Theoretical Aspects and Applications*, pages 217–260. D.Reidel Publishing Company, 1984.

[321] W. Schempp. Radar ambiguity functions, nilpotent harmonic analysis, and holomorphic theta series. In R.A. Askey et al., editors, *Special Functions: Group Theoretical Aspects and Applications*, pages 217–260. D.Reidel Publishing Company, 1984.

[322] I.E. Segal. Transforms for operators and symplectic automorphisms over a locally compact abelian group. *Math. Scand.*, 13:31–43, 1963.

[323] K. Seip. Density theorems for sampling and interpolation in the Bargmann-Fock space i. *Journ. reine ang. Math.*, 429:91–106, 1992.

[324] K. Seip and R. Wallsten. Density theorems for sampling and interpolation in the Bargmann-Fock space II. *J. reine angewandte Mathematik*, 429:107–113, 1992.

[325] J.S. Shamma and M. Athans. Guaranteed properties of gain scheduled control for linear parameter–varying plants. *Automatica*, 27(3):559–564, 1991.

[326] L.L. Sharf. *Statistical Signal Processing.* Addison–Wesley, Reading (Massachusetts), 1991.

[327] R.G. Shenoy and T.W. Parks. The Weyl correspondence and time–frequency analysis. *IEEE T-SP*, 42(2):318–331, 1994.

[328] S. Shokoohi, L.M. Silverman, and P. Van Dooren. Linear time–variable systems: Stability of reduced models. *Automatica*, 20(1):59–67, 1984.

[329] R.A. Silverman. Locally stationary random processes. *IRE Trans. Info. Theory*, 3:182–187, 1957.

[330] D. Slepian. Prolate spheroidal wave functions, Fourier analysis and uncertainty—v: The discrete case. *Bell System Technical Journal*, 57(5):1371–1430, 1978.

[331] D. Slepian and H.O. Pollak. Prolate spheroidal wave functions, Fourier analysis and uncertainty—i. *Bell System Technical Journal*, 40(1):43–64, 1961.

[332] S.S. Soliman. Tracking loop for fading dispersive channels. *IEEE T-C*, 38(3):292–299, 1990.

[333] S.S. Soliman and R.A. Scholtz. Spread ambiguity functions. *IEEE T-IT*, 34(2):343–347, 1988.

[334] K.A. Sostrand. Mathematics of the time–varying channel. *Proceedings NATO Advanced Study Institute on Signal Procesing with Emphasis on Underwater Acoustics*, 2:25–1—25–20, 1968.

[335] D. Stanhill and Y.Y. Zeevi. Finding the optimal wavelet function for image and signal representation. In *SPIE Symp. VCIP'93*, pages 1691–1700, Boston, 1993.

[336] R. Steele. *Mobile Radio Propagation.* Pentech Press, London, 1992.

[337] R. Strichartz. *A Guide to Distribution Theory and Fourier Transforms.* CRC Press, Boca Raton, 1993.

[338] C.A. Stutt. Some results on real–part/imaginary–part and magnitude–phase relations in ambiguity functions. *IEEE T-IT*, pages 321–327, Oct. 1964.

[339] S.M. Sussman. Least–squares synthesis of radar ambiguity functions. *IRE Trans. Information Theory*, pages 246–254, 1962.

[340] K. Tachizawa. The boundedness of pseudodifferential operators on modulation spaces. *Math. Nachr.*, 168:263–277, 1994.

[341] K. Tachizawa. Some estimates for eigenvalues of Schrödinger operators. *Proc. Japan Academy*, 70(4):85–87, 1994.

[342] G. Tadmore. Uncertain feedback loops and robustness in general linear systems. *Automatica*, 27(6):1039–1042, 1991.

[343] M. Taylor. *Noncommutative Harmonic Analysis.* Number 22 in Math. Surveys and Monographs. AMS, Providence, 1986.

