A Wavelet Tour of Signal Processing

Stéphane Mallat

September 3, 2008

À la mémoire de mon père, Alexandre.
Pour ma mère, Francine.

Preface to the Sparse Edition

I can not help but find striking resemblances between scientific communities and schools of fish. We interact in conferences and through articles, we move together while a global trajectory emerges from individual contributions. Some of us like to be at the center of the school, others prefer to wander around, and few swim in multiple directions in front. To avoid dying by starvation in a progressively narrower and specialized domain, a scientific community needs to move on. Computational harmonic analysis is still well alive because it goes beyond wavelets. Writing such a book is about decoding the trajectory of the school, and gathering the pearls that have been uncovered on the way. Wavelets are not any more the central topic, despite the original title. It is just an important tool, as the Fourier transform is. Sparse representation and processing are now at the core.

In the 80's, many researchers were focused on building time-frequency decompositions, trying to avoid the uncertainty barrier, and hoping to discover the ultimate representation. Along the way came the construction of wavelet orthogonal bases, which opened new perspectives through collaborations with physicists and mathematicians. Designing orthogonal bases with Xlets became a popular sport, with compression and noise reduction applications. Connections with approximations and sparsity also became more apparent. The search for sparsity has taken over, leading to new grounds, where orthonormal bases are replaced by redundant dictionaries of waveforms. Geometry is now also becoming more apparent through sparse approximation supports in dictionaries.

During these last 7 years, I also encountered the industrial world. With a lot of naiveness, some bandlets and more mathematics, we created a start-up with Christophe Bernard, Jérome Kalifa and Erwan Le Pennec. It took us some time to learn that in 3 months good engineering should produce robust algorithms that operate in real time, as opposed to the 3 years we were used to have for writing new ideas with promissing perspectives. Yet, we survived because mathematics is a major source of industrial innovations for signal processing. Semi-conductor technology offers amazing computational power and flexibility. However, ad-hoc algorithms often do not scale easily and mathematics accelerates the trial and error development process. Sparsity decreases computations, memory and data communications. Although it brings beauty, mathematical understanding is not a luxury. It is required by increasingly sophisticated information processing devices.

New Additions Putting sparsity at the center of the book implied rewriting many parts and adding sections. Chapter 12 is new and introduces sparse representations in redundant dictionaries, with inverse problems, super-resolution and compressive sensing. Here is a small catalogue of new elements in this third edition.

- Radon transform and tomography.
- Lifting for wavelets on surfaces, bounded domains and fast computations.
- JPEG-2000 image compression.
- Block thresholding for denoising.
- Geometric representations with adaptive triangulations, curvelets and bandlets.
- Sparse approximations in redundant dictionaries with pursuits algorithms.
- Noise reduction with model selection, in redundant dictionaries.
- Recovery of sparse approximation supports in incoherent dictionaries.
- Inverse problems and super-resolution.
- Compressive sensing.
- Source separation.

Teaching This book is intended as a graduate textbook. Its evolution is also the result of teaching courses in electrical engineering and applied mathematics. A new web site provides softwares for reproducible experimentations, exercise solutions, together with teaching material such as slides with figures, and Matlab softwares for numerical classes: *http://wavelet-tour.com*.

More exercises have been added at the end of each chapter, ordered by level of difficulty. Level$^{\mathbf{1}}$ exercises are direct applications of the course. Level$^{\mathbf{2}}$ requires more thinking. Level$^{\mathbf{3}}$ includes some technical derivations. Level$^{\mathbf{4}}$ are projects at the interface of research, that are possible topics for a final course project or an independent study. More exercises and projects can be found in the web site.

Sparse Course Programs The Fourier transform and analog to digital conversion through linear sampling approximations provide a common ground for all courses (Chapters 2 and 3). It introduces basic signal representations, and reviews important mathematical and algorithmic tools needed afterwards. Many trajectories are then possible to explore and teach sparse signal processing. The following list gives several topics that can orient the course structure, with elements that can be covered along the way.

Sparse representations with bases and applications
- Principles of linear and non-linear approximations in bases (Chapter 9).
- Lipschitz regularity and wavelet coefficients decay (Chapter 6).
- Wavelet bases (Chapter 7).
- Properties of linear and non-linear wavelet basis approximations (Chapter 9).
- Image wavelet compression (Chapter 10).
- Linear and non-linear diagonal denoising (Chapter 11).

Sparse time-frequency representations
- Time-frequency wavelet and windowed Fourier ridges for audio processing (Chapter 4).
- Local cosine bases (Chapter 8).
- Linear and non-linear approximations in bases (Chapter 9).
- Audio compression (Chapter 10).
- Audio denoising and block thresholding (Chapter 11).
- Compression and denoising in redundant time-frequency dictionaries, with best bases or pursuit algorithms (Chapter 12).

Sparse signal estimation
- Bayes versus minimax, and linear versus non-linear estimations (Chapter 11).
- Wavelet bases (Chapter 7).
- Linear and non-linear approximations in bases (Chapter 9).
- Thresholding estimation (Chapter 11).
- Minimax optimality (Chapter 11).
- Model selection for denoising in redundant dictionaries (Chapter 12).
- Compressive sensing (Chapter 12).

Sparse compression and information theory
- Wavelet orthonormal bases (Chapter 7).
- Linear and non-linear approximations in bases (Chapter 9).
- Compression and sparse transform codes in bases (Chapter 10).
- Compression in redundant dictionaries (Chapter 12).
- Compressive sensing (Chapter 12).
- Source separation (Chapter 12).

Dictionary representations and inverse problems
- Frames and Riesz bases (Chapter 5).

- Linear and non-linear approximations in bases (Chapter 9).
- Ideal redundant dictionary approximations (Chapter 12).
- Pursuit algorithms and dictionary incoherence (Chapter 12).
- Linear and thresholding inverse estimators (Chapter 12).
- Super-resolution and source separation (Chapter 12).
- Compressive sensing (Chapter 12).

Geometric sparse processing

- Time-frequency spectral lines and ridges (Chapter 4).
- Frames and Riesz bases (Chapter 5).
- Multiscale edge representations with wavelet maxima (Chapter 6).
- Sparse approximation supports in bases (Chapter 9).
- Approximations with geometric regularity, curvelets and bandlets (Chapters 9 and 12).
- Sparse signal compression and geometric bit budget (Chapters 10 and 12).
- Recovery of sparse approximation supports in incoherent dictionaries, and super-resolution (Chapter 12).

Acknowledgments Some things do not change with new editions, in particular the traces left by the ones that were, and remain for me wonderful references. As always, I am deeply grateful to Ruzena Bajcsy and Yves Meyer.

I spent the last few years, with three brilliant and kind colleagues, Christophe Bernard, Jérome Kalifa, and Erwan Le Pennec, in a pressure cooker called a start-up. Pressure means stress, despite very good moments. The resulting sauce was a blend of what all of us could provide, and which brought new flavors to our personalities. I am thankful to them for the ones I got, some of which I am still discovering.

This new edition is the result of a collaboration with Gabriel Peyré, who made these changes not only possible, but also very interesting to do. I thank him for his remarkable work and help.

Stéphane Mallat

Contents

11 Denoising 383

12 Sparse in Redundant Dictionaries 441

A Mathematical Complements 525

Notations

$\langle f, g\rangle$	Inner product (A.6).
$\|f\|$	Euclidean or Hilbert space norm.
$\|f\|_1$	$\mathbf{L^1}$ or $\mathbf{l^1}$ norm.
$\|f\|_\infty$	$\mathbf{L}^\infty$ norm.
$f[n] = O(g[n])$	Order of: there exists K such that $f[n] \leqslant K g[n]$.
$f[n] = o(g[n])$	Small order of: $\lim_{n\to+\infty} \frac{f[n]}{g[n]} = 0$.
$f[n] \sim g[n]$	Equivalent to: $f[n] = O(g[n])$ and $g[n] = O(f[n])$.
$A < +\infty$	A is finite.
$A \gg B$	A is much bigger than B.
z^*	Complex conjugate of $z \in \mathbb{C}$.
$\lfloor x\rfloor$	Largest integer $n \leqslant x$.
$\lceil x\rceil$	Smallest integer $n \geqslant x$.
$(x)_+$	$\max(x, 0)$.
$n \bmod N$	Remainder of the integer division of n modulo N.

Sets

$\mathbb{N}$	Positive integers including 0.
$\mathbb{Z}$	Integers.
$\mathbb{R}$	Real numbers.
$\mathbb{R}^+$	Positive real numbers.
$\mathbb{C}$	Complex numbers.
$\|\Lambda\|$	Number of elements in a set Λ.

Signals

$f(t)$	Continuous time signal.
$f[n]$	Discrete signal.
$\delta(t)$	Dirac distribution (A.30).
$\delta[n]$	Discrete Dirac (3.32).
$\mathbf{1}_{[a,b]}$	Indicator function which is 1 in $[a, b]$ and 0 outside.

Spaces

$\mathbf{C}_0$	Uniformly continuous functions (7.207).
$\mathbf{C}^p$	p times continuously differentiable functions.
$\mathbf{C}^\infty$	Infinitely differentiable functions.
$\mathbf{W}^s(\mathbb{R})$	Sobolev s times differentiable functions (9.8).
$\mathbf{L^2}(\mathbb{R})$	Finite energy functions $\int \|f(t)\|^2\, dt < +\infty$.
$\mathbf{L^p}(\mathbb{R})$	Functions such that $\int \|f(t)\|^p\, dt < +\infty$.
$\boldsymbol{\ell}^2(\mathbb{Z})$	Finite energy discrete signals $\sum_{n=-\infty}^{+\infty} \|f[n]\|^2 < +\infty$.
$\boldsymbol{\ell}^p(\mathbb{Z})$	Discrete signals such that $\sum_{n=-\infty}^{+\infty} \|f[n]\|^p < +\infty$.
$\mathbb{C}^N$	Complex signals of size N.
$\mathbf{U} \oplus \mathbf{V}$	Direct sum of two vector spaces.
$\mathbf{U} \otimes \mathbf{V}$	Tensor product of two vector spaces (A.19).

NullU	Null space of an operator U.
ImU	Image space of an operator U.

Operators

Id	Identity.
$f'(t)$	Derivative $\frac{df(t)}{dt}$.
$f^{(p)}(t)$	Derivative $\frac{d^p f(t)}{dt^p}$ of order p .
$\vec{\nabla} f(x,y)$	Gradient vector (6.51).
$f \star g(t)$	Continuous time convolution (2.2).
$f \star g[n]$	Discrete convolution (3.33).
$f \circledast g[n]$	Circular convolution (3.73)

Transforms

$\hat{f}(\omega)$	Fourier transform (2.6), (3.39).
$\hat{f}[k]$	Discrete Fourier transform (3.49).
$Sf(u,s)$	Short-time windowed Fourier transform (4.11).
$P_S f(u,\xi)$	Spectrogram (4.12).
$Wf(u,s)$	Wavelet transform (4.31).
$P_W f(u,\xi)$	Scalogram (4.55).
$P_V f(u,\xi)$	Wigner-Ville distribution (4.120).

Probability

X	Random variable.
$\mathsf{E}\{X\}$	Expected value.
$\mathcal{H}(X)$	Entropy (10.4).
$\mathcal{H}_d(X)$	Differential entropy (10.20).
$\mathrm{Cov}(X_1, X_2)$	Covariance (A.22).
$F[n]$	Random vector.
$R_F[k]$	Autocovariance of a stationary process (A.26).

I

Sparse Representations

Signals carry overwhelming amounts of data in which relevant information is often harder to find than a needle in a haystack. Processing is faster and simpler in a sparse representation where few coefficients reveal the information we are looking for. Such representations can be constructed by decomposing signals over elementary waveforms chosen in a family called a dictionary. But the search for the Holy Grail of an ideal sparse transform adapted to all signals is a hopeless quest. The discovery of wavelet orthogonal bases and local time-frequency dictionaries has opened the door to a huge jungle of new transforms. Adapting sparse representations to signal properties, and deriving efficient processing operators, is therefore a necessary survival strategy.

An orthogonal basis is a dictionary of minimum size, that can yield a sparse representation if designed to concentrate the signal energy over a set of few vectors. This set gives a geometric signal description. Efficient signal compression and noise reduction algorithms are then implemented with diagonal operators, computed with fast algorithms. But this is not always optimal.

In natural languages, a richer dictionary helps to build shorter and more precise sentences. Similarly, dictionaries of vectors that are larger than bases are needed to build sparse representations of complex signals. But choosing is difficult, and requires more complex algorithms. Sparse representations in redundant dictionaries can improve pattern recognition, compression and noise reduction, but also the resolution of new inverse problems. This includes super-resolution, source separation and compressive sensing.

This first chapter is a sparse book representation, providing the story line and main ideas. It gives a sense of orientation, to choose a path for traveling in the book.

1.1 Computational Harmonic Analysis

Fourier and wavelet bases are the starting point of our journey. They decompose signals over oscillatory waveforms that reveal many signal properties, and provide a path to sparse representations. Discretized signals often have a very large size $N \geqslant 10^6$, and thus can only be processed by fast algorithms, typically implemented with $O(N \log N)$ operations and memories. Fourier and wavelet transforms illustrate the deep connection between well structured mathematical tools and fast algorithms.

1.1.1 Fourier Kingdom

The Fourier transform is everywhere in physics and mathematics, because it diagonalizes time-invariant convolution operators. It rules over linear time-invariant signal processing, whose building blocks are *frequency filtering* operators.

Fourier analysis represents any finite energy function $f(t)$ as a sum of sinusoidal waves $\mathrm{e}^{i\omega t}$:

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \hat{f}(\omega)\, \mathrm{e}^{i\omega t}\, d\omega. \tag{1.1}$$

The amplitude $\hat{f}(\omega)$ of each sinusoidal wave $\mathrm{e}^{i\omega t}$ is equal to its correlation with f, also called Fourier transform:

$$\hat{f}(\omega) = \int_{-\infty}^{+\infty} f(t)\,\mathrm{e}^{-i\omega t}\,dt. \tag{1.2}$$

The more regular $f(t)$ the faster the decay of the sinusoidal wave amplitude $|\hat{f}(\omega)|$ when the frequency ω increases.

When $f(t)$ is defined only on an interval, say $[0,1]$, then the Fourier transform becomes a decomposition in a Fourier orthonormal basis $\{\mathrm{e}^{i2\pi mt}\}_{m\in\mathbb{Z}}$ of $\mathbf{L^2}[0,1]$. If $f(t)$ is uniformly regular then its Fourier transform coefficients also have a fast decay when the frequency $2\pi m$ increases, so it can be well approximated with few low-frequency Fourier coefficients. The Fourier transform therefore defines a sparse representation of uniformly regular functions.

Over discrete signals, the Fourier transform is a decomposition in a discrete orthogonal Fourier basis $\{\mathrm{e}^{i2\pi kn/N}\}_{0\leqslant k<N}$ of $\mathbb{C}^N$, which has properties similar to a Fourier transform on functions. Its embedded structure leads to a Fast Fourier Transform algorithms, which computes discrete Fourier coefficients with $O(N\log N)$ instead of N^2. This FFT algorithm is a corner stone of discrete signal processing.

As long as we are satisfied with linear time-invariant operators or uniformly regular signals, the Fourier transform provides simple answers to most questions. Its richness makes it suitable for a wide range of applications such as signal transmissions or stationary signal processing. However, to represent a transient phenomena—a word pronounced at a particular time, an apple located in the left corner of an image—the Fourier transform becomes a cumbersome tool that requires many coefficients to represent a localized event. Indeed, the support of $\mathrm{e}^{i\omega t}$ covers the whole real line, so $\hat{f}(\omega)$ depends on the values $f(t)$ for all times $t\in\mathbb{R}$. This global "mix" of information makes it difficult to analyze or represent any local property of $f(t)$ from $\hat{f}(\omega)$.

1.1.2 Wavelet Bases

Wavelet bases, like Fourier bases, reveal the signal regularity through the amplitude of coefficients, and their structure leads to a fast computational algorithm. However, wavelets are well localized and few coefficients are needed to represent local transient structures. As opposed to a Fourier basis, a wavelet basis defines a sparse representation of piecewise regular signals, which may include transients and singularities. In images, large wavelet coefficients are located in the neighborhood of edges and irregular textures.

The story begins in 1910, when Haar [265] constructed a piecewise constant function

$$\psi(t) = \begin{cases} 1 & \text{if } 0\leqslant t<1/2 \\ -1 & \text{if } 1/2\leqslant t<1 \\ 0 & \text{otherwise} \end{cases}$$

whose dilations and translations generate an orthonormal basis

$$\left\{\psi_{j,n}(t) = \frac{1}{\sqrt{2^j}}\,\psi\left(\frac{t-2^j n}{2^j}\right)\right\}_{(j,n)\in\mathbb{Z}^2}$$

of the space $\mathbf{L^2}(\mathbb{R})$ of signals having a finite energy

$$\|f\|^2 = \int_{-\infty}^{+\infty} |f(t)|^2\,dt < +\infty.$$

Let us write $\langle f,g\rangle = \int_{-\infty}^{+\infty} f(t)\,g^*(t)\,dt$ the inner product in $\mathbf{L^2}(\mathbb{R})$. Any finite energy signal f can thus represented by its wavelet inner-product coefficients

$$\langle f,\psi_{j,n}\rangle = \int_{-\infty}^{+\infty} f(t)\,\psi_{j,n}(t)\,dt$$

and recovered by summing them in this wavelet orthonormal basis:

$$f = \sum_{j=-\infty}^{+\infty}\sum_{n=-\infty}^{+\infty} \langle f,\psi_{j,n}\rangle\,\psi_{j,n}. \tag{1.3}$$

Each Haar wavelet $\psi_{j,n}(t)$ has a zero average over its support $[2^j n, 2^j(n+1)]$. If f is locally regular and 2^j is small then it is nearly constant over this interval and the wavelet coefficient $\langle f, \psi_{j,n} \rangle$ is thus nearly zero. This means that large wavelet coefficients are located at sharp signal transitions only.

With a jump in time, the story continues in 1980, when Strömberg [409] found a piecewise linear function ψ that also generates an orthonormal basis and gives better approximations of smooth functions. Meyer was not aware of this result, and motivated by the work of Morlet and Grossmann over continuous wavelet transform, he tried to prove that there exists no regular wavelet ψ that generates an orthonormal basis. This attempt was a failure since he ended up constructing a whole family of orthonormal wavelet bases, with functions ψ that are infinitely continuously differentiable [342]. This was the fundamental impulse that lead to a widespread search for new orthonormal wavelet bases, which culminated in the celebrated Daubechies wavelets of compact support [181].

The systematic theory for constructing orthonormal wavelet bases was established by Meyer and Mallat through the elaboration of multiresolution signal approximations [329] presented in Chapter 7. It was inspired by original ideas developed in computer vision by Burt and Adelson [119] to analyze images at several resolutions. Digging more into the properties of orthogonal wavelets and multiresolution approximations brought to light a surprising relation with filter banks constructed with conjugate mirror filters, and a fast wavelet transform algorithm decomposing signals of size N with $O(N)$ operations [328].

Filter Banks Motivated by speech compression, in 1976 Croisier, Esteban and Galand [176] introduced an invertible filter bank, which decomposes a discrete signal $f[n]$ in two signals of half its size, using a filtering and subsampling procedure. They showed that $f[n]$ can be recovered from these subsampled signals by canceling the aliasing terms with a particular class of filters called *conjugate mirror filters*. This breakthrough led to a 10-year research effort to build a complete filter bank theory. Necessary and sufficient conditions for decomposing a signal in subsampled components with a filtering scheme, and recovering the same signal with an inverse transform, were established by Smith and Barnwell [405], Vaidyanathan [430] and Vetterli [432].

The multiresolution theory of Mallat [329] and Meyer [42] proves that any conjugate mirror filter characterizes a wavelet ψ that generates an orthonormal basis of $\mathbf{L^2}(\mathbb{R})$, and that a fast discrete wavelet transform is implemented by cascading these conjugate mirror filters [328]. The equivalence between this continuous time wavelet theory and discrete filter banks led to a new fruitful interface between digital signal processing and harmonic analysis, first creating an interesting culture shock that is now well resolved.

Continuous Versus Discrete and Finite Originally, many signal processing engineers were wondering what is the point of considering wavelets and signals as functions, since all computations are performed over discrete signals, with conjugate mirror filters. Why bother with the convergence of infinite convolution cascades if in practice we only compute a finite number of convolutions? Answering these important questions is necessary in order to understand why this book alternates between theorems on continuous time functions and discrete algorithms applied to finite sequences.

A short answer would be "simplicity". In $\mathbf{L^2}(\mathbb{R})$, a wavelet basis is constructed by dilating and translating a single function ψ. Several important theorems relate the amplitude of wavelet coefficients to the local regularity of the signal f. Dilations are not defined over discrete sequences, and discrete wavelet bases are therefore more complex to describe. The regularity of a discrete sequence is not well defined either, which makes it more difficult to interpret the amplitude of wavelet coefficients. A theory of continuous time functions gives asymptotic results for discrete sequences with sampling intervals decreasing to zero. This theory is useful because these asymptotic results are precise enough to understand the behavior of discrete algorithms.

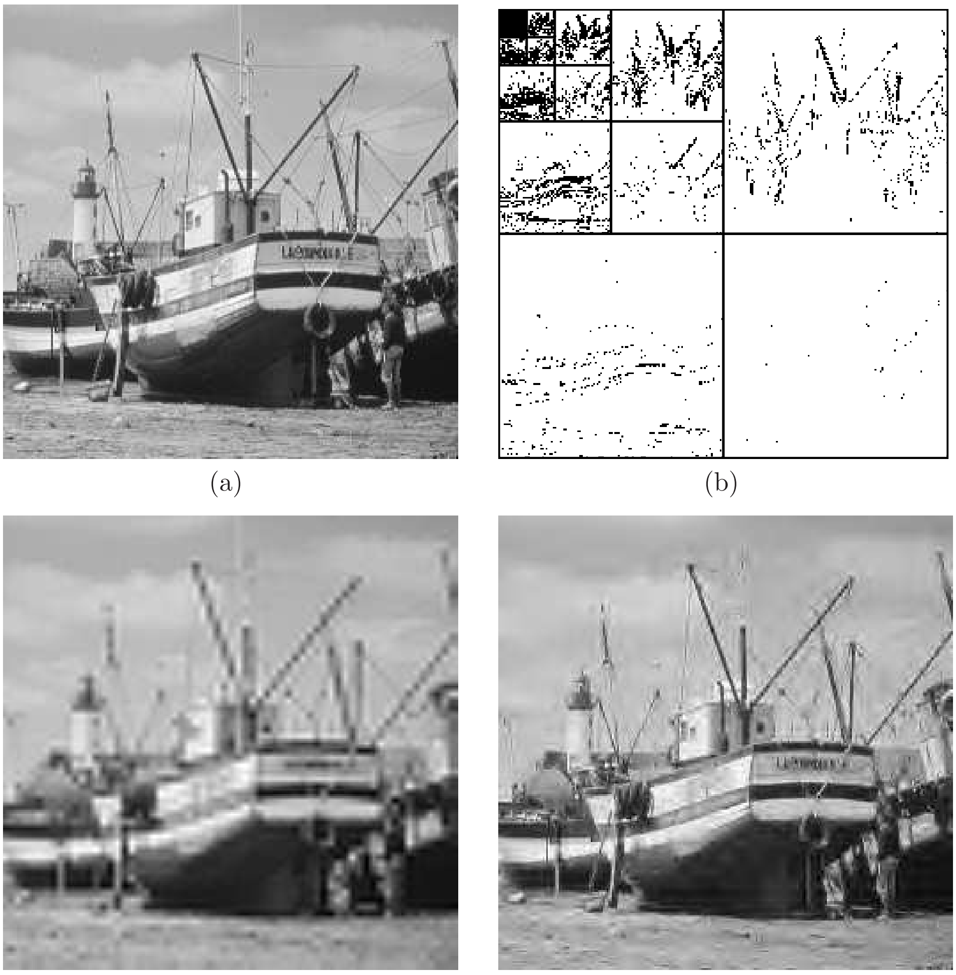
But continuous time or space models are not sufficient for elaborating discrete signal processing algorithms. The transition between continuous and discrete signals must be done with great care, to maintain important properties such as orthogonality. Restricting the constructions to finite discrete signals adds another layer of complexity because of border problems. How these border issues affect numerical implementations is carefully addressed once the properties of the bases are well understood.

Wavelets for Images Wavelet orthonormal bases of images can be constructed from wavelet orthonormal bases of one-dimensional signals. Three mother wavelets $\psi^1(x)$, $\psi^2(x)$ and $\psi^3(x)$, with $x = (x_1, x_2) \in \mathbb{R}^2$, are dilated by 2^j and translated by $2^j n$ with $n = (n_1, n_2) \in \mathbb{Z}^2$. It yields an orthonormal basis of the space $\mathbf{L^2}(\mathbb{R}^2)$ of finite energy functions $f(x) = f(x_1, x_2)$:

$$\left\{ \psi_{j,n}^k(x) = \frac{1}{2^j} \psi^k \left(\frac{x - 2^j n}{2^j} \right) \right\}_{j \in \mathbb{Z}, n \in \mathbb{Z}^2, 1 \leqslant k \leqslant 3} .$$

The support of a wavelet $\psi_{j,n}^k$ is a square of width proportional to the scale 2^j. Two dimensional wavelet bases are discretized to define orthonormal bases of images including N pixels. Wavelet coefficients are calculated with a fast $O(N)$ algorithm described in Chapter 7.

Like in one dimension, a wavelet coefficient $\langle f, \psi_{j,n}^k \rangle$ has a small amplitude if $f(x)$ is regular over the support of $\psi_{j,n}^k$. It is large near sharp transitions such as edges. Figure 1.1(b) is the array of N wavelet coefficients. Each direction k and scale 2^j corresponds to a sub-image, which shows in black the position of the largest coefficients above a threshold: $|\langle f, \psi_{j,n}^k \rangle| \geqslant T$.



(a) (b)

(c) (d)

Figure 1.1: (a): Discrete image $f[n]$ of $N = 256^2$ pixels. (b): Array of N orthogonal wavelet coefficients $\langle f, \psi_{j,n}^k \rangle$ for $k = 1, 2, 3$ and 4 scales 2^j. Black points correspond to $|\langle f, \psi_{j,n}^k \rangle| > T$. (c): Linear approximation from the $N/16$ wavelet coefficients at the 3 largest scales. (d): Non-linear approximation from the $M = N/16$ wavelet coefficients of largest amplitude shown in (b).

1.2 Approximation and Processing in Bases

Analog to digital signal conversion is the first step of digital signal processing. Chapter 3 explains that it amounts to projecting the signal over a basis of an approximation space. Most often, the resulting digital representation remains much too large and needs to be further reduced. A digital image includes typically more than 10^6 samples and a CD music recording has $40\,10^3$ samples per second. Sparse representations that reduce the number of parameters can be obtained by thresholding coefficients in an appropriate orthogonal basis. Efficient compression and noise reduction algorithms are then implemented with simple operators in this basis.

Stochastic versus Deterministic Signal Models A representation is optimized relatively to a signal class, corresponding to all potential signals encountered in an application. This requires building signal models that carries available prior information.

A signal f can be modeled as a realization of a random process F, whose probability distribution is known a priori. A Bayesian approach then tries to minimize the expected approximation error. Linear approximations are simpler because they only depend upon the covariance. Chapter 9 shows that optimal linear approximations are obtained in a basis of principal components that are the eigenvectors of the covariance matrix. However, the expected error of non-linear approximations depends upon the full probability distribution of F. This distribution is most often not know for complex signals such as images or sounds, because their transient structures are not well modeled as realizations of known processes such as Gaussian processes.

To optimize non-linear representations, weaker but sufficiently powerful deterministic models can be elaborated. A deterministic model specifies a set Θ where the signal belongs. This set is defined by any prior information, for example on the time-frequency localization of transients in musical recordings or on the geometric regularity of edges in images. Simple models can also define Θ as a ball in a functional space, with a specific regularity norm such as a total variation norm. A stochastic model is richer because it provides the probability distribution in Θ. When this distribution is not available, the average error can not be calculated and is replaced by the maximum error over Θ. Optimizing the representation then amounts to minimize this maximum error, which is called a *minimax* optimization.

1.2.1 Sampling with Linear Approximations

Analog to digital signal conversion is most often implemented with a linear approximation operator that filters and samples the input analog signal. From these samples, a linear digital to analog converter recovers a projection of the original analog signal over an approximation space whose dimension depends upon the sampling density. Linear approximations project signals in spaces of lowest possible dimensions to reduce computations and storage cost, while controlling the resulting error.

Sampling Theorems Let us consider finite energy signals $\|\bar f\|^2 = \int |\bar f(x)|^2\,dx$ of finite support, which is normalized to $[0,1]$ or $[0,1]^2$ for images. A sampling process implements a filtering of $\bar f(x)$ with a low-pass impulse response $\bar\phi_s(x)$, and a uniform sampling to output a discrete signal:

$$f[n] = \bar f \star \bar\phi_s(ns) \quad \text{for} \quad 0 \leqslant n < N.$$

In two dimensions, $n = (n_1, n_2)$ and $x = (x_1, x_2)$. These filtered samples can also be written as inner products:

$$\bar f \star \bar\phi_s(ns) = \int f(u)\,\bar\phi_s(ns-u)\,du = \langle f(x), \phi_s(x-ns)\rangle$$

with $\phi_s(x) = \bar\phi_s(-x)$. Chapter 3 explains that ϕ_s are chosen, like in the classical Shannon-Whittaker sampling theorem, so that family of functions $\{\phi_s(x-ns)\}_{1\leqslant n\leqslant N}$ is a basis of an appropriate approximation space $\mathbf{U}_N$. The best linear approximation of $\bar f$ in $\mathbf{U}_N$ recovered from these samples is the orthogonal projection $\bar f_N$ of f in $\mathbf{U}_N$, and if the basis is orthonormal then

$$\bar f_N(x) = \sum_{n=0}^{N-1} f[n]\,\phi_s(x-ns). \tag{1.4}$$

A sampling theorem states that if $\bar{f} \in \mathbf{U}_N$ then $\bar{f} = \bar{f}_N$ so (1.4) recovers $\bar{f}(x)$ from the measured samples. Most often, $\bar{f}$ does not belong to this approximation space. It is called *aliasing* in the context of the Shannon-Whittaker sampling, where $\mathbf{U}_N$ is the space of functions having a frequency support restricted to the N lower frequencies. The approximation error $\|\bar{f} - \bar{f}_N\|^2$ must then be controled.

Linear Approximation Error The approximation error is computed by finding an orthogonal basis $\mathcal{B} = \{\bar{g}_m(x)\}_{0 \leqslant m < +\infty}$ of the whole analog signal space $\mathbf{L^2}[0,1]^2$, whose first N vectors $\{\bar{g}_m(x)\}_{0 \leqslant m < N}$ is an orthogonal basis of $\mathbf{U}_N$. The orthogonal projection on $\mathbf{U}_N$ can thus be rewritten:

$$\bar{f}_N(x) = \sum_{m=0}^{N-1} \langle \bar{f}, \bar{g}_m \rangle \, \bar{g}_m(x) \ .$$

Since $\bar{f} = \sum_{m=0}^{+\infty} \langle \bar{f}, \bar{g}_m \rangle \, \bar{g}_m$, the approximation error is the energy of the removed inner products:

$$\varepsilon_l[N] = \|\bar{f} - \bar{f}_N\|^2 = \sum_{m=N}^{+\infty} |\langle \bar{f}, \bar{g}_m \rangle|^2.$$

This error decreases quickly when N increases if the coefficient amplitudes $|\langle \bar{f}, \bar{g}_m \rangle|$ have a fast decay when the index m increases. The dimension N is adjusted to the desired approximation error. Figure 1.1(a) shows a discrete image $f[n]$ approximated with $N = 256^2$ pixels. Figure 1.1(c) displays a lower resolution image $f_{N/16}$ projected on a space $\mathbf{U}_{N/16}$ of dimension $N/16$, generated by $N/16$ large scale wavelets. It is calculated by setting to zero all the wavelet coefficients at the first two smaller scales. The approximation error is $\|f - f_{N/16}\|^2 / \|f\|^2 = 14\,10^{-3}$. Reducing the resolution introduces more blur and errors. A linear approximation space $\mathbf{U}_N$ corresponds to a uniform grid that approximates precisely uniformly regular signals. Since images $\bar{f}$ are often not uniformly regular, it is necessary to measure it at a high resolution N. This is why digital camera have a resolution that increases as technology improves.

1.2.2 Sparse Non-linear Approximations

Linear approximations reduce the space dimensionality but can introduce important errors when reducing the resolution, if the signal is not uniformly regular, as shown by Figure 1.1(c). To improve such approximations, more coefficients should be kept where it is needed, not in regular regions but near sharp transitions and edges. This requires defining an irregular sampling adapted to the local signal regularity. This optimized irregular sampling has a simple equivalent solution through non-linear approximations in wavelet bases.

Non-linear approximations operate in two stages. First a linear approximation approximates the analog signal $\bar{f}$ with N samples that are written $f[n] = \bar{f} \star \bar{\phi}_s(ns)$. Then a non-linear approximation of $f[n]$ is computed to reduce the N coefficients $f[n]$ to $M \ll N$ coefficients in a sparse representation.

The discrete signal f can be considered as a vector of $\mathbb{C}^N$. Inner products and norms in $\mathbb{C}^N$ are written

$$\langle f, g \rangle = \sum_{n=0}^{N-1} f[n] \, g^*[n] \quad \text{and} \quad \|f\|^2 = \sum_{n=0}^{N-1} |f[n]|^2 \ .$$

To obtain a sparse representation with a non-linear approximation, we choose a new orthonormal basis $\mathcal{B} = \{g_m[n]\}_{m \in \Gamma}$ of $\mathbb{C}^N$, which concentrates as much as possible the signal energy over few coefficients. Signal coefficients $\{\langle f, g_m \rangle\}_{m \in \Gamma}$ are computed from the N input sample values $f[n]$ with an orthogonal change of basis that takes N^2 operations in non-structured bases. In a wavelet or Fourier bases, fast algorithms require respectively $O(N)$ and $O(N \log_2 N)$ operations.

Approximation by Thresholding For $M < N$, an approximation f_M is computed by selecting the "best" $M < N$ vectors within $\mathcal{B}$. The orthogonal projection of f on the space $\mathbf{V}_\Lambda$ generated by M vectors $\{g_m\}_{m \in \Lambda}$ in $\mathcal{B}$ is

$$f_\Lambda = \sum_{m \in \Lambda} \langle f, g_m \rangle \, g_m \ . \tag{1.5}$$

Since $f = \sum_{m\in\Gamma} \langle f, g_m\rangle\, g_m$, the resulting error is

$$\|f - f_\Lambda\|^2 = \sum_{m\notin\Lambda} |\langle f, g_m\rangle|^2 \ . \tag{1.6}$$

We write $|\Lambda|$ the size of the set Λ. The best $M = |\Lambda|$ term approximation which minimizes $\|f - f_\Lambda\|^2$ is thus obtained by selecting the M coefficients of largest amplitude. These coefficients are above a threshold T that depends on M:

$$f_M = f_{\Lambda_T} = \sum_{m\in\Lambda_T} \langle f, g_m\rangle\, g_m \quad \text{with} \quad \Lambda_T = \{m \in \Gamma \ : \ |\langle f, g_m\rangle| \geqslant T\} \ . \tag{1.7}$$

This approximation is non-linear because the approximation set Λ_T changes with f. The resulting approximation error is:

$$\varepsilon_n[M] = \|f - f_M\|^2 = \sum_{m\notin\Lambda_T} |\langle f, g_m\rangle|^2 \ . \tag{1.8}$$

Figure 1.1(b) shows that the approximation support Λ_T of an image in a wavelet orthonormal basis, which depends upon the geometry of edges and textures. Keeping large wavelet coefficients is equivalent to constructing an adaptive approximation grid specified by the scale-space support Λ_T. It increases the approximation resolution where the signal is irregular. The geometry of Λ_T gives the spatial distribution on sharp image transitions and edges, and shows how they propagate across scales. Chapter 6 proves that it gives important information on their sharpness and local Lipschitz regularity. This example illustrates how an approximation support provides "geometric" information on f, relatively to a dictionary, which is a wavelet basis in this example.

Figure 1.1(d) gives the non-linear wavelet approximation f_M recovered from the $M = N/16$ largest amplitude wavelet coefficients, with an error $\|f - f_M\|^2/\|f\|^2 = 5\,10^{-3}$. This error is nearly 3 times smaller than the the linear approximation error obtained with the same number of wavelet coefficients, and the image quality is much better.

An analog signal can be recovered from the discrete non-linear approximation f_M:

$$\bar{f}_M(x) = \sum_{n=0}^{N-1} f_M[n]\, \phi_s(x - ns).$$

Since all projections are orthogonal, the overall approximation error on the original analog signal $\bar{f}(x)$ is the sum of the analog sampling error and the discrete non-linear error:

$$\|\bar{f} - \bar{f}_M\|^2 = \|\bar{f} - \bar{f}_N\|^2 + \|f - f_M\|^2 = \varepsilon_l[N] + \varepsilon_n[M].$$

In practice, N is imposed by the resolution of the signal acquisition hardware, and M is typically adjusted so that $\varepsilon_n[M] \geqslant \varepsilon_l[N]$.

Sparsity with Regularity Sparse representations are obtained in a basis that takes advantage of some form of regularity of the input signals, which creates many small amplitude coefficients. Since wavelets have a localized support, functions with isolated singularities produce few large amplitude wavelet coefficients in the neighborhood of these singularities. Non-linear wavelet approximation produce a small error over spaces of functions that do not have "too many" sharp transitions and singularities. Chapter 9 shows that functions having a bounded total variation norm are useful models for images with non-fractal (finite length) edges.

Edges often define regular geometric curves. Wavelets detect the location of edges but their square support can not take advantage of their potential geometric regularity. More sparse representations are defined in dictionaries of curvelets or bandlets, that have elongated support in multiple directions that can be adapted to this geometrical regularity. In such dictionaries, the approximation support Λ_T is smaller but provides explicit information about the local geometrical properties of edges, such as their orientation. In this context, geometry does not just apply to multidimensional signals. Audio signals such as musical recordings also have a complex geometric regularity in time-frequency dictionaries.

1.2.3 Compression

Storage limitations and fast transmission through narrow band-width channels requires to compress signals while minimizing the degradation. Transform codes compress signals by coding a sparse representation. Chapter 10 introduces the information theory needed to understand these codes and optimize their performance.

In a compression framework, the analog signal has already been discretized into a signal $f[n]$ of size N. This discrete signal is decomposed in an orthonormal basis $\mathcal{B} = \{g_m\}_{m\in\Gamma}$ of $\mathbb{C}^N$:

$$f = \sum_{m\in\Gamma} \langle f, g_m \rangle \, g_m \, .$$

Coefficients $\langle f, g_m\rangle$ are approximated by quantized values $Q(\langle f, g_m\rangle)$. If Q is a uniform quantizer of step Δ then $|x - Q(x)| \leqslant \Delta/2$, and if $|x| < \Delta/2$ then $Q(x) = 0$. The signal $\tilde{f}$ restored from quantized coefficients is

$$\tilde{f} = \sum_{m\in\Gamma} Q(\langle f, g_m\rangle) \, g_m \, .$$

An entropy code records these coefficients with R bits. The goal is to minimize the signal distortion-rate $d(R, f) = \|\tilde{f} - f\|^2$.

The coefficients not quantized to zero correspond to the set $\Lambda_T = \{m \in \Gamma \ : \ |\langle f, g_m\rangle| \geqslant T\}$ with $T = \Delta/2$. For sparse signals, Chapter 10 shows that the bit budget R is dominated by the number of bits to code Λ_T in Γ, which is nearly proportional to its size $|\Lambda_T|$. This means that the "information" of a sparse representation is mostly geometric. Moreover, the distortion is dominated by the non-linear approximation error $\|f - f_{\Lambda_T}\|^2$, for $f_{\Lambda_T} = \sum_{m\in\Lambda_T} \langle f, g_m\rangle g_m$. Compression is thus a sparse approximation problem. For a given distortion $d(R, f)$, minimizing R requires to reduce $|\Lambda_T|$ and optimize the sparsity.

The number of bits to code Λ_T can take advantage of any prior information on the geometry. Figure 1.1(b) shows that large wavelet coefficients are not randomly distributed. They have a tendency to be aggregated towards larger scales, and at fine scales they are regrouped along edge curves or in texture regions. Using such prior geometric model is an source of gain in coders such as JPEG-2000.

Chapter 10 describes the implementation of audio transform codes. Image transform codes in block cosine bases and wavelet bases are introduced, together with the JPEG and JPEG-2000 compression standards.

1.2.4 Denoising

Signal acquisition devices add noises, that can be reduced by estimators using prior information on signal properties. Signal processing has long remained mostly Bayesian and linear. Non-linear smoothing algorithms existed in statistics, but these procedures were often ad-hoc and complex. Two statisticians, Donoho and Johnstone [208], changed the game by proving that simple thresholding in sparse representations can yield nearly optimal non-linear estimators. This was the beginning of a considerable refinement of non-linear estimation algorithms that is still on-going.

Let us consider digital measurements that add a random noise $W[n]$ to the original signal $f[n]$

$$X[n] = f[n] + W[n] \quad \text{for} \quad 0 \leqslant n < N.$$

The signal f is estimated by transforming the noisy data X with an operator D:

$$\tilde{F} = D\,X \ .$$

The risk of the estimator $\tilde{F}$ of f is the average error, calculated with respect to the probability distribution of the noise W:

$$r(D, f) = \mathsf{E}\{\|f - DX\|^2\} \ .$$

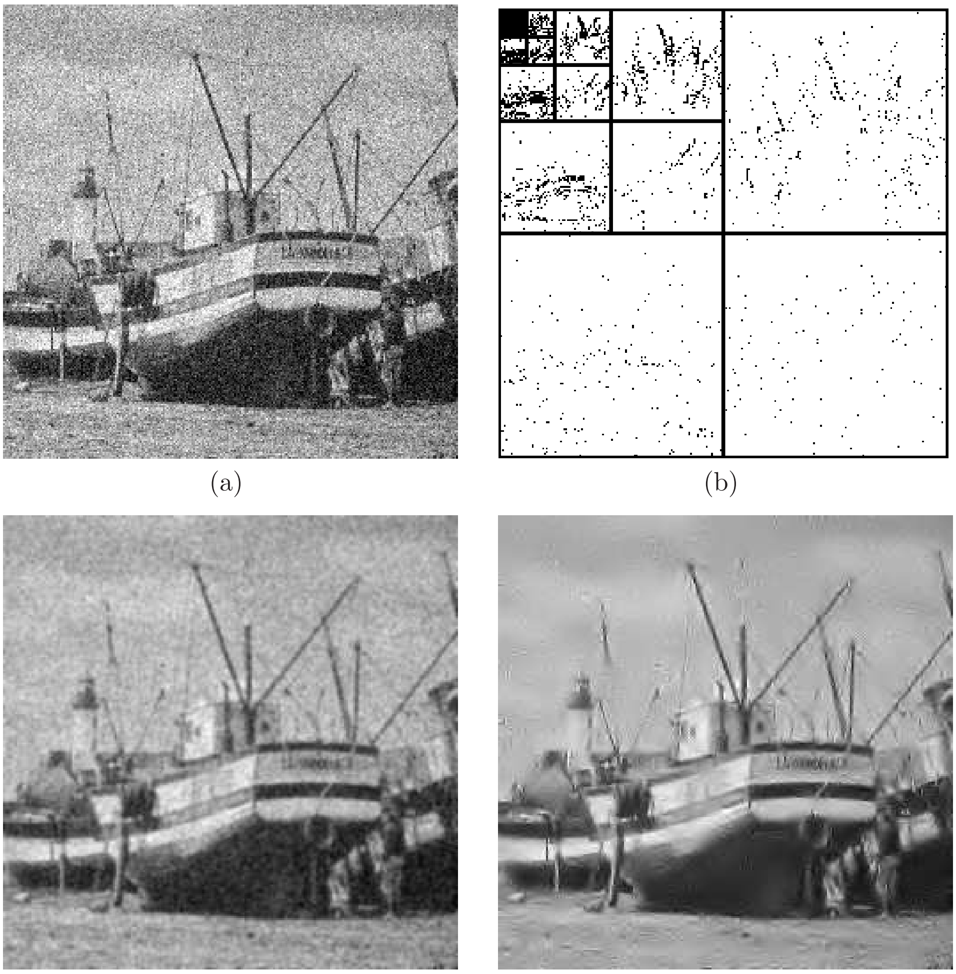

(a) (b)
(c) (d)

Figure 1.2: (a): Noisy image X. (b): Noisy wavelet coefficients above threshold, $|\langle X, \psi_{j,n}\rangle| \geqslant T$. (c): Linear estimation $X \star h$. (d): Non-linear estimator recovered from thresholded wavelet coefficients, over several translated bases.

Bayes Versus Minimax To optimize the estimation operator D, one must take advantage of prior information available about the signal f. In a Bayes framework, f is considered as a realization of a random vector F and the Bayes risk is the expected risk calculated with respect to the prior probability distribution π of the random signal model F:

$$r(D,\pi) = \mathsf{E}_\pi\{r(D,F)\} \ .$$

Optimizing D among all possible operators yields the *minimum Bayes risk*:

$$r_n(\pi) = \inf_{all\ D} r(D,\pi) \ .$$

In the 1940's, Wald brought a new perspective on statistics, through a decision theory partly imported from the theory of games. This point of view uses deterministic models, where signals are elements of a set Θ, without specifying their probability distribution in this set. To control the risk for any $f \in \Theta$, we compute the maximum risk

$$r(D,\Theta) = \sup_{f\in\Theta} r(D,f) \ .$$

The *minimax risk* is the lower bound computed over all operators D:

$$r_n(\Theta) = \inf_{all\ D} r(D,\Theta).$$

In practice, the goal is to find an operator D that is simple to implement and which yields a risk close the minimax lower bound.

Thresholding estimators It is tempting to restrict calculations to linear operators D, because of their simplicity. Optimal linear Wiener estimators are introduced in Chapter 11. Figure 1.2(a) is an image contaminated by a Gaussian white noise. Figure 1.2(b) shows an optimized linear filtering estimation $\tilde{F} = X \star h[n]$, which is therefore diagonal in a Fourier basis $\mathcal{B}$. This convolution operator averages the noise but also blurs the image and keeps low-frequency noise by keeping the image low-frequencies.

If f has a sparse representation in a dictionary, then projecting X on the vectors of this sparse support can considerably improve linear estimators. The difficulty is to identify the sparse support of f from the noisy data X. Donoho and Johnstone [208] proved that in an orthonormal basis, a simple thresholding of noisy coefficients does the trick. Noisy signal coefficients in an orthonormal basis $\mathcal{B} = \{g_m\}_{m\in\Gamma}$ are

$$\langle X, g_m\rangle = \langle f, g_m\rangle + \langle W, g_m\rangle \quad \text{for} \quad m \in \Gamma \ .$$

Thresholding these noisy coefficients yields an orthogonal projection estimator

$$\tilde{F} = X_{\tilde{\Lambda}_T} = \sum_{m\in\tilde{\Lambda}_T} \langle X, g_m\rangle\, g_m \quad \text{with} \quad \tilde{\Lambda}_T = \{m \in \Gamma \ : \ |\langle X, g_m\rangle| \geqslant T\} \ . \tag{1.9}$$

The set $\tilde{\Lambda}_T$ is an estimate of an approximation support of f. It is hopefully close to the optimal approximation support $\Lambda_T = \{m \in \Gamma \ : \ |\langle f, g_m\rangle| \geqslant T\}$. Figure 1.2(b) shows the estimated approximation set $\tilde{\Lambda}_T$ of noisy wavelet coefficients $|\langle X, \psi_{j,n}| \geqslant T$, that can be compared to the optimal approximation support Λ_T shown in Figure 1.1(b). The estimation in Figure 1.2(d) from wavelet coefficients in $\tilde{\Lambda}_T$ has considerably reduced the noise in regular regions while keeping the sharpness of edges by preserving large wavelet coefficients. This estimation is improved with a translation invariant procedure that averages this estimator over several translated wavelet bases. Thresholding wavelet coefficients implements an adaptive smoothing, which averages the data X with a kernel that depends on the estimated regularity of the original signal f.

Donoho and Johnstone proved that for a Gaussian white noise of variance σ^2, choosing $T = \sigma\sqrt{2\log_e N}$ yields a risk $\mathsf{E}\{\|f - \tilde{F}\|^2\}$ of the order of $\|f - f_{\Lambda_T}\|^2$, up to a $\log_e N$ factor. This spectacular result shows that the estimated support $\tilde{\Lambda}_T$ does nearly as well as the optimal unknown support Λ_T. The resulting risk is small if the representation is sparse and precise.

The set $\tilde{\Lambda}_T$ in Figure 1.2(b) "looks" different from the Λ_T in Figure 1.1(b) because it has more isolated points. This indicates that some prior information on the geometry of Λ_T could be used to improve the estimation. For audio noise reduction, thresholding estimators are applied in sparse representations provided by time-frequency bases. Similar isolated time-frequency coefficients produce a highly annoying "musical noise". Musical noise is removed with a block thresholding that regularizes the geometry of the estimated support $\tilde{\Lambda}_T$, and avoids leaving isolated points. Block thresholding also improves wavelet estimators.

If W is a Gaussian noise and signals in Θ have a sparse representation in $\mathcal{B}$, then Chapter 11 proves that thresholding estimators can produce a nearly minimax risk. In particular, wavelet thresholding estimators have a nearly minimax risk for large classes of piecewise smooth signals, including bounded variation images.

1.3 Time-Frequency Dictionaries

Motivated by quantum mechanics, in 1946 the physicist Gabor [246] proposed to decompose signals over dictionaries of elementary waveforms, that he called time-frequency atoms, which have a minimal spread in a time-frequency plane. By showing that such decompositions are closely related to our perception of sounds, and that they exhibit important structures in speech and music recordings, Gabor demonstrated the importance of localized time-frequency signal processing. Beyond sounds, large classes of signals have sparse decompositions as sums of time-frequency atoms selected from appropriate dictionaries. The key issue is to understand how to construct dictionaries with time-frequency atoms adapted to signal properties.

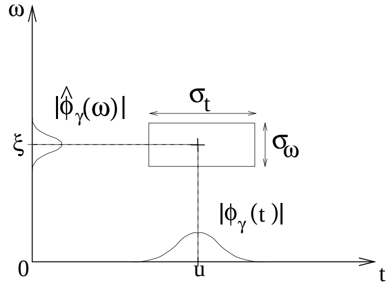


Figure 1.3: Heisenberg box representing an atom ϕ_γ.

1.3.1 Heisenberg Uncertainty

A time-frequency dictionary $\mathcal{D} = \{\phi_\gamma\}_{\gamma\in\Gamma}$ is composed of waveforms of unit norm $\|\phi_\gamma\| = 1$, which are well localized in time and in frequency. The time localization u of ϕ_γ and its spread around u are defined by

$$u = \int t\,|\phi_\gamma(t)|^2\,dt \quad \text{and} \quad \sigma^2_{t,\gamma} = \int |t-u|^2\,|\phi_\gamma(t)|^2\,dt\ .$$

Similarly, the frequency localization and spread of $\hat{\phi}_\gamma$ are defined by

$$\xi = (2\pi)^{-1}\int \omega\,|\hat{\phi}_\gamma(\omega)|^2\,d\omega \quad \text{and} \quad \sigma^2_{\omega,\gamma} = (2\pi)^{-1}\int |\omega-\xi|^2\,|\hat{\phi}_\gamma(\omega)|^2\,d\omega\ .$$

The Fourier Parseval formula

$$\langle f, \phi_\gamma\rangle = \int_{-\infty}^{+\infty} f(t)\,\phi^*_\gamma(t)\,dt = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \hat{f}(\omega)\,\hat{\phi}^*_\gamma(\omega)\,d\omega\,. \tag{1.10}$$

shows that $\langle f, \phi_\gamma\rangle$ depends mostly on the values $f(t)$ and $\hat{f}(\omega)$ where $\phi_\gamma(t)$ and $\hat{\phi}_\gamma(\omega)$ are non-negligible, and hence for (t, ω) in a rectangle centered at (u, ξ), of size $\sigma_{t,\gamma} \times \sigma_{\omega,\gamma}$. This rectangle is illustrated by Figure 1.3 in this time-frequency plane (t, ω). It can be interpreted as a "quantum of information" over an elementary resolution cell. The uncertainty principle theorem proves in Chapter 2 that this rectangle has a minimum surface which limits the joint time-frequency resolution:

$$\sigma_{t,\gamma}\,\sigma_{\omega,\gamma} \geqslant \frac{1}{2}\,. \tag{1.11}$$

Constructing a dictionary of time-frequency atoms can thus be thought as covering the time-frequency plan with resolution cells whose time width $\sigma_{t,\gamma}$ and frequency width $\sigma_{\omega,\gamma}$ may vary, but with a surface larger than 1/2. Windowed Fourier and wavelet transforms are two important examples.

1.3.2 Windowed Fourier Transform

A windowed Fourier dictionary is constructed by translating in time and frequency a time window $g(t)$, of unit norm $\|g\| = 1$, centered at $t = 0$:

$$\mathcal{D} = \left\{ g_{u,\xi}(t) = g(t-u)\,\mathrm{e}^{i\xi t} \right\}_{(u,\xi)\in\mathbb{R}^2}\,.$$

The atom $g_{u,\xi}$ is translated by u in time and by ξ in frequency. The time and frequency spread of $g_{u,\xi}$ is independent of u and ξ. This means that each atom $g_{u,\xi}$ corresponds to a Heisenberg rectangle whose size $\sigma_t \times \sigma_\omega$ is independent from its position (u,ξ), as shown by Figure 1.4.

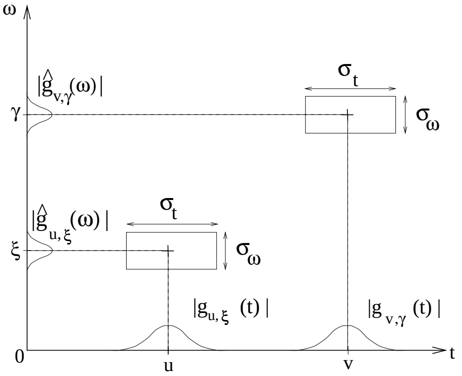


Figure 1.4: Time-frequency boxes ("Heisenberg rectangles") representing the energy spread of two windowed Fourier atoms.

The windowed Fourier transform projects f on each dictionary atom $g_{u,\xi}$:

$$Sf(u,\xi) = \langle f, g_{u,\xi}\rangle = \int_{-\infty}^{+\infty} f(t)\,g(t-u)\,\mathrm{e}^{-i\xi t}\,dt. \tag{1.12}$$

It can be interpreted as a Fourier transform of f at the frequency ξ, localized by the window $g(t-u)$ in the neighborhood of u. This windowed Fourier transform is highly redundant and represents one-dimensional signals by a time-frequency image in (u,ξ). It is thus necessary to understand how to select much fewer time-frequency coefficients that represent the signal efficiently.

When listening to music, we perceive sounds that have a frequency that varies in time. Chapter 4 shows that a spectral line of f creates high amplitude windowed Fourier coefficients $Sf(u,\xi)$ at frequencies $\xi(u)$ that depend on the time u. These spectral components are detected and characterized by ridge points, which are local maxima in this time-frequency plane. Ridge points define a time-frequency approximation support Λ of f, whose geometry depends on the time-frequency evolution of the signal spectral components. Modifying the sound duration or audio transpositions are implemented by modifying the geometry of the ridge support in time-frequency.

A windowed Fourier transform decomposes signals over waveforms that have the same time and frequency resolution. It is thus effective as long as the signal does not include structures having different time-frequency resolution, some being very localized in time and others very localized in frequency. Wavelets address this issue by changing the time and frequency resolution.

1.3.3 Continuous Wavelet Transform

In reflection seismology, Morlet knew that the waveforms sent underground have a duration that is too long at high frequencies to separate the returns of fine, closely-spaced geophysical layers. These waveforms are called *wavelets* in geophysics. Instead of emitting pulses of equal duration, he thus thought of sending shorter waveforms at high frequencies. Such waveforms could simply be obtained by scaling the mother wavelet, hence the name of this transform. Although Grossmann was working in theoretical physics, he recognized in Morlet's approach some ideas that were close to his own work on coherent quantum states. Nearly forty years after Gabor, Morlet and Grossmann reactivated a fundamental collaboration between theoretical physics and signal processing, which led to the formalization of the continuous wavelet transform [263]. Yet, these ideas were not totally new to mathematicians working in harmonic analysis, or to computer vision researchers studying multiscale image processing. It was thus only the beginning of a rapid catalysis that brought together scientists with very different backgrounds.

A wavelet dictionary is constructed from a mother wavelet ψ of zero average

$$\int_{-\infty}^{+\infty} \psi(t)\, dt = 0,$$

which is dilated with a scale parameter s, and translated by u:

$$\mathcal{D} = \left\{ \psi_{u,s}(t) = \frac{1}{\sqrt{s}}\, \psi\left(\frac{t-u}{s}\right) \right\}_{u \in \mathbb{R}, s>0}. \tag{1.13}$$

The continuous wavelet transform of f at any scale s and position u is the projection of f on the corresponding wavelet atom:

$$Wf(u,s) = \langle f, \psi_{u,s} \rangle = \int_{-\infty}^{+\infty} f(t)\, \frac{1}{\sqrt{s}}\, \psi^*\left(\frac{t-u}{s}\right) dt. \tag{1.14}$$

It represents one-dimensional signals by highly redundant time-scale images in (u, s).

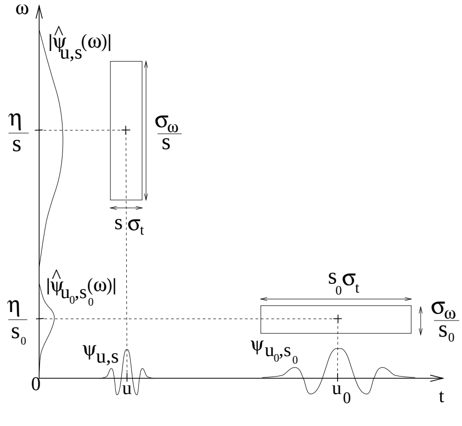


Figure 1.5: Heisenberg time-frequency boxes of two wavelets $\psi_{u,s}$ and ψ_{u_0,s_0}. When the scale s decreases, the time support is reduced but the frequency spread increases and covers an interval that is shifted towards high frequencies.

Varying Time-Frequency Resolution As opposed to windowed Fourier atoms, wavelets have a time-frequency resolution that change. The wavelet $\psi_{u,s}$ has a time support centered at u whose size is proportional to s. Let us choose a wavelet ψ whose Fourier transform $\hat\psi(\omega)$ is non-zero in a positive frequency interval centered at η. The Fourier transform $\hat\psi_{u,s}(\omega)$ is dilated by $1/s$ and is hence localized in a positive frequency interval centered at $\xi = \eta/s$, whose size is scaled by $1/s$. In the time-frequency plane, the Heisenberg box of a wavelet atom $\psi_{u,s}$ is therefore a rectangle centered at $(u, \eta/s)$, with time and frequency widths respectively proportional to s and $1/s$. When

s varies, the time and frequency width of this time-frequency resolution cell changes but its area remains constant, as illustrated by Figure 1.5.

Large amplitude wavelet coefficients can detect and measure short high frequency variations because they are well localized in time at high frequencies. At low frequencies their time resolution is lower but they have a better frequency resolution. This modification of time and frequency resolution is well adapted to represent sounds with sharp attacks, or radar signals whose frequency may vary quickly at high frequencies.

Multiscale Zooming A wavelet dictionary is also well adapted to analyze the scaling evolution of transients with zooming procedures across scales. Suppose now that ψ is real. Since it has a zero average, a wavelet coefficient $Wf(u,s)$ measures the variation of f in a neighborhood of u whose size is proportional to s. Sharp signal transitions create large amplitude wavelet coefficients.

Signal singularities have specific scaling invariance characterized by Lipschitz exponents. Chapter 6 relates the pointwise regularity of f to the asymptotic decay of the wavelet transform amplitude $|Wf(u,s)|$, when s goes to zero. Singularities are detected by following across scales the local maxima of the wavelet transform.

In images, wavelet local maxima indicate the position of edges, which are sharp variations of the image intensity. It defines a scale-space approximation support of f from which precise image approximations are reconstructed. At different scales, the geometry of this local maxima support provides contours of image structures of varying sizes. This multiscale edge detection is particularly effective for pattern recognition in computer vision [138].

The zooming capability of the wavelet transform not only locates isolated singular events, but can also characterize more complex multifractal signals having non-isolated singularities. Mandelbrot [39] was the first to recognize the existence of multifractals in most corners of nature. Scaling one part of a multifractal produces a signal that is statistically similar to the whole. This self-similarity appears in the continuous wavelet transform, which modifies the analyzing scale. From global measurements of the wavelet transform decay, Chapter 6 measures the singularity distribution of multifractals. This is particularly important in analyzing their properties and testing multifractal models in physics or in financial time series.

1.3.4 Time-Frequency Orthonormal Bases

Orthonormal bases of time-frequency atoms remove all redundancy and define stable representations. A wavelet orthonormal basis is an example of time-frequency basis obtained by scaling a wavelet ψ with dyadic scales $s = 2^j$ and translating it by $2^j n$, which is written $\psi_{j,n}$. In the time-frequency plane, the Heisenberg resolution box of $\psi_{j,n}$ is a dilation by 2^j and translation by $2^j n$ of the Heisenberg box of ψ. A wavelet orthonormal is thus a subdictionary of the continuous wavelet transform dictionary, which yields a perfect tiling of the time-frequency plane illustrated in Figure 1.6.

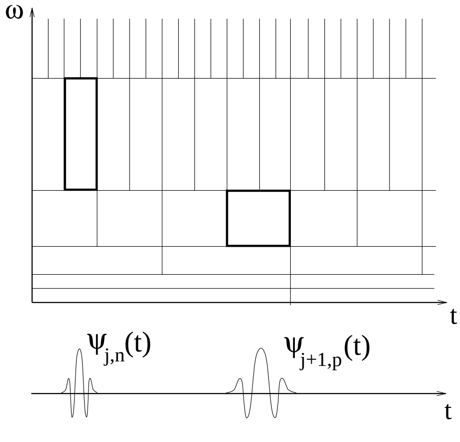


Figure 1.6: The time-frequency boxes of a wavelet basis define a tiling of the time-frequency plane.

One can construct many other orthonormal bases of time-frequency atoms, corresponding to different tilings of the time-frequency plane. Wavelet packet and local cosine bases are two important examples constructed in Chapter 8, with time-frequency atoms that split respectively the frequency and the time axis in intervals of varying sizes.

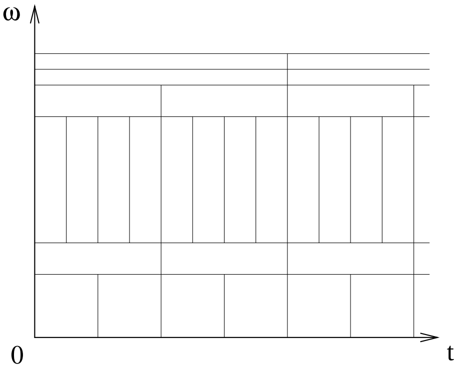


Figure 1.7: A wavelet packet basis divides the frequency axis in separate intervals of varying sizes. A tiling is obtained by translating in time the wavelet packets covering each frequency interval.

Wavelet Packet Bases Wavelet bases divide the frequency axis into intervals of 1 octave bandwidth. Coifman, Meyer and Wickerhauser [172] have generalized this construction with bases that split the frequency axis in intervals whose bandwidths may be adjusted. Each frequency interval is covered by the Heisenberg time-frequency boxes of wavelet packet functions translated in time, in order to cover the whole plane, as shown by Figure 1.7.

Like for wavelets, wavelet packet coefficients are obtained with a filter bank of conjugate mirror filters, that splits the frequency axis in several frequency intervals. Different frequency segmentations correspond to different wavelet packet bases. For images, a filter bank divides the image frequency support in squares of dyadic sizes, that can be adjusted.

Local Cosine Bases Local cosine orthonormal bases are constructed by dividing the time axis instead of the frequency axis. The time axis is segmented in successive intervals $[a_p, a_{p+1}]$. The local cosine bases of Malvar [335] are obtained by designing smooth windows $g_p(t)$ that cover each interval $[a_p, a_{p+1}]$, and multiplying them by cosine functions $\cos(\xi t+\phi)$ of different frequencies. This is yet another idea that was independently studied in physics, signal processing and mathematics. Malvar's original construction was for discrete signals. At the same time, the physicist Wilson [445] was designing a local cosine basis with smooth windows of infinite support, to analyze the properties of quantum coherent states. Malvar bases were also rediscovered and generalized by the harmonic analysts Coifman and Meyer [171]. These different views of the same bases brought to light mathematical and algorithmic properties that opened new applications.

A multiplication by $\cos(\xi t + \phi)$ translates the Fourier transform $\hat{g}_p(\omega)$ of $g_p(t)$ by $\pm\xi$. Over positive frequencies, the time-frequency box of the modulated window $g_p(t)\cos(\xi t + \phi)$ is therefore equal to the time-frequency box of g_p translated by ξ along frequencies. Figure 1.8 shows the time-frequency tiling corresponding to such a local cosine basis. For images, a two-dimensional cosine basis is constructed by dividing the image support in squares of varying sizes.

1.4 Sparsity in Redundant Dictionaries

In natural languages, large dictionaries are needed to refine ideas with short sentences, and they evolve with their usage. Eskimos have 8 different words to describe snow quality where as a single word is used in a Parisian dictionary. Similarly, large signal dictionaries of vectors are needed to construct sparse representations of complex signals. However, computing and optimizing a signal approximation by choosing the best M dictionary vectors is much more difficult.

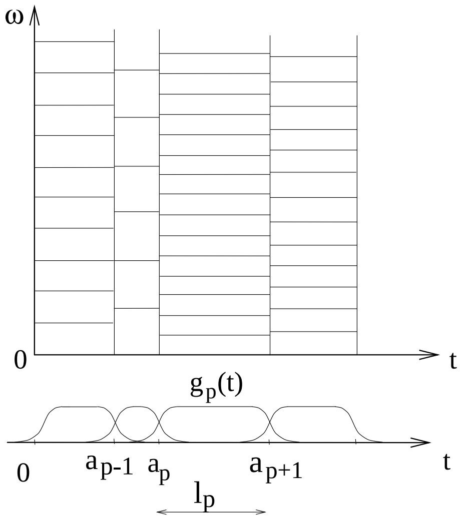


Figure 1.8: A local cosine basis divides the time axis with smooth windows $g_p(t)$, and translates these windows in frequency.

1.4.1 Frame Analysis and Synthesis

Suppose that a sparse family of vectors $\{\phi_p\}_{p\in\Lambda}$ has been selected to approximate a signal f. An approximation can be recovered as an orthogonal projection in the space $\mathbf{V}_\Lambda$ generated by these vectors. We are then facing one of the following two problems.

In a *dual synthesis* problem, the orthogonal projection f_Λ of f in $\mathbf{V}_\Lambda$ must be computed from dictionary coefficients $\{\langle f,\phi_p\rangle\}_{p\in\Lambda}$, provided by an analysis operator. This is the case when a signal transform $\{\langle f,\phi_p\rangle\}_{p\in\Gamma}$ is calculated in some large dictionary, and a subset of inner products are selected. Such inner products may correspond to coefficients above a threshold or local maxima values.

In a *dual analysis* problem, the decomposition coefficients of f_Λ must be computed on a family of selected vectors $\{\phi_p\}_{p\in\Lambda}$. This problem appears when sparse representation algorithms select vectors as opposed to inner products. This is the case for pursuit algorithms, which compute approximation supports in highly redundant dictionaries.

The frame theory gives energy equivalence conditions to solve both problems with stable operators. A family $\{\phi_p\}_{p\in\Lambda}$ is a frame of the space $\mathbf{V}$ it generates if there exists $B \geqslant A > 0$ such that

$$\forall h \in \mathbf{V} \quad , \quad A\|h\|^2 \leqslant \sum_{m\in\Lambda} |\langle h,\phi_p\rangle|^2 \leqslant B\|h\|^2 \ .$$

The representation is stable since any perturbation of frame coefficients implies a modification of similar magnitude on h. Chapter 5 derives the existence of a dual frame $\{\tilde{\phi}_p\}_{p\in\Lambda}$ that solves both the dual synthesis and analysis problems:

$$f_\Lambda = \sum_{p\in\Lambda