

Important Notice

This copy may be used only for the purposes of research and private study, and any use of the copy for a purpose other than research or private study may require the authorization of the copyright owner of the work in question. Responsibility regarding questions of copyright that may arise in the use of this copy is assumed by the recipient.

UNIVERSITY OF CALGARY

Theory of adaptive, nonstationary filtering in the Gabor domain

with applications to seismic inversion

by

Jeff P. Grossman

A DISSERTATION

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR

THE

DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF GEOLOGY AND GEOPHYSICS

AND

DEPARTMENT OF MATHEMATICS AND STATISTICS

CALGARY, ALBERTA

JUNE, 2005

© Jeff P. Grossman 2005

UNIVERSITY OF CALGARY

FACULTY OF GRADUATE STUDIES

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies for acceptance, a dissertation entitled "Theory of adaptive, non-stationary filtering in the Gabor domain with applications to seismic inversion" submitted by Jeff P. Grossman in partial fulfillment of the requirements for the degree of DOCTOR OF PHILOSOPHY.

Supervisor, Dr. G. F. Margrave
Department of Geology and Geophysics

Co-supervisor, Dr. M. P. Lamoureux
Department of Mathematics and Statistics

Dr. L. Lines
Department of Geology and Geophysics

Internal/External Examiner, Dr. T. Ware
Department of Mathematics and Statistics

External Examiner, Dr. M. Yedlin
Department of Electrical and Computer
Engineering
University of British Columbia

Date

Dedication

For my wonderfully supportive wife, Elena, and for our son, Evan, who has brought us great joy while teaching us how to function with a lot less sleep.

Abstract

Many physical processes can be modelled by a linear partial differential operator. Classically, this corresponds to a convolution operation, or equivalently, the application of a Fourier filter. The scope of such operators is limited to the stationary setting, but physical processes are never stationary. Part of our objective is to consider two ways to extend the Fourier theory: (1) via Gabor analysis, and (2) via pseudodifferential operator theory. The main emphasis is on the former, which we deem general enough to handle most, if not all, linear physical processes.

Two important seismic inverse problems, each modelled by a pseudodifferential operator, can be solved successfully via a Gabor filtering approach: (1) seismic deconvolution for a dissipative earth; and (2) wavefield extrapolation in an inhomogeneous earth.

The first solution relies on the existence of a well-defined concept of nonstationary minimum phase in the analog setting; the second can be overly computationally intensive, owing to excessive redundancies in the Gabor frame. Both of these issues can be addressed, respectively, by: (1) extending the idea of minimum phase to the class of tempered distributions; and (2) developing an adaptive, nonuniform sampling scheme to minimize the redundancy of the Gabor transform, while simultaneously respecting the inherent nonstationarity of the problem.

Acknowledgements

I would first like to thank my advisors, Gary Margrave and Michael Lamoureaux, for their outstanding guidance, collaboration, motivation, and support throughout my Ph.D. program. It has been a great pleasure learning from both of you—thanks for sharing all that knowledge and wisdom.

My lovely wife Elena had the determination, faith, and patience to put up with my many ups and downs. Bonnie Nasim volunteered her expert technical-writing skills, and managed to make time during her precious Christmas holidays to edit this work in painful detail. Thanks for a great job Bonnie.

I thank the following organizations for their generous financial support: the CREWES Project and its sponsors, the POTSI Project and its sponsors, the Pacific Institute for the Mathematical Sciences (PIMS), Mathematics of Information Technology and Complex Systems (MITACS), Natural Sciences and Engineering Research Council of Canada (NSERC), and the University of Calgary.

I had the pleasure of attending two very rewarding workshops at the inspirational Banff International Research Station (BIRS)—I highly recommend a visit to BIRS, and hope that it stays afloat for many years to come.

Table of Contents

Approval Page	iii
Dedication	v
Abstract	vii
Acknowledgements	ix
Table of Contents	xi
1 Introduction	1
2 Linear systems	9
2.1 Spaces for representing signals and images	10
2.1.1 The Hilbert space of finite-energy signals	10
2.1.2 The Lebesgue spaces	12
2.1.3 The Schwartz space of rapidly decreasing signals	13
2.2 The space of tempered distributions	14
2.2.1 Distributional derivatives	17
2.2.2 The Fourier transform of a distribution	18
2.3 Linear operators in Hilbert space	19
3 Minimum phase for nonstationary analog systems	22
3.1 Causality and the Hilbert transform	24
3.1.1 Extension of the Hilbert transform to distributions	26
3.2 Minimum phase	31
3.2.1 Examples	35
3.2.2 Some characteristics of minimum phase	37
3.3 Nonstationary minimum phase filters	41
4 Adaptive time-frequency analysis with nonuniform Gabor frames	44
4.1 Preliminaries	46
4.1.1 Partitions of unity	46
4.1.2 Partitions of unity in higher dimensions	47
4.1.3 Translation and modulation operators in Hilbert space	48
4.2 The Gabor transform	49
4.2.1 The short time Fourier transform	49

4.2.2	Gabor frames and the discrete Gabor transform	50
4.2.3	Implementation of Gabor transforms with FFTs	54
4.2.4	Window selection	54
4.3	Nonuniform frames from maximal partition-of-unity frames	55
4.3.1	Maximally redundant partition-of-unity Gabor frames	55
4.3.2	One-parameter families of Gabor frames	57
4.4	Adaptive, nonuniform Gabor frames	62
5	Theory of Gabor filtering	65
5.1	Gabor filters	66
5.1.1	The analog theory	66
5.1.2	A functional calculus for Gabor multipliers	71
5.1.3	On the discrete theory	74
5.2	Applications to Mathematical Physics	75
5.2.1	General linear partial differential operators	75
5.2.2	Pseudodifferential operators and Gabor filters	77
6	Deconvolution and Q-inversion in dissipative media	80
6.1	Introduction	82
6.2	Theory	83
6.3	Method	84
6.4	A synthetic example	87
6.5	Summary	95
7	Wavefield extrapolation in inhomogeneous media	98
7.1	Introduction	100
7.2	Overview of the AGPS method	101
7.3	Synthetic examples	106
7.4	Summary	113
7.5	Future work	114
8	Summary	116
	Bibliography	121

List of Tables

7.1	Comparison of computation times for various extrapolation algorithms. In contrast to the first three methods, the cost of AGPS is proportional to the complexity of the velocity model. The cost of AGPS spans an order of magnitude between the two extremes, namely, the velocity models depicted in figure 7.2.	107
-----	--	-----

List of Figures

- 4.1 Example of a Gabor amplitude spectrum, top-right. The input time series on the left is a constant-Q-attenuated synthetic seismic trace. Its Fourier amplitude spectrum is depicted below the Gabor amplitude spectrum. The time series and its amplitude spectrum can be thought of as projections of the Gabor spectrum onto the time and frequency axes, respectively. 53
- 4.2 A Gaussian atom of halfwidth σ , and the family of Gaussians, of halfwidth σ/\sqrt{p} , obtained by raising the initial Gaussian to the power $p \in [0, 1]$. As p decreases, the atoms gain mass, and their mass distribution spreads. In the limit as $p \rightarrow 0$ the family converges to constant function, 1. 58
- 4.3 A Lamoureux atom and the family of windows obtained by raising the initial atom to the power $p \in [0, 1]$. As p decreases, the atoms gain mass and their spreading mass-distributions are confined to the original domain. At $p = 1/4$, the window is no longer differentiable at the two boundaries. However, this can be resolved for any given p by starting with a sufficiently high-order atom. In the limit as $p \rightarrow 0$, the family converges to a boxcar. 61
- 4.4 A demonstration of how an adaptive, nonuniform partition of unity Gabor frame (lower, solid) can be obtained from a maximal partition of unity, (the result of translating any of the smallest windows to each offset sample-point) which reflects the local nonstationarity of a velocity model. The piecewise-defined velocity model (upper, solid) represents all types of velocity variation: random, constant, a jump discontinuity between two constants, and a transition from smooth to random. The smoothed velocity, (dashed) obtained by convolving the velocity model with a fundamental atom, (any one of the smallest windows) serves as a visual aid in comparing the molecules to the local velocity variations. Specifically, where there is more variation in the velocity function, there are correspondingly more molecules, and conversely. 63

5.1	Implementation of Gabor filters which are too sharp can lead to disastrous results. Two sharp Gabor filters are displayed at (a) and (d). The middle column shows their corresponding inverse Gabor transforms, (synthesis window of unity) and the right column shows the magnitude of the forward Gabor transforms (Gaussian analysis windows) of the middle column, that is, the magnitude of the orthogonal projections of the original filters onto the range of the Gabor transform. The phase spectrum (shown in figure 5.3) for (c) is rather exotic, while that for (f) is zero. The signal (b) is a sinc function (with some Fourier wraparound effect), and (e) is a spike at time zero. Each frequency slice of the output (f) is just the Gaussian analysis atom, centered at the origin (see figure 5.2).	70
5.2	A frequency slice of the plot in figure 5.1(f).	71
5.3	The phase spectrum corresponding to the amplitude spectrum of figure 5.1(c).	72
6.1	A pseudo-random reflectivity (lower) and a nonstationary synthetic (upper) generated with an attenuation factor of 25.	88
6.2	Magnitude of the Gabor transform of the pseudo-random reflectivity of figure 6.1.	89
6.3	Magnitude of the Gabor transform of the synthetic trace of figure 6.1.	90
6.4	Time-variant filter for weighting the Gabor spectrum of the synthetic trace according to the available dynamic range, and hence the available numerical precision of the data.	91
6.5	Amplitude spectra of original and output signatures. The output was smoothed to remove small residual oscillations.	92
6.6	The magnitude of the smoothed spectrum, which models the data (from figure 6.3) in a least-squares sense. The pointwise inverse (plus a small constant) defines the magnitude of the deconvolution operator.	93
6.7	Fourier amplitude spectra of original reflectivity and Gabor deconvolution result.	94
6.8	Comparison of the bandpass-filtered Gabor deconvolution result (lower) with a bandpass-filtered (DC to half Nyquist) version of the reflectivity of figure 6.1 (upper).	95
6.9	The time-variant, bandlimited amplitude spectrum of the reflectivity estimate. Compare with the Gabor amplitude spectrum of the true reflectivity in figure 6.2.	96

7.1	An adaptive partition-of-unity frame and its correspondence to the local stationarity of a velocity model. The velocity function has been rescaled for display purposes. The smoothed velocity is displayed as a visual aid in comparing the molecules to the local velocity variations. The larger the local variation in velocity, the smaller and more abundant the molecules become.	103
7.2	Adaptive partition-of-unity frames and their correspondence to the local stationarity of the given velocity models. The velocity functions have been rescaled for display purposes. The smoothed velocities are displayed as a visual aid in comparing the molecules to the local velocity variations. The velocity model on the top jumps from 2250 m/s to 3750 m/s. The velocity model on the bottom fluctuates randomly between 1500 m/s and 2500 m/s.	105
7.3	Input wavefield (top) for two upward extrapolations by 200m, using the exact extrapolator; the two bottom plots show the resulting wavefields, using the step and random velocity models of figure 7.2, respectively.	109
7.4	Inversion of the corresponding two wavefields at the bottom of figure 7.3, using the exact extrapolator: (top) using the step velocity model of figure 7.2, and (bottom) using the random velocity model of figure 7.2.	110
7.5	Results of inverse extrapolating the wavefield at the bottom of figure 7.3, using the step velocity model of figure 7.2. The corresponding computation times are listed in Table 7.1. The rows contain pairs of results for PSPI, NSPS, and AGPS, respectively; the first column is for a single depth step of 200m, and the second is for five steps of 40m. These results should be compared with the exact results of figure 7.4.	111
7.6	Results of inverse extrapolating the wavefield at the right of figure 7.3, using the random velocity model of figure 7.2. The corresponding computation times are listed in Table 7.1. The rows contain pairs of results for PSPI, NSPS, and AGPS, respectively; the first column is for a single depth step of 200m, and the second is for five steps of 40m. These results should be compared with the exact results of figure 7.4.	112

Chapter 1

Introduction

Once you study the wave equation for the first time, you will continue to study it for the rest of your life. Well, unless you become an accountant, or a lion tamer, or something like that.

—*D. Hobill, while substitute-teaching physics class*

Our experience with sound—as in music—calls for a unified mathematical treatment of time and frequency analysis. Indeed, the human ear, combined with the brain’s processing skills, interprets acoustic amplitude information as temporally localized packets of bandlimited spectral information. Conversely, in following a musical score, a musician transforms a time-frequency representation of a signal into temporally varying acoustic amplitude data.

Classical Fourier theory fails to explain what we intuitively know: that “frequency content” changes with time. “The reason is that the Fourier-integral method considers phenomena in an infinite interval, and this is very far from

our everyday point of view” [25].

These fundamental observations imply that we ought to be modelling the localized time and frequency characteristics of a signal simultaneously; and indeed, Gabor’s milestone theory sets the stage for nonstationary analysis of a signal, “in which time and frequency play symmetrical parts, and which contains ‘time analysis’ and ‘frequency analysis’ as special cases” [25].

The subject of time-frequency analysis¹ is largely based on the short-time Fourier transform (STFT), namely

$$\text{STFT}(f)(x, \xi) = \int_{\mathbf{R}^d} f(y) \overline{g(y-x)} e^{-2\pi i \xi \cdot y} dy, \quad (1.1)$$

since “most other time-frequency representations can be expressed in terms of the STFT” [36] (the overline in equation 1.1 denotes complex conjugation). In particular, the discrete Gabor transform of a signal or image is a discrete sampling—in the time-frequency plane—of the STFT of the signal.

In our applications, the windowing function, g , is real-valued and symmetric about the origin. The x -dependence of the STFT of f thus corresponds to the central coordinate of the translated atom, $g(t-x)$, which in turn acts by windowing the signal before the Fourier transform is applied. Thus for each x , we obtain a localized spectrum of the signal. The suite of these spectra, $\text{STFT}(f)$, is a representation of f in the time-frequency plane.

The structure of the STFT of f depends on the choice of window: larger windows achieve better frequency resolution at the cost of reduced time resolu-

¹Much of what is often called “time-frequency analysis” (e.g., [36]) is actually carried out on functions of d variables, and so its scope encompasses both signals and images. (The term “space-wavenumber analysis” is not quite as elegant!)

tion, while smaller windows achieve better time resolution at the cost of reduced frequency resolution. This is a fundamental obstruction to the concept of instantaneous frequency, and, more generally, to time-frequency analysis [36]. Indeed, it would be impossible to determine in an infinitesimal amount of time, for example, which note is plucked on a violin at any particular instant. The act of plucking the string takes a finite amount of time itself.

The formal expression of this obstruction is known as *Heisenberg's uncertainty principle*, and it is well known that the window which balances this resolution trade-off is the Gaussian.

Under broad conditions on the choice of window and the time-frequency lattice on which the discrete Gabor transform is sampled, the corresponding representation of a signal is lossless—in fact it is typically redundant—and a perfect reconstruction of the signal is always possible. In other words, given a suitable choice of Gabor frame, the Gabor transform is guaranteed to be invertible.

If we interpret the Gabor transform for what it essentially is—simply a natural, nonstationary extension of the Fourier transform—then our intuition leads us to conjecture that the whole of classical Fourier analysis can be carried over to the nonstationary setting, simply by replacing the Fourier transform with an appropriate Gabor transform.

In particular, this guiding principle leads us to consider the idea of a Gabor filter. Suppose we want to perform a generalized convolution operation on a signal, by applying a time-variant impulse response. A Fourier transform over the time variable yields a representation of this time-variant impulse response as a time-variant transfer function on the time-frequency plane. The natural temptation now

is to multiply the Gabor transform of the original signal by this transfer function, and then apply an inverse Gabor transform to reconstruct a filtered version of the signal.

An equally compelling approach to time-frequency analysis—known as microlocal analysis—comes from the study of differential equations in mathematical physics. Here, the analysis tool of choice is a pseudodifferential operator. Formally, a pseudodifferential operator looks much like a classical Fourier multiplier; that is, a Fourier filter, except that its transfer function is no longer required to be shift-invariant over time.

As we shall see, every linear partial differential equation with functional coefficients has associated with it a Kohn-Nirenberg pseudodifferential operator.² In fact, if we allow distributional symbols, then every linear operator can be represented as a pseudodifferential operator ([33], [49]). In particular, every Gabor filter corresponds to a (possibly non-classical) pseudodifferential operator.

For a certain class of analysis and synthesis windows, (for example, gaussians) Gibson, Lamoureux and Margrave ([40]) have determined exactly which pseudodifferential operators can be represented as a Gabor filter, and explicit formulas for computing the Gabor symbol when it exists. Their important technical paper ([40]) is concerned exclusively with the continuous version of the Gabor transform.

On a more practical level, we can rest assured that as long as the symbol of the pseudodifferential operator of interest is reasonably smooth compared to the analysis and synthesis atoms, the Gabor filter defined by the same symbol will be

²Albeit, the associated operator may have a non-classical, or distributional symbol, depending on the regularity of the coefficients.

a good approximation to the original operator.

These results help to justify our approach to the two nonstationary seismic inversion problems considered in Chapters 6 and 7 of this thesis. Both of these inverse problems are based on linear wave-propagation models, with mathematical descriptions that are most readily formulated as Kohn-Nirenberg pseudodifferential operators. As one might have guessed by now, we have developed solutions to these inverse problems within the framework of Gabor filter theory.

In Chapter 6 we consider a nonstationary seismic deconvolution problem. Margrave, et al. have proposed a general framework for solving this problem—called Gabor deconvolution—which first appeared in [60], (a related method appeared previously in [85]) and is now rapidly gaining popularity within the Geophysics community and beyond.

Current deconvolution methods, based on a stationary principle known as “the convolutional model”, attempt to estimate and remove the embedded causal wavelet from the temporal seismic data. We propose a more realistic, nonstationary seismic model, expressed in the time-frequency domain as a pseudodifferential operator, in which (1) the embedded causal wavelet factors as the product of a stationary source signature and a nonstationary exponential decay; and (2) a nonstationary impulse response of the earth is tractable.

Analytic expressions are derived for least-squares fitting the model to the Gabor-transform of the seismic trace, which yield estimates of Q (a dimensionless quantity controlling the dissipation of the medium) and the source signature, hence an estimate of the travelling wavelet.

These estimates lead to a smoothed version of the magnitude of the Gabor

spectrum of the seismic trace, from which a least-squares nonstationary minimum-phase deconvolution filter is easily constructed. Our results on synthetic data are very promising.

The assumption of a minimum phase seismic wavelet is central to the Wiener and Gabor deconvolution algorithms. The former assumes a stationary wavelet, while the latter explicitly involves a nonstationary wavelet, and so an extension of the concept of minimum phase to this setting is required. Moreover, although the concept of minimum phase is well understood by the geophysics community (and the sound engineering, electrical engineering, and other applied physics communities) for the case of digital signals, (e.g., Robinson, [77] and [76]) the extension of minimum phase to analog signals has not received sufficient attention in the literature. The resolution of these issues forms the main content of Chapter 3.

Certain well-known properties of digital minimum phase signals, such as convolutional invertibility, stability, and Hilbert transform relationships between Fourier amplitude and phase spectra, do not carry over so easily to the continuous setting. In fact, there are some commonly held misconceptions surrounding the theory, and it is a primary objective of Chapter 3 to overcome these issues by presenting a logically consistent and mathematically rigorous theory of minimum phase.

We have found it necessary to develop the analog theory within the rather general framework of tempered distributions. This includes the derivation of the Hilbert transform from the assumption of causality, and its subsequent extension to the space of tempered distributions.

The stage is then set for the basic development of the theory of nonstationary, minimum-phase filters. As we shall see, these are the pseudodifferential operators

which preserve minimum phase; that is, the ones that transform a minimum phase signal into another minimum phase signal. As mentioned above, such a filter is used explicitly in the theoretical development of Gabor deconvolution.

The second application of Gabor filtering theory to seismic inversion appears in Chapter 7. It is based on the premise that wavefield extrapolation for a laterally varying velocity model can be achieved by applying a nonstationary phase-shift filter to an adaptive, nonuniform Gabor transform over the lateral coordinate.

The resulting algorithm, called adaptive Gabor phase-shift (AGPS), uses results from Chapter 4, which introduces a new method for constructing adaptive, nonuniform Gabor frames from maximally redundant, uniform partitions of unity. The main purpose of the construction is to minimize the redundancy of the Gabor transform. The result is a significant decrease in computational intensity. A preliminary discussion of Gabor frames, culminating in a criterion (Theorem 4.2.1) for generating such frames, is also provided in Chapter 4.

AGPS has a computational cost that is proportional to the complexity of the velocity model, while its accuracy is comparable to the NSPS (nonstationary phase-shift, based on an adjoint-form pseudodifferential operator) and generalized PSPI (phase-shift plus interpolation, based on a pseudodifferential operator) algorithms (see [56] and [27], respectively). AGPS includes NSPS and PSPI as complementary limiting cases, yet the cost of AGPS can be an order of magnitude less than that of either of these two methods.

The NSPS and PSPI filters differ to the extent that the functional dependence of the velocity field of the medium is either on the input coordinates (NSPS) or the output coordinates (PSPI)—in fact, they are spatial transposes of each other.

This is why both methods are approximate: indeed, any accurate phase-shift operator in a $v(x)$ medium must allow for velocity variations along the trajectory of a ray. AGPS attempts to address this problem by representing the input and output wavefields as superpositions of windowed components, each of which is approximately stationary with respect to the velocity field.

A general overview of analog linear systems from the perspective of functional analysis is provided in Chapter 2. This serves to fix the notation and provide the standard results required for the remainder of the thesis. Topics covered there include function spaces for representing signals, the generalized function space of tempered distributions, linear operators acting in Hilbert spaces, the Fourier transform, and distributional derivatives.

Generally, the interdependence of the chapters is as follows: Chapter 2 is a prerequisite for the entire work; Chapter 6 depends on Chapters 3 and 2; Chapter 7 relies on Chapters 4 and 2. A basic index is also provided, which contains references to most of the notation and key concepts.

This work was typeset with $\text{\LaTeX}2\text{e}$, with figures generated by MATLAB and encoded as EPS Level 2 Color.

Chapter 2

Linear systems

It is unfortunate for the world we live in that all of the operators that arise naturally are not bounded.
—*John B. Conway*

The mathematical topics considered in this chapter fall under the broad heading of functional analysis. The books by Rudin, [81], Conway, [12], Reed and Simon, [75], and Cheney, [8], are all excellent resources. These texts generally assume an elementary background in real analysis (see Rudin's book, [82], or [80]) and some experience with complex analysis (see [52], or [9]). Halmos' book, [41], has many interesting problems on Hilbert space, complete with hints and solutions. The two volumes by Dunford and Schwartz, [18], are the classic treatises on linear operator theory. Finally, the volumes by Kadison and Ringrose, [45], provide excellent coverage of the theory of operator algebras.

2.1 Spaces for representing signals and images

A signal is a real-valued function of a single time variable, while an image is a real-valued function of up to three spatial variables. By considering classes of functions defined on d -dimensional Euclidean space, \mathbf{R}^d , the many properties that signals and images have in common can be studied simultaneously. For this reason and for brevity, we adopt a common convention and often refer to both signals and images as signals. Similarly, much of what is called “time-frequency analysis” (e.g., [36]) is carried out on functions of d variables, and so its scope encompasses both signals and images.

Unless otherwise stated, by a function we shall mean a mapping $f : \mathbf{R}^d \rightarrow \mathbf{C}$ from real d -dimensional Euclidean space to the set of complex numbers, \mathbf{C} . The support of a function f , denoted by $\text{Supp}(f)$, is the closure in \mathbf{R}^d of the set of points at which f is nonzero.

We will also consider occasionally functions $f : \mathbf{C} \rightarrow \mathbf{C}$ of a complex variable (see e.g., [52], [1], or [9] for more on this topic). Such a function is analytic at a point z if it is complex differentiable at that point. If f is analytic at every point $z \in \mathbf{C}$, then f is called entire. A function of a complex variable has the remarkable property of being infinitely complex differentiable at each point at which it is analytic.

2.1.1 The Hilbert space of finite-energy signals

The functions we shall consider for representing signals and images will have finite, measurable energy. These belong to the class of the so-called \mathcal{L}^2 -functions, a space

of complex-valued, measurable (see e.g., [82] or [80] for more on measurability) functions on \mathbf{R}^d , given by

$$\mathcal{L}^2(\mathbf{R}^d) = \{f : \mathbf{R}^d \rightarrow \mathbf{C} : \|f\|_2 < \infty\}, \quad (2.1)$$

where the energy, or norm, of f is defined by

$$\|f\|_2 = \left(\int_{\mathbf{R}^d} |f(x)|^2 dx \right)^{1/2}. \quad (2.2)$$

The reason for the subscript “2” in the notation for the norm will become clear in the next subsection. An inner product is defined on $\mathcal{L}^2(\mathbf{R}^d)$ by

$$\langle f, g \rangle = \int_{\mathbf{R}^d} f(x) \overline{g(x)} dx, \quad (2.3)$$

where the overbar denotes complex conjugation. It induces the norm on $\mathcal{L}^2(\mathbf{R}^d)$, since $\|f\|_2 = \sqrt{\langle f, f \rangle}$. The inner product space $\mathcal{L}^2(\mathbf{R}^d)$ is complete (all Cauchy sequences converge within the space) with respect to the distance function induced by the norm,

$$\text{dist}(f, g) = \|f - g\|_2, \quad (2.4)$$

and thus it forms a Hilbert space (that is, a vector space with an inner product, which is also complete with respect to the distance function induced by the inner product). Indeed, $\mathcal{L}^2(\mathbf{R}^d)$ is the prototypical Hilbert space.

A sequence of vectors $f_n \in \mathcal{L}^2(\mathbf{R}^d)$ is a Cauchy sequence if $\|f_m - f_n\|_2 \rightarrow 0$ as $m, n \rightarrow \infty$. Completeness means that the sequence converges in norm to a unique function $f \in \mathcal{L}^2(\mathbf{R}^d)$: $\|f - f_n\|_2 \rightarrow 0$ as $n \rightarrow \infty$.

Two vectors f, g in a Hilbert space are orthogonal to each other if $\langle f, g \rangle = 0$. Similarly, a pair of subspaces K, K_1 of a Hilbert space are orthogonal if every pair $f \in K$ and $g \in K_1$ is orthogonal.

Although a realizable signal is real-valued, its Fourier transform is typically complex-valued. Also, it is well known that the Fourier transform maps $\mathcal{L}^2(\mathbf{R}^d)$ onto itself, which motivates us to consider signals in this more general setting.

2.1.2 The Lebesgue spaces

For $1 \leq p < \infty$, a measurable function f (see e.g., [82] or [80] for more on measurability) is in $\mathcal{L}^p(\mathbf{R}^d)$ if its p -norm is finite; that is, if

$$\|f\|_p = \left(\int_{\mathbf{R}^d} |f(x)|^p dx \right)^{1/p} < \infty. \quad (2.5)$$

Thus, for $1 \leq p < \infty$, a measurable function f is in $\mathcal{L}^p(\mathbf{R}^d)$ if and only if $|f|^p \in \mathcal{L}^1(\mathbf{R}^d)$. For $p = \infty$, $f \in \mathcal{L}^\infty(\mathbf{R}^d)$ if f is measurable and

$$\|f\|_\infty = \text{ess sup}|f(x)| < \infty. \quad (2.6)$$

Here, “ess sup” means “essential supremum” (see [82], [51] or [80] for details). For continuous functions, this is the same as the supremum.

The most interesting \mathcal{L}^p -spaces for our purposes are for $p = 1, 2$, or ∞ . Elements of $\mathcal{L}^p(\mathbf{R}^d)$ are actually equivalence classes of measurable functions which are identified if and only if they agree “almost everywhere”; that is, except on a set of measure zero (the integral of the indicator function of the set of points at which the two functions disagree vanishes). For example, the zero function can be perturbed at countably many points, and still remain in the same equivalence class. Generally, two measurable functions f and g are equivalent if and only if $f - g = 0$ almost everywhere. It is customary to use the same notation both for a function and the equivalence class it represents.

2.1.3 The Schwartz space of rapidly decreasing signals

The \mathcal{L}^p -spaces contain a host of rather exotic functions, lacking basic regularity properties such as differentiability or even continuity. However, there is a space of C^∞ , (smooth, or infinitely differentiable) rapidly-decreasing functions, called the Schwartz class, which is dense in \mathcal{L}^p for $1 \leq p < \infty$, (and weak-* dense in \mathcal{L}^∞ , but we shall not need this fact) (see [12] or [81]). Thus we can approximate an \mathcal{L}^p -function arbitrarily well by a smooth function in the \mathcal{L}^p -norm.

A complex-valued function $\phi(x)$ on \mathbf{R}^d is called “rapidly decreasing” if it is infinitely differentiable and it and each of its derivatives decays faster than the inverse of any polynomial; that is, if for all multiindices α, β , the functions $x^\alpha \partial^\beta \phi(x)$ are bounded ([84]). A multiindex, α , is a d -tuple of positive integers, $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_d) \in \mathbf{Z}_+^d$ (\mathbf{Z} denotes the set of all integers). Thus, with $x = (x_1, x_2, \dots, x_d) \in \mathbf{R}^d$, the polynomial x^α is defined as

$$x^\alpha = x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_d^{\alpha_d}, \quad (2.7)$$

and ∂^α means

$$\partial^\alpha = \partial_1^{\alpha_1} \partial_2^{\alpha_2} \cdots \partial_d^{\alpha_d}, \quad (2.8)$$

where $\partial_k^{\alpha_k}$ is the α_k th partial derivative with respect to the k th coordinate x_k .

The set of all rapidly decreasing functions is called Schwartz space and is denoted by $\mathcal{S}(\mathbf{R}^d)$. The Fourier transform, (formula (2.10) below) gives a one-to-one and onto correspondence between $\mathcal{S}(\mathbf{R}^d)$ and $\mathcal{S}(\mathbf{R}^d)$, with pointwise multiplication taken into convolution and vice versa (see [12] or [81]).

By Plancherel's theorem,

$$\|f\|_2 = \left(\frac{1}{2\pi}\right)^{\frac{d}{2}} \|\hat{f}\|_2 \quad (2.9)$$

for all $f \in \mathcal{S}(\mathbf{R}^d)$, and since $\mathcal{S}(\mathbf{R}^d)$ is dense in $\mathcal{L}^2(\mathbf{R}^d)$, the Fourier transform extends to a unitary operator on $\mathcal{L}^2(\mathbf{R}^d)$. We emphasize that the formula

$$\hat{f}(\xi) = \int_{\mathbf{R}^d} f(x) e^{-i\xi \cdot x} dx \quad (2.10)$$

for the Fourier transform of f only makes sense when f is in $\mathcal{L}^1(\mathbf{R}^d)$ —otherwise the integral is not well-defined. The point is that although the Fourier transform extends to an isometry on $\mathcal{L}^2(\mathbf{R}^d)$, its action on \mathcal{L}^2 -functions outside $\mathcal{L}^1(\mathbf{R}^d)$ is not given by formula (2.10).

An important feature of a function $f \in \mathcal{L}^1(\mathbf{R}^d)$ is that its Fourier transform is a continuous function that vanishes at infinity: $|\hat{f}(\xi)| \rightarrow 0$ as $|\xi| \rightarrow \infty$ (see [12] or [81]). This decay property is not obvious by inspection of 2.10—the proof involves the density of $\mathcal{S}(\mathbf{R}^d)$ in $\mathcal{L}^1(\mathbf{R}^d)$, and the completeness of $C_o(\mathbf{R}^d)$ (the space of continuous functions that vanish at infinity) in the \mathcal{L}^∞ -norm (see [81], Theorem 7.5).

2.2 The space of tempered distributions

Distributions, or generalized functions, arise as a generalization of the idea of a bounded function. Motivated by the desire to understand Dirac's delta-function, δ , Laurent Schwartz developed his theory of distributions in 1944—and received the Fields Medal in 1950 for this work (see e.g., [87]). Schwartz actually studied generalized functions later than many others (e.g., Hadamard, Riesz, Sobolev,

Bochner), but he was the first to systematize the theory, while relating all the earlier approaches and establishing many important results ([29]).

It would be difficult, if not impossible, to avoid this subject in any rigorous account of inverse-filtering theory. For a relevant example—the relevance of which will become apparent below—suppose we are given $f \in \mathcal{L}^1(\mathbf{R}^d) \cap \mathcal{L}^2(\mathbf{R}^d)$, and we want to solve the following problem for g :

$$f * g = \delta, \tag{2.11}$$

subject to the condition that g is in the domain of the Fourier transform. Here the $*$ denotes the operation of convolution. Then the problem can be transformed to the Fourier domain, as

$$\widehat{f}\widehat{g} = 1. \tag{2.12}$$

Formally, then, we expect a solution of the form

$$\widehat{g} = 1/\widehat{f}, \text{ or } g = \mathcal{F}^{-1}\{1/\widehat{f}\}, \tag{2.13}$$

where \mathcal{F}^{-1} denotes the inverse Fourier transform. However, given any $g \in \mathcal{L}^p(\mathbf{R}^d)$ with $1 \leq p \leq \infty$, we have $f * g \in \mathcal{L}^p(\mathbf{R}^d)$, while $\delta \notin \mathcal{L}^p(\mathbf{R}^d)$. Thus $g \notin \mathcal{L}^p(\mathbf{R}^d)$ for any p . Another way to see this is since $f \in \mathcal{L}^1(\mathbf{R}^d)$, \widehat{f} vanishes at infinity, and so any “function” g whose Fourier transform satisfies (2.12) must have a spectrum that grows without bound at infinity. The point is that the problem (2.11) has no classical solution; that is, one that can properly be called a function.

For an authoritative account of generalized function theory and its applications, the English translation of the Russian 5-volume series, *Generalized Functions*, by Gel’fand, et al., is highly recommended (see [28] through to [32]).

Definition A tempered distribution T is a linear functional which is continuous on Schwartz space; that is, a continuous linear mapping $T : \mathcal{S}(\mathbf{R}^d) \rightarrow \mathbf{C}$.

The continuity is with respect to the family of norms ([84]):

$$|\phi|_k = \sup\{|x^\alpha \partial^\beta \phi(x)| : x \in \mathbf{R}^d \text{ and } |\alpha + \beta| \leq k\}, \quad (2.14)$$

indexed by the positive integers, $k \in \mathbf{Z}_+$, and where α and β in \mathbf{Z}_+^d are multiindices.

Given a multiindex α , the notation $|\alpha|$ means the sum $|\alpha| = \alpha_1 + \alpha_2 + \cdots + \alpha_d$.

The space of tempered distributions is denoted by $\mathcal{S}'(\mathbf{R}^d)$. A very large, and possibly the most important class of tempered distributions comes from the tempered functions on \mathbf{R}^d (see [108]). A measurable function f on \mathbf{R}^d is said to be tempered if

$$\int_{\mathbf{R}^d} \frac{|f(x)|}{(1 + |x|)^N} dx < \infty \quad (2.15)$$

for some positive integer N . Convergence of the integral (2.15) implies f has at most polynomial growth:

$$|f(x)| \leq C(1 + |x|)^N. \quad (2.16)$$

Given any tempered function f on \mathbf{R}^d , the linear functional T_f on $\mathcal{S}(\mathbf{R}^d)$ given as

$$T_f(\phi) = \int_{\mathbf{R}^d} f(x)\phi(x) dx, \quad \phi \in \mathcal{S}(\mathbf{R}^d) \quad (2.17)$$

is a tempered distribution. For example, every \mathcal{L}^p -function is tempered, (for all p with $1 \leq p \leq \infty$), and so (2.17) defines a tempered distribution whenever $f \in \mathcal{L}^p(\mathbf{R}^d)$.

Definition The support of a distribution T , denoted by $\text{Supp}(T)$, is the smallest closed set $F \subseteq \mathbf{R}^d$ satisfying the following property:

$$T(\phi) = 0 \text{ for all } \phi \in \mathcal{S}(\mathbf{R}^d) \text{ with } \text{Supp}(\phi) \subseteq \mathbf{R}^d \setminus F. \quad (2.18)$$

If Γ is the collection of all closed sets $F \subseteq \mathbf{R}^d$ satisfying property (2.18), then ([8])

$$\text{Supp } (\phi) = \bigcap \{F : F \in \Gamma\}. \quad (2.19)$$

The definition of support for a distribution is compatible with the definition of support for a function, in the sense that if T_g is the tempered distribution corresponding to a suitably behaved function g , (e.g., any tempered function) then its support agrees with the usual definition of support for a function (see [8]). This is why the same notation is conventionally used for both definitions.

2.2.1 Distributional derivatives

The concept of differentiation of functions can be extended to the entire class of tempered distributions, as follows.

Definition (e.g., [51], [84]) Let $D^\alpha = (-i)^{|\alpha|} \partial^\alpha$, ($\alpha \in \mathbf{Z}_+^d$) and let $T \in \mathcal{S}'(\mathbf{R}^d)$.

Then the derivative of T is defined as

$$D^\alpha T = T D^\alpha. \quad (2.20)$$

The use of the differential operator $D^\alpha = (-i)^{|\alpha|} \partial^\alpha$ helps to simplify expressions involving partial derivatives (D is then a self-adjoint operator, as we shall see below).

It is a remarkable fact that any tempered distribution supported at the origin is a finite linear combination of derivatives of the Dirac distribution ([51], Theorem 5, page 387). More generally, any tempered distribution with discrete (consisting of isolated points) support is just a countable superposition of translates of distributions supported at the origin ([51]). One can also show (e.g., [51])

that a distribution with compact (closed and bounded) support is $\mathcal{D}T_f$, for some differential operator \mathcal{D} and continuous function f .

2.2.2 The Fourier transform of a distribution

Definition The Fourier transform, \hat{T} , of a tempered distribution $T \in \mathcal{S}'(\mathbf{R}^d)$ is defined by its action on Schwartz functions, (see e.g., [51], [8], or [82]) as

$$\hat{T}(\phi) = T(\hat{\phi}). \quad (2.21)$$

An important property of the class of tempered distributions is the following.

Proposition 2.2.1 *The Fourier transform defines a continuous linear bijection of $\mathcal{S}'(\mathbf{R}^d)$ onto itself.*

Proof The Fourier transform \mathcal{F} is a continuous linear bijection of $\mathcal{S}(\mathbf{R}^d)$ onto itself. Since the Fourier transform \hat{T} of a tempered distribution T is defined by a composition of continuous linear maps, namely

$$\hat{T} = T \circ \mathcal{F}, \quad (2.22)$$

the mapping $T \mapsto \hat{T}$ is continuous and linear. In particular, $\hat{T} \in \mathcal{S}'(\mathbf{R}^d)$, and so the Fourier transform maps $\mathcal{S}'(\mathbf{R}^d)$ into itself.

Given any $T \in \mathcal{S}'(\mathbf{R}^d)$, set

$$S = T \circ \mathcal{F}^3 \in \mathcal{S}'(\mathbf{R}^d). \quad (2.23)$$

Then since \mathcal{F}^4 is the identity mapping on Schwartz space, we have that $T = \hat{S}$, and so the Fourier transform maps $\mathcal{S}'(\mathbf{R}^d)$ onto itself. It is one-to-one since if $\hat{T} = \hat{S}$, then applying the Fourier transform to both sides three times gives back $T = S$. ■

2.3 Linear operators in Hilbert space

Let \mathcal{H} and \mathcal{K} be separable Hilbert spaces. Separable means having a countable dense subset, and by applying the Gram-Schmidt orthogonalization algorithm to this set, one can always form an orthonormal basis. For our purposes, it is enough to consider the cases $\mathcal{H} = \mathcal{L}^2(\mathbf{R}^d)$, with $\mathcal{K} = \mathcal{H}$ or $\mathcal{K} = \mathbf{C}$.

Definition A linear operator is a mapping $K : \mathcal{H} \rightarrow \mathcal{K}$ that preserves the linear structure of its domain, $\text{Dom } K$. Specifically, for all $\lambda \in \mathbf{C}$ and for all $f, g \in \text{Dom } K \subseteq \mathcal{H}$,

$$K(\lambda f + g) = \lambda Kf + Kg \in \mathcal{K}. \quad (2.24)$$

We require $\text{Dom } K$ to be a linear subset of \mathcal{H} , but we do not require it to be closed (topologically closed, meaning that it need not contain all its limit points).

If there is a $C > 0$ such that for all $f \in \text{Dom } K$

$$\|Kf\| \leq C\|f\|, \quad (2.25)$$

then K is called a “bounded operator”. If K is bounded, then its norm, $\|K\|$ is defined as the smallest constant C satisfying (2.25):

$$\|K\| = \inf\{C \geq 0 : \|Kf\| \leq C\|f\|, \forall f \in \mathcal{H}\}. \quad (2.26)$$

Otherwise, K is unbounded. There is a standard abuse of notation for the norms on either side of expression (2.25): the left side refers to the norm on \mathcal{K} , while the right refers to the norm on \mathcal{H} .

Bounded operators

The collection of bounded linear operators $\mathcal{H} \rightarrow \mathcal{K}$ is denoted by $\mathcal{B}(\mathcal{H}, \mathcal{K})$. When $\mathcal{H} = \mathcal{K}$, we simply write $\mathcal{B}(\mathcal{H})$. We state a few basic results from the theory of

bounded operators on Hilbert space (see e.g., [12], [18], [41], and [8]; for more advanced results, see [66], [15], [17], [71], and [45]).

The space $\mathcal{B}(\mathcal{H})$ is complete with respect to the operator norm (2.26). It forms a unital algebra over \mathbf{C} under composition and addition of bounded operators. The mapping from $\mathcal{B}(\mathcal{H})$ to itself defined by $K \mapsto K^*$, which carries an operator to its adjoint, satisfies

$$\begin{aligned}(K^*)^* &= K, & (KH)^* &= H^*K^*, \\ (\lambda K + H)^* &= \bar{\lambda}K^* + H^*, & \text{and} \\ \|K^*\| &= \|K\|,\end{aligned}$$

for all $\lambda \in \mathbf{C}$ and $K, H \in \mathcal{H}$. A linear operator $K : \mathcal{H} \rightarrow \mathcal{K}$ is continuous if and only if it is bounded.

Let $K \in \mathcal{B}(\mathcal{H}, \mathcal{K})$. Then by continuity, K can be extended uniquely to a bounded operator on the closure of its domain, $\text{cl}[\text{Dom } K]$, and hence to all of \mathcal{H} by letting K vanish outside $\text{cl}[\text{Dom } K]$. Thus, it is normally assumed that $\text{Dom } K = \mathcal{H}$ whenever K is bounded.

A bounded operator $K \in \mathcal{B}(\mathcal{H})$ is self adjoint if $K = K^*$, unitary if $K^* = K^{-1}$, idempotent if $K^2 = K$, and a projection if and only if it is a self adjoint idempotent.

Every unitary operator

$$U : \mathcal{H} \rightarrow \mathcal{H} \tag{2.27}$$

on a Hilbert space \mathcal{H} is an isometric isomorphism (or isometry); that is, it preserves norms (which implies boundedness and hence continuity) and defines a one-to-one correspondence from \mathcal{H} onto itself.

A partial isometry is an operator $W \in \mathcal{B}(\mathcal{H})$ for which $\|Wf\| = \|f\|$ for all $f \in (\text{Ker}W)^\perp$, where $(\text{Ker}W)^\perp$ is the closed linear subspace of \mathcal{H} consisting of all $f \in \mathcal{H}$ which are orthogonal to every $k \in \text{Ker}W$, and $\text{Ker}W$ is the closed linear subspace of \mathcal{H} consisting of all $k \in \mathcal{H}$ with $Wk = 0$.

Chapter 3

Minimum phase for nonstationary analog systems

If the facts don't fit the theory, change the facts.
—*Albert Einstein*

The assumption of a minimum phase seismic wavelet is central to the Wiener and Gabor deconvolution algorithms. Although the concept of minimum phase is well understood by the geophysics community (and the sound engineering, electrical engineering, and other applied physics communities) for the case of digital signals, (e.g., Robinson, [77] and [76]) the extension of minimum phase to analog signals has not received sufficient attention in the literature.

Certain well-known properties of digital minimum phase signals, such as convolutional invertibility, stability, and Hilbert transform relationships between Fourier

amplitude and phase spectra, do not carry over so easily to the continuous setting. In fact, there are some commonly held misconceptions surrounding the theory. This chapter addresses these issues by presenting a logically consistent and mathematically rigorous theory of minimum phase.

We have found it necessary to develop the analog theory within the rather general framework of tempered distributions. This includes the derivation of the Hilbert transform from the assumption of causality, and its subsequent extension to the space of tempered distributions.

The stage is then set for the development of the theory of nonstationary, minimum-phase filters. As we shall see, these are the pseudodifferential (and Fourier integral) operators which preserve minimum phase; that is, the ones which carry minimum phase signals to minimum phase signals. Such a filter is used explicitly in the theoretical development of Gabor deconvolution.

3.1 Causality and the Hilbert transform

Definition A tempered distribution (or function) T is causal if $\text{Supp}(T) \subseteq [0, \infty)$.

More generally, T can be any generalized function for which $\text{Supp}(T)$ makes sense and $\text{Supp}(T) \subseteq [0, \infty)$. The reason for working in such a general setting will become clear as we proceed, particularly when we extend the concept of minimum phase to the analog setting.

We shall make use of the Hilbert transform, \mathcal{H} , a bounded linear operator on $\mathcal{L}^2(\mathbf{R})$. We first define its action on Schwartz functions as

$$(\mathcal{H}s)(\omega) = \text{P.V.} \frac{1}{\pi} \int_{\mathbf{R}} \frac{s}{\omega - \xi} d\xi$$

$$= \frac{1}{\pi} \lim_{R \rightarrow \infty} \lim_{\epsilon \rightarrow 0} \left[\int_{-R}^{\omega - \epsilon} \frac{s}{\omega - \xi} d\xi + \int_{\omega + \epsilon}^R \frac{s}{\omega - \xi} d\xi \right], \quad (3.1)$$

where P.V. denotes the Cauchy principal value of the integral (see e.g., [1], [52]). As we shall see, the Hilbert transform extends as a bounded linear operator on $\mathcal{L}^2(\mathbf{R})$, and satisfies the following properties on $\mathcal{L}^2(\mathbf{R})$: $\mathcal{H}^2 = -I$, $\mathcal{H}^* = -\mathcal{H}$, and thus $\mathcal{H}^{-1} = \mathcal{H}^*$; that is, \mathcal{H} is unitary. Also, if s is real-valued then so is $\mathcal{H}s$. Some authors (e.g., [11]) have a factor of $1/i$ in the definition of \mathcal{H} , which makes \mathcal{H} a self-adjoint unitary.

Proposition 3.1.1 *If s is a causal Schwartz function, then the real and imaginary parts of its Fourier transform are a Hilbert transform pair, namely*

$$\Re \{ \hat{s}(\omega) \} = (\mathcal{H} \Im \{ \hat{s} \})(\omega), \quad (3.2)$$

where \Re and \Im are used to denote the real and imaginary parts, respectively.

Proof Suppose $s \in \mathcal{S}(\mathbf{R})$ is causal. Then $s = hs$, where h is the Heaviside function; that is, h is the indicator function of $[0, \infty)$. The Fourier transform of h is given by (e.g., [70], or [72])

$$\hat{h}(\omega) = \pi\delta(\omega) - \frac{i}{\omega}, \quad (3.3)$$

and so

$$\hat{s}(\omega) = (\hat{s} * \hat{h})(\omega) = \frac{1}{2} \left(\hat{s}(\omega) - \frac{i}{\pi} \int_{\mathbf{R}} \frac{\hat{s}(\xi)}{\omega - \xi} d\xi \right). \quad (3.4)$$

Thus

$$\hat{s}(\omega) = -\frac{i}{\pi} \int_{\mathbf{R}} \frac{\hat{s}(\xi)}{\omega - \xi} d\xi, \quad (3.5)$$

that is,

$$\Re\{\hat{s}(\omega)\} = \Im\left\{\frac{1}{\pi}\int_{\mathbf{R}}\frac{\hat{s}(\xi)}{\omega-\xi}d\xi\right\} = \frac{1}{\pi}\int_{\mathbf{R}}\frac{\Im\{\hat{s}(\xi)\}}{\omega-\xi}d\xi = (\mathcal{H}\Im\{\hat{s}\})(\omega), \quad (3.6)$$

which is what we wanted to show. ■

Larry Lines has kindly pointed out that Jon Claerbout has a nice method for performing this calculation digitally (see [10]).

3.1.1 Extension of the Hilbert transform to distributions

Expression (3.5) in the proof of proposition 3.1.1 shows that the action of the Hilbert transform on the spectrum of a causal Schwartz function is rotation by 90° in the complex plane: $\mathcal{H}\hat{s} = i\hat{s}$. From the relationship $\widehat{\tilde{s}} = \bar{\hat{s}}$, where $\tilde{s}(t) = s(-t)$ is the time reversal of s , we see that the Hilbert transform rotates the spectrum of an anti-causal signal by -90° .

Given any $f \in \mathcal{L}^2(\mathbf{R})$, f can be represented uniquely (up to a set of measure zero, hence uniquely as an \mathcal{L}^2 -function) in the form

$$f = f_+ + f_-, \quad (3.7)$$

where f_+ and f_- are the causal and anti-causal parts of f , respectively. That is, $\mathcal{L}^2(\mathbf{R})$ can be identified with the direct sum

$$\mathcal{L}^2(\mathbf{R}) = \mathcal{L}^2(\mathbf{R})_+ \oplus \mathcal{L}^2(\mathbf{R})_-, \quad (3.8)$$

where $\mathcal{L}^2(\mathbf{R})_+$ and $\mathcal{L}^2(\mathbf{R})_-$ are the orthogonal subspaces of $\mathcal{L}^2(\mathbf{R})$ consisting of the causal and anti-causal \mathcal{L}^2 -functions, respectively. Applying the Fourier transform, we can then identify $\mathcal{FL}^2(\mathbf{R})$ with the direct sum

$$\mathcal{FL}^2(\mathbf{R}) \cong \mathcal{FL}^2(\mathbf{R})_+ \oplus \mathcal{FL}^2(\mathbf{R})_-. \quad (3.9)$$

Consider now the subspace $\tilde{\mathcal{S}}$ of Schwartz space given as the direct sum of the causal and the anti-causal Schwartz functions:

$$\tilde{\mathcal{S}}(\mathbf{R}) = \mathcal{S}(\mathbf{R})_+ \oplus \mathcal{S}(\mathbf{R})_-. \quad (3.10)$$

This is just Schwartz space with $\{s \in \mathcal{S}(\mathbf{R}) : 0 \in \text{Supp}(s)\}$ removed. In particular, $\tilde{\mathcal{S}}(\mathbf{R})$ is still dense in $\mathcal{L}^2(\mathbf{R})$ (because each factor $\mathcal{S}(\mathbf{R})_{\pm}$ is dense in $\mathcal{L}^2(\mathbf{R})_{\pm}$, respectively), and moreover

$$\mathcal{F}\tilde{\mathcal{S}}(\mathbf{R}) = \mathcal{F}\mathcal{S}(\mathbf{R})_+ \oplus \mathcal{F}\mathcal{S}(\mathbf{R})_- \quad (3.11)$$

is dense in $\mathcal{F}\mathcal{L}^2(\mathbf{R}) \cong \mathcal{L}^2(\mathbf{R})$. Knowing how the Hilbert transform acts on the spectra of causal and anti-causal Schwartz functions, we see that \mathcal{H} acts diagonally on $\mathcal{F}\tilde{\mathcal{S}}(\mathbf{R})$:

$$\mathcal{H}(\hat{f}) = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix} \begin{pmatrix} \hat{f}_+ \\ \hat{f}_- \end{pmatrix} = i \begin{pmatrix} \hat{f}_+ \\ -\hat{f}_- \end{pmatrix}. \quad (3.12)$$

Evidently, expression (3.12) defines a continuous linear mapping

$$\mathcal{H} : \tilde{\mathcal{S}}(\mathbf{R}) \rightarrow \mathcal{S}(\mathbf{R}). \quad (3.13)$$

Since $\tilde{\mathcal{S}}(\mathbf{R})$ is dense in $\mathcal{L}^2(\mathbf{R})$, and $\mathcal{F}\mathcal{S}(\mathbf{R})_{\pm} \subset \mathcal{F}\mathcal{L}^2(\mathbf{R})_{\pm}$, we have the following result.

Proposition 3.1.2 *The action defined by expression (3.12) extends the Hilbert transform to a unitary operator on $\mathcal{L}^2(\mathbf{R})$. The eigenvalues of the Hilbert transform are exactly i and $-i$, and the corresponding eigenspaces are the Fourier spectra of the causal and anti-causal signals, respectively.*

The fact that \mathcal{H} is unitary can be seen directly from the following calculation:

$$\begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix} \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}^* = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix} \begin{pmatrix} -i & 0 \\ 0 & i \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \quad (3.14)$$

Actually, the same definition (3.12) extends the domain of the Hilbert transform into the space of tempered distributions. The only caveat is that the inverse Fourier transform of a tempered distribution, being also a distribution, can be both causal and anti-causal without being equivalent to the zero distribution, so the sum

$$\mathcal{S}'(\mathbf{R}) = \mathcal{S}'(\mathbf{R})_+ + \mathcal{S}'(\mathbf{R})_- \quad (3.15)$$

will no longer be direct (some nonzero objects belong to both factors, e.g., the Dirac delta distribution). This means some choices may have to be made at the origin.

Example Let us see what is involved in computing the Hilbert transform of the Dirac delta distribution, δ . First, define δ_- and δ_+ as

$$\delta_+ = \frac{1}{2\pi} \widehat{h} = \frac{1}{2} \delta - \frac{i}{2\pi\omega}, \quad (3.16)$$

(from 3.3) and

$$\delta_- = \frac{1}{2\pi} \widehat{\tilde{h}} = \frac{1}{2} \delta + \frac{i}{2\pi\omega} = \bar{\delta}_+. \quad (3.17)$$

Then δ_- and δ_+ are the Fourier spectra of the causal and the anti-causal parts of the constant function, $1 = h + \tilde{h}$, respectively, where the heaviside function h and its time reversal, \tilde{h} , are taken to have value 1/2 at the origin. We have

$$\mathcal{H}(\delta) = \frac{1}{2\pi} \mathcal{H}(\widehat{1}) = \frac{1}{2\pi} \mathcal{H}(\widehat{h} + \widehat{\tilde{h}})$$

$$\begin{aligned}
&= \mathcal{H}(\delta_+ + \delta_-) \\
&= \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix} \begin{pmatrix} \delta_+ \\ \delta_- \end{pmatrix} \\
&= (i\delta_+ - i\bar{\delta}_+) \\
&= \frac{1}{\pi\omega}.
\end{aligned}$$

Causal \mathcal{L}^1 -functions

The following property of the causal \mathcal{L}^1 -functions in terms of their Fourier spectra is used in the next section, where we discuss the topic of minimum phase.

Theorem 3.1.3 *Let $f \in \mathcal{L}^1(\mathbf{R})$. If f is causal, then its Fourier transform extends to an analytic function in the open lower half of the complex plane.*

Proof Let $f \in \mathcal{L}^1(\mathbf{R})$. Then its Fourier transform is computable using formula (2.10). Writing $\xi = u + iv$, with $u, v \in \mathbf{R}$, we have

$$\hat{f}(\xi) = \int_{\mathbf{R}} f(x)e^{-i\xi \cdot x} dx = \int_{\mathbf{R}} f(x)e^{v \cdot x} e^{-iu \cdot x} dx, \quad (3.18)$$

which, for each fixed v where the integral converges, is the Fourier transform of $f(x)e^{v \cdot x}$, evaluated at u . If f is causal, then for $v \leq 0$,

$$|f(x)e^{v \cdot x}| \leq |f(x)|, \quad (3.19)$$

and the integral converges. In fact, if f is causal and $v < 0$, then for any $N \in \mathbf{N}$, (\mathbf{N} denotes the natural numbers) the function $(-ix)^N f(x)e^{v \cdot x}$ is absolutely integrable: this follows from the assumption that f is absolutely integrable, and since $e^{v \cdot x}$ is rapidly decreasing, so is the product $(-ix)^N f(x)e^{v \cdot x}$. It now follows that \hat{f} has derivatives of all orders in the lower half of the complex plane, where the N^{th} derivative is given as the Fourier transform of $(-ix)^N f(x)$. ■

More generally, we can see from the proof that if f is measurable and the integral in (3.18) converges for some $v = v_0$, then it converges in the closed half plane $\Im(\xi) \leq v_0$, and $\hat{f}(\xi)$ is analytic in the open halfspace $\{\xi \in \mathbf{C} : \Im(\xi) < v_0\}$ ([106], or [70]). If f vanishes at infinity, then the more rapidly it decays, the larger the open halfspace becomes in which \hat{f} is analytic. Conversely, if f grows without bound at infinity, (so $v_0 < 0$) then the more rapidly f increases, the smaller the open half plane of analyticity of \hat{f} .

Theorem 3.1.4 *Let f be a measurable function such that*

$$f(x)e^{Ax} \in \mathcal{L}^1(\mathbf{R}) \quad (3.20)$$

for some fixed constant $A \in \mathbf{R}$. Then for each fixed $\nu \leq A$, the function f can be recovered as

$$f(x) = \frac{1}{2\pi} \int_{L_\nu} \hat{f}(\xi) e^{i\xi} d\xi, \quad (3.21)$$

where L_ν is the horizontal line passing through $i\nu$ in the complex plane.

Proof Fix any $\nu \leq A$. Then $g_\nu(x) \equiv f(x)e^{\nu x} \in \mathcal{L}^1(\mathbf{R})$, and its Fourier transform is given as

$$\hat{g}_\nu(\omega) = \int_{\mathbf{R}} f(x) e^{-ix(\omega+i\nu)} dx \equiv \hat{f}(\omega + i\nu). \quad (3.22)$$

Applying the inverse Fourier transform yields

$$\begin{aligned} f(x)e^{\nu x} &= g_\nu(x) \\ &= \frac{1}{2\pi} \int_{\mathbf{R}} \hat{f}(\omega + i\nu) e^{i\omega x} d\omega, \end{aligned}$$

or

$$\begin{aligned} f(x) &= \frac{1}{2\pi} \int_{\mathbf{R}} \hat{f}(\omega + i\nu) e^{ix(\omega+i\nu)} d\omega \\ &= \frac{1}{2\pi} \int_{L_\nu} \hat{f}(\xi) e^{i\xi} d\xi, \end{aligned}$$

where $\xi = \omega + i\nu$, $d\xi = d\omega$, and L_ν is the horizontal line passing through $i\nu$ in the complex plane. ■

This result is a Fourier-domain analogue to Bromwich's¹ method for recovering the inverse Laplace transform of $F(s)$ by integrating F along any vertical line, or Bromwich path, [70] to the right of all singularities of F .

3.2 Minimum phase

Definition A tempered distribution (or function) $s \in \mathcal{S}'(\mathbf{R})$ is minimum phase if \hat{s} and $1/\hat{s}$ correspond to analytic functions on the open lower half of the complex plane,

$$\mathbf{C}_- = \{u + iv \in \mathbf{C} : v < 0\}. \quad (3.23)$$

This definition requires some explanation. Indeed, it is not obvious at first exactly how one should go about extending the Fourier transform of a tempered distribution to a *function* on the open lower half of the complex plane. This process was made clear for the causal \mathcal{L}^1 -functions, as we saw in the proof of Theorem 3.1.3. To see how we can accomplish this task for a causal tempered distribution, we first look at how a Schwartz function, say $s \in \mathcal{S}(\mathbf{R})$, can be extended to the complex plane (following advice of M. Lamoureux).

The Fourier transform of s is $\hat{s}(u) = \int_{\mathbf{R}} s(x)e^{-iux} dx$, and we can define its extension to the complex plane, where it makes sense, by formally replacing u with $z = u + iv$:

$$\hat{s}(z) = \int_{\mathbf{R}} s(x)e^{vx}e^{-iux} dx. \quad (3.24)$$

¹Thomas John l'Anson Bromwich, 1875-1929

Certainly if s has compact support then this integral makes perfect sense, since the exponential growth is cut off outside the support of s . If s is causal, then for any $z \in \mathbf{C}_-$, the exponential growth is cut off, while the exponential decay makes it even easier for the integral to converge. Similarly, if s is anti-causal, the integral makes sense on the open upper half plane. The main point here is that wherever the integral converges, it defines a function.

Now for tempered distributions, it is convenient to use the inner product notation to extend the definition 3.24, namely

$$\hat{s}(z) = \langle s, e^{iz(\cdot)} \rangle, \quad (3.25)$$

where the (\cdot) is just a placeholder for the argument x . Even though $e^{iz(\cdot)}$ is not a Schwartz class function, the inner product makes sense, for instance, for any $z \in \mathbf{C}_-$ whenever s is a causal distribution. Indeed, take any smooth cutoff function ϕ , with $\phi(x) = 1$ on $[0, \infty)$, and vanishing on $(-\infty, -\epsilon]$, for some $\epsilon > 0$. Then the product $\phi(x)e^{izx}$ is Schwartz class for any $z \in \mathbf{C}_-$, and we can define

$$\hat{s}(z) = \langle s, \phi(\cdot)e^{iz(\cdot)} \rangle. \quad (3.26)$$

Now this definition is independent of the choice of cutoff function, for if ψ is another cutoff function which is 1 on the support of s , then $\phi - \psi$ vanishes on the support of s , so that

$$0 = \langle s, (\phi - \psi)(\cdot)e^{iz(\cdot)} \rangle = \langle s, \phi(\cdot)e^{iz(\cdot)} \rangle - \langle s, \psi(\cdot)e^{iz(\cdot)} \rangle; \quad (3.27)$$

that is, $\langle s, \phi(\cdot)e^{iz(\cdot)} \rangle = \langle s, \psi(\cdot)e^{iz(\cdot)} \rangle$. Thus the prescription 3.25 extends the Fourier transform of a causal tempered distribution to a well-defined function

on the open lower half plane. So long as the tempered distribution has one-sided support, we can extend its Fourier transform to a function on an open half plane in \mathbf{C} , and to the entire complex plane in the case of a compactly supported distribution.

Example Consider the extension of the Fourier transform of the (compactly supported) Dirac distribution, δ , to the whole complex plane, as follows. This is just given as

$$\hat{\delta}(z) = \langle \delta, e^{iz(\cdot)} \rangle = e^{iz(0)} = 1, \quad (3.28)$$

which incidentally is the only natural way to analytically continue the constant function 1 from the real line to the complex plane. This also shows that the Dirac distribution is minimum phase.

Example Consider the Fourier transform of the distributional derivative, δ' , of the Dirac distribution, δ . We have

$$\hat{\delta}'(z) = \langle \delta', e^{iz(\cdot)} \rangle = \langle \delta, iz e^{iz(\cdot)} \rangle = \langle -iz\delta, e^{iz(\cdot)} \rangle = -iz, \quad (3.29)$$

which is entire. Also, since $-1/iz$ is analytic in the lower half plane, we see that δ' is minimum phase.

An immediate consequence of our definition of minimum phase is the following.

Proposition 3.2.1 *If $f, g \in \mathcal{S}'(\mathbf{R})$ are minimum phase tempered distributions for which their convolution $f * g$ is defined, then $f * g$ is also minimum phase.*

Proof The Fourier transform of $f * g$ is the product FG of the Fourier transforms of f and g . Since both factors and their pointwise inverses are analytic in the open lower half of the complex plane, the same is true of their product. ■

The definition is also independent of the choice of scale, as it ought to be. That is, given any $\lambda > 0$ and $g \in \mathcal{S}'(\mathbf{R})$, it follows immediately that g is minimum phase if and only if λg is minimum phase.

Actually, the term “minimum phase” more appropriately relates to the Fourier spectrum \hat{s} , whereas s is more accurately described as being “minimum delay”. The two terms have come to have the same meaning in the geophysics community (e.g., [88]), reasonably so due to the identification of $\mathcal{L}^2(\mathbf{R})$ with itself effected by the Fourier transform. We shall continue with this convention. We choose to forego a discussion of the physical implications of the definition, as it is beyond the scope of our present objectives. The interested reader should consult (with a grain of salt!) the appendices in Robinson’s book [77]. We mention also Futterman’s work on frequency-dependent attenuation for dispersive waves, and its connection with the minimum phase concept ([24]). The acoustic transmission response of layered media, and the seismic signature of a dynamite source, are two more examples of known minimum phase phenomena (see [78] and [16]). We should remark at this point (as suggested by Larry Lines) that not all seismic sources are minimum phase. For example, it is generally accepted that vibroseis sources are not minimum phase.

According to our definition, \hat{s} is minimum phase if and only if $1/\hat{s}$ is minimum phase. This is the main reason for demanding such generality of s . Indeed, if s is a signal, then its Fourier transform vanishes at infinity: thus $1/\hat{s}$ is unbounded, and its Fourier transform, $\mathcal{F}^{-1}\{1/\hat{s}\}$, provided it exists, cannot be a function. Quite naturally, we want to preserve the symmetry implied by the definition, namely, that s should be minimum phase if and only if $\mathcal{F}^{-1}\{1/\hat{s}\}$ is minimum phase.

3.2.1 Examples

Example The function $f(x) = h(x)e^{-ax}$, for each $a > 0$, is a minimum phase signal. Indeed, $f \in \mathcal{L}^2(\mathbf{R}) \cap \mathcal{L}^1(\mathbf{R})$, so f is a signal and its Fourier transform can be computed using formula (2.10) as

$$\hat{f}(\xi) = \int_{\mathbf{R}} h(x)e^{-x(a+i\xi)} dx = \int_0^\infty e^{-x(a+i\xi)} dx = \frac{1}{a+i\xi}, \quad (3.30)$$

provided $a > 0$. Certainly, \hat{f} and $1/\hat{f}$ are tempered functions on \mathbf{R} , and so they correspond to tempered distributions defined as in (2.17). Now \hat{f} is analytic everywhere except at the point $\xi = ai$ in the upper half plane, and $1/\hat{f}$ is entire, which proves the claim. ■

Example It is instructive to go through the case $a = 0$, namely, to show that the Heaviside function is minimum phase. As in (3.3), we have

$$\hat{h}(\omega) = \pi\delta(\omega) - \frac{i}{\omega}. \quad (3.31)$$

Now δ is supported at the origin, and so on the open lower half of the complex plane, \hat{h} agrees with the function

$$\hat{h}'(\xi) = -\frac{i}{\xi}, \quad \Im(\xi) < 0. \quad (3.32)$$

This function and its pointwise inverse are analytic for $\Im(\xi) < 0$, and the claim follows. ■

The previous example illustrates why, in our definition of minimum phase, we do not require \hat{s} or its inverse to be analytic on the real line.

Example As we have already seen, the Dirac distribution is minimum phase, since its Fourier transform is the distribution corresponding to the constant function, 1, whose extension to the whole complex plane is entire. More generally, for any complex numbers μ and $\epsilon \neq 0$, the tempered distribution $\mu + \epsilon\delta$ is minimum phase. Indeed, its Fourier transform is $2\pi\mu\delta + \epsilon$, which agrees with the constant function ϵ everywhere except at the origin of \mathbf{C} . The condition $\epsilon \neq 0$ is necessary to ensure $1/\epsilon$ is analytic. ■

Example Suppose g is a causal Schwartz function, and consider the function of frequency obtained by exponentiating its Fourier transform (recall that \hat{g} is also a Schwartz function):

$$\exp(\hat{g}(\omega)) = \exp(\Re(\hat{g}(\omega))) \exp(i \Im(\hat{g}(\omega))). \quad (3.33)$$

Since g is causal and certainly in $\mathcal{L}^1(\mathbf{R})$, Proposition 3.1.3 asserts that $\pm\hat{g}$ are analytic functions on the open lower half plane; hence so are

$$\exp(\hat{g}) \text{ and } 1/\exp(\hat{g}) = \exp(-\hat{g}), \quad (3.34)$$

since the exponential function is entire. In particular, the tempered (actually bounded) function $\exp(\hat{g})$ is the Fourier spectrum of a minimum-phase tempered distribution $T \in \mathcal{S}'(\mathbf{R})$.

By Proposition 3.1.1, since $g \in \mathcal{S}(\mathbf{R}) \subset \mathcal{L}^1(\mathbf{R})$ is causal, the real and imaginary parts of its Fourier transform are related by the Hilbert transform:

$$\Re(\hat{g}) = \mathcal{H}\Im(\hat{g}), \text{ or } \Im(\hat{g}) = -\mathcal{H}\Re(\hat{g}). \quad (3.35)$$

In particular, we see that the phase and magnitude spectra of T are related by

$$\phi_T = \Im(\hat{g}) = \mathcal{H} \ln(|\hat{T}|). \quad \blacksquare \quad (3.36)$$

Example The zero function is not minimum phase. Indeed, its Fourier transform is again zero, which is entire, but nowhere invertible. ■

3.2.2 Some characteristics of minimum phase

We should expect the property of minimum phase—for a good definition of minimum phase—to be translation invariant; that is, to be independent of the choice of origin for the time axis.

Proposition 3.2.2 *A tempered distribution f is minimum phase if and only if each of its translates $T_\tau f$, ($\tau \in \mathbf{R}$) is minimum phase.*

Proof Suppose f is minimum phase, so \hat{f} and $1/\hat{f}$ are analytic on the open lower half of the complex plane. Since $(\widehat{T_\tau f})(\omega) = \exp(i\omega\tau)\hat{f}(\omega)$, and $\exp(i\omega\tau)$ and its inverse $\exp(-i\omega\tau)$ are entire functions (that is, analytic everywhere on the complex plane), it follows that $\widehat{T_\tau f}$ and $1/\widehat{T_\tau f}$ are analytic on the open lower half plane, whence $T_\tau f$ is minimum phase for all $\tau \in \mathbf{R}$. The converse follows obviously by setting $\tau = 0$. ■

The following result offers an alternate characterization of minimum-phase tempered distributions.

Proposition 3.2.3 *A tempered distribution s is minimum phase if and only if $\ln \hat{s}$ is analytic on the open lower half of the complex plane.*

Proof The complex logarithm is an analytic function everywhere except at the origin. Thus $\ln \hat{s}$ is analytic if and only if \hat{s} is analytic and non-zero. Also, $1/\hat{s}$ is analytic if and only if \hat{s} is analytic and non-zero. It follows that s is minimum

phase if and only if $\ln \hat{s}$ is an analytic function on the open lower half of the complex plane. ■

We next recall a few preliminary results from complex analysis that will be used in the remainder of this section. The main and final result of this brief excursion is Jordan's lemma (3.2.6).

A few results from complex analysis

Theorem 3.2.4 (*Cauchy-Goursat Theorem, e.g., Churchill and Brown, 1984*) *If the function F is analytic at all points interior to and on a simple closed curve Γ , then*

$$\int_{\Gamma} F(z) dz = 0. \quad (3.37)$$

Lemma 3.2.5 (*Jordan's Inequality, adapted from Churchill and Brown, 1984*) *For each $R > 0$,*

$$\int_{-\pi/2}^0 e^{R \sin \theta} d\theta < \frac{\pi}{2R}. \quad (3.38)$$

Proof (Churchill and Brown, 1984) For $-\pi/2 \leq \theta \leq 0$, we have $0 \geq 2\theta/\pi \geq \sin \theta$. Thus, for any $R > 0$ and $\theta \in [-\pi/2, 0]$, $e^{R \sin \theta} \leq e^{2R\theta/\pi}$, so that

$$\begin{aligned} \int_{-\pi/2}^0 e^{R \sin \theta} d\theta &\leq \int_{-\pi/2}^0 e^{2R\theta/\pi} d\theta \\ &= \frac{\pi}{2R} (1 - e^{-R}) \\ &< \frac{\pi}{2R}. \quad \blacksquare \end{aligned}$$

Proposition 3.2.6 (*Jordan's Lemma, adapted from (Papoulis, 1977)*) *Let $\gamma = \gamma_R$ be the semicircle of radius R in the lower half of \mathbf{C} , centered at the origin. If $F(\omega)$*

is a measurable function such that $F(\omega) \rightarrow 0$ uniformly on γ as $R \rightarrow \infty$, then for all $t < 0$,

$$\int_{\gamma} F(\omega) e^{i\omega t} d\omega \rightarrow 0 \quad \text{as } R \rightarrow \infty. \quad (3.39)$$

Proof Let $\epsilon > 0$ be given. Since $F(\omega) \rightarrow 0$ as $R \rightarrow \infty$, (for $\Im(\omega) \leq 0$) we can choose $R_0 > 0$ such that $|F(\omega)| < \epsilon$ for $|\omega| > R_0$. With the change of variables $\omega = Re^{i\theta}$, $d\omega = iRe^{i\theta}d\theta$, we have that for all $R > R_0$ and $t < 0$,

$$\begin{aligned} \left| \int_{\Gamma} F(\omega) e^{i\omega t} d\omega \right| &\leq R \int_{-\pi}^0 \left| e^{iRt(\cos\theta + i\sin\theta)} F(Re^{i\theta}) \right| d\theta \\ &\leq \epsilon R \int_{-\pi}^0 \left| e^{-Rt(\sin\theta)} \right| d\theta \\ &< \frac{\epsilon\pi}{-t}, \end{aligned}$$

where the last inequality follows from Lemma 3.2.5, together with the fact that $\sin\theta$ is symmetric about the vertical line through $-\pi/2$. Hence the integral over $[-\pi, 0]$ is twice that over $[-\pi/2, 0]$. ■

We now in a position to establish some useful properties concerning the minimum phase concept.

Theorem 3.2.7 *If $F(\omega)$ is analytic for $\Im(\omega) \leq 0$ and satisfies the hypotheses of Proposition 3.2.6, then $F(\omega)$ is the Fourier transform of a causal tempered distribution $f \in \mathcal{S}'(\mathbf{R})$.*

Proof First, F is bounded on \mathbf{R} , since $F(\omega) \rightarrow 0$ as $R \rightarrow \infty$. Thus F is tempered, so that $F \in \mathcal{S}'(\mathbf{R})$, and we have that $f \in \mathcal{S}'(\mathbf{R})$, where f is the inverse Fourier transform of F . For each $R > 0$, let $\gamma = \gamma_R$ be the semicircle of radius R in the closed lower half of \mathbf{C} , centered at the origin, and let $\Gamma = \Gamma_R$ be the simple closed

curve connecting γ with the interval $[-R, R]$. Since $F(\omega)$ is analytic for $\Im(\omega) \leq 0$, and $e^{i\omega t}$ is entire, Cauchy's Theorem 3.2.4 asserts that

$$\int_{\Gamma} F(\omega)e^{i\omega t} d\omega = 0. \quad (3.40)$$

By Proposition 3.2.6, for each $t < 0$ the integrals $\int_{\gamma} F(\omega)e^{i\omega t} d\omega$ converge to 0 as $R \rightarrow \infty$. In particular, for each $t < 0$, the integrals defined by

$$\int_{-R}^R F(\omega)e^{i\omega t} d\omega = \int_{\Gamma} F(\omega)e^{i\omega t} d\omega - \int_{\gamma} F(\omega)e^{i\omega t} d\omega = - \int_{\gamma} F(\omega)e^{i\omega t} d\omega \quad (3.41)$$

form a bounded net, indexed by $R > 0$, which converges to 0 as $R \rightarrow \infty$. Thus, for all $t < 0$, passing to the limit as $R \rightarrow \infty$ shows that

$$\int_{\mathbf{R}} F(\omega)e^{i\omega t} d\omega = 0, \quad (3.42)$$

and it follows that the distribution f agrees with the zero function on $(-\infty, 0)$, that is, f is causal. ■

Corollary 3.2.8 *If $f \in \mathcal{S}'(\mathbf{R})$ is minimum phase and its Fourier transform, $F \in \mathcal{S}'(\mathbf{R})$, satisfies the hypotheses of Proposition 3.2.6, then $\mathcal{H}F = iF$.*

Proof Since f is minimum phase, it is analytic in the lower half plane. By hypothesis, it follows from Theorem 3.2.7 that f is causal. By definition of the Hilbert transform of a causal distribution, $\mathcal{H}F = iF$. ■

Theorem 3.2.9 *If $f \in \mathcal{S}'(\mathbf{R})$ is minimum phase, such that \hat{f} is a function on the real line, and if $\ln \hat{f}$ satisfies the hypotheses of Proposition 3.2.6, then $\mathcal{H}(\ln \hat{f}) = i \ln \hat{f}$.*

Proof By Theorem 3.2.7, $\ln \hat{f}$ corresponds to a causal tempered distribution.

Theorem 3.2.10 (*extension of Signal Front result in [70]*) *If $\hat{f} \in \mathcal{S}'(\mathbf{R})$ is analytic in the open lower half plane, and if for some $\tau \in \mathbf{R}$,*

$$e^{i\tau\xi}\hat{f}(\xi) \rightarrow 0 \text{ as } \xi \rightarrow \infty \text{ on } \gamma_R, \quad (3.43)$$

(as in Proposition 3.2.6) then $\text{Supp}(f) \subseteq [\tau, \infty)$.

Proof Consider the translate, $T_{-\tau}f$ of f by $-\tau$. Then $\{T_{-\tau}f\}^\wedge(\xi) = e^{i\tau\xi}\hat{f}(\xi)$ is analytic in the open lower half plane and satisfies the hypotheses of Proposition 3.2.6. Thus, by Theorem 3.2.7, $\text{Supp}(f) \subseteq [0, \infty)$, and the result follows.

Definition (*extension of Signal Front concept [70] to tempered distributions*) Given $f \in \mathcal{S}'(\mathbf{R})$ satisfying (3.43) for some $\tau \in \mathbf{R}$, the supremum, τ_0 , over the set of all τ satisfying (3.43) is called the signal front of f .

Of course, f need not be a signal in the classical sense. Also, $\tau_0 = \infty$ if and only if $f \equiv 0$.

3.3 Nonstationary minimum phase filters

Definition A linear operator L is minimum phase if it preserves minimum phase; that is, if for each minimum phase tempered distribution $s \in \mathcal{S}'(\mathbf{R})$ for which Ls defines a tempered distribution, Ls is again minimum phase.

The domain of the operator L is intentionally left unspecified, since it depends on the regularity of the Schwartz kernel of L . We thus consider L as a mapping from the preimage, $L^{-1}\{\mathcal{S}'(\mathbf{R})\}$, of $\mathcal{S}'(\mathbf{R})$ into $\mathcal{S}'(\mathbf{R})$.

Example Given any minimum phase Schwartz class function $f \in \mathcal{S}(\mathbf{R})$, consider the convolution operator

$$L_f : \mathcal{S}'(\mathbf{R}) \rightarrow \mathcal{S}'(\mathbf{R}) \quad (3.44)$$

defined as

$$L_f s = f * s. \quad (3.45)$$

The product $\hat{f}(\xi)\hat{s}(\xi)$ and its pointwise inverse are analytic functions in the open lower half plane, since this is true of both \hat{f} and \hat{s} by assumption. In particular, $L_f s \in \mathcal{S}'(\mathbf{R})$ is minimum phase, and so L_f is a minimum phase linear operator.

If the kernel f were in $\mathcal{L}^1(\mathbf{R})$, then we would have a bounded (continuous) minimum phase linear operator on $\mathcal{L}^2(\mathbf{R})$. Since $\mathcal{L}^2(\mathbf{R}) \subset \mathcal{S}'(\mathbf{R})$, L_f can be viewed—in a trivial way—as an operator $L_f : \mathcal{L}^2(\mathbf{R}) \rightarrow \mathcal{S}'(\mathbf{R})$. ■

The convolution operators are commonly referred to as “filters” in signal processing and imaging. They represent time-invariant, or stationary, linear systems. Of much greater interest to us in our research is *nonstationary* filtering. Indeed many, if not all, linear physical systems exhibit nonstationarity and thus cannot properly be described by convolution operators.

To this end, consider a collection

$$f = \{f_\tau : \tau \in \mathbf{R}\} \quad (3.46)$$

of minimum phase tempered distributions, $f_\tau \in \mathcal{S}'(\mathbf{R})$, parameterized by $\tau \in \mathbf{R}$. This just describes a function $f = f(\tau, t)$ of two variables whenever each member of f is itself a function. Alternatively, we can interpret the Fourier transform of f as a family of filter kernels, $\hat{f} = \{\hat{f}_\tau : \tau \in \mathbf{R}\}$, also parameterized by $\tau \in \mathbf{R}$.

Specifically, the action of the resulting family of filters can be represented (at least formally) by the following inverse Fourier transform (see Kohn and Nirenberg, 1965, and Stein, 1993):

$$(L_f s)(\tau) = \frac{1}{2\pi} \int_{\mathbf{R}} \hat{f}(\tau, \omega) \hat{s}(\omega) e^{i\tau\omega} d\omega, \quad \text{where } s \in L_f^{-1}\{\mathcal{S}'(\mathbf{R})\}. \quad (3.47)$$

For each fixed τ , and for each minimum phase $s \in L_f^{-1}\{\mathcal{S}'(\mathbf{R})\}$, the product $\hat{f}_\tau \hat{s}$ and its pointwise inverse are analytic functions in the open lower half of the complex plane, and thus represent (via an inverse Fourier transform) a minimum phase tempered distribution: $L_f s$. Thus L_f defines a nonstationary, minimum phase linear operator.

Chapter 4

Adaptive time-frequency analysis with nonuniform Gabor frames

Mathematical rigor...all too often leads to rigor mortis.

—*R. W. Hamming*

Yes indeed, but rigor mortis is also a necessary condition for reincarnation. We introduce here a new method for constructing adaptive, nonuniform Gabor frames. The adaptive frame results are subsequently utilized in Chapter 7, where we introduce the new “adaptive Gabor phase-shift” (AGPS) algorithm for wavefield extrapolation. The motivation for the construction is to minimize the redundancy of the Gabor transform, and hence obtain a potentially significant decrease in computational intensity.

We begin by fixing some notation and introducing the relevant background material from Gabor analysis. This is followed by a discussion of Gabor frames, culminating in a criterion (Theorem 4.2.1) for generating such frames. The remainder of the chapter is mainly concerned with the construction of a one-parameter family of Gabor frames from a maximally redundant uniform partition of unity. These preliminaries set the stage for our ultimate objective of constructing a family of adaptive Gabor frames.

4.1 Preliminaries

4.1.1 Partitions of unity

A partition of unity is a collection $\{\psi_j : j \in \mathbf{Z}\}$ of functions that sums to 1:

$$\sum_{j \in \mathbf{Z}} \psi_j(x) = 1, \text{ for each } x \in \mathbf{R}. \quad (4.1)$$

With respect to such a partition of unity, any function can be represented as the superposition of its windowed components,

$$f_j(x) = \psi_j(x)f(x), \quad (4.2)$$

as follows:

$$f(x) = \sum_{j \in \mathbf{Z}} \psi_j(x)f(x) = \sum_{j \in \mathbf{Z}} f_j(x). \quad (4.3)$$

Given any collection of functions, $\{\phi_j\}$ satisfying

$$0 < A \leq \sum_{j \in \mathbf{Z}} \phi_j(x) \leq B < \infty, \text{ (for each } x \in \mathbf{R}), \quad (4.4)$$

for some constants, $A, B > 0$, the collection can always be normalized to give a partition of unity. Indeed, by setting

$$\psi_j(x) = \frac{1}{\sum_{i \in \mathbf{Z}} \phi_i(x)} \phi_j(x), \quad (4.5)$$

we have

$$\sum_{j \in \mathbf{Z}} \psi_j(x) = 1, \text{ for each } x \in \mathbf{R}. \quad (4.6)$$

4.1.2 Partitions of unity in higher dimensions

A d -dimensional partition of unity can easily be constructed from any set of d one-dimensional partitions of unity as follows. We show how to do this in two dimensions, since the procedure is similar in any number of dimensions.

Given a pair of one-dimensional partitions of unity, say $\{\phi_n\}$ and $\{\psi_m\}$, consider the product (actually, the tensor product—see the next paragraph) $\{\gamma_{m,n} = \phi_n \psi_m\}$. By assumption, we have

$$\sum_{n \in \mathbf{Z}} \phi_n = \sum_{m \in \mathbf{Z}} \psi_m = 1, \quad (4.7)$$

so that

$$\sum_{m,n \in \mathbf{Z}} \gamma_{m,n} = \sum_{n,m \in \mathbf{Z}} \phi_n \psi_m = \sum_{n \in \mathbf{Z}} \phi_n \sum_{m \in \mathbf{Z}} \psi_m = 1. \quad (4.8)$$

Thus $\{\gamma_{m,n}\}$ forms a partition of unity.

A partition of unity constructed in this fashion will not generally inherit the regularity properties of its factors. If an approximate partition of unity is acceptable, however, the product of two approximate partitions of unity consisting of Gaussians will always yield another approximate partition of unity, again consisting of Gaussians ([5]). The tensor product, $g \otimes g$; that is,

$$(g \otimes g)(x, y) = g(x)g(y), \quad (4.9)$$

of two identical Gaussians has the further property of being radially symmetric.

4.1.3 Translation and modulation operators in Hilbert space

Given a (possibly complex-valued) function $g(x)$ on the real line, and constants $a, b \in \mathbf{Z}$, the translation operator, T_{na} , and the modulation operator, M_{mb} are defined by

$$(T_{na}f)(x) = f(x - na) \text{ and } (M_{mb}f)(x) = e^{2\pi imbx} f(x) \quad (4.10)$$

for any $f \in \mathcal{L}^2(\mathbf{R}^d)$. Here, $\mathcal{L}^2(\mathbf{R}^d)$ is the Hilbert space of square-integrable functions, with inner product

$$\langle f, h \rangle = \int_{\mathbf{R}^d} f(x) \overline{h(x)} dx \quad (4.11)$$

and norm (or energy)

$$\|f\| = \langle f, f \rangle^{1/2} = \left(\int_{\mathbf{R}^d} |f(x)|^2 dx \right)^{1/2}. \quad (4.12)$$

A function belongs to $\mathcal{L}^2(\mathbf{R}^d)$ if and only if it has finite measurable energy: $\|f\| < \infty$. Functions in $\mathcal{L}^2(\mathbf{R}^d)$ will be referred to as signals, (where $d = 1$) although the theory we develop applies equally well to images ($d = 1, 2$, or 3). In fact, the generalization to any dimension is straightforward, but we consider only a single dimension for simplicity. So the real variable x represents time or space, depending on the desired application, and the corresponding Fourier dual variable, ξ , accordingly represents temporal or spatial frequency. The time-frequency plane is the continuum of all pairs (x, ξ) ; in quantum mechanics, these are the position and momentum coordinates of phase space \mathbf{R}^{2d} .

We consider also the Hilbert space $\ell^2(\mathbf{Z} \times \mathbf{Z})$, consisting of square-summable sequences of complex numbers, $\{c_{m,n}\}$, indexed by the integers $m, n \in \mathbf{Z}$. A sequence belongs to $\ell^2(\mathbf{Z} \times \mathbf{Z})$ if and only if

$$\sum_{m,n \in \mathbf{Z}} |c_{m,n}|^2 < \infty. \quad (4.13)$$

4.2 The Gabor transform

4.2.1 The short time Fourier transform

Further details pertaining to the current section may be found in [36], [20], [73], and [25].

The continuous Gabor transform is an invertible linear mapping from $\mathcal{L}^2(\mathbf{R}^d)$ into $\mathcal{L}^2(\mathbf{R}^{2d})$, defined as follows: fix a nonzero function g , called a window function, or atom. The continuous Gabor transform of a signal $f \in \mathcal{L}^2(\mathbf{R}^d)$ with respect to the window g is then denoted as

$$V_g f(x, \omega) = \int_{\mathbf{R}^d} f(t) \overline{g(t-x)} e^{-2\pi i t \omega} dt. \quad (4.14)$$

Here, the overline denotes complex conjugation. In our applications, g is usually a symmetric, real-valued function (usually a Gaussian window, or some finitely supported bump function). The continuous Gabor transform is also known as the windowed Fourier transform, or short-time Fourier transform, for obvious reasons.

4.2.2 Gabor frames and the discrete Gabor transform

The discrete Gabor transform is the partially discretized version of the windowed Fourier transform, expressed as

$$V_g f(n, m) = \int_{\mathbf{R}^d} f(t) \overline{g(t - na)} e^{-2\pi i m b t} dt, \quad (4.15)$$

where a and b are fixed positive constants, and n and m are d -tuples of integers. This is just a sampled version of the windowed Fourier transform. The constant a controls the spacing between windows, while b determines the sample rate for the Fourier dual variable.

The range of the Gabor transform is contained in $\ell^2(\mathbf{Z}^{2n})$, the space of complex-valued, square-summable sequences $\{c_{n,m}\}$ for which

$$\sum_{n,m \in \mathbf{Z}^n} |c_{n,m}|^2 < \infty. \quad (4.16)$$

We often use the shorthand notation, $g_{n,m}(x) = g(t - na)e^{-2\pi i m b t}$, for the so-called translated and modulated atoms. Given a triple $\{g, a, b\}$ the corresponding collection $\{g_{n,m}\}$ is called a Gabor frame if there exist constants $A, B > 0$ for which

$$A\|f\|^2 \leq \sum_{n,m \in \mathbf{Z}^n} |\langle f, g_{n,m} \rangle|^2 \leq B\|f\|^2, \quad (f \in \mathcal{L}^2(\mathbf{R}^d)). \quad (4.17)$$

A and B are called frame bounds, and a Gabor frame is called a tight frame if $A = B$. Any orthonormal basis of $\mathcal{L}^2(\mathbf{R}^d)$, for example, is a tight frame, with frame bounds $A = B = 1$.

Any Gabor frame $\{g_{n,m}\}$ has the property that the Gabor transform, or analysis mapping,

$$V_g : \mathcal{L}^2(\mathbf{R}^d) \rightarrow \ell^2(\mathbf{Z}^n \times \mathbf{Z}^n), \quad (4.18)$$

$$(V_g f)(n, m) = \langle f, g_{n,m} \rangle = \int_{\mathbf{R}^d} f(t) g(t - na) e^{-2\pi i m b t} dt, \quad (4.19)$$

and its adjoint, the synthesis mapping, or Gabor expansion,

$$V_g^* : \ell^2(\mathbf{Z}^n \times \mathbf{Z}^n) \rightarrow \mathcal{L}^2(\mathbf{R}^{2n}) \quad (4.20)$$

given by

$$V_g^* \{c_{n,m}\}(t) = \sum_{n,m \in \mathbf{Z}^n} c_{n,m} g_{n,m}(t), \quad (4.21)$$

are bounded linear operators. Consequently, the Gabor frame operator, S_g , defined by

$$S_g = V_g^* V_g : \mathcal{L}^2(\mathbf{R}^d) \rightarrow \mathcal{L}^2(\mathbf{R}^d) \quad (4.22)$$

$$S_g f = \sum_{n,m \in \mathbf{Z}^n} \langle f, g_{n,m} \rangle g_{n,m}, \quad (4.23)$$

being a composition of bounded linear operators, is itself a bounded linear operator. Given any Gabor frame, $\{g_{n,m}\}$, the associated frame operator (4.23) is always invertible: indeed, as implied by condition (4.17), S_g is bounded away from zero, meaning S_g is injective, hence invertible on its range. The most important consequence of this fact is that any $f \in \mathcal{L}^2(\mathbf{R}^d)$ can be reconstructed from its Gabor coefficients as

$$f = \sum_{n,m \in \mathbf{Z}^n} \langle f, g_{n,m} \rangle S_g^{-1} g_{n,m}, \quad (4.24)$$

which follows from applying S_g^{-1} to both sides of equation (4.23). The resulting system, $\{\gamma_{n,m} = S_g^{-1} g_{n,m}\}$, is also a Gabor frame, called the dual frame. It is generated by translations and modulations of the dual atom, or synthesis window, $\gamma = S_g^{-1} g$. The dual frame is a synthesis frame: it has the property that $V_\gamma^* V_g = I$, as one can easily check from the definitions and expression 4.24. The choice of synthesis window is not unique in general. Many other windows can be substituted

for γ that will allow reconstruction of the signal from its Gabor coefficients, but the dual atom happens to be the synthesis window of least energy.

The Gabor frame is commonly called an analysis frame, while its dual is called a synthesis frame. Precise information about how to choose analysis and synthesis frames will be discussed in the next section, (also in Grossman, et al., [40]) where 1-parameter families of Gabor frames will be constructed from any partition of unity.

Figure 4.1 shows an example of a Gabor amplitude spectrum, (magnitude of the Gabor transform) where the analysis frame is an approximate partition of unity consisting of Gaussian windows, and the corresponding synthesis frame is unity. The left side is a synthetic seismic trace modelled with constant-Q attenuation, and the bottom its Fourier amplitude spectrum. The Gabor amplitude spectrum clearly demonstrates the continuous amplitude decay and loss of bandwidth with time.

There has been a great deal of research on the existence of Gabor frames, particularly in finding sufficient conditions on g , a , and b for which the frame inequalities (4.17) can be satisfied (see [36] and [20]). However, as the following theorem will show, this question of existence is not an issue in the applied setting.

Theorem 4.2.1 (Daubechies et al., [14]) *Suppose that $g \in \mathcal{L}^\infty(\mathbf{R}^n)$ is supported on the cube $Q_L = [0, L]^n$. If $a \leq L$ and $b \leq 1/L$, then the frame operator is the multiplication operator*

$$Sf(x) = \left(\frac{1}{b^n} \sum_{m,n \in \mathbf{Z}} |g_n(x)|^2 \right) f(x), \quad (4.25)$$

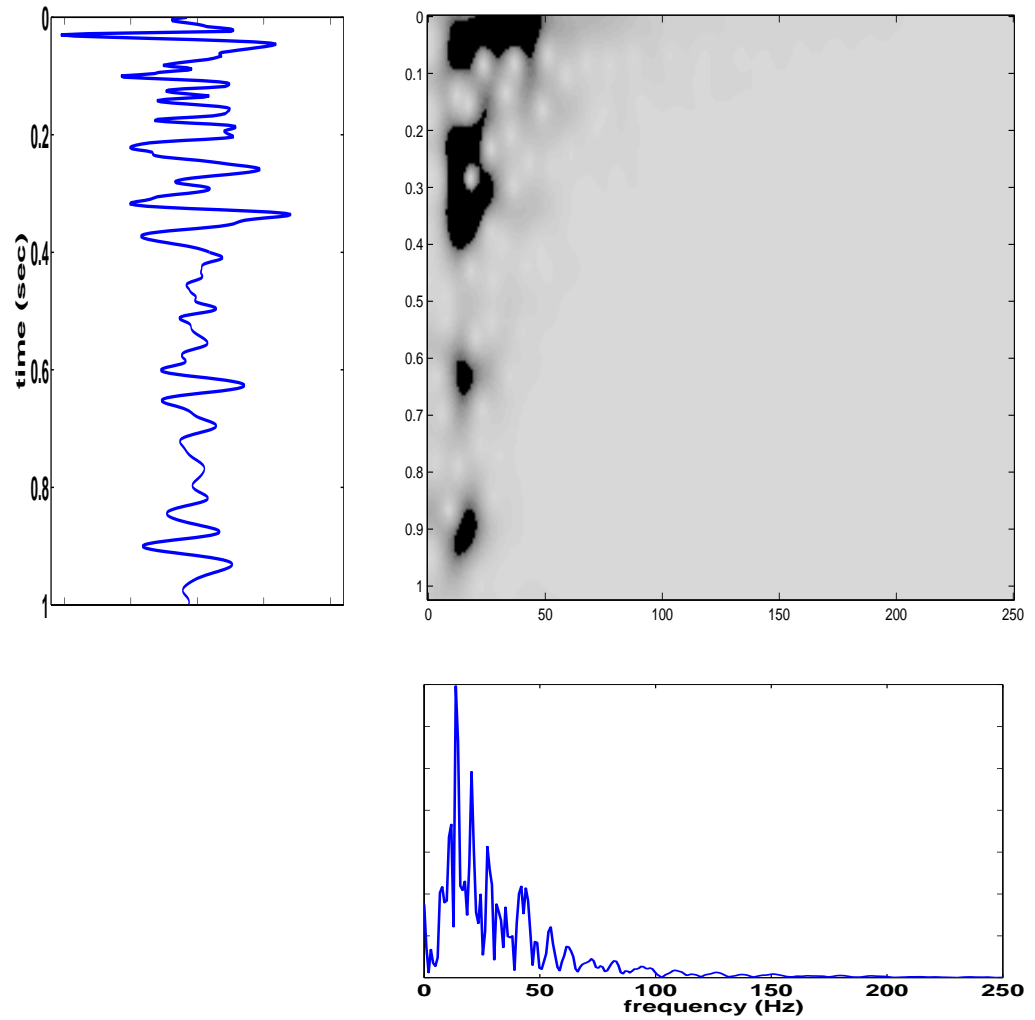


Figure 4.1: Example of a Gabor amplitude spectrum, top-right. The input time series on the left is a constant-Q-attenuated synthetic seismic trace. Its Fourier amplitude spectrum is depicted below the Gabor amplitude spectrum. The time series and its amplitude spectrum can be thought of as projections of the Gabor spectrum onto the time and frequency axes, respectively.

where $g_n(x) = (T_{na}g)(x)$. Consequently, $\{g_{m,n}\}$ is a Gabor frame, with frame bounds $A = \alpha/a^n$ and $B = \beta/b^n$, if and only if

$$0 < \alpha \leq \sum_{n \in \mathbf{Z}} |g_n(x)|^2 \leq \beta \text{ a.e.} \quad (4.26)$$

Moreover, $\{g_{m,n}\}$ is a tight frame if and only if $\sum_{n \in \mathbf{Z}} |g_n(x)|^2 = \text{constant a.e.}$

The sufficient conditions for $g_{m,n}$ to be a frame are easily satisfied in practice. Indeed, any continuous function g on a closed interval is bounded, so it belongs to $\mathcal{L}^\infty(\mathbf{R}^d)$. Also, the distance a between window centers cannot exceed the width, L , of the window; otherwise, part of the signal being analyzed would be neglected. Moreover, since we will be using the FFT (fast Fourier transform), we have that $b = 1/L$. These remarks, and the criterion (4.26), will be further discussed below.

4.2.3 Implementation of Gabor transforms with FFTs

In applications, efficiency is gained by implementing the Gabor transform with the FFT. This specializes the definition of the Gabor transform—to the extent that the sampling interval, b , (of the frequency coordinate) is predetermined by the FFT and the window's length.

4.2.4 Window selection

As we have mentioned, the Gaussian window is optimal in the sense that it achieves the best possible time-frequency resolution. For numerical efficiency, however, we recommend using a finite-length window that has a similar shape to the Gaussian. The Gaussian decays to below machine precision after only several halfwidths, so a suitably truncated Gaussian could be used. The window should also be sufficiently

smooth to avoid artifacts such as spectral ringing; and it should be at least wide enough so that its sampled version faithfully represents it.

The Gaussian, as promising as it may seem, is symmetric and hence zero phase. It is conceivable that there may be situations for which non-zero phase windows would be advantageous. For example, as we shall see when we introduce Gabor filters below, locally minimum phase filters might be more accurately implemented with locally minimum phase analysis and synthesis frames.

4.3 Nonuniform frames from maximal partition-of-unity frames

4.3.1 Maximally redundant partition-of-unity Gabor frames

Consider any signal $f \in \mathcal{L}^2(\mathbf{R})$, and select a window $\tilde{g} \geq 0$ according to the prescription above (see subsection 4.2.4). Suppose the signal is sampled at M points, with a uniform sample spacing of Δx units, and that \tilde{g} has N points, with the same sample spacing. Then the sampling interval in the frequency domain is given by

$$b = \frac{2}{N-1} f_{\text{Nyquist}} = \frac{1}{\Delta x(N-1)}. \quad (4.27)$$

Next, consider the family of translates:

$$\{\tilde{g}_n = T_{na}\tilde{g} : n = 0, 1, 2, \dots, M-1\}, \quad (4.28)$$

with $a = \Delta x$. This is a suite of windows, one centered at each sample point. By construction, it satisfies the inequalities (4.4). Thus, by normalizing if necessary,

we can replace the family with a partition of unity, namely:

$$\{g_n = T_{na}\tilde{g} : n = 0, 1, 2, \dots, M - 1\}, \text{ where } g_n = \frac{\tilde{g}_n}{\sum_{m=0}^{M-1} \tilde{g}_m}. \quad (4.29)$$

Remark Although this defines a partition of unity, which would be uniform on the whole real line, it will not be uniform near the edges of the finite-length domain, $[x_0, x_{M-1}]$, because the sum is over a finite set. One way around this problem is to pad the input data with zeros, to displace these edge effects away from the domain of interest.

The area of a cell in the corresponding sampling grid in the time-frequency plane is thus

$$ab = \frac{1}{N - 1}. \quad (4.30)$$

In this case, we chose the distance a between windows to be Δx , but in general, we would have

$$\Delta x \leq a \leq (N - 1)\Delta x. \quad (4.31)$$

Multiplying (4.31) by expression (4.27) for b , we see that the area, ab is bounded by

$$\frac{1}{N - 1} \leq ab \leq 1. \quad (4.32)$$

The redundancy is defined as $1/ab$, and the system is said to be oversampled by a factor of $1/ab$, or, in this case, $N - 1$. This is the maximum possible oversampling rate, and for this reason we call the partition of unity maximally redundant. Evidently, the hypotheses of the first part of Theorem 4.2.1 are satisfied for this maximally redundant system (4.29). To prove that our system is in fact a Gabor frame, it remains to verify condition (4.26). First, $|g(x)| \leq 1$, and so we have that

$[g(x)]^2 \leq g(x)$. Thus

$$\sum_{n=1}^M |g_n(x)|^2 \leq \sum_{n=1}^M |g_n(x)| = 1 = \beta. \quad (4.33)$$

For the lower bound, we need only check that $\sum_{n=1}^M |g_n(x)|^2 \neq 0$ for all x . This is clear, since otherwise there would be some $x_0 \in \mathbf{R}$, for which

$$|g_n(x_0)|^2 = 0 = g_n(x_0) \quad (4.34)$$

for all n , implying the following contradiction:

$$1 = \sum_{n=1}^M |g_n(x_0)| = 0. \quad (4.35)$$

Thus (4.29) represents a maximally redundant partition-of-unity Gabor frame.

4.3.2 One-parameter families of Gabor frames

Theorem 4.3.1 *Suppose*

$$\{\phi_n = T_{na}\phi : n \in \mathbf{Z}\} \quad (4.36)$$

is a partition of unity such that $\phi_n \geq 0$, and choose any $p \in [0, 1]$. If we set $g(x) = [\phi(x)]^p$, then the collection

$$\{g_{m,n} = T_{na}M_{mb}g : m, n \in \mathbf{Z}\} \quad (4.37)$$

evidently defines a Gabor analysis frame. We claim that a corresponding synthesis frame is given by

$$\{\gamma_{m,n} = T_{na}M_{mb}\gamma : m, n \in \mathbf{Z}\} \quad (4.38)$$

where γ is defined as $\gamma(x) = [\phi(x)]^{1-p}$.

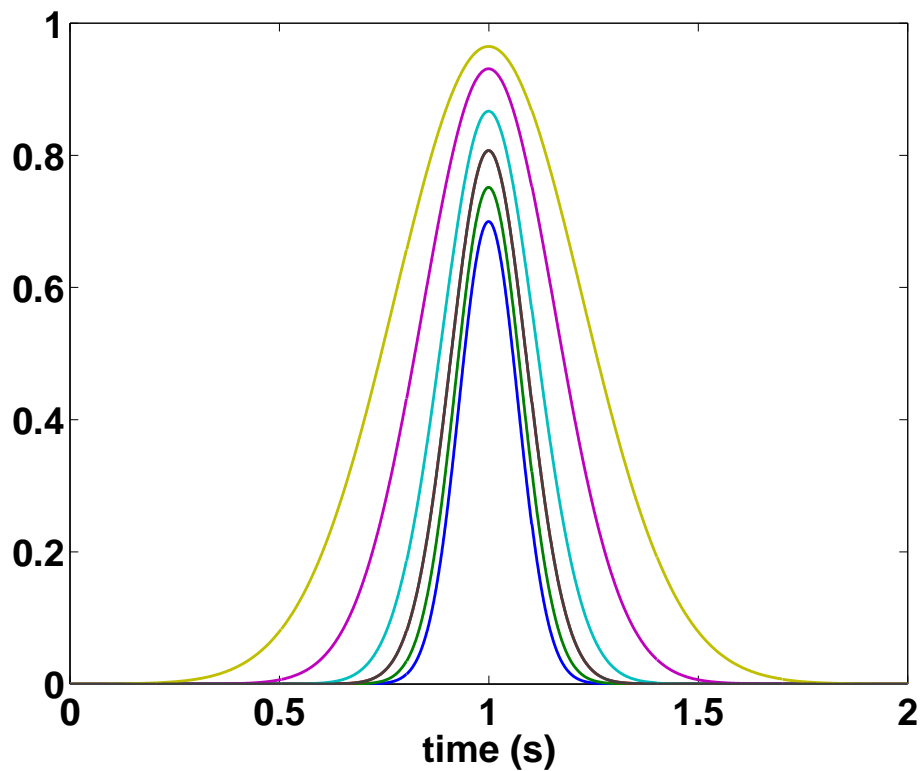


Figure 4.2: A Gaussian atom of halfwidth σ , and the family of Gaussians, of halfwidth σ/\sqrt{p} , obtained by raising the initial Gaussian to the power $p \in [0, 1]$. As p decreases, the atoms gain mass, and their mass distribution spreads. In the limit as $p \rightarrow 0$ the family converges to constant function, 1.

Proof Indeed, let $f \in \mathcal{L}^2(\mathbf{R}^d)$ be given, and consider the Gabor coefficients of f , namely:

$$c_{m,n} = \int f(\tilde{x})g(\tilde{x} - na)e^{-2\pi imb\tilde{x}} d\tilde{x}. \quad (4.39)$$

Summing these coefficients against the system (4.38) yields:

$$\begin{aligned} & \sum_{m,n \in \mathbf{Z}} c_{m,n} \gamma(x - na) e^{2\pi imb x} = \dots \\ \dots &= \sum_{m,n \in \mathbf{Z}} \int f(\tilde{x})g(\tilde{x} - na)e^{-2\pi imb\tilde{x}} d\tilde{x} \gamma(x - na) e^{2\pi imb x} \\ &= \int f(\tilde{x}) \sum_{n \in \mathbf{Z}} g(\tilde{x} - na) \gamma(x - na) \sum_{m \in \mathbf{Z}} e^{-2\pi imb(x - \tilde{x})} d\tilde{x} \\ &= \int f(\tilde{x}) \sum_{n \in \mathbf{Z}} g(\tilde{x} - na) \gamma(x - na) \delta(x - \tilde{x}) d\tilde{x} \\ &= f(x) \sum_{n \in \mathbf{Z}} g(x - na) \gamma(x - na) \\ &= f(x) \sum_{n \in \mathbf{Z}} \phi^p(x - na) \phi^{1-p}(x - na) \\ &= f(x), \end{aligned}$$

which is what we set out to prove, i.e., that (4.37) and (4.38) form a dual pair of Gabor frames. \blacksquare

We have used the fact that the sum, $\sum_{m \in \mathbf{Z}} e^{-2\pi imb(x - \tilde{x})}$ under the integral in (4.40), has the same action as the Dirac-delta distribution concentrated at $\tilde{x} = x$. Indeed, the sum represents the inverse DFT, evaluated at $x - \tilde{x}$, of the constant function, 1.

Remark The choice of p assigns a relative balance between the mass distributions of the analysis and synthesis atoms. Since the atoms are non-negative members of a partition of unity, their amplitudes are bounded above by one. Thus, as the exponent $p \in [0, 1]$ decreases, the atomic mass increases and spreads

simultaneously.

From Theorem 4.2.1, it is easy to see that for $p = 1/2$, the resulting frame is tight; that the converse also holds in this setting remains a conjecture. Evidently, this is also the only p for which the analysis and synthesis frames are identical.

In the case of a Gaussian of halfwidth σ , it is easily seen that exponentiation by p results in a new Gaussian, of halfwidth σ/\sqrt{p} . A suite of such Gaussians, parameterized by $p \in [0, 1]$, is displayed in figure 4.2. In the limit as $p \rightarrow 0$, the family degenerates to the constant function, $g(x) = 1$.

As we have mentioned, the Gaussian can only form an approximate partition of unity. However, the error can be made as small as desired by increasing the ratio σ/a of the halfwidth to the spacing between translates. Margrave and Lamoureux [58] showed that this error is dominated by a sinusoid with amplitude that decreases exponentially with the ratio σ/a . Practically speaking, the error can be made negligible: for example, by setting the spacing between Gaussians to about 10 percent of the halfwidth. Alternatively, the suite of Gaussians can be normalized, as in (4.5), at the cost of a deviation from optimal time-frequency resolution.

In the case of a compactly supported atom, for example a Lamoureux window of order k , (a bell-shaped polynomial spline of order C^k) a little caution is required. Figure 4.3 displays a suite of atoms generated by exponentiating a Lamoureux window of order four. As is evident by inspection of the endpoints of the windows, the order of differentiability of the resulting atom decreases as the exponent $p \in [0, 1]$ decreases (any $p \leq 1/k$). However, in practice this is easily resolved by suitably adjusting the order of the Lamoureux window for a given choice of p .

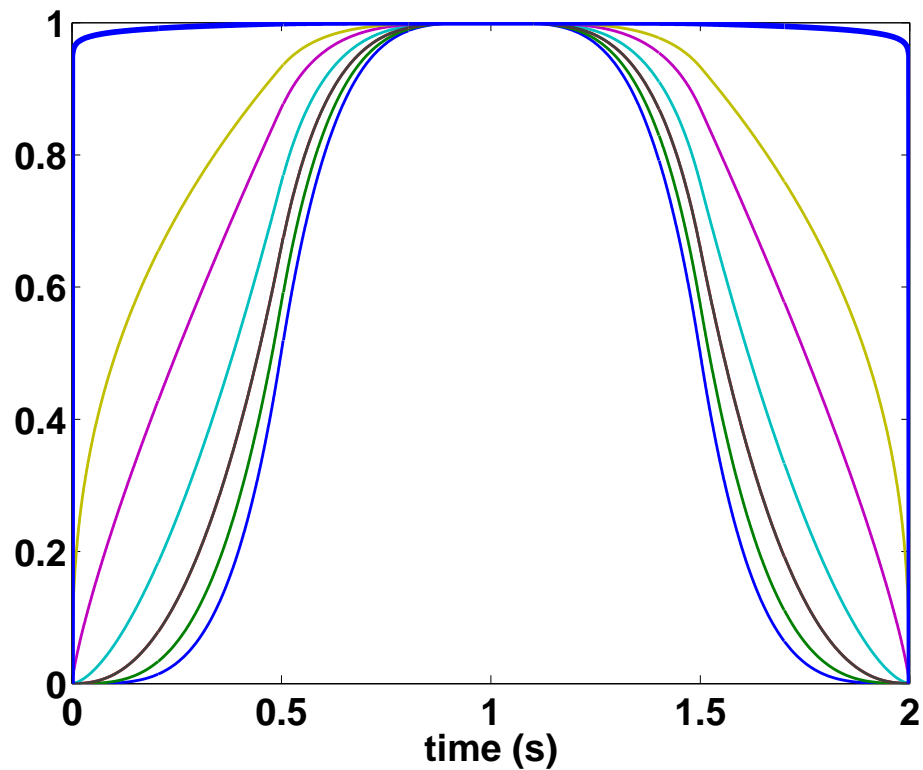


Figure 4.3: A Lamoureux atom and the family of windows obtained by raising the initial atom to the power $p \in [0, 1]$. As p decreases, the atoms gain mass and their spreading mass-distributions are confined to the original domain. At $p = 1/4$, the window is no longer differentiable at the two boundaries. However, this can be resolved for any given p by starting with a sufficiently high-order atom. In the limit as $p \rightarrow 0$, the family converges to a boxcar.

The Lamoureux window has the advantage of being compactly supported, in contrast to the Gaussian window. Although the Gaussian is effectively compact when digitized, the extent to which it is so depends on the available dynamic range of the computing system.

4.4 Adaptive, nonuniform Gabor frames

Any maximal partition of unity can be adapted to yield a nonuniform partition of unity, in such a way that the essential nonstationarity of the particular problem is respected. The idea is to form molecules, or macro-windows, by summing neighbouring atoms over regions that meet a local stationarity measure. The result is an adaptive, nonuniform partition of unity, or molecular decomposition. These molecules can then be used to generate a one-parameter family of nonuniform Gabor analysis and synthesis frames, in precisely the same way that they were generated above from atomic partitions of unity (equations (4.37) and (4.38)).

For example, a stationarity measure can be specified by a fixed, user-assigned threshold, against which the deviation of a given velocity field from its mean over the current molecule is compared. Thus, if this deviation is less than the threshold, then this region is deemed approximately stationary, and the current atom is conjoined to the current molecule (the first molecule starts off as an atom). Thus each molecule continues to grow by gathering neighbouring atoms until it encounters a large enough velocity anomaly; then a new molecule is constructed, and so on until the atoms are exhausted.

A 1-D example of such a molecular decomposition is shown in figure 4.4. The illustrated velocity model is meant to be as general as possible, ranging from a simple constant to a nowhere-continuous random function. As desired, the atoms cluster near the larger local variations in the velocity. Everywhere else, the atoms ‘bond’ to form molecules, and their size varies inversely with the magnitude of the local variation of v . In addition, the molecules sum to unity, as they should. The

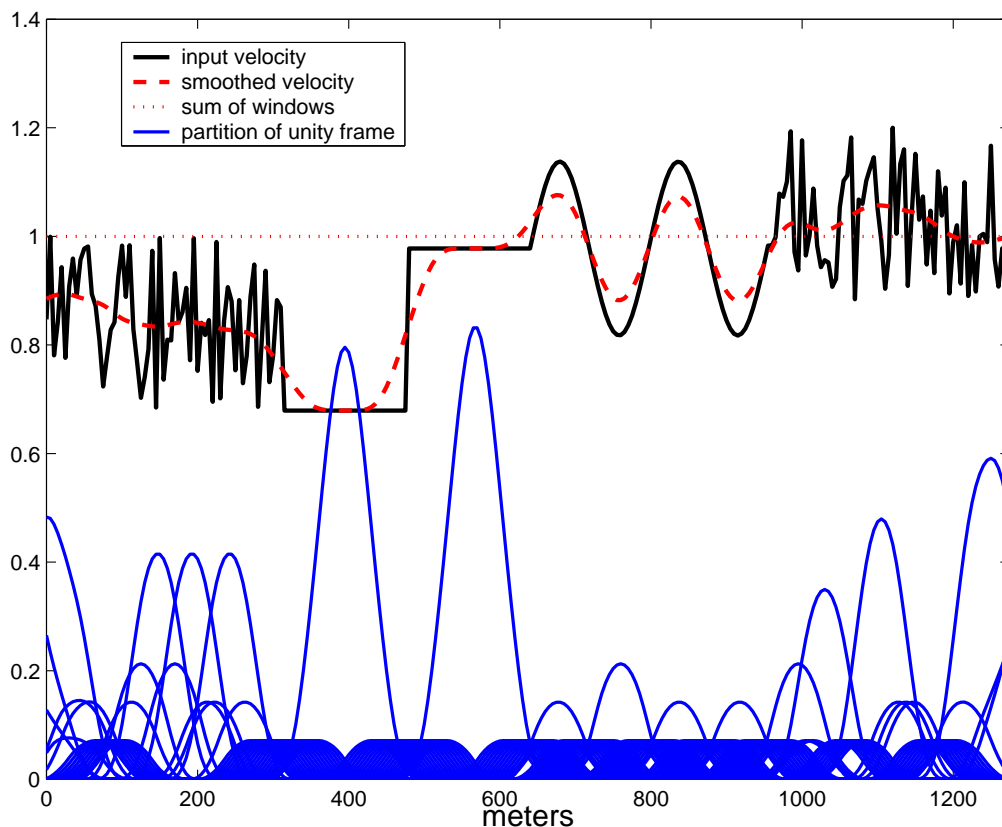


Figure 4.4: A demonstration of how an adaptive, nonuniform partition of unity Gabor frame (lower, solid) can be obtained from a maximal partition of unity, (the result of translating any of the smallest windows to each offset sample-point) which reflects the local nonstationarity of a velocity model. The piecewise-defined velocity model (upper, solid) represents all types of velocity variation: random, constant, a jump discontinuity between two constants, and a transition from smooth to random. The smoothed velocity, (dashed) obtained by convolving the velocity model with a fundamental atom, (any one of the smallest windows) serves as a visual aid in comparing the molecules to the local velocity variations. Specifically, where there is more variation in the velocity function, there are correspondingly more molecules, and conversely.

amount of molecular bonding is controlled by the degree of stationarity as well as the choice of thresholding parameter, the latter being user-defined.

Chapter 5

Theory of Gabor filtering

The more precisely the position is determined,
the less precisely the momentum is known.¹
— *W. Heisenberg*

We begin by developing an approach to time-variant filtering from a signal processing perspective using Gabor multipliers. This provides a particular way to generalize linear time-invariant filters to the nonstationary setting. Essentially, this is accomplished by localizing the Fourier transform in a time-variant fashion, and hence by the use of the Gabor transform.

Another way to generalize linear time-invariant filters to the nonstationary setting is to consider the class of linear operators corresponding to linear partial differential equations with non-constant coefficients. More generally, we shall consider nonstationary filters within the framework of the pseudodifferential and

¹The uncertainty principle of Werner Heisenberg, (1901-1976) founder of quantum mechanics.

Fourier integral operator theories that arose from this point of view in the first place.

On the surface, these two approaches to nonstationary filtering are quite distinct: the first is guided by intuition in a signal processing context,² while the second is based on hard mathematical physics. Which is the better approach? As we shall see, the resulting operators are very closely related, and for all practical purposes they are virtually identical.

5.1 Gabor filters

5.1.1 The analog theory

By a Gabor filter function, we mean any complex-valued, finite-energy function $F(t, \omega)$ on the time-frequency plane, or phase space, \mathbf{R}^{2d} , that is,

$$F \in \mathcal{L}^2(\mathbf{R}^{2d}) = \mathcal{L}^2(\mathbf{R}^d \times \mathbf{R}^d). \quad (5.1)$$

Given such a function F , we define the action of the corresponding Gabor filter on a signal $s \in \mathcal{L}^2(\mathbf{R}^d)$ by

$$\tilde{s}(t) = V_\gamma^*[F(x, \omega)V_g s(x, \omega)](t), \quad (5.2)$$

where \tilde{s} denotes the filtered output signal. Since V_γ^* maps $\mathcal{L}^2(\mathbf{R}^{2d})$ back into $\mathcal{L}^2(\mathbf{R}^d)$, we see that the filtered output of the signal is again a signal.

This restriction of F to $\mathcal{L}^2(\mathbf{R}^{2d})$ is not necessary to ensure that the filtered output of a signal is a signal. Indeed, we can take F to be any function that leaves

²This is but one approach to Gabor analysis. Essentially the same theory comes from quantum mechanics, where what we call a ‘‘Gabor frame’’ is called a ‘‘Weyl-Heisenberg frame’’ (see [36] and [105]).

$\mathcal{L}^2(\mathbf{R}^{2d})$ invariant under pointwise multiplication; that is, any F for which

$$F \cdot S \in \mathcal{L}^2(\mathbf{R}^{2d}) \quad (5.3)$$

for all $S \in \mathcal{L}^2(\mathbf{R}^{2d})$. Thus, for example, F can be any measurable, essentially bounded function on \mathbf{R}^{2d} . In practice, we might only be dealing with a small subset of $\mathcal{L}^2(\mathbf{R}^{2d})$, which would admit much more general energy-preserving Gabor filters.

A Gabor multiplier, in our sense, (see [33]) is the operator

$$M_F^{g,\gamma} = V_\gamma^* N_F V_g, \quad (5.4)$$

associated with the triple $\{g, \gamma, F\}$, where N_F denotes the linear operation of pointwise multiplication by F . A Gabor multiplier is then just the linear operator associated with the action of a Gabor filter, precisely as in (5.2). We also refer to the Gabor filter function, F , as the Gabor symbol of the multiplier.

We have introduced Gabor symbols as \mathcal{L}^2 -functions, but this restriction is not necessary. In fact, the theory generalizes—and as we shall see, it must in order to be useful for solving inverse problems—to the setting of tempered distributions (see e.g., [33]). We sometimes use the adjectives “classical” and “non-classical”, respectively, when we want to make a distinction between functions and distributions.

With respect to any given atom, the range of the Gabor transform is a proper subspace of $\mathcal{L}^2(\mathbf{R}^{2d})$. In particular, a Gabor filter need not be the Gabor transform of a signal. This is a consequence of the fact that “the STFT of a tempered distribution is a continuous function on the time-frequency plane” ([36], p. 227).

Thus, for example, any discontinuous function on the time-frequency plane cannot come from the Gabor transform of even a non-classical signal (albeit, one could probably determine a suitable space of test functions such that the corresponding distributions would be general enough for this purpose).

The Gabor transform operator, V_g , and its (left) inverse, V_γ^* , have the property that $V_\gamma^*V_g = I$, the identity mapping on $\mathcal{L}^2(\mathbf{R}^d)$. We warn that composing these operators in the reverse order does not give back the identity on $\mathcal{L}^2(\mathbf{R}^{2d})$; however, the latter composition defines the unique orthogonal projection onto the range of the Gabor transform. So if a given symbol, F , happens to also be the Gabor transform of some signal, say $F = V_g s$, then applying the projection $V_g V_\gamma^*$ to F leaves F unchanged: formally, we have

$$[V_g V_\gamma^*]F = [V_g V_\gamma^*]V_g s = V_g [V_\gamma^* V_g]s = V_g s = F. \quad (5.5)$$

We thus see that the restriction of the projection, $V_g V_\gamma^*$, to the range of V_g is the identity operator on the range of V_g . For symbols F outside the range of V_g , applying the projection $V_g V_\gamma^*$ tells us, in the least-squares sense, its nearest neighbour inside the range of V_g . We warn that the property of being the “nearest neighbour” in the least squares sense can be misleading—it need not be “close” in the common sense of the word.

These observations are illustrated in figure 5.1. Two sharp filters, each consisting entirely of ones and zeros, (black = 1 and white = 0) are displayed in the left column. As we have mentioned, since both of these functions are discontinuous, they must lie outside the range of the Gabor transform. Indeed, their orthogonal projections onto the range of V_g , which are displayed in the right column of figure

5.1, are clearly distinguished from the original filter functions. The middle column shows the inverse Gabor transform of the first column.

Are these results to be expected? The Gabor symbol in figure 5.1 (a) describes a boxcar in the frequency direction, so we expect a superposition of corresponding sinc functions in the time domain, which is (up to Fourier wrap-around) what we see in figure 5.1 (b). There is relatively little energy appearing away from the edges in figure 5.1 (b), so we do not expect to see much energy at intermediate time shifts in the Gabor amplitude spectrum, as in figure 5.1 (c). On the other hand, the Gabor symbol of figure 5.1 (d) contains a temporal segment consisting of effectively infinite-bandwidth pieces, so we ought to expect a superposition of delta functions for the inverse Gabor transform. This matches the result shown in figure 5.1 (e). Now figure 5.1 (f) is the result of Gabor transforming the spike in figure 5.1 (e), and a quick calculation reveals that each frequency slice is just the Gaussian analysis atom (except that we have only displayed the positive-temporal half of it—see figure 5.2). Under the inverse Gabor transform, both figures 5.1 (a) and 5.1 (c) map to the signal of 5.1 (b); similarly, both figures 5.1 (d) and 5.1(f) map to figure 5.1 (e). In particular, V_γ^* is not one-to-one on $\mathcal{L}^2(\mathbf{R}^{2d})$, at least for a synthesis window of unity.

Remark The behaviour of the inverse Gabor transform V_γ^* would be quite different had we chosen a non-trivial synthesis atom, γ (we used a synthesis atom of unity). For example, the projection of the stationary Gabor filter of figure 5.1 (a) onto the range of the Gabor transform would not have collapsed the latter to the frequency axes near the time origin as in figure 5.1 (c). This is because, unlike the previous examples, the synthesis frame would preserve the localized

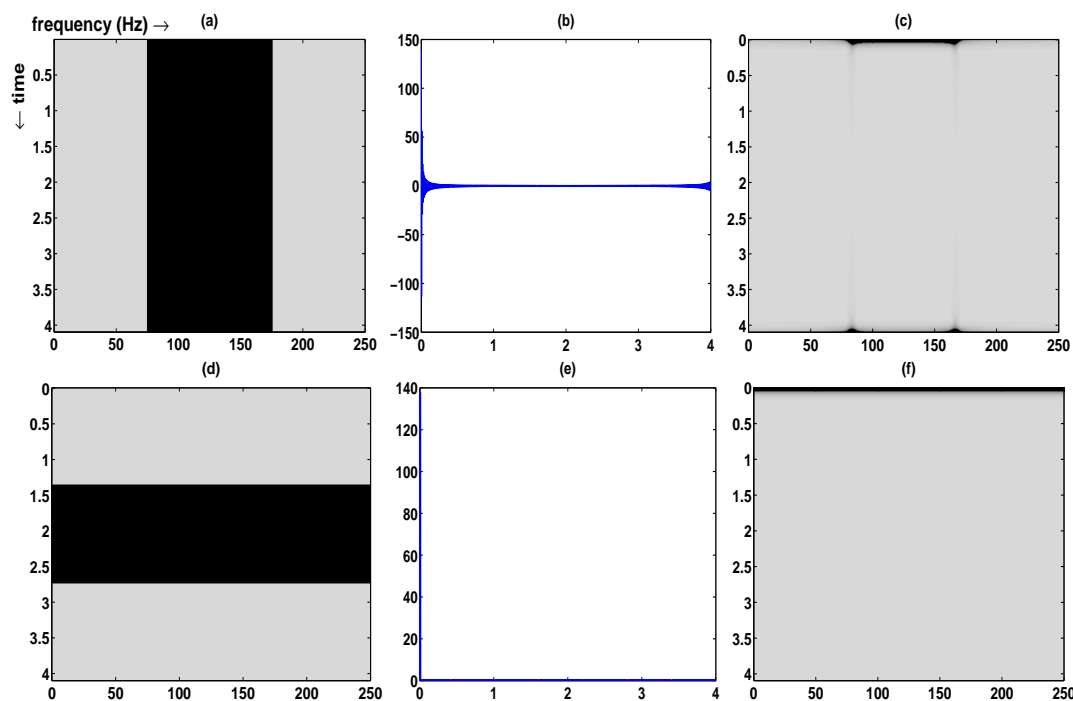


Figure 5.1: Implementation of Gabor filters which are too sharp can lead to disastrous results. Two sharp Gabor filters are displayed at (a) and (d). The middle column shows their corresponding inverse Gabor transforms, (synthesis window of unity) and the right column shows the magnitude of the forward Gabor transforms (Gaussian analysis windows) of the middle column, that is, the magnitude of the orthogonal projections of the original filters onto the range of the Gabor transform. The phase spectrum (shown in figure 5.3) for (c) is rather exotic, while that for (f) is zero. The signal (b) is a sinc function (with some Fourier wraparound effect), and (e) is a spike at time zero. Each frequency slice of the output (f) is just the Gaussian analysis atom, centered at the origin (see figure 5.2).

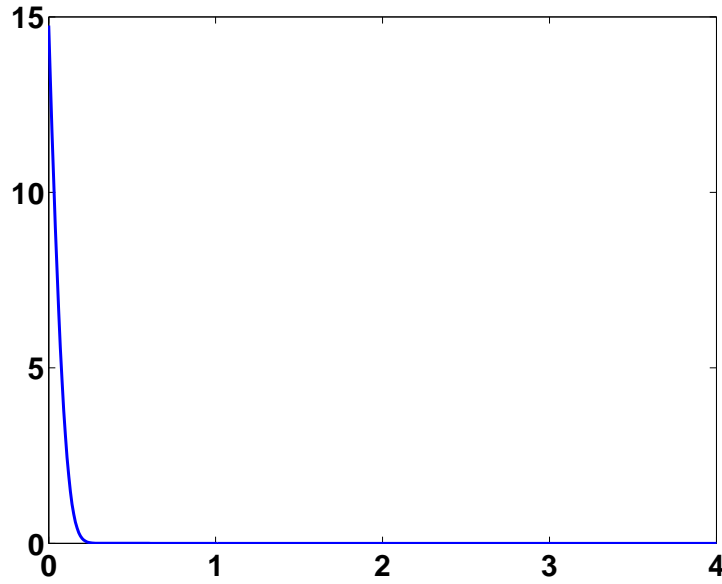


Figure 5.2: A frequency slice of the plot in figure 5.1(f).

information content at the correct time-shifts.

5.1.2 A functional calculus for Gabor multipliers

A more interesting question about Gabor filters from a practical perspective is the following: if we have two Gabor symbols F and G in $\mathcal{L}^2(\mathbf{R}^{2d})$, and we want to compose their corresponding Gabor filters, is this the same as applying the single Gabor filter with multiplier FG ?

A sufficient condition for this to hold is that the symbol G is in the range of the Gabor transform, V_g . Indeed, $G \in \text{Ran}V_g$ if and only if $V_gV_\gamma^*G = G$, and if so, then we have that

$$M_{FG}^{g,\gamma} = V_\gamma^*N_{FG}V_G = V_\gamma^*N_FN_GV_G = V_\gamma^*N_FV_gV_\gamma^*N_GV_G = M_F^{g,\gamma}M_G^{g,\gamma}. \quad (5.6)$$

It follows that the set of all finite-energy Gabor multipliers with symbol in $\text{Ran}V_g$

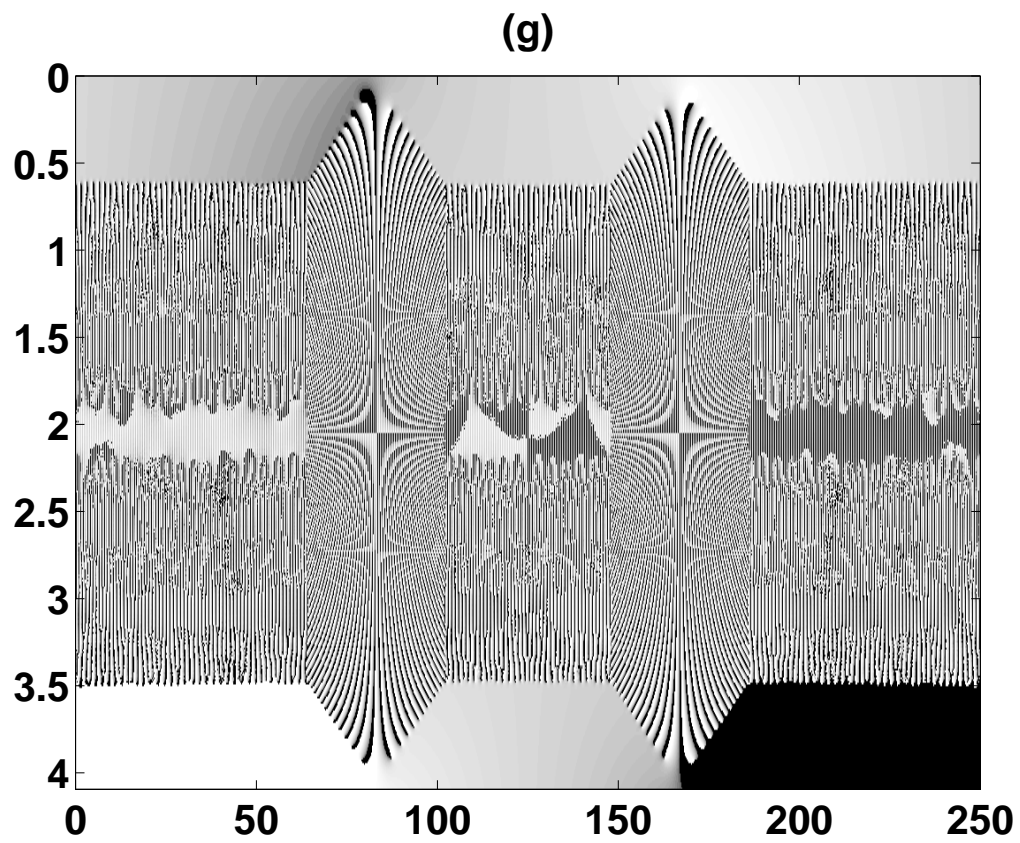


Figure 5.3: The phase spectrum corresponding to the amplitude spectrum of figure 5.1(c).

forms a semigroup, or monoid (it would be a group if it had an identity element, and if each of its members were invertible, with the inverse symbol being also in $\text{Ran}V_g$, but this is asking far too much!). This semigroup property might remain true for weaker conditions, but our examples above (see figure 5.1) strongly suggest the use of caution when cascading Gabor filters.

The most important practical consequence of expression 5.6 is the following: if a physical process can be modelled by the action of a Gabor filter, and if that process can be reversed, then the reverse process can be modelled by the Gabor filter corresponding to the pointwise inverse of the original symbol. This is exactly the guiding principle for solving the inverse problems of Chapters 6 and 7.

It can often be the case, as it is for the former problem, that the symbol of the forward model either has zeros or it decreases in magnitude so rapidly that when sampled in a computer, the dynamic range of the system is exhausted. Either way, the pointwise inverse of the symbol is simply unbounded. A standard way around this type of problem is to add a small constant—called a stability factor—to the symbol before inverting it. Other methods will also work, but the point is that an approximate inverse can always be constructed.

Evidently, the multiplier defined by the scaled sum of two Gabor symbols is identical to the corresponding scaled sum of the individual multipliers defined by the two symbols: that is,

$$M_{\alpha(F+G)}^{g,\gamma} = \alpha (M_F^{g,\gamma} + M_G^{g,\gamma}), \quad (\alpha \in \mathbf{C}). \quad (5.7)$$

In other words, a Gabor multiplier is a linear function of its symbol.

Is the Gabor multiplier a one-to-one function of its symbol? That is, if

$$M_F^{g,\gamma} = M_G^{g,\gamma}, \quad (5.8)$$

can we conclude that $F = G$?

The answer to this question again depends on the choice of window pair, $\{g, \gamma\}$. Whenever $\{g, \gamma\}$ is a compatible pair (see below, and [33]) of nowhere-zero functions, with nowhere-zero Fourier spectra—for example, a pair of Gaussians—then the answer is yes. As shown in [33], the identical Schwartz kernels, $k(M_F^{g,\gamma})$ and $k(M_G^{g,\gamma})$, of the two multipliers can be expressed in terms of their Gabor symbols via a cascade of linear operators. Under the current assumptions, it turns out that each of these operators is one-to-one; hence so is their product. In particular, $F = G$; and so we see that for good window choices, distinct Gabor symbols define distinct Gabor filters. This result—perhaps not a very surprising one from an intuitive perspective—involves some challenging mathematics.

5.1.3 On the discrete theory

We briefly discuss the implementation of the analog theory within the discrete setting. We start with a particular Gabor frame, $\{g_{m,n}\}$, its dual, $\{\gamma_{m,n}\}$, and some discretized Gabor filter function,

$$F : a\mathbf{R}^d \times b\mathbf{R}^d \rightarrow \mathbf{C}, \quad (5.9)$$

sampled on the corresponding time-frequency lattice, $a\mathbf{R}^d \times b\mathbf{R}^d$. In practice, this is just a matrix with complex-valued entries. The definition of a Gabor multiplier, translated from expression (5.2), becomes

$$\tilde{s}(t) = V_\gamma^*[F(an, bm) \cdot V_g s(an, bm)](t), \quad (5.10)$$

where we have kept the same notation from (5.2) to denote the forward and inverse Gabor operators. The product inside the square brackets is taken pointwise. The effectiveness of the filtering process depends on: (1) how well the desired filter is interpolated onto the time-frequency lattice; and (2) the choice of time-frequency lattice; and (3) the choice of analysis and synthesis atoms. Each of these three areas are still current topics of research (see e.g., [21], [20], [36], [49], [43], [14], [38], [73], [102]).

5.2 Applications to Mathematical Physics

A key motivation for the study of pseudodifferential operators comes from their connection with partial differential equations. When represented in the Kohn-Nirenberg form, it becomes clear that pseudodifferential operators are a natural extension of the translation-invariant Fourier filters, or convolution operators, to the nonstationary setting.

5.2.1 General linear partial differential operators

We present a brief heuristic development similar to that of Stein in [93], chapter 6, section 1. We choose to present the theory of pseudodifferential operators only formally. Passing beyond this scope in any meaningful way would require a rather lengthy exposition. The literature is very rich in this topic, and the following list should provide a more-than-sufficient starting point: [47], [55], [57], [93], [84], [36], [96], [23], [105], [108], [43], [48], [63],[62], [64], and [102].

Let L be a linear partial differential operator:

$$L = \sum_{|\alpha| \leq m} a_\alpha(x) \left(\frac{\partial}{\partial x} \right)^\alpha, \quad (5.11)$$

where α is a multiindex (see Section 2.1.3), and $x \in \mathbf{R}^d$. The characteristic polynomial of L , $a(x, \xi)$, is defined by formally replacing the partial derivative operator $\partial/\partial x$ with its Fourier dual, $2\pi i\xi$:

$$a(x, \xi) = \sum_{|\alpha| \leq m} a_\alpha(x) (2\pi i\xi)^\alpha. \quad (5.12)$$

The action of L can be expressed in the Fourier domain in terms of $a(x, \xi)$ as follows:

$$(Lf)(x) = \int_{\mathbf{R}^d} a(x, \xi) \hat{f}(\xi) e^{2\pi i x \cdot \xi} d\xi. \quad (5.13)$$

Equation 5.13 is a formal representation of the operator L as a Kohn-Nirenberg pseudodifferential operator, with Kohn-Nirenberg symbol $a(x, \xi)$. The case of a partial differential operator is the prototype for a pseudodifferential operator. The only difference is that the symbol in the former can always be expressed as a superposition of tensor products, (see page 47) each of which involves a pair of functions: a function of x alone, the other a function of the dual variable, ξ , alone.

A class of symbols that leads to approximately translation-invariant operators ([93]) is given by the conditions

$$|\partial_x^\beta \partial_\xi^\alpha a(x, \xi)| \leq A_{\alpha, \beta} (1 + |\xi|)^{m - |\alpha|} \quad (5.14)$$

for all multiindices α and β .

For a general symbol, a , the associated pseudodifferential operator T_a can be thought of as a limit of linear combinations of product-convolution operators

([93]). A product-convolution operator is the composition of a multiplication operator with a convolution operator, or equivalently, a Fourier multiplier operator: specifically, if $a(x, \xi) = a_1(x)$, is a function of x alone, then

$$(T_a f)(x) = a_1(x)f(x) \quad (5.15)$$

defines a multiplication operator; while if $a(x, \xi) = a_2(\xi)$ depends only on ξ , then

$$(T_a f)(x) = \left\{ \left(\mathcal{F}^{-1} a_2 \right) * f \right\} (x) \quad (5.16)$$

defines a convolution operator. These operators are dual to each other—either can be represented in the form of the other simply by changing domains via the Fourier transform.

5.2.2 Pseudodifferential operators and Gabor filters

Following a conjecture of Margrave, which goes back to unpublished ideas from 1998, Gibson, et al., ([33]) established an important connection between the (possibly distributional) Kohn-Nirenberg symbol of a pseudodifferential operator and Gabor multipliers. More specifically, for any given pair of “compatible” analysis and synthesis windows, (e.g., any pair of Gaussians, or any Schwartz function combined with the constant function, 1) the authors determine precisely which Kohn-Nirenberg operators can be represented as a Gabor multiplier, and also provide an explicit formula for computing the exact Gabor symbol—if it exists—from the Kohn-Nirenberg symbol.

The Gaussian serves as a prototype for the definition of compatible analysis and synthesis windows, introduced by Gibson, et al., [33]. They showed that the

class of operators which are representable as Gabor multipliers based on Gaussian atoms is a proper subset of the corresponding class based on “extreme value windows”. These extreme value windows are compatible, analytic, rapidly decreasing (Schwartz class) windows, with the property that their Fourier transform is also analytic and rapidly decreasing.

If we allow distributional atoms, (say 1 and δ) we can construct completely general Gabor multipliers (see [33], Proposition 20).

For a given compatible window pair, we intuitively expect a classical pseudodifferential operator to be representable as the Gabor multiplier having the same symbol, just in case its symbol is sufficiently smooth. Indeed, this intuition was validated in [33], Theorem 4, for the case of compatible windows (see Table 3 in [33] for a list of compatible windows).

This connection between classical Kohn-Nirenberg symbols and Gabor filters leads us to conclude what is already known in pseudodifferential operator theory: that the operator corresponding to the pointwise product of two Kohn-Nirenberg symbols is approximately equal to the composition of the operators corresponding to the individual symbols, and that the approximation tends to equality as the symbols increase in smoothness. We simply use the given symbols as Gabor symbols, after choosing an appropriate window pair, then compose the resulting operators according to the functional calculus presented in section 5.1.2, then use the resulting Gabor symbol as the Kohn-Nirenberg symbol for the approximate operator.

For a great variety of realistic physics problems, then, the associated pseudodifferential operator and Gabor filter will have virtually the same symbols. Data

collected from real experiments will always have a certain degree of random noise present; applying an “exact” filter, which was designed for a noise-free case, will tend to deteriorate the signal-to-noise level. Finally, the ability to accurately represent the data, even in a noise-free environment, is highly limited by the sampling process; the fidelity of the “real” signal gets degraded, (bandlimited) so in essence the data represents a smoothed version of the signal (plus random noise). Naturally then, a probability distribution is associated with the degree of certainty in our measurements—just as Heisenberg, et al. have taught us.³

A few suggestions for further reading in this area: [36], [19], [48], [63],[62], [64], and [102].

³His 1927 paper on the uncertainty principle, *Über die Grundprinzipien der Quantenmechanik*: FF 3, no. 11, 83, was translated to English in 1984: *The physical content of quantum kinematics and mechanics*, John Archibald Wheeler and Wojciek Hubert Zurek, eds. *Quantum theory and measurement*: Princeton University Press, 62–84.

Chapter 6

Deconvolution and Q -inversion in dissipative media

Seismic attenuation can be modelled macroscopically via an exponential amplitude decay in both time and frequency, at a rate determined by a single dimensionless quantity, Q .

Current deconvolution methods, based on the convolutional model, attempt to estimate and remove the embedded causal wavelet. We propose a more generalized, nonstationary seismic model, expressed as a pseudodifferential operator in the time-frequency domain. The model has the properties that (1) the embedded causal wavelet factors as the product of a stationary seismic signature with a nonstationary exponential decay, and (2) a nonstationary impulse response for the earth is tractable.

Analytic expressions are derived for least-squares fitting our model to the Gabor-transformed seismic trace, thus yielding both a Q -value and an estimate of the source signature, and hence an estimate of the nonstationary travelling wavelet. These estimates are then used to compute a smoothed version of the magnitude of the Gabor spectrum of the seismic trace, from which a least-squares nonstationary minimum-phase deconvolution filter is easily constructed.

Our results on randomly generated synthetic seismic data are very promising.

6.1 Introduction

This chapter concerns the application of Gabor filter theory to the modelling, analysis, and subsequent deconvolution of a nonstationary, constant- Q -attenuated seismic signal. Gabor deconvolution is discussed in a broader context by Margrave, et al., [60].

We outline the derivation of a time-variant spectral model for the Gabor transform of a constant- Q -attenuated seismic trace. Remarkably, in addition to being intuitively plausible, it generalizes the standard convolutional model. Next, we derive explicit least-squares, model-based estimates for both Q and the stationary part of the wavelet, and thus an estimate of the nonstationary, Q -attenuated wavelet. The theory is then evaluated via an algorithm for the deconvolution of a Q -attenuated synthetic seismogram.

6.2 Theory

Let $r(t)$ be a reflectivity, $w(t)$ a source signature, and $\alpha(t, f)$ the time-frequency symbol of a constant- Q operator, namely

$$\alpha(t, f) = \exp\left(-\frac{\pi}{Q} \{ft + i\mathcal{H}(ft)\}\right), \quad (6.1)$$

where \mathcal{H} is the Hilbert transform. We assume that the source signature is a causal, stationary wavelet, and hence assume that its time-frequency decomposition is equivalent to its Fourier transform, $\hat{w}(f)$. We further assume that $\hat{w}(f)$ is smooth; by definition, $\alpha(t, f)$ is also smooth. A synthetic seismic trace, $s(t)$, is built by nonstationary convolution of the Q -operator with the reflectivity, (see [54] for more detail) followed by stationary convolution with the source signature. This yields

$$\hat{s}(f) = \hat{w}(f) \int_{-\infty}^{\infty} \alpha(t, f)r(t)e^{-2\pi ift} dt, \quad (6.2)$$

or

$$s(t) = \int_{-\infty}^{\infty} \hat{w}(f) \left[\int_{-\infty}^{\infty} \alpha(u, f)r(u)e^{-2\pi ifu} \right] e^{2\pi ift} df, \quad (6.3)$$

which can be expressed more compactly as

$$s(t) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \hat{w}(f)\alpha(u, f)r(u)e^{2\pi if(t-u)} df du. \quad (6.4)$$

Now, the continuous Gabor transform (also called the short-time Fourier transform) of $s(t)$ is defined as

$$V_g s(\tau, f) = \int_{-\infty}^{\infty} s(t)g(t - \tau)e^{-2\pi ift} dt, \quad (6.5)$$

where $g(t)$ is called the analysis window, e.g., a Gaussian. Margrave et al., [60] show that the Gabor transform of $s(t)$ factorizes, to first order, as

$$V_g s(\tau, \nu) = \hat{w}(\nu)\alpha(\tau, \nu)V_g r(\tau, \nu). \quad (6.6)$$

To first order, our Gabor-spectral model is given by (6.6). The problem is simplified by first considering only the magnitudes

$$|V_g s(\tau, \nu)| = |\hat{w}(\nu)| |\alpha(\tau, \nu)| |V_g r(\tau, \nu)|. \quad (6.7)$$

The Gabor transform of the reflectivity is assumed white, with unit mean. By “white”, we mean that $|\hat{w}(\nu)| |\alpha(\tau, \nu)|$ provides the general spectral shape, while the Gabor spectrum of the reflectivity provides only details.¹ Thus, we drop the term $|V_g r(\tau, \nu)|$ from (6.7), and seek a (smoothed) time-frequency trace model as

$$S(\tau, \nu) = W(\nu) e^{-\pi\tau\nu/Q}, \quad (6.8)$$

where $W(\nu) = |\hat{w}(\nu)|$ and $e^{-\pi\tau\nu/Q} = |\alpha(\tau, \nu)|$. The equality in (6.8) is taken in the least-squares sense, meaning that a residual error with minimized \mathcal{L}^2 -norm is assumed. For a further simplification, we consider the logarithm of both sides of (6.8):

$$\ln S(\tau, \nu) = \ln W(\nu) - \pi\tau\nu/Q. \quad (6.9)$$

6.3 Method

We first minimize the function $\alpha = \alpha(W, Q)$ given by

$$\alpha(W, Q) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left[\ln \frac{S(\tau, \nu)}{W(\nu)} + \frac{\pi\tau\nu}{Q} \right]^2 d\tau d\nu \quad (6.10)$$

with respect to Q . Equation (6.10) is the square of the \mathcal{L}^2 -norm of the difference between both sides of (6.9). The domain of integration is finite in practice, since

¹Seismic data is not necessarily white; in fact, it can be what has been called “blue”. This is a quality which can be inferred from well logs, when available, and which can subsequently be used to whiten the data. See e.g., [16].

seismic data are bandlimited and have finite duration. The region of integration, Ω , is selected to encompass only the numerically significant part of the time-variant spectrum of the signal, thus avoiding large errors due to division by excessively small numbers. To this end, a characteristic weighting function, χ_Ω , defined as

$$\chi_\Omega(\tau, \nu) = \begin{cases} 1 & \text{if } (\tau, \nu) \in \Omega \\ 0 & \text{if } (\tau, \nu) \notin \Omega \end{cases}, \quad (6.11)$$

is included as a factor in the integrand. For example,

$$\int_\Omega f(\tau, \nu) d\tau d\nu = \int_{\nu_0}^{\nu_1} \int_{\tau_0}^{\tau_1} f(\tau, \nu) \chi_\Omega(\tau, \nu) d\tau d\nu, \quad (6.12)$$

where f is any integrand with numerically stable support Ω , contained in the rectangle $[\tau_0, \tau_1] \times [\nu_0, \nu_1]$. Any positive weighting function can be substituted for χ_Ω , provided it decays to zero smoothly enough to avoid spectral ringing. Given such a region Ω , (or weighting function) expression (6.10) becomes

$$\alpha(W, Q) = \int_\Omega \left[\ln \frac{S(\tau, \nu)}{W(\nu)} + \frac{\pi\tau\nu}{Q} \right]^2 d\tau d\nu \quad (6.13)$$

which can be computed according to the prescription (6.12). Minimizing (6.13) with respect to Q amounts to solving

$$0 = \frac{\partial \alpha}{\partial Q} = 2 \int_\Omega \left[\ln \frac{S(\tau, \nu)}{W(\nu)} + \frac{\pi\tau\nu}{Q} \right] \left[-\frac{\pi\tau\nu}{Q^2} \right] d\tau d\nu \quad (6.14)$$

for Q , which leads to the finite estimate:

$$Q = \pi \frac{\int_\Omega \tau^2 \nu^2 d\tau d\nu}{\int_\Omega \tau \nu \ln \frac{W(\nu)}{S(\tau, \nu)} d\tau d\nu} \quad (6.15)$$

in terms of the unknown signature, $W(\nu)$. The variational calculus yields an expression for the least-squares-optimal function, $W(\nu)$: write $s = \ln S$, $w = \ln W$,

and let $\delta w = \delta w(\nu)$ be an arbitrary variation. Incrementing $w(\nu)$ by $\delta w(\nu)$, (6.13)

becomes

$$\alpha(w + \delta w, Q) = \int_{\Omega} \left[s(\tau, \nu) - w(\nu) - \delta w(\nu) + \frac{\pi\tau\nu}{Q} \right]^2 d\tau d\nu, \quad (6.16)$$

or

$$\alpha(w + \delta w, Q) = \alpha(w, Q) - 2 \int_{\Omega} \left[s(\tau, \nu) - w(\nu) + \frac{\pi\tau\nu}{Q} \right] \delta w(\nu) + \int_{\Omega} [\delta w(\nu)]^2. \quad (6.17)$$

Any function $w(\nu)$ for which the middle term in (6.17) vanishes will minimize $\alpha(w, Q)$ with respect to w , so we need only solve the following equation for $w(\nu)$:

$$\int_{\Omega} \left[s(\tau, \nu) - w(\nu) + \frac{\pi\tau\nu}{Q} \right] \delta w(\nu) d\tau d\nu = 0. \quad (6.18)$$

Since $\delta w(\nu)$ is independent of time, we have

$$\int_{\nu_0}^{\nu_1} \left\{ \int_{\tau_0}^{\tau_1} s(\tau, \nu) \chi_{\Omega}(\tau, \nu) d\tau - w(\nu) \int_{\tau_0}^{\tau_1} \chi_{\Omega}(\tau, \nu) d\tau + \frac{\pi\nu}{Q} \int_{\tau_0}^{\tau_1} \tau \chi_{\Omega}(\tau, \nu) d\tau \right\} \delta w(\nu) d\nu = 0. \quad (6.19)$$

The ν -integral vanishes for arbitrary $\delta w(\nu)$, so the function of ν inside the braces must vanish. This leads to

$$w(\nu) = \frac{\int_{\tau_0}^{\tau_1} s(\tau, \nu) \chi_{\Omega}(\tau, \nu) d\tau}{\int_{\tau_0}^{\tau_1} \chi_{\Omega}(\tau, \nu) d\tau} + \frac{\pi\nu \int_{\tau_0}^{\tau_1} \tau \chi_{\Omega}(\tau, \nu) d\tau}{Q \int_{\tau_0}^{\tau_1} \chi_{\Omega}(\tau, \nu) d\tau}. \quad (6.20)$$

Converting back to the logarithmic notation, we arrive at

$$W(\nu) = \exp \left\{ \frac{\int_{\tau_0}^{\tau_1} \ln[S(\tau, \nu)] \chi_{\Omega}(\tau, \nu) d\tau}{\int_{\tau_0}^{\tau_1} \chi_{\Omega}(\tau, \nu) d\tau} + \frac{\pi\nu \int_{\tau_0}^{\tau_1} \tau \chi_{\Omega}(\tau, \nu) d\tau}{Q \int_{\tau_0}^{\tau_1} \chi_{\Omega}(\tau, \nu) d\tau} \right\}. \quad (6.21)$$

The first term in the exponential is the time average of $\ln S(\tau, \nu)$, and the second involves the average time over each frequency. Using an overline to denote a time average, (6.21) becomes

$$W(\nu) = \exp \left\{ \overline{\ln S}(\nu) + \frac{\pi\nu}{Q} \bar{\tau}(\nu) \right\}. \quad (6.22)$$

Substitution of (6.22) into the expression (6.15) for Q yields

$$Q = \pi \frac{\int_{\Omega} \tau^2 \nu^2 d\tau d\nu}{\int_{\Omega} \tau \nu \left[\ln \bar{S}(\nu) + \frac{\pi\nu}{Q} \bar{\tau}(\nu) - \ln S(\tau, \nu) \right] d\tau d\nu} \quad (6.23)$$

Finally, solving for Q leads to

$$Q = \pi \frac{\int_{\Omega} \tau \nu^2 [\tau - \bar{\tau}(\nu)] d\tau d\nu}{\int_{\Omega} \tau \nu \left[\ln \bar{S}(\nu) - \ln S(\tau, \nu) \right] d\tau d\nu}. \quad (6.24)$$

6.4 A synthetic example

MATLAB functions from the CREWES toolboxes were used extensively to generate the least-squares estimates (6.24) and (6.22) of Q and $W(\nu)$, and to perform a Gabor deconvolution. Figure 6.1 displays a pseudo-random reflectivity and a non-stationary synthetic trace. The latter was built by applying a Q -operator to the reflectivity, followed by convolution with a 20Hz minimum phase source signature.

Figure 6.2 displays the magnitude of the Gabor transform of the reflectivity. The sample points in the time direction (row number) correspond to successive translations by 0.01s of Gaussian window centers. These Gaussian windows have the property that their sum is approximately unity over the duration of the synthetic trace, and their half-width (0.1s) has been selected to meet this criterion. Each row is computed as the discrete Fourier transform of the windowed trace that is centered at the corresponding offset time. Each sequence of coherent peaks, e.g., those at about 1s, corresponds to a zone of high amplitude in the reflectivity of figure 6.1.

Since the Q -operator represents an exponential decay surface in time and frequency, its logarithm is a surface with hyperbolic contours, decaying according

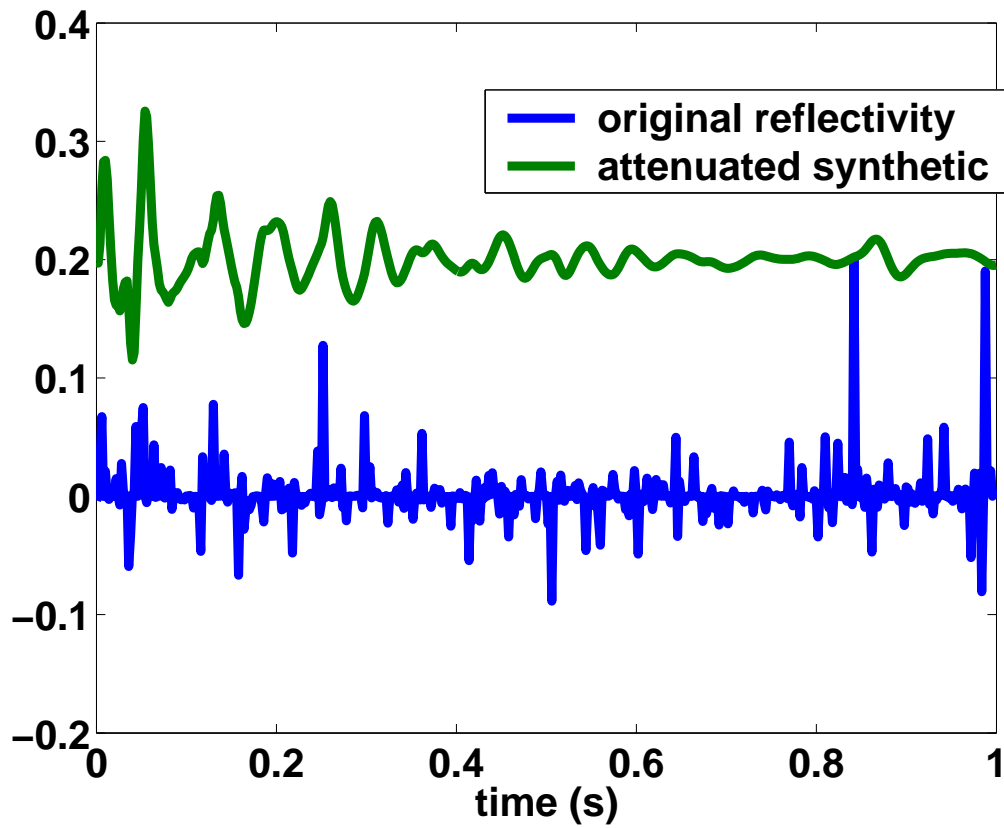


Figure 6.1: A pseudo-random reflectivity (lower) and a nonstationary synthetic (upper) generated with an attenuation factor of 25.

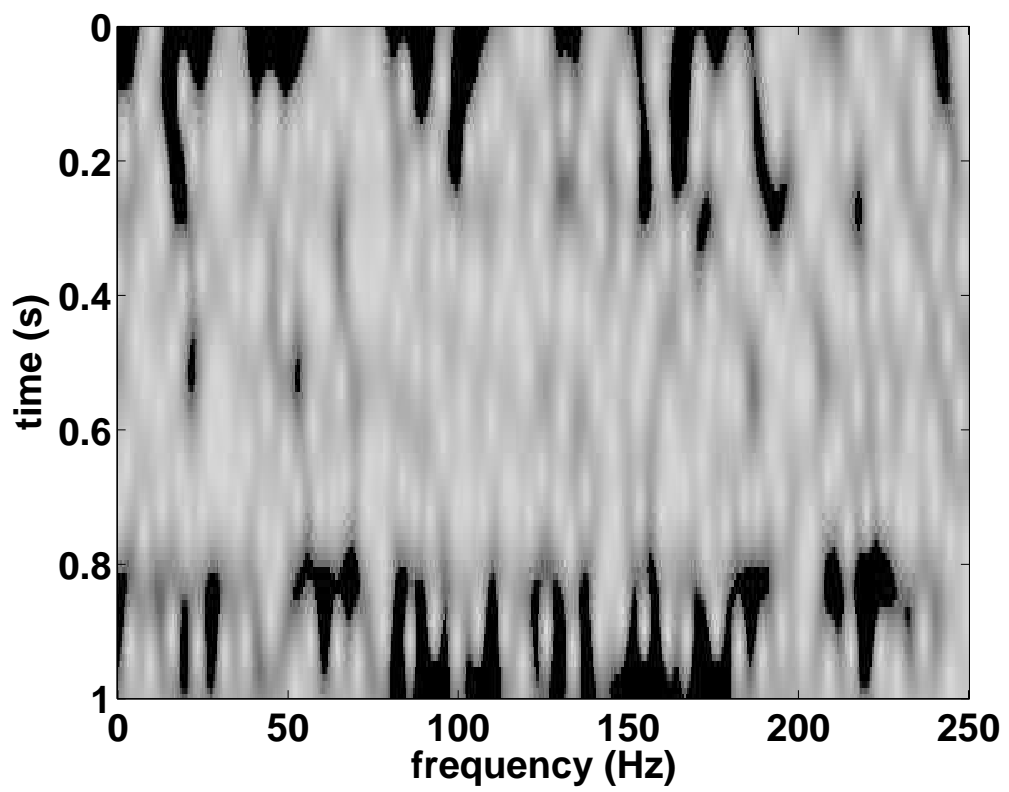


Figure 6.2: Magnitude of the Gabor transform of the pseudo-random reflectivity of figure 6.1.

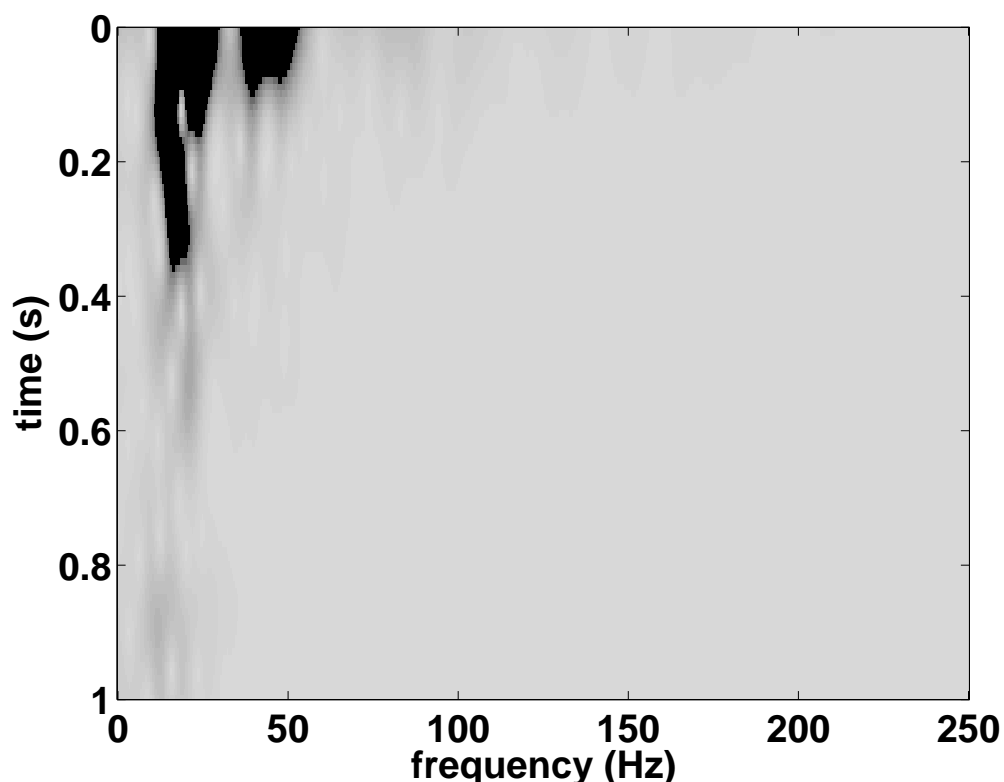


Figure 6.3: Magnitude of the Gabor transform of the synthetic trace of figure 6.1.

to the product of the time and frequency values. This explains the general decay pattern from the top left toward the bottom right in the magnitude of the Gabor transform of the synthetic trace (figure 6.3). Note the progressive loss of bandwidth and mean amplitude over time.

The filter in figure 6.4 was used to weight the integrands in (6.24) and (6.22) according to the prescription (6.12). This produced a Q -estimate of 28.3, (the correct value was 25, which, as is well known in the seismic community, is an extremely low value for Q , hence a good test for the inversion algorithm) and the source signature estimate, whose smoothed Fourier spectrum is plotted with that of the original signature in figure 6.5. Small residual oscillations in the signature

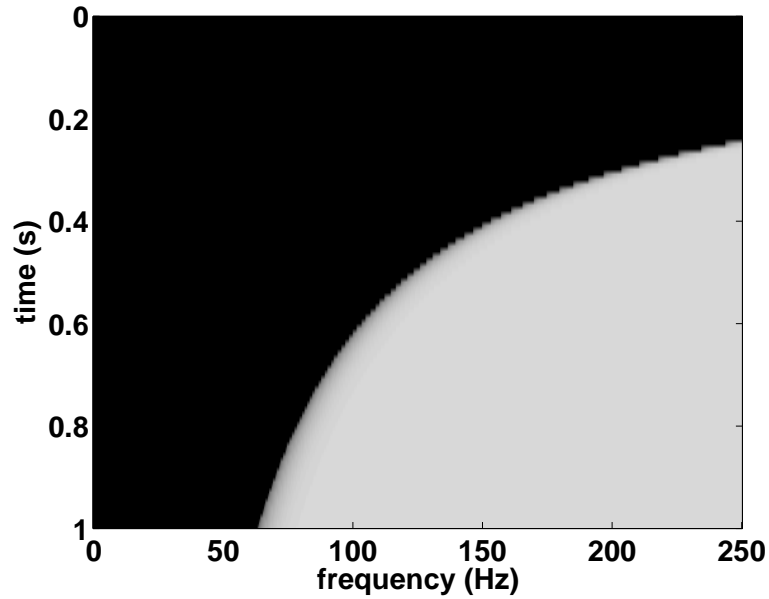


Figure 6.4: Time-variant filter for weighting the Gabor spectrum of the synthetic trace according to the available dynamic range, and hence the available numerical precision of the data.

spectrum, inherited from the reflectivity, were smoothed by convolution with a boxcar.

Figure 6.6 displays the magnitude of the least-squares model of the data, calculated as

$$|w(\nu)|_{est} e^{-\pi\tau\nu/Q_{est}}, \quad (6.25)$$

a smoothed version of the display in figure 6.3 (*est* refers to estimated quantities).

The deconvolution operator is the pointwise inverse of (6.25),² combined with its nonstationary minimum phase spectrum.

The deconvolved result is compared with the original reflectivity in figure 6.8;

²A small constant, known as a “stability factor” by seismologists, was added to avoid “division by zero”. The size of this constant can be related to the available dynamic range, as well as resolution and noise suppression factors as discussed in [100].

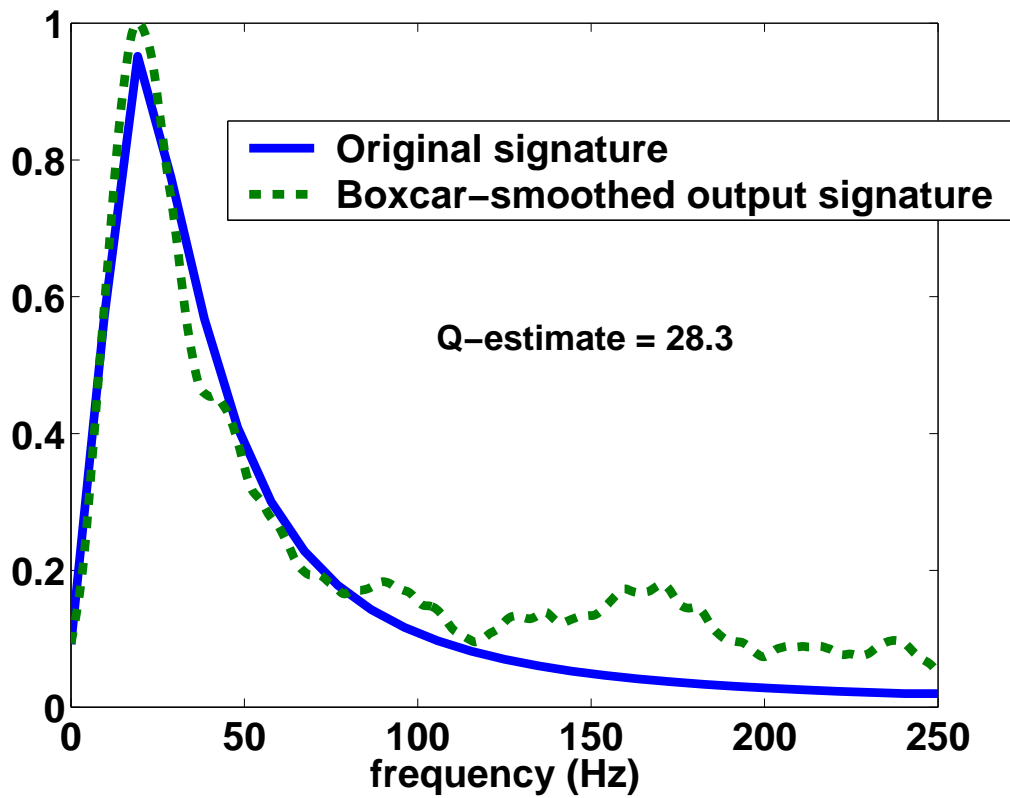


Figure 6.5: Amplitude spectra of original and output signatures. The output was smoothed to remove small residual oscillations.

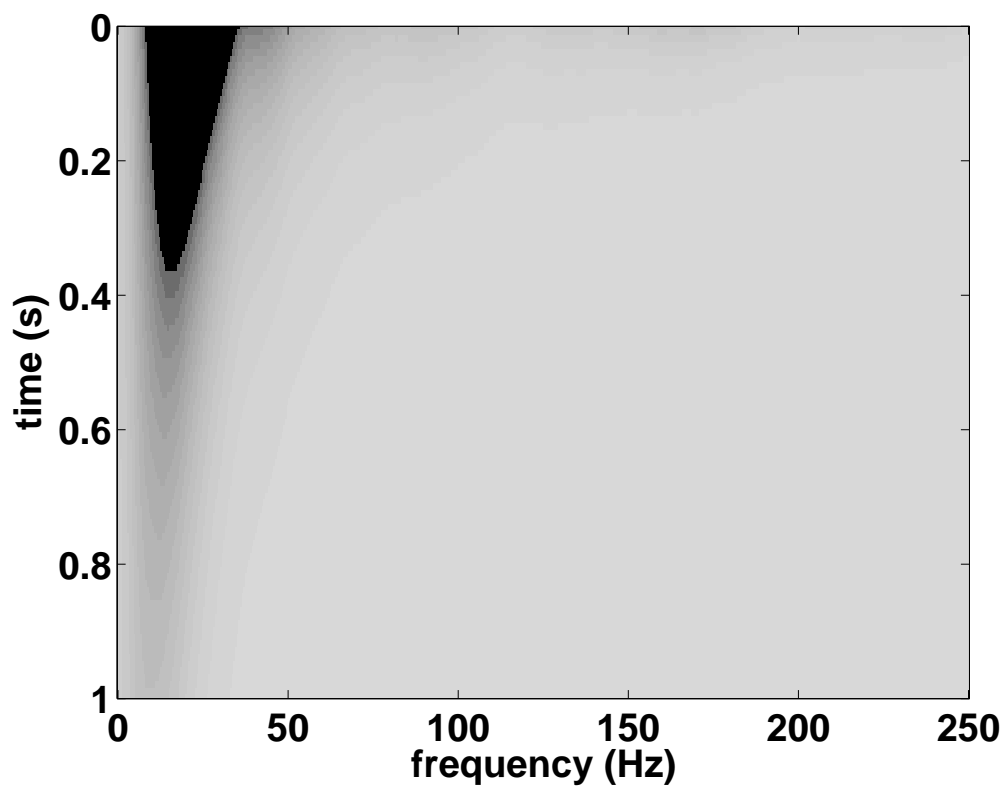


Figure 6.6: The magnitude of the smoothed spectrum, which models the data (from figure 6.3) in a least-squares sense. The pointwise inverse (plus a small constant) defines the magnitude of the deconvolution operator.

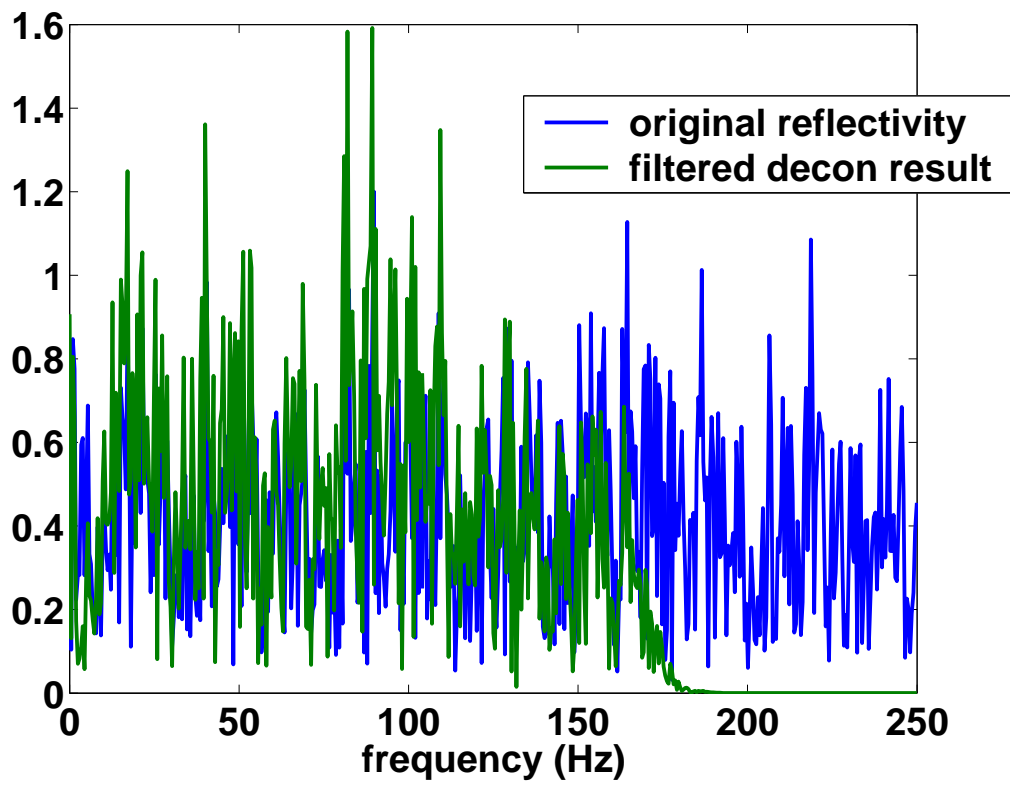


Figure 6.7: Fourier amplitude spectra of original reflectivity and Gabor deconvolution result.

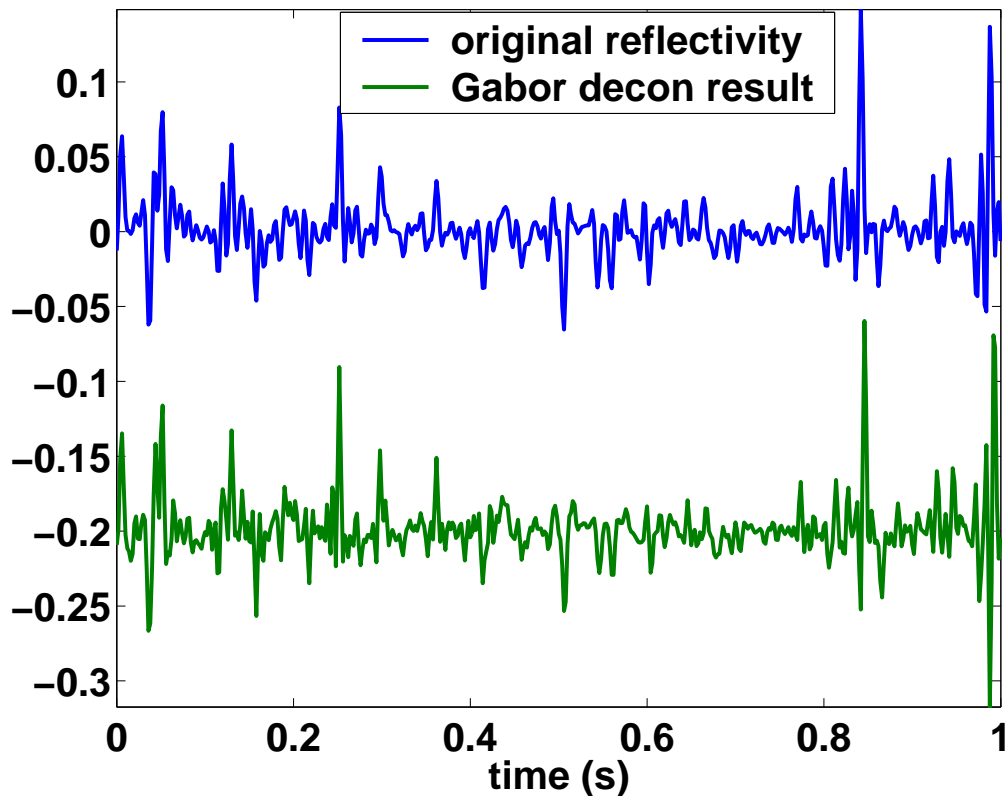


Figure 6.8: Comparison of the bandpass-filtered Gabor deconvolution result (lower) with a bandpass-filtered (DC to half Nyquist) version of the reflectivity of figure 6.1 (upper).

both are bandpass filtered from D.C. to Nyquist. This bandlimited version of the reflectivity is a good benchmark by which to compare our results, since it is the best that one could hope to obtain. As is clearly shown, the reflector locations are highly correlated and the amplitudes agree very well.

6.5 Summary

A time-variant spectral model for the Gabor transform of a constant- Q -attenuated trace was proposed. Using differential calculus and the calculus of variations, the

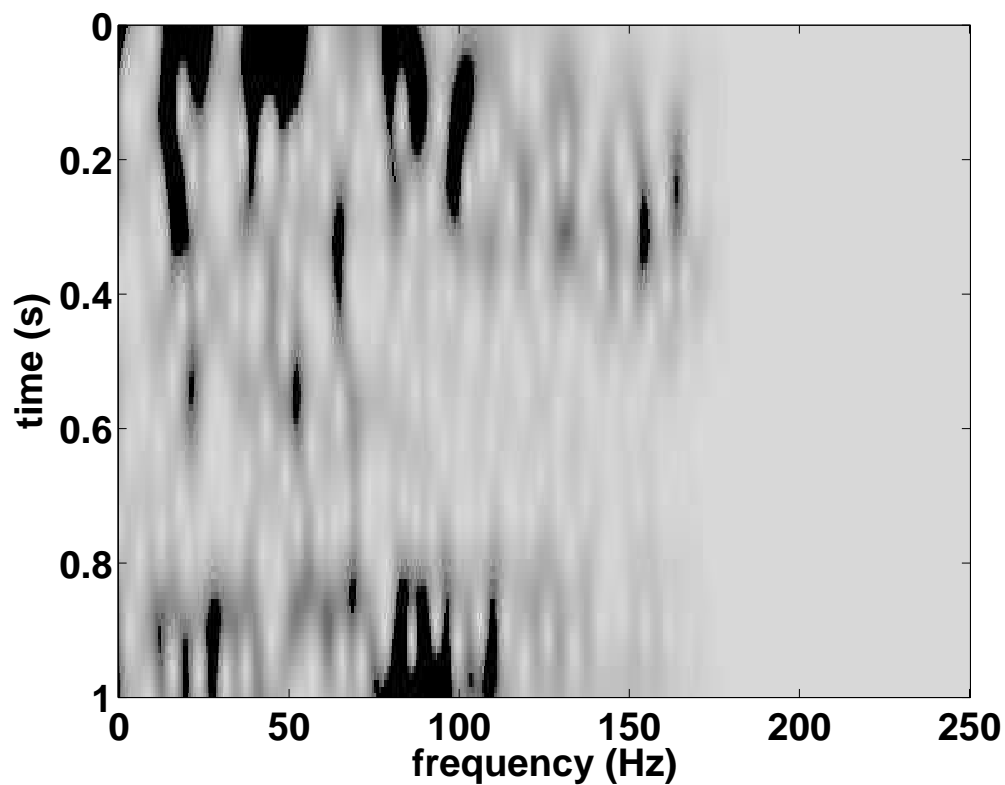


Figure 6.9: The time-variant, bandlimited amplitude spectrum of the reflectivity estimate. Compare with the Gabor amplitude spectrum of the true reflectivity in figure 6.2.

model was fitted to the data in the least squares sense. This yielded estimates for Q and the source signature, hence an estimate of the nonstationary, Q -attenuated wavelet. The theory was numerically evaluated via an algorithm for the Gabor deconvolution of a synthetic, Q -attenuated trace. The illustrated example used a very low input Q -value, yet returned highly promising initial deconvolution results. These results, reinforced by the fact that our model includes the convolutional model as a special case, strongly motivate further research in this direction.

Chapter 7

Wavefield extrapolation in inhomogeneous media

Wavefield extrapolation for a laterally varying velocity model can be achieved by applying a nonstationary phase-shift filter to an adaptive, nonuniform Gabor transform over the lateral coordinate. A family of adaptive Gabor frames can be constructed from a molecular decomposition of unity, each molecule of the latter being built by conjoining neighbouring atoms from a uniform partition of unity—consisting of translates of a single atom along the lateral coordinate—according to a local stationarity criterion derived from the velocity model.

The resulting extrapolation algorithm—called AGPS (adaptive Gabor phase-shift)—has a computational cost that is proportional to the complexity of the velocity model, while its accuracy is comparable to both NSPS (nonstationary

phase-shift) and generalized PSPI (phase-shift plus interpolation). AGPS includes NSPS and PSPI as complementary limiting cases, yet the cost of AGPS ranges from an order of magnitude less to about the same order.

7.1 Introduction

This chapter has a predominantly qualitative scope. An investigation into the underlying mathematical and theoretical background for the new AGPS algorithm forms the core of Chapter 4, and appears in a more compact form in [38]. For a derivation of the new and exact wavefield extrapolator—used here to forward model the propagation of an impulsive wavefield—see [59]. Apart from its remarkable accuracy, a distinguishing feature of this extrapolator is its transitivity through any $v(x)$ medium. By transitivity, we mean that a single depth step yields the same result as the iteration of any number of intermediate steps does. We use this algorithm as a benchmark to test the performance of three approximate wavefield extrapolation methods: the NSPS and generalized PSPI methods of Margrave and Ferguson, (see [56]) and AGPS.

The AGPS algorithm includes NSPS and PSPI as complementary limiting cases. Each of these methods applies a velocity-dependent, nonstationary phase-shift filter over each temporal frequency slice of the input data. The two latter filters differ to the extent that the functional dependence of the velocity is either on the input coordinates (NSPS) or the output coordinates (PSPI)—in fact, they are spatial transposes of each other. This is the key reason why both methods are approximate: indeed, any accurate phase-shift operator in a $v(x)$ medium

has to account for the fact that the velocity can vary along the trajectory of a ray. AGPS attempts to address this problem by representing the input and the output wavefields as a superposition of windowed components, each of which is approximately stationary with respect to the velocity.

We begin with an overview of the AGPS algorithm, and refer the reader to Chapter 4, and references therein, for the mathematical details. We then provide a performance analysis for the AGPS, PSPI, and NSPS extrapolators, based on forward-modelled data generated by Margrave’s exact extrapolator.

7.2 Overview of the AGPS method

AGPS propagates a wavefield from a depth level z to a depth level $z + \Delta z$ by applying nonstationary phase-shift filters to an adaptive spatial Gabor transform of the input data. Given a velocity model, $v(x)$, the first step is to choose a suitable window, or atom. The choice of atom is in itself a current topic of research, (e.g., [20] or [36]) but we recommend using a Gaussian or Gaussian-like window, with a halfwidth that is at least large enough to ensure that its sampled version faithfully represents it. This atom is then translated along the discrete x -coordinate, one atom centered at each sample point. The resulting suite of windows is rescaled, if necessary, to ensure that its superposition equals one: thus, the translated atoms form a maximal, uniform partition of unity (see Chapter 4).

Next, this maximal partition of unity is adapted to yield a nonuniform partition of unity, in such a way that the essential nonstationarity of the velocity function is respected. This procedure involves the formation of what we call “molecules”,

or macro-windows, by summing neighbouring atoms over regions that satisfy a local stationarity measure. The stationarity measure in the current context is simply an acceptable threshold—chosen empirically—against which the deviation of the velocity¹ from its mean over the current molecule is compared. Thus, if this deviation is less than the threshold value, the current atom is conjoined to the current molecule. Each molecule gathers atoms until it encounters a large enough velocity anomaly. The result is an adaptive partition of unity, or what we call a “molecular decomposition.”

In our numerical implementation of the AGPS algorithm, the molecules are defined recursively as follows. Let $\lambda \geq 0$ be a constant, called a threshold, and let the current state of the n^{th} molecule be denoted by \mathcal{M}_n^N , where N is the current number of atoms it contains. If x_{n_1} denotes the center coordinate of the first atom in \mathcal{M}_n^N , we may write

$$\mathcal{M}_n^N(x) = \sum_{j=n_1}^{n_N} g(x - x_j) \quad (7.1)$$

where $n_N = N + n_1 - 1$. Next, define the mean of the velocity over the current molecule by

$$v_{\mathcal{M}_n^N} = \frac{\sum_{j=n_1}^{n_N} v(x_j) \mathcal{M}_n^N(x_j)}{\sum_{j=n_1}^{n_N} \mathcal{M}_n^N(x_j)}. \quad (7.2)$$

The stationarity condition with respect to the given threshold λ is as follows:

if $|v(x_{n_{N+1}}) - v_{\mathcal{M}_n^N}| < \lambda$, then

$$\mathcal{M}_n^N \rightarrow \mathcal{M}_n^{N+1} = \mathcal{M}_n^N + g_{n_{N+1}};$$

otherwise $\mathcal{M}_n^N \rightarrow \mathcal{M}_n^N$ and $\mathcal{M}_{n+1}^1 = g_{n_{N+1}}$.

¹M. Yedlin has suggested using not the velocity variation, but the variation of the ratio of the velocity and its derivative, v/v' . This quantity is related to WKBJ-type length scales, and hence offers a nice bridge to ray theory.

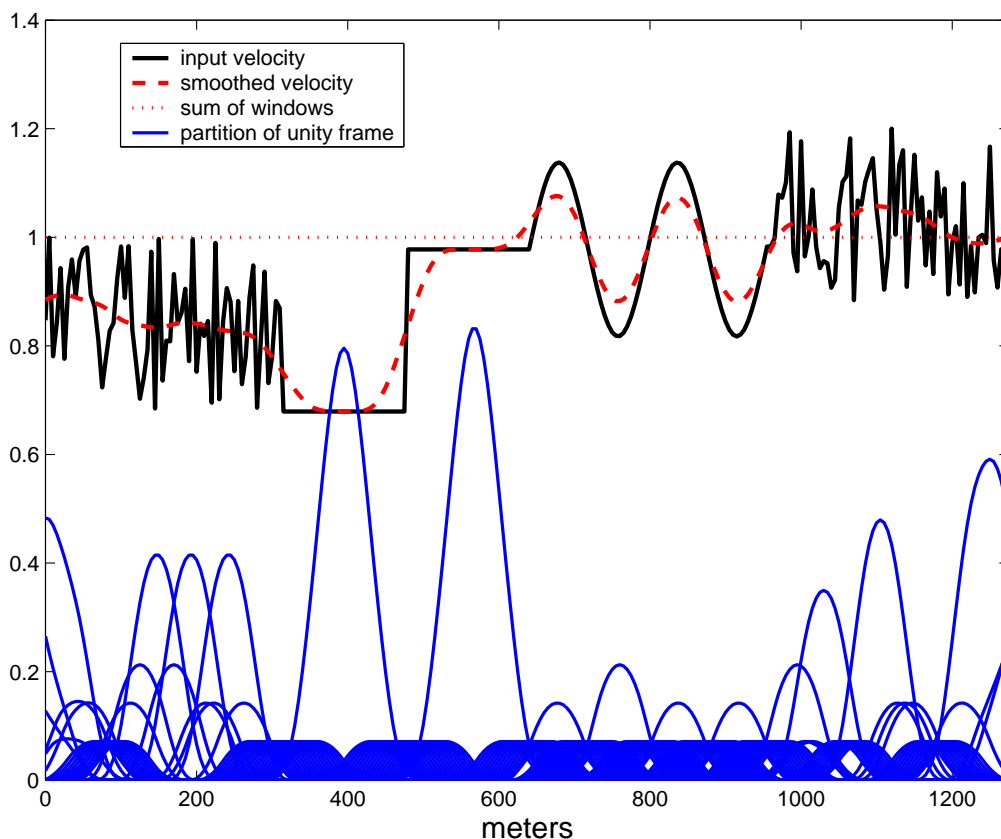


Figure 7.1: An adaptive partition-of-unity frame and its correspondence to the local stationarity of a velocity model. The velocity function has been rescaled for display purposes. The smoothed velocity is displayed as a visual aid in comparing the molecules to the local velocity variations. The larger the local variation in velocity, the smaller and more abundant the molecules become.

Such a molecular decomposition is shown in figure 7.1. The illustrated velocity model is representative of all types of variations, ranging from a simple constant to an everywhere discontinuous random function. The atoms cluster near the larger local variations in the velocity, while everywhere else, the atoms ‘bond’ to form molecules. Their size varies inversely with the magnitude of the local variation of v . The molecules also sum to unity, as desired. The next step is to

use this molecular decomposition to construct a one-parameter family of analysis and synthesis frames for the Gabor transform, indexed by $p \in [0, 1]$. For each fixed p , the analysis windows are defined by raising each molecule to the power p . The corresponding dual frame, consisting of synthesis windows, is defined by raising each molecule in the decomposition to the complementary power, $1 - p$. For example, if the fundamental atom is supported on a finite interval, and $p = 1$, then the molecular decomposition itself forms the analysis frame, and the synthesis windows are boxcars. For $p = 0$, these roles are reversed. However, if the atom is supported on the whole of \mathbf{R} , (for example, a Gaussian) then no synthesis (analysis) windows are used if $p = 1$ ($p = 0$). In either case, the analysis and synthesis frames are identical for $p = 1/2$.

The analysis and synthesis frames, constructed in this way, ensure an overall energy and amplitude preserving transformation to the Gabor domain and back (see [38]). Moreover, with respect to the choice of atom and stationarity measure, the redundancy of the Gabor transform is minimized; hence so is its computation time. Each frequency slice of the input data, $\psi(x, \omega, z = z_0)$ is then Gabor transformed, $(x \rightarrow (\tilde{x}, k_x))$, phase-shifted by $e^{i\Delta z\phi}$, where

$$\phi(\tilde{x}, k_x, \omega) = \sqrt{\frac{\omega^2}{\bar{v}(\tilde{x})^2} - k_x^2}, \quad (7.3)$$

then inverse Gabor transformed. This phase-shift operation results in an estimate of the wavefield, $\psi(x, \omega, z = z + \Delta z)$, at the new depth $z + \Delta z$ and frequency ω . After repeating this process for each ω , an inverse Fourier transform over ω gives the desired estimate in space-time coordinates.

The velocity function $\bar{v}(\tilde{x})$ in the dispersion relation (7.3) is defined, for each

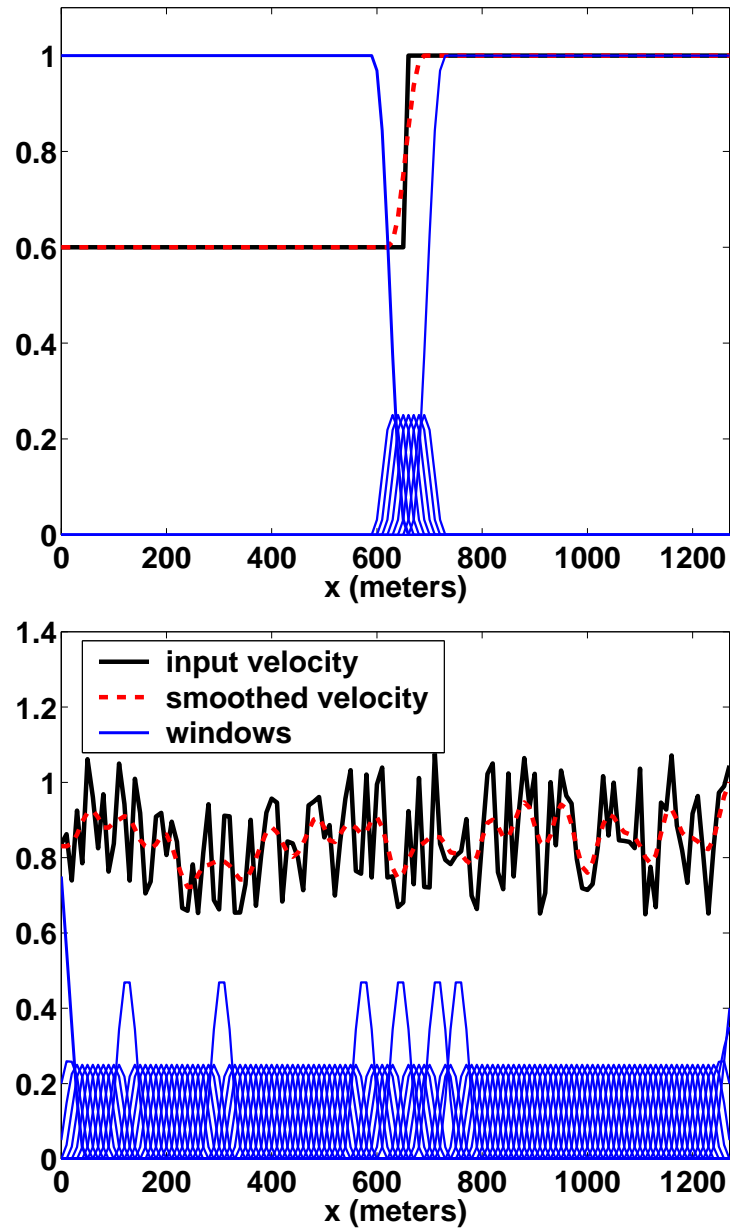


Figure 7.2: Adaptive partition-of-unity frames and their correspondence to the local stationarity of the given velocity models. The velocity functions have been rescaled for display purposes. The smoothed velocities are displayed as a visual aid in comparing the molecules to the local velocity variations. The velocity model on the top jumps from 2250 m/s to 3750 m/s. The velocity model on the bottom fluctuates randomly between 1500 m/s and 2500 m/s.

\tilde{x} , as the mean of the velocity over the molecule centered at \tilde{x} . It thus assigns a constant velocity to each molecule in the Gabor decomposition. A key feature that controls the accuracy of the algorithm is the choice of stationarity measure: a large threshold can lead to large molecules, hence a more coarsely sampled velocity. On the other hand, too strict a measure could lead to higher cost than necessary for the desired accuracy.

Roughly speaking, the balance between the dependence on the input and the output coordinates can be shifted by adjusting the parameter $p \in [0, 1]$. However, the situation is more complicated than this, and has yet to be fully understood. In the limit as the window size shrinks to zero, the cases $p = 1$ and $p = 0$ reduce to the NSPS and PSPI algorithms, respectively. It turns out that for the examples below, which use compactly supported windows, the best choice is $p = 1$. Intuitively, we might expect $p = 1/2$ to be optimal, since this implies a symmetric dependence on the input and output coordinates. Experiment suggests that the optimal value of p depends on the choice of window, but this remains an open question.

7.3 Synthetic examples

Margrave's exact one-way extrapolation algorithm was used as a forward model to upward propagate an impulsive wavefield through a laterally varying velocity field. This impulse response was then inverted, by reverse-extrapolation, using the exact, NSPS, PSPI, and AGPS extrapolators. The relative costs for two extreme-case velocity models appear in Table 1. The first is a step between two constant velocities, (figure 7.2, (top)) and the second involves a velocity function that varies

Table 7.1: Comparison of computation times for various extrapolation algorithms. In contrast to the first three methods, the cost of AGPS is proportional to the complexity of the velocity model. The cost of AGPS spans an order of magnitude between the two extremes, namely, the velocity models depicted in figure 7.2.

Extrapolation algorithm	Absolute cost step/random	Relative cost step/random
Exact - 1 step	47.1/46.3	100/100
NSPS - 1 step	4.0/3.7	8.2/7.9
NSPS - 5 steps	20.6/17.5	42.5/37.9
PSPI - 1 step	4.0/3.7	8.2/8.1
PSPI - 5 steps	20.4/19.2	41.9/41.5
AGPS - 1 step	0.6/3.9	1.2/8.5
AGPS - 5 steps	3.2/19.4	6.4/42.0

randomly at each offset (figure 7.2, (bottom)). In the first case, AGPS outperforms NSPS and PSPI by a factor of six, while all three have similar costs in the second case. The extra cost for the exact algorithm in the random case is small, so its use in complex media may be justified. Figure 7.2 shows the velocity models used in our examples, and the corresponding molecular decompositions of unity. Both decompositions are used as analysis frames (so $p = 1$) for the nonuniform Gabor transform. The fundamental atoms (black) are Lamoureux windows of order two, (twice differentiable polynomial splines) sampled at 7 points, and the spacing between atoms is 10m.

The input wavefield for both examples, shown in figure 7.3 (top), consists of eight bandlimited impulses. Figures 7.3 (middle) and 7.3 (bottom) show the exact upward extrapolations of the input wavefield by 200m, using the velocity models of figures 7.2 (top) and 7.2 (bottom), respectively.

Figure 7.4 shows the results of reverse extrapolating the last two wavefields with the exact operator. These last two results will serve as benchmarks for comparing the quality of the remaining methods. Aside from some minor numerical artifacts, these images are the best one can hope to obtain (see [59]).

Figure 7.5 displays the various results of inverse extrapolating the wavefield of figure 7.3 (middle)—using the step velocity model of figure 7.2 (top)—which should be compared to figure 7.4 (top). The corresponding computation times are listed in Table 7.1. The rows contain pairs of results for PSPI, NSPS, and AGPS, respectively; the first column is for a single depth step of 200m, and the second is for five steps of 40m. PSPI gives a reasonable result in both cases, but it introduces an artificial discontinuity in the wavefield at the interface between the two constant velocity blocks. It is well known that PSPI produces this artefact for large depth steps. The results for NSPS are better, particularly since the fifth impulse is more focused. AGPS yields essentially the same result as NSPS, but shows marginally fewer artefacts in the case of five steps. This makes AGPS a bargain—at least for simple velocity models. Figure 7.6 displays the results, in exactly the same format as for figure 7.5, using the input wavefield of figure 7.3 (bottom) and random velocity model of figure 7.2 (bottom). These should be compared to figure 7.4 (bottom). The corresponding computation times are listed in Table 7.1. The PSPI and NSPS results 7.6 (top left) and 7.6 (middle left) are remarkably different. PSPI again introduces discontinuities in the wavefield at every discontinuity in the velocity—hence the randomly scattered appearance of the extrapolation. The result of 7.6 (middle, left) for NSPS shows “migration smiles” that arise due to the use of incorrect velocities. Increasing the number of

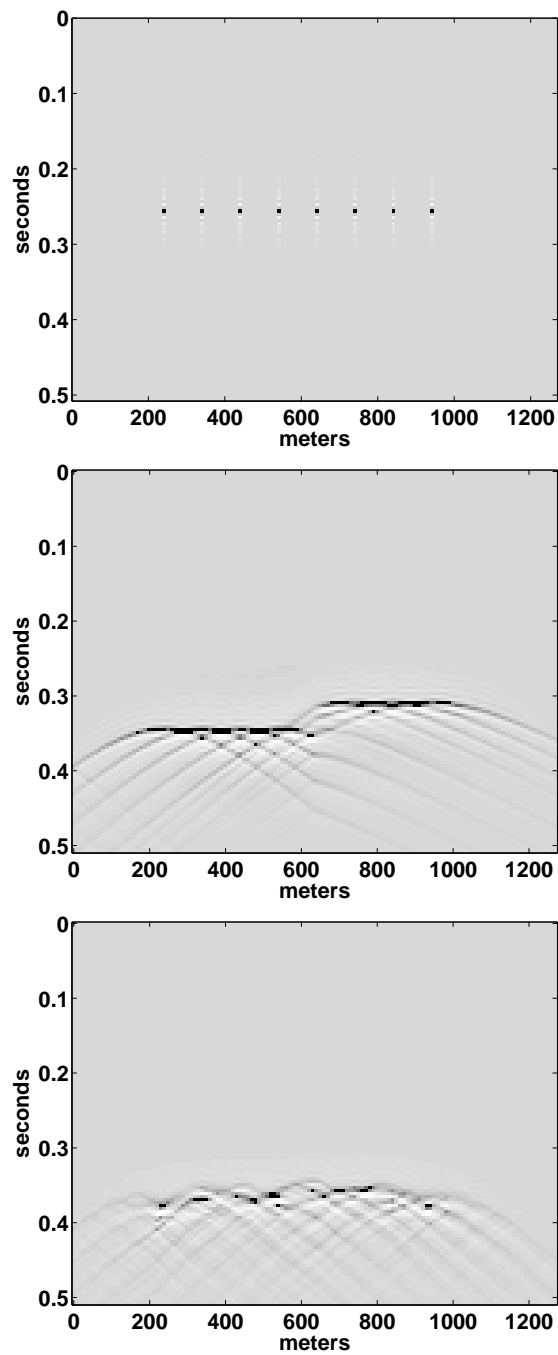


Figure 7.3: Input wavefield (top) for two upward extrapolations by 200m, using the exact extrapolator; the two bottom plots show the resulting wavefields, using the step and random velocity models of figure 7.2, respectively.

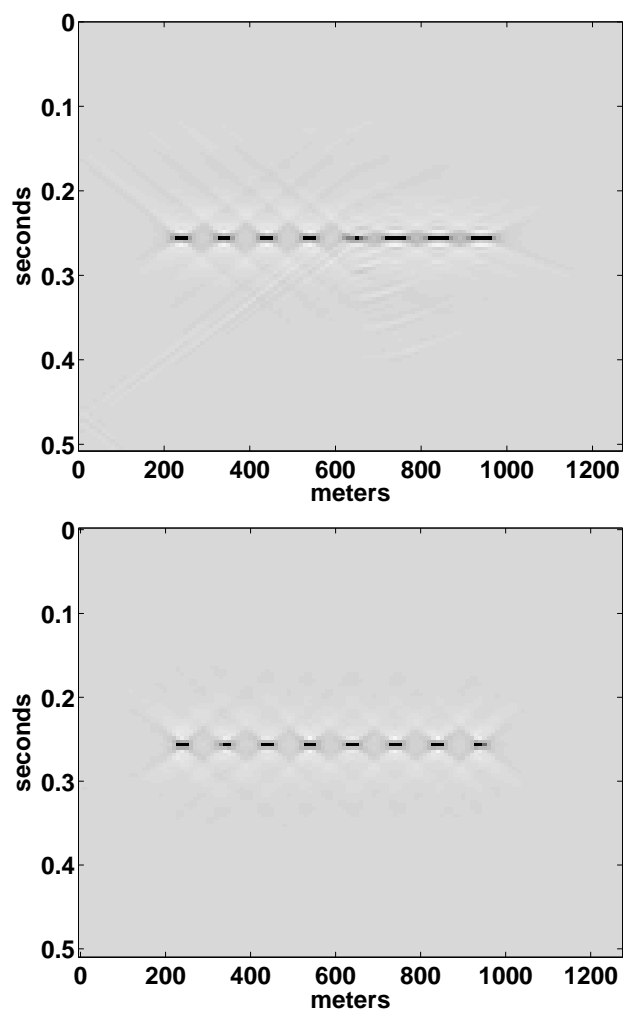


Figure 7.4: Inversion of the corresponding two wavefields at the bottom of figure 7.3, using the exact extrapolator: (top) using the step velocity model of figure 7.2, and (bottom) using the random velocity model of figure 7.2.

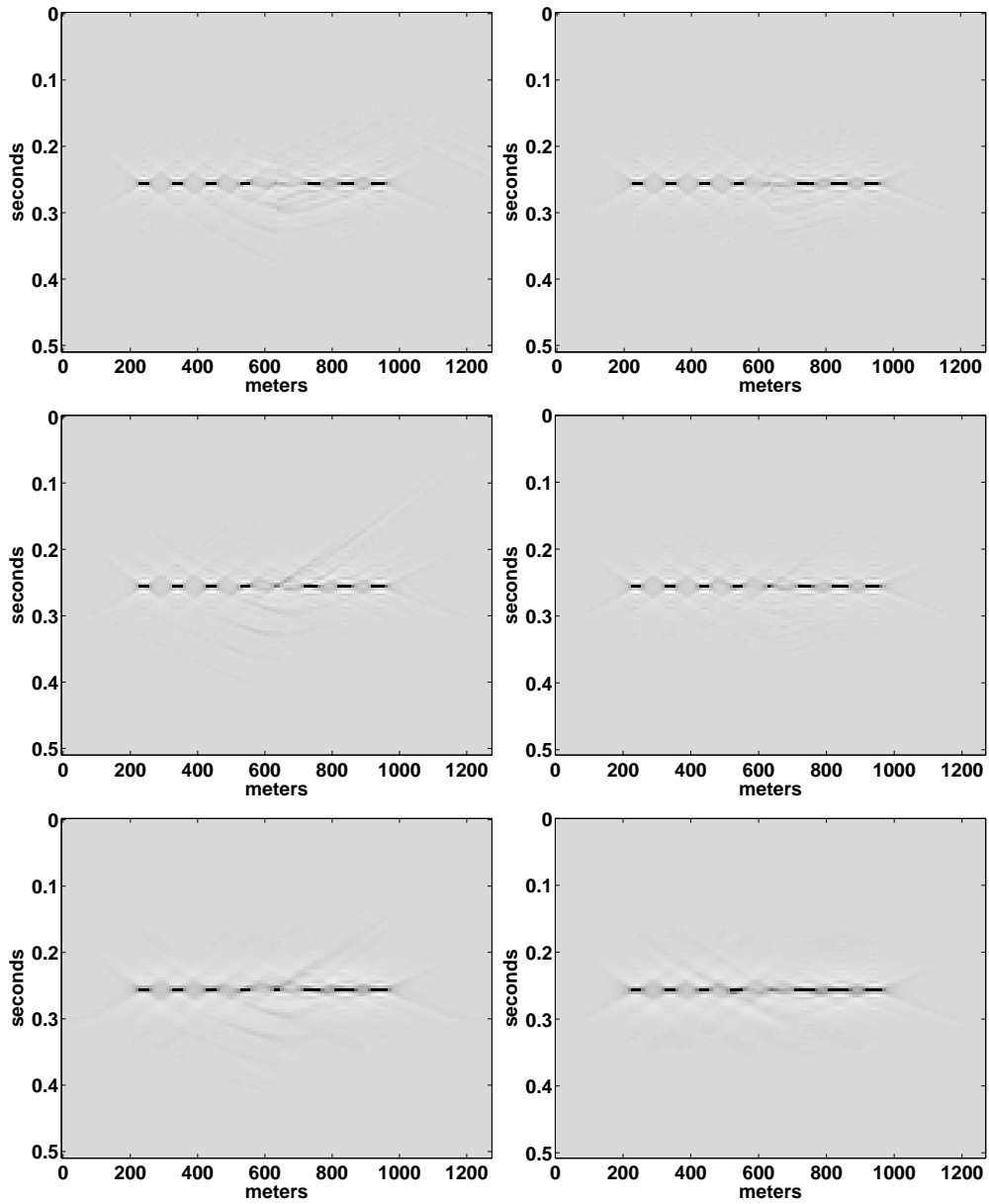


Figure 7.5: Results of inverse extrapolating the wavefield at the bottom of figure 7.3, using the step velocity model of figure 7.2. The corresponding computation times are listed in Table 7.1. The rows contain pairs of results for PSPI, NSPS, and AGPS, respectively; the first column is for a single depth step of 200m, and the second is for five steps of 40m. These results should be compared with the exact results of figure 7.4.

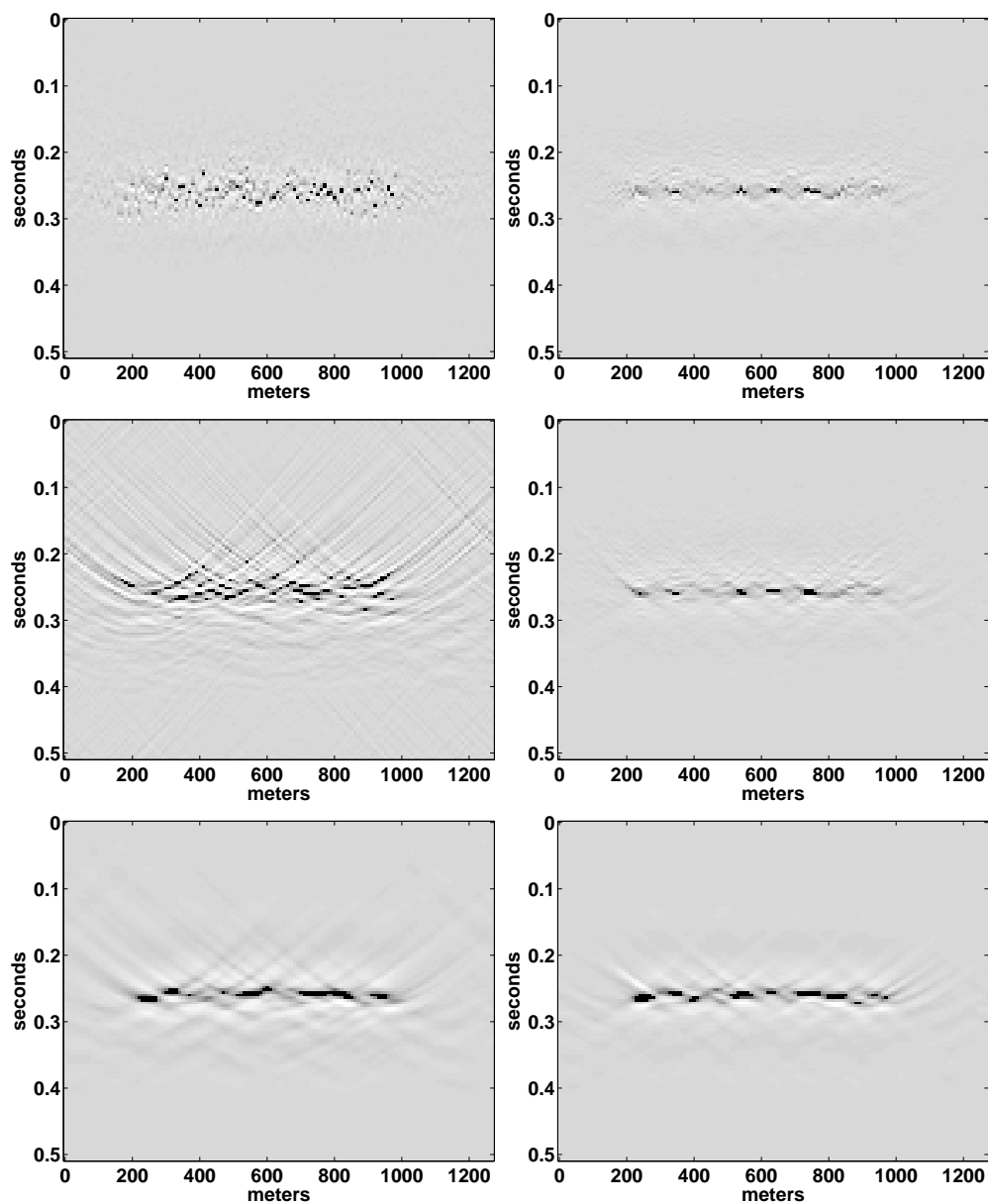


Figure 7.6: Results of inverse extrapolating the wavefield at the right of figure 7.3, using the random velocity model of figure 7.2. The corresponding computation times are listed in Table 7.1. The rows contain pairs of results for PSPI, NSPS, and AGPS, respectively; the first column is for a single depth step of 200m, and the second is for five steps of 40m. These results should be compared with the exact results of figure 7.4.

steps, as in 7.6 (top right) and 7.6 (middle right), helps to cure both of these issues. Again, AGPS yields very similar results as NSPS in both cases, although AGPS is slightly better focused, and contains fewer artefacts. Since they use approximate velocities, all three methods tend to misplace the wavefield in time. This effect is most easily observed near discontinuities in the velocity function.

7.4 Summary

We showed that AGPS propagates a wavefield from a depth level z to a depth level $z + \Delta z$ by applying nonstationary phase-shift filters to an adaptive spatial Gabor transform of the input data.

A family of adaptive Gabor frames was constructed from a molecular decomposition of unity. Each molecule of the frame was built by conjoining neighbouring atoms from a uniform partition of unity—consisting of translates of a single atom along the lateral coordinate—according to a stationarity criterion derived from the velocity model.

Each molecule in the resulting Gabor frame was assigned a mean velocity, and then the phase-shift operator was applied, using these velocities, in the nonuniform Gabor domain.

This process was repeated for each temporal frequency slice, and the desired extrapolation was represented in the space-time domain by applying an inverse Gabor transform, followed by an inverse Fourier transform over frequency.

The cost of AGPS is proportional to both the complexity of the velocity model and the desired degree of accuracy. Its accuracy is comparable to that of NSPS

and PSPI, but its cost is much lower, especially for simple velocity models. For random media, the cost of Margrave's exact algorithm is only marginally greater than the latter three, so its superior quality warrants its use.

7.5 Future work

Each of the approximate wavefield extrapolators suffers from a misplacement of the wavefield in time, hence also in depth, especially near vertical discontinuities in the velocity field. However, this can likely be healed somewhat by using a processing technique similar to that used by Stoffa et al., ([94]). We first rewrite the phase-shift, $\phi\Delta z$, (see expression (7.3)) as the sum of a focusing term, $\phi_f\Delta z$, and a shifting term, $\phi_s\Delta z$, where

$$\phi_f(x, k_x, \omega) = \frac{\omega}{v(x)} \left(\sqrt{1 - \frac{k_x^2}{\omega^2}} - 1 \right) \quad (7.4)$$

and

$$\phi_s(x, \omega) = \frac{\omega}{v(x)}. \quad (7.5)$$

The shifting term is independent of k_x , so it does not vary with propagation direction. Physically, this term is responsible only for a vertical depth shift of the data, while the focusing term serves as an angle-dependent correction. The idea is to precondition the data by applying the phase-shift operator

$$e^{i\phi_s\Delta z} \quad (7.6)$$

in the (x, ω) -domain, before implementing the focusing phase-shift filter in the Gabor domain. The second step amounts to replacing expression (7.3) with

$$e^{i\phi_f\Delta z} \quad (7.7)$$

in the AGPS algorithm.

Chapter 8

Summary

We interpreted the Gabor transform as a natural, nonstationary extension of the Fourier transform. This was the basis for conjecturing that classical Fourier analysis could be adapted to handle nonstationary systems, simply by replacing the Fourier transform with the appropriate Gabor transform. Following this principle motivated the idea of a Gabor filter, the corresponding analogue to a Fourier filter.

We saw that every linear partial differential equation with functional coefficients has associated with it a (non-classical) Kohn-Nirenberg pseudodifferential operator. Moreover, we saw that every linear operator, which includes every Gabor filter, could be represented as a pseudodifferential operator. We observed that a pseudodifferential operator looks much like a classical Fourier filter, but with a transfer function that is not necessarily translation invariant.

Provided that the symbol of the pseudodifferential operator of interest was

reasonably smooth compared to the analysis and synthesis atoms, the Gabor filter defined by the same symbol was deemed to be a good approximation to the original operator.

These arguments helped to justify our approach to the two nonstationary seismic inversion problems considered in this thesis. Both of these inverse problems were based on linear wave-propagation models, with mathematical descriptions given as Kohn-Nirenberg pseudodifferential operators. We then developed algorithms to solve these inverse problems within the framework of Gabor filter theory.

We considered the nonstationary Gabor deconvolution problem, based on a pseudodifferential operator generalization of the convolutional model for a seismic trace—in which the embedded causal wavelet factored as the product of a stationary source signature and a nonstationary exponential decay.

Analytic expressions were derived for least-squares fitting the model to the Gabor-transform of the seismic trace, which yielded excellent estimates of Q and the source signature, hence an estimate of the travelling wavelet.

These estimates lead to a smoothed version of the magnitude of the Gabor spectrum of the seismic trace, from which a least-squares nonstationary minimum-phase deconvolution filter was easily constructed. Our results on synthetic data are very promising.

The assumption of a minimum phase seismic wavelet is central to the Wiener and Gabor deconvolution algorithms. The former assumes a stationary wavelet, while the latter explicitly involves a nonstationary wavelet, and for this reason an extension of the concept of minimum phase to this setting was developed.

Although the concept of minimum phase is well understood for the case of

digital signals, we felt its extension to analog signals had not received sufficient attention in the literature. We showed that certain well-known properties of digital minimum phase signals, such as convolutional invertibility, stability, and the Hilbert transform relationships between Fourier amplitude and phase spectra, only hold under certain conditions in the analog setting.

We attempted to overcome these issues by developing a mathematical theory of minimum phase in the necessarily general context of tempered distributions. This in turn allowed us to develop a theory of nonstationary, minimum-phase filters—those pseudodifferential operators which preserve minimum phase.

Our second application of Gabor filtering theory to seismic inversion was based on the premise that wavefield extrapolation for a laterally varying velocity model could be achieved by applying a nonstationary phase-shift filter to an adaptive, nonuniform Gabor transform over the lateral coordinate.

The resulting algorithm, called adaptive Gabor phase-shift, (AGPS) made use of a new method—which we also presented—for constructing adaptive, nonuniform Gabor frames from maximally redundant, uniform partitions of unity. The main purpose of the construction was to minimize the redundancy of the Gabor transform, and thus significantly reduce its computational intensity.

We found that AGPS had a computational cost that was proportional to the complexity of the velocity model, while its accuracy was comparable to the NSPS and generalized PSPI algorithms. Also, AGPS included NSPS and PSPI as limiting cases, yet the cost of AGPS ranged from an order of magnitude less to about the same order.

The NSPS and PSPI filters differ to the extent that the functional dependence

of the velocity field of the medium is either on the input coordinates (NSPS) or the output coordinates (PSPI). However, we pointed out that any accurate phase-shift operator in a $v(x)$ medium must allow for velocity variations along the trajectory of a ray. AGPS attempted to address this problem by representing the input and output wavefields as superpositions of windowed components, each of which was approximately stationary with respect to the velocity.

A general overview of analog linear systems from the perspective of functional analysis was provided to support the remainder of the thesis. Topics covered included function spaces for representing signals, the generalized function space of tempered distributions, linear operators acting in Hilbert spaces, the Fourier transform, and distributional derivatives.

Bibliography

- [1] Ahlfors, L. V., 1979, Complex analysis: an introduction to the theory of analytic functions of one complex variable: McGraw-Hill.
- [2] Aki, K. and Richards, P.G., 1980, Quantitative seismology, theory and methods: W.H. Freeman and Co., Vol. 1 and 2.
- [3] Arnold, V. I., 1989, Mathematical methods of classical mechanics: Springer-Verlag.
- [4] Averbuch, A., Braverman, E., Coifman, R., Israeli, M., and Sidi, A., 2000, Efficient computation of oscillatory integrals via adaptive multiscale local Fourier bases, Applied and computational harmonic analysis, **9**, 19-53.
- [5] Bale, R. A., Grossman, J. P., Margrave, G. F., and Lamoureux, M. P., 2002, Multidimensional partitions of unity and Gaussian terrain: CREWES Research Report, **14**.
- [6] Batzle, M., Han, D.-H. and Castagna, J., 1996, Attenuation and velocity dispersion at seismic frequencies, 66th Ann. Internat. Mtg: Soc. of Expl. Geophys., 1687-1690.
- [7] Busby, R. C. and Smith, H. A., 1981, Product-convolution operators and mixed-norm spaces: Trans. Amer. Math. Soc., **263**, 309-341.
- [8] Cheney, W., 2001, Analysis for applied mathematics: Springer-Verlag.

- [9] Churchill, R. V., and Brown, J. W., 1984, Complex variables and applications: McGraw-Hill.
- [10] Claerbout, J. F., 1985, Fundamentals of geophysical data processing: Blackwell.
- [11] Connes, A., 1994, Noncommutative Geometry: Academic Press.
- [12] Conway, J. B., 1990, A course in functional analysis, 2nd Ed.: Springer-Verlag.
- [13] Dales, H. G., 1983, Convolution algebras on the real line: in Bachar, J. M., et al., (Eds) Radical Banach algebras and automatic continuity, Lecture Notes in Mathematics, **975**, Springer-Verlag, 180-209.
- [14] Daubechies, I., Grossmann, A., and Meyer, Y., 1986, Painless nonorthogonal expansions: J. Math. Phys., **27**, No. 5, 1271-1283.
- [15] Davidson, K., 1996, C^* -algebras by example, Fields Inst. Mono., **6**: American Mathematical Society.
- [16] Dey, A. K., 1999, An analysis of seismic wavelet estimation, M. Sc. thesis: University of Calgary.
- [17] Dixmier, J., 1977, C^* -algebras, North-Holland Mathematical Library, Vol. **15**, North-Holland.
- [18] Dunford, N., and Schwartz, J. T., 1958, Linear operators, Vol. I and II: Interscience.

- [19] Feichtinger, H. G. and Nowak, K., 2002, A first survey of Gabor multipliers, in *Advances in Gabor analysis*, H. G. Feichtinger and T. Strohmer: Birkhäuser.
- [20] Feichtinger, H. G. and Strohmer, T., eds., 1998, *Gabor analysis and algorithms, theory and applications*: Birkhäuser.
- [21] Feichtinger, H. G. and Strohmer, T., eds., 2002, *Advances in Gabor analysis*: Birkhäuser.
- [22] Flanagan, J. L., and Golden, R. M., 1966, Phase Vocoder: *Bell System Technical Journal*, November, 1493-1509.
- [23] Folland, G. B., 1989, *Harmonic analysis in phase space*: Princeton University Press.
- [24] Futterman, W. I., 1962, Dispersive body waves: *J. Geophys. Res.*, **67**, 5279-5291.
- [25] Gabor, D., 1946, Theory of communication: *J. IEEE*, **93**, 429–457.
- [26] Gazdag, J., 1978, Wave equation migration with the phase-shift method: *Geophysics, Soc. of Expl. Geophys.*, **43**, 1342-1351.
- [27] Gazdag, J., and Sguazero, P., 1984, Migration of seismic data by phase shift plus interpolation: *Geophysics*, **49**, 124-131.
- [28] Gel'fand, I. M., Graev, M. I., and Vilenkin, N. Ya., 1966, *Generalized functions*, Vol. 5, *Integral geometry and representation theory*, E. Saletan, translator: Academic Press.

- [29] Gel'fand, I. M., and Shilov, G. E., 1964, Generalized functions, Vol. 1: Properties and operations, E. Saletan, translator: Academic Press.
- [30] Gel'fand, I. M., and Shilov, G. E., 1967, Generalized functions, Vol. 3: Theory of differential equations, M. E. Mayer, translator: Academic Press.
- [31] Gel'fand, I. M., and Shilov, G. E., 1968, Generalized functions, Vol. 2: Spaces of fundamental and generalized functions, M. D. Friedman, A. Feinstein, and C. P. Peltzer, translators: Academic Press.
- [32] Gel'fand, I. M., and Vilenkin, N. Ya., 1964, Generalized functions, Vol. 4, Applications of harmonic analysis, A. Feinstein, translator: Academic Press.
- [33] Gibson, P. C., Lamoureux, M. P., and Margrave, M. F., 2003, Representation of linear operators by Gabor multipliers: CREWES Research Report, **15**.
- [34] Grant, F. S., and West, G. F., 1965, Interpretation theory in applied geophysics: McGraw-Hill.
- [35] Griffiths, L. J., Smolka, F. R. and Trembly, L. D., 1977, Adaptive deconvolution - A new technique for processing time-varying seismic data: Geophysics, Soc. of Expl. Geophys., **42**, 742-759.
- [36] Gröchenig, K., 2001, Foundations of time-frequency analysis: Birkhäuser.
- [37] Grossman, J. P., 1998, An introduction to the spectral calculus of noncommutative geometry, M.Sc. Thesis: University of Calgary.

- [38] Grossman, J. P., Margrave, G. F., and Lamoureux, M. P., 2002, Constructing adaptive nonuniform Gabor frames from partitions of unity: CREWES Research Report, **14**.
- [39] Grossman, J. P., Margrave, G. F., and Lamoureux, M. P., 2002, Fast wavefield extrapolation by phase-shift in the nonuniform Gabor domain: CREWES Research Report, **14**.
- [40] Grossman, J. P., Margrave, G. F., Lamoureux, M. P., 2003, Adaptive, nonuniform Gabor frames from partitions of unity: in review, submitted to The Journal of Computational and Applied Mathematics.
- [41] Halmos, P. R., 1967, A Hilbert space problem book: Van Nostrand.
- [42] Hayes, M. H., 1987, The unique reconstruction of multidimensional sequences from Fourier transform magnitude or phase, in Image recovery: theory and application, H. Stark, ed.: Academic Press.
- [43] Heil, C., 2003, Integral operators, pseudodifferential operators, and Gabor frames: in Advances in Gabor analysis, Feichtinger, H. G., and Strohmer, T., Editors: Birkhäuser.
- [44] Heisenberg, W., 1927, Über die Grundprinzipien der Quantenmechanik: FF 3, no. 11, **83**, translated to English in 1984: The physical content of quantum kinematics and mechanics, John Archibald Wheeler and Wojciek Hubert Zurek, eds. Quantum theory and measurement: Princeton University Press, 62–84.

- [45] Kadison, R. V., and Ringrose, J. R., 1983, Fundamentals of the theory of operator algebras, volume I and II: Academic Press.
- [46] Kjartansson, E., 1979, Constant Q – wave propagation and attenuation, *J. Geophys. Res.*, **84**, 4737-4748. Also, reprinted in *Seismic wave attenuation*, 1981, M. N. Toksöz and D. H. Johnston eds.: Society of Exploration Geophysicists, 448-459.
- [47] Kohn, J. J., and Nirenberg, L., 1965, An algebra of pseudo-differential operators: *Comm. Pure Appl. Math.*, **18**, 269-305.
- [48] Kozek, W., 1997, On the transfer function calculus for underspread LTV channels, *IEEE Trans. Signal Processing*, **45**, 219-223.
- [49] Lamoureux, M. P., Gibson, P. C., Grossman, J. P., and Margrave, G. F., 2003, A fast, discrete Gabor transform via a partition of unity: CREWES Research Report, **15**.
- [50] Lange, S., 1968, *Analysis I*: Addison-Wesley.
- [51] Lange, S., 1969, *Analysis II*: Addison-Wesley.
- [52] Lang, S., 1977, *Complex analysis*: Addison-Wesley.
- [53] Li, X. G. and Ulrych, T. J., 1996, Coherent noise filtering using a 2-D Gabor transform, 66th Ann. Internat. Mtg: Soc. of Expl. Geophys., 1180-1183.
- [54] Margrave, G. F., 1998, Theory of nonstationary linear filtering in the Fourier domain with application to time-variant filtering: *Geophysics*, **63**, 244-259.

- [55] Margrave, G. F. and Ferguson, R.J., 1998, Nonstationary filters, pseudodifferential operators, and their inverses: CREWES Research Report, **10**.
- [56] Margrave, G. F. and Ferguson, R.J., 1999, Wavefield extrapolation by nonstationary phase shift: *Geophysics*, **64**, 1067-1078.
- [57] Margrave, G. F., Gibson, P., Grossman, J. P., Henley, D. C., Illiescu, V., and Lamoureux, M. P., 2004, The Gabor transform, pseudodifferential operators, and seismic deconvolution: *Integrated Computer-Aided Engineering*, **9**, 1–13.
- [58] Margrave, G. F. and Lamoureux, M. P., 2001, Gabor deconvolution: CREWES Research Report, **13**, 241–276.
- [59] Margrave, G. F., Lamoureux, M. P., Gibson, P., Bale, R. A., and Grossman, J. P., 2002, Exact wavefield extrapolation in 2D for $v(x)$: CREWES Research Report, **14**.
- [60] Margrave, G. F., Lamoureux, M.P., Grossman, J.P., and Illiescu, V. 2002, Gabor deconvolution of seismic data for source waveform and Q correction: SEG Convention expanded abstracts, 2190-2193.
- [61] Marion, J. B., and Thornton, S. T., 1988, *Classical dynamics of particles and systems*: Harcourt Brace Jovanovich.
- [62] Matz, G., and Hlawatsch, F., 1998, Extending the transfer function calculus of time-varying linear systems: a generalized underspread theory, *Proc. IEEE ICASSP-98*, 2189-2192.

- [63] Matz, G., and Hlawatsch, F., 1998, Time-frequency transfer function calculus (symbolic calculus) of linear time-varying systems (linear operators) based on a generalized underspread theory: *J. Math. Phys.*, Special issue on wavelet and time-frequency analysis, **39**, 4041-4071.
- [64] Matz, G., and Hlawatsch, F., 2002, Linear time-frequency filters: online algorithms and applications, *in* Papandreou-Suppappola, A., ed., *Applications in time-frequency signal processing*: CRC Press.
- [65] McGillem, C. D., and Cooper, G. R. 1974, *Continuous and discrete signal and system analysis*: Holt, Rinehart and Winston.
- [66] Mosak, R. D., 1975, *Banach algebras*: Chicago Press.
- [67] Oppenheim, A. V., and Schaffer, R. W., 1975, *Digital Signal Processing*: Prentice-Hall.
- [68] Paley, R. E. A. C, and Wiener, N., 1967, *Fourier transforms in the complex domain*: Am. Math. Soc. Colloquium publications, **19**. Originally published in 1934.
- [69] Papoulis, A., 1962, *The Fourier integral and its applications*: McGraw-Hill.
- [70] Papoulis, A., 1977, *Signal analysis*: McGraw-Hill.
- [71] Pedersen, G. K., 1979, *C*-algebras and their automorphism groups*: Academic Press.
- [72] Pujol, J., 2003, *Elastic wave propagation and generation in seismology*: Cambridge University Press.

- [73] Qian, S., 2002, Introduction to time-frequency and wavelet transforms: Prentice Hall.
- [74] Reed, M., 1993, Adaptive least-squares filtering: A method for deconvolution of nonstationary seismic data, 63rd Ann. Internat. Mtg: Soc. of Expl. Geophys., 1112-1115.
- [75] Reed, M., and Simon, B., 1980, Methods of modern mathematical physics: Academic Press.
- [76] Robinson, E. A., 1962, Random wavelets and cybernetic systems: Hafner.
- [77] Robinson, E. A., 1967, Statistical communication and detection: with special reference to digital data processing of radar and seismic signals: Hafner.
- [78] Robinson, E. A., and Treitel, S., 1980, Geophysical signal analysis: Prentice-Hall.
- [79] Rochberg, R., and Tachizawa, K., 1998, Pseudodifferential operators, Gabor frames, and local trigonometric bases, in Advances in Gabor analysis, H. G. Feichtinger and T. Strohmer, editors: Birkhäuser.
- [80] Royden, H. L., 1988, Real analysis: Prentice Hall.
- [81] Rudin, W., 1973, Functional analysis: McGraw-Hill.
- [82] Rudin, W., 1976, Principles of mathematical analysis, 3rd Ed.: McGraw-Hill.
- [83] Rushforth, C. K., 1987, Signal restoration, functional analysis, and Fredholm integrals of the first kind, in Image recovery: theory and application, H. Stark,

ed.: Academic Press.

- [84] Saint Raymond, X., 1991, Elementary introduction to the theory of pseudodifferential operators: CRC Press.
- [85] Schoepf, A. R. and Margrave, G. F., 1998, Improving seismic resolution with nonstationary deconvolution, 68th Ann. Internat. Mtg: Soc. of Expl. Geophys., 1096-1099.
- [86] Schwartz, L., 1950, Théorie des noyaux: Proceedings of the International Congress of Mathematicians, **1**, 220-230.
- [87] Schwartz, L., 1957, Théorie des Distributions, 2nd Ed.: Hermann.
- [88] Sheriff, R. E., 1991, Encyclopedic dictionary of exploration geophysics: Society of Exploration Geophysicists.
- [89] Simon, B., 1979, Trace ideals and their applications: London Mathematical Society Lecture Notes, Vol. 35: Cambridge University Press.
- [90] Soutas-Little, R.W., 1973, Elasticity: Dover.
- [91] Spratt, S., 1988, Adaptive processing using Gabor transformation, 58th Ann. Internat. Mtg: Soc. of Expl. Geophys., Session: S22.2.
- [92] Steeghs, T. P. H. and Drijkoningen, G. G., 1994, Joint time-frequency analysis of seismic data, 56th Mtg.: Eur. Assn. of Expl. Geophys., Session: P163.
- [93] Stein, E. M., 1993, Harmonic analysis: real-variable methods, orthogonality, and oscillatory integrals: Princeton University Press.

- [94] Stoffa, P. L., Fokkema, J. T., de Luna Freire, R. M., and Kessinger, W. P., 1990, Split-step Fourier migration: *Geophysics*, **55**, 410-421.
- [95] Stolt, R.H., 1978, Migration by Fourier transform: *Geophysics*, **43**, 23-48.
- [96] Taylor, M. E., 1981, *Pseudodifferential operators*: Princeton University Press.
- [97] Titchmarsh, E. C., 1937, *Introduction to the theory of Fourier integrals*:, Oxford University Press.
- [98] Tobback, T., Steeghs, P., Drijkoningen, G. G. and Fokkema, J. T., 1996, Decomposition of seismic signals via time-frequency representations, 66th Ann. Internat. Mtg: Soc. of Expl. Geophys., 1638-1641.
- [99] Toll, J., 1956, Causality and the dispersion relation: logical foundations: *Phys. Rev.*, **104**, 1760-1770.
- [100] Treitel, S., and Lines, L. R., 1982, Linear inverse theory and deconvolution: *Geophysics*, **47**, 1153-1159.
- [101] Tyler, D. W., and Matson, C. L., 1997, Reduction of nonstationary noise in telescope imagery using a support constraint: *Optics Express*, Vol. 1, No. 11.
- [102] Vio, R., Feichtinger, H.G., and Wamsteker, W., 2001, Joint time-frequency analysis: a valuable tool for practical signal processing: preprint.
- [103] Wang, R. J., 1969, The determination of optimum gate lengths for time-varying Wiener filtering : *Geophysics*, Soc. of Expl. Geophys., **34**, 683-695. (Errata in GEO-35-1-0190)

- [104] Wang, R. J., 1977, Adaptive predictive deconvolution of seismic data: Geophys. Prosp., Eur. Assn. Geosci. Eng., **25**, 342-381.
- [105] Weyl, H., 1950, The theory of groups and quantum mechanics: Dover.
- [106] Widder, D. V., 1971, An introduction to transform theory: Academic Press.
- [107] Williams, D. F., and Alpert, B. K., 2001, Causality and wave-guide circuit theory: IEEE Microwave Theory and Techniques, **49**, 615-623.
- [108] Wong, M. W., 1998, Weyl transforms: Springer-Verlag.
- [109] Wu, R-S. and Chen, L., 2001, Beamlet Migration Using Gabor-Daubechies Frame Propagator, 63rd Mtg.: Eur. Assn. of Expl. Geophys., Session: P074.
- [110] Yilmaz, Ö., 1987, Seismic data processing: Society of Exploration Geophysicists.
- [111] Zadeh, L. A., 1950, The determination of the impulsive response of variable networks: J. Applied Physics, **21**, 642-645.
- [112] Zadeh, L. A., 1950, Frequency analysis of variable networks: Proc. of IRE, **38**, 291-299.

Index

- adaptive frame, 98
- adaptive partition of unity, 96
- K^* , 20
- almost everywhere, 12
- analysis mapping, 48
- analytic, 10
- atom, 2
- $\mathcal{B}(\mathcal{H})$, 19
- Bromwich path, 30
- \mathbf{C} , 10
- causal distribution, 24
- C^∞ , 13
- \mathbf{C}_- , 30
- continuous Gabor transform, 47
- convolution operator, 41
- D^α , 17
- δ_+ , δ_- , 27
- discrete Gabor transform, 2, 48
- Dom K , 19
- energy, 11
- entire, 10
- ess sup, 12
- filter, 41
- \mathcal{F}^{-1} , 15
- Fourier transform, 14
- f_+ , f_- , 26
- Gabor frame, 48
- Gabor transform, 48
- h, 25
- \hat{f} , 14
- \hat{T} , 18
- \mathcal{H} , 24
- Hilbert transform, 6, 24, 27
- idempotent, 20
- \mathfrak{S} , 25
- \langle, \rangle , 11
- isometry, 20
- Ker W , 21
- Kohn-Nirenberg, 73
- Kohn-Nirenberg symbol, 73
- $\mathcal{L}^2(\mathbf{R}^d)$, 11

$\mathcal{L}^2(\mathbf{R})_+, \mathcal{L}^2(\mathbf{R})_-, 26$ $\ell^2, 46$ $\mathcal{L}^p(\mathbf{R}^d), 12$

measure zero, 12

minimum delay, 31

minimum phase, 30

 $M_{mb}, 45$

modulation operator, 45

molecular decomposition, 59, 96, 102

molecule, 59

 $M_F^{g,\gamma}, 65$

multiindex, 13

multiplication operator, 73

 $\mathbf{N}, 28$ $\|f\|_2, 11$ $|\alpha|, 16$ $|\phi|_k, 16$ $\|f\|_p, 12$ $N_F, 65$ $\|K\|, 19$

orthogonal, 11

 $\overline{g(x)}, \bar{g}, 11$

partial isometry, 21

 $\partial^\alpha, 13$ $\partial_k^{\alpha_k}, 13$

partition of unity, 44

 p -norm, 12

polynomial growth, 16

product-convolution operator, 73

projection, 20

pseudodifferential operator, 73

 $\mathbf{R}^d, 10$

rapidly decreasing function, 13

 $\Re, 25$ $\mathcal{S}(\mathbf{R}^d), 13$

Schwartz class, 13

self adjoint, 20

short-time Fourier transform, 2

signal, 10

 $\mathcal{S}'(\mathbf{R}^d), 16$ $S_g, 49$

*, 15

stationary, 41

STFT, 2

 $\text{Supp}(f), 10$ $\text{Supp}(T), 16$

synthesis mapping, 48

tempered distribution, 16

tempered function, 16

tensor product, 45

\tilde{s} , 26

time reversal, 26

\otimes , 45

T_{na} , 45

translation operator, 45

unitary, 20

V_γ^* , 49

V_g , 47

V_g^* , 48

x^α , 13

$\mathbf{Z}, \mathbf{Z}^d, \mathbf{Z}_+^d$, 13