[344] C. Thiele and L.F. Villemoes. A fast algorithm for adapted time–frequency tilings. *J. Appl. Comp. Harm. Anal.*, 3(2):91–99, 1996.

[345] D.J. Thomson. Spectrum estimation and harmonic analysis. *IEEE Proc.*, 70(9):1055–1096, 1982.

[346] D. Tjostheim. Spectral generating operators for nonstationary processes. *Adv. Appl. Prob.* , 8:831–846, 1976.

[347] R. Tolimieri. Characterization of Weyl-Heisenberg frames via Poisson summation relationships. In *Proc. IEEE ICASSP-92*, volume 4, pages 277–280, 1992.

[348] R. Tolimieri. Problems in Gabor representation. In *Wavelets and their applications. Proceedings of the NATO ASI Conference, 16-29 August 1992*, NATO ASI Ser., Ser. C, pages 95–120, Il Ciocco, Italy, 1994. Kluwer Academic Publishers.

[349] R. Tolimieri and R. Orr. Poisson summation, the ambiguity function and the theory of Weyl-Heisenberg frames. *J. Four. Anal. Appl.*, 1(3):233–247, 1995.

[350] R. Tolimieri and S. Winograd. Computing the ambiguity surface. *IEEE T-ASSP*, 33(4):1239–1245, 1985.

[351] Y.C. Trivedi and L. Kurz. Image restoration using recursive estimators. *IEEE T-SMC*, 25(11):1470–1482, 1995.

[352] M.K. Tsatsanis and G.B. Giannakis. Time–varying system identification and model validation using wavelets. *IEEE T-SP*, 41(12):3512–3523, 1993.

[353] M. Unser. On the approximation of the discrete Karhunen–Loeve transform for stationary processes. *Sig. Proc.*, 7(3):231–249, 1984.

[354] M. Unser. An extension of the Karhunen–Loeve transform for wavelets and perfect reconstruction filter banks. In *Proc. SPIE Mathematical Imaging: Wavelet Applications in Signal and Image Processing*, volume 2034, pages 45–56, July 1993.

[355] H.L. Van Trees. *Detection, Estimation and Modulation Theory.* Wiley, New York, 1971.

[356] V. Venkatasubramanian, H. Schättler, and J. Zaborsky. Fast time–varying phasor analysis in the balanced three–phase large electric power system. *IEEE T-AC*, 40(11):1975–1982, 1995.

[357] M. Vetterli and J. Kovacevic. *Wavelets and Subband Coding.* Signal Processing Series. Prentice Hall, Englewood Cliffs, NJ, 1995.

[358] J. Ville. Theorie et applications de la notion de signal analytique. *Cables et Transmission*, 2A:61–74, 1948. translated into English by I.Selin, RAND Corp. Report T-92, Santa Monica, CA, Aug. 1958.

[359] A. Voros. An algebra of pseudodifferential operators and the asymptotics of quantum mechanics. *J. Funct. Anal.*, pages 104–132, 1978.

[360] D. Walnut. Lattice size estimates for Gabor decompositions. *Monatshefte für Mathematik*, 115(3):245–256, 1993.

[361] D.F Walnut. Application of Gabor and wavelet expansions to the Radon transform. In J.S. Byrens, editor, *Probablistic and Stochastic Methods in Analysis, with Applicatios*, pages 187–205. Kluwer Acad. Publ., 1992.

[362] D.F. Walnut. Continuity properties of the Gabor frame operator. *J. Math. Anal. Appl.*, 165(2):479–504, 1992.

[363] S. Watanabe. Karhunen–Loeve expansion and factor analysis, theoretical remarks and applications. In *Trans.Fourth Prague Conf.Inform.Theory*, pages 635–660, Prague, 1965. Statist.Decision Functions and Random Processes.

[364] U. Wellens and A. Baier. Empfaengertechniken bei digitaler Funkuebertragung ueber Mobilfunkkanaele unterschiedlicher Bandbreite. *e&i*, 106(4):150–157, 1989.

[365] M. Werner. *Modellierung und Bewertung von Mobilfunkkanälen.* PhD thesis, Erlangen–Nuernberg University of Technology, Germany, (in German), 1991.

[366] J. Wexler and S. Raz. Discrete Gabor expansions. *Signal Processing*, 21(3):207–221, November 1990.

[367] H. Weyl. Das asymptotische Verteilungsgesetz der Eigenwerte linearer partieller Differentialgleichungen (mit einer Anwendung auf die Theorie der Hohlraumstrahlung. *Mathematische Annalen*, 71:441–479, 1911.

[368] H. Weyl. Quantenmechanik und Gruppentheorie. *Zeitschrift f.Physik*, 46:1–46, 1927.

[369] H. Weyl. *The Theory of Groups and Quantum Mechanics.* Methuen (London) 1931, reprinted by Dover Puplications, New York, 1950.

[370] E.P. Wigner. On the quantum correction for thermo–dynamic equilibrium. *Phys. Rev. Lett.*, 40:749–759, 1932.

[371] C.H. Wilcox. The synthesis problem for radar ambiguity functions. Technical Report 157, Mathematics Research Center, United States Army, University Wisconsin, Madison (Wisconsin), 1960.

[372] R.M. Wilcox. Exponential operators and parameter differentiation in quantum physics. *J. Math. Phys.*, 8(4):962–982, 1967.

[373] N.P. Willis and Y. Bresler. Optimal scan for time–varying tomography i: Theoretical analysis and fundamental limitations. *IEEE T-IP*, 4(5):642–653, 1995.

[374] K.G. Wilson and W. Zimmermann. Operator product expansions and composite field operators in the general framework of quantum field theory. *Comm. Math. Phys.*, 24:87–106, 1972.

[375] P.W. Wong. Wavelet decomposition of harmonizable random processes. *IEEE T-IT*, 39(1):7–18, 1993.

[376] S. Wood, M.G. Larimore, and J.R. Treichler. The impact of an adaptive equalizer's update behavior symbol error rate in a nonstationary environment. In *Proc. IEEE ICASSP'88*, pages 1608–1611, 1988.

[377] P.M. Woodward. *Probability and Information Theory with Application to Radar.* Pergamon Press, London, 1953.

[378] G. Wornell. A Karhunen–Loeve like expansion for 1/f processes via wavelets. *IEEE T-IT*, 36(4):1534–1543, 1990.

[379] G. Wornell. Spread–signature CDMA: Efficient multiuser communication in the presence of fading. *IEEE T-IT*, 41(5):1418–1438, 1995.

[380] H. Yamaguchi. Adaptive DCT coding of video signals. *IEEE T-C*, 41(10):1534–1543, 1993.

[381] K. Ylinen. Fourier transforms of noncommutative analogues of vector measures and bimeasures with applications to stochastic processes. *Ann. Acad. Sci. Fenn. Ser. A. I.*, 1:249–275, 1975.

[382] D.C. Youla. The synthesis of linear dynamical systems from prescribed weighting patterns. *J. SIAM Appl. Math.*, 14(3):526–549, 1966.

[383] L. Zadeh. Correlation functions and power spectra in variable networks. *Proc. of IRE*, 38:1342–1345, November 1950.

[384] L. Zadeh. The determination of the impulsive response of variable networks. *J. Applied Physics*, 21:642–645, July 1950.

[385] L. Zadeh. Frequency analysis of variable networks. *Proc. of IRE*, 38:291–299, March 1950.

[386] L. Zadeh. Time–varying networks — i. *Proc. of IRE*, pages 1488–1503, 1961.

[387] L.A. Zadeh and C.A. Desoer. *Linear System Theory.* McGraw Hill, New York, 1963.

[388] J. Zak. Finite translations in solid state physics. *Phys. Rev. Lett.*, 19:1385–1397, 1967.

[389] Y.Y. Zeevi and E. Shlomot. Nonuniform sampling and antialiasing in image representation. *IEEE T-SP*, 3:1223–1236, 1993.

[390] A.H. Zemanian. Realizability conditions for time–varying and time–invariant Hilbert ports. *SIAM J. Appl. Math.*, 22(4):612–628, 1972.

[391] M. Zibulski and Y.Y. Zeevi. Frame analysis of the discrete Gabor scheme. *IEEE Trans. SP*, 42(8):942–945, 1993.

[392] M. Zibulski and Y.Y. Zeevi. Oversampling in the Gabor scheme. *IEEE Trans. SP*, 41(8):2679–2687, 1993.

[393] M. Zibulski and Y.Y. Zeevi. Analysis of multi-window Gabor-type schemes by frame methods. Technical report, Technical Report CC PUB 101, Technion–Israel Institute of Technology, Center for Communications, Haifa, Israel, 1995.

[394] L.J. Ziomek. A scattering function approach to underwater acoustic detection and signal design. Technical Report TM 81-144, Pennsylvania State University, 1968.

[395] L.J. Ziomek. *Underwater Acoustics: A Linear System Theory Approach.* Academic Press, Orlando (FL), 1985.

[396] L.J. Ziomek. *Fundamentals of Acoustic Field Theory.* CRC Press, Boca Raton (FL), 1995.

Notation

List of Abbreviations

EAF ..... Expected Ambiguity Function.

GSF ..... Generalized Spreading Function.

GWD ..... Generalized Wigner Distribution.

GWS ..... Generalized Weyl Symbol.

GWVS ..... Generalized Wigner–Ville Spectrum.

HS ..... Hilbert–Schmidt.

KL ..... Karhunen–Loeve.

LFI ..... Linear Frequency–Invariant.

LTI ..... Linear Time–Invariant (Linear Translation–Invariant).

LTV ..... Linear Time–Varying.

MMSE ..... Minimum Mean–Squared Error.

MVUB ..... Minimum Variance Unbiased.

STFT ..... Short–Time Fourier Transform.

TFDMA ..... Time–Frequency Division Multiple Access.

WH ..... Weyl–Heisenberg.

WVS ..... Wigner–Ville Spectrum.

WSSUS ..... Wide–Sense Stationary Uncorrelated Scattering.

List of Important Symbols

$(\ )^*$ ..... Complex conjugation, adjoint operator.

$(\ )^{(\tau,\nu)}$ ..... Time–frequency superscript with different meaning for
(i) a signal as $x^{(\tau,\nu)}(t) = x(t-\tau)e^{j2\pi\nu t}$,
(ii) the time–frequency shift operator $\mathbf{S}^{(\tau,\nu)}$ by $\mathbf{S}^{(\tau,\nu)}x = x^{(\tau,\nu)}$, and
(iii) for a general operator in the sense of shifting its symbol: $\mathbf{H}^{(\tau,\nu)} \stackrel{\text{def}}{=} \mathbf{S}^{(\tau,\nu)}\mathbf{H}\mathbf{S}^{(\tau,\nu)*}$.

α ..... Parameter of the generalized Weyl correspondence and the associated definitions, always $|\alpha| < 1/2$.

$A_x^{(\alpha)}(\tau,\nu), A_{x,y}^{(\alpha)}(\tau,\nu)$ ..... Generalized auto and cross ambiguity function, respectively, see (F.19).

$B_H(f,s)$ ..... Bifrequency function of LTV system (see(4.4)).

$\delta(t)$ ..... Delta distribution.

δ_{kl} ..... Kronecker's delta function.

$E\{\ \}$ ..... Expectation operator.

$EA_x^{(\alpha)}(t,f), EA_{x,y}^{(\alpha)}(\tau,\nu)$ ..... Expected generalized auto and cross ambiguity function, respectively, see (2.60)

$ES_x(t,f)$ ..... Evolutionary spectrum of a nonstationary process, see (2.36).

$ESPEC_x^{(\gamma)}(t,f)$ ..... Physical spectrum of a nonstationary process, see (2.16).

$EW_x^{(\alpha)}(t,f), EW_{x,y}^{(\alpha)}(t,f)$ ..... Expected generalized auto and cross Wigner distribution, respectively, see (2.62).

f, ν ..... Frequency.

F ..... Frequency Period.

F_x ..... Square root of spectral moment of a signal $x(t)$ (see p. 43).

$\mathcal{F}$ ..... Fourier transform defined as $(\mathcal{F}_{t\to f}x)(f) = \int_t x(t)e^{-j2\pi ft}dt$.

$\gamma(t)$ ..... STFT/Gabor analysis window.

$g(t)$ ..... STFT/Gabor synthesis window.

$G_x^{(\gamma)}(m,k)$ ..... Gabor coefficient, with signal $x(t)$ based on a normalized analysis window $\gamma(t)$, see (3.29).

$h(t,s)$ ..... Impulse response of a linear system, kernel of a linear integral operator, see (4.3).

$h_2(t,\tau)$ ..... Alternative form of the impulse response of a linear system, see (4.2).

$\mathbf{H}$ ..... Linear operator, usually Hilbert–Schmidt unless otherwise specified.

$\mathbf{I}$ ..... Identity operator.

$M(t,f)$ ..... Multiplier function for STFT–based linear time–varying systems.

$\mathbf{M}_\nu$ ..... Modulation operator acting as $(\mathbf{M}_\nu x)(t) = x(t)e^{j2\pi\nu t}$.

$\mathcal{O}(\sigma)$ ..... Bachmann–Landau O–notation (used only for asymptotical results valid for $\sigma \to 0$, where $q(\sigma) = \mathcal{O}(\sigma)$ means that $q(\sigma) \le K\sigma$ for sufficiently small σ where K is a σ–independent constant).

$\mathbf{P}$ ..... Orthogonal projection operator, see (A.13) or, more generally, a prototype operator.

$\mathbf{P}_\gamma$ ..... Rank–one projection onto the function $\gamma \in L_2(\mathbb{R})$.

$\mathbf{P}_{\gamma,g}$ ..... Rank–one operator with the singular functions $\gamma, g \in L_2(\mathbb{R})$ (the kernel is given by $(\mathbf{P})(t,t') = \gamma(t)g^*(t')$).

$L_H^{(\alpha)}(t,f)$ ..... Generalized Weyl symbol of a linear operator; without superscript: Weyl symbol, see (C.1)

$\chi_{[-T,T]}(t)$ ..... Indicator function (0/1–valued) of the interval $[-T,T] \subset \mathbb{R}$

$R_x(t,f), R_{x,y}(t,f)$ ..... Auto and cross Rihaczek distribution, respectively, see (F.10).

$\mathbf{R}_x$ ..... Correlation operator, its kernel is the autocorrelation function of a nonstationary process, see the footnote on page 6.

σ_H ..... Total spread of underspread operator, see (4.14)

σ_x ..... (i) Total spread of underspread process (2.66), (ii) In Appendix E: σ_x^2 denotes the power spectral density of the input process.

$\mathbf{S}^{(\tau,\nu)}$ ..... Time–frequency shift operator, see Section 4.2.1.

$S_{x,y}(f)$ ..... Cross power density spectrum of two stationary processes.

$S_H^{(\alpha)}(\tau,\nu)$ ..... Generalized spreading function of a linear operator, see (B.10).

$SPEC_x^{(\gamma)}(t,f)$ ..... Spectrogram of $x(t)$ with window $\gamma(t)$ (the magnitude–squared STFT).

$STFT_x^{(\gamma)}(t,f)$ ..... Short–time Fourier transform of $x(t)$ with window $\gamma(t)$, see (F.1).

t,τ ..... Time.

T ..... Time Period.

$T_H^{(\gamma)}(t,f)$ ..... Short–time transfer function (lower WH symbol, see (4.38)).

$\mathbf{T}_\tau$ ..... Translation operator acting as $(\mathbf{T}x)(t) = x(t-\tau)$

T_x ..... Square root of temporal moment of a signal $x(t)$ (see p. 43).

tr$\mathbf{H}$ ..... Trace of the operator $\mathbf{H}$, see (A.27).

$W_x^{(\alpha)}(t,f), W_{x,y}^{(\alpha)}(t,f)$ ..... Generalized auto and cross Wigner distribution, respectively; without superscript: Wigner distribution in particular, see (F.11).

$x(t)$ ..... Signal, usually a function $\in L_2(\mathbb{R})$.

$X(f)$ ..... Fourier transform (spectrum) of signal $x(t)$.

$Z_H(t,f)$ ..... Zadeh's time–varying transfer function or, equivalently, Kohn–Nirenberg symbol of a linear operator, $Z_H(t,f) = L_H^{(1/2)}(t,f)$, see (4.22).

- All sums and integrals go from $-\infty$ to ∞ unless otherwise specified.
- The inner product is defined as usual

$$\langle x,y\rangle \stackrel{\text{def}}{=} \int_t x(t)y^*(t)dt \qquad \text{and} \qquad \langle h,g\rangle \stackrel{\text{def}}{=} \int_t\int_s h(t,s)g^*(t,s)dt\,ds$$

- The signal norm is defined as

$$\|x\|^2 \stackrel{\text{def}}{=} \langle x,x\rangle.$$

- The (Hilbert–Schmidt) operator inner product is defined as usual via the kernels

$$\langle \mathbf{H},\mathbf{G}\rangle = \int_t\int_s (\mathbf{H})(t,s)(\mathbf{G})^*(t,s)dtds.$$

- The operator norm is (unusually) defined as the Hilbert–Schmidt norm

$$\|\mathbf{H}\|^2 \stackrel{\text{def}}{=} \langle \mathbf{H},\mathbf{H}\rangle.$$

The standard operator norm (which induces the uniform operator topology) is denoted by

$$\|\mathbf{H}\|_\infty \stackrel{\text{def}}{=} \sup\left\{\frac{\|\mathbf{H}x\|}{\|x\|} : x \neq 0\right\}.$$

Index

Curriculum Vitae

Der Verfasser dieser Arbeit wurde am 26.7.1964 in Steyr, Oberösterreich als Sohn von Ing. Siegmund Kozek, Prokurist and Helga Kozek, Lehrerin geboren. Er besuchte die Volkschule in Wien, danach das Bundesrealgymnasium in Salzburg, Akademiestrasse, wo er 1982 mit ausgezeichnetem Erfolg maturierte. Nach dem Studium der Elektrotechnik, Studienzweig Nachrichtentechnik an der Technischen Universität Wien war er von 1990 bis 1993 als Vertragsassistent am Institut für Nachrichtentechnik und Hochfrequenztechnik an ebendieser Universität im Rahmen des FWF Projektes "Zeit–Frequenz Methoden zur Signalverarbeitung" tätig. Seit 1994 ist er wissenschaftlicher Mitarbeiter am Institut für Mathematik der Universität Wien im Rahmen des FWF–Forschungsschwerpunktes "Bildverarbeitung", Teilprojekt "Mathematische Methoden zur Bildverarbeitung". 1995 erfolgte die Ableistung des ordentlichen Zivildienstes im sozialmedizinischen Zentrum Ost der Stadt Wien.

Der Verfasser ist seit 1983 mit Marion Raab verheiratet und ist Vater dreier Kinder im Alter von 7, 10 und 13 Jahren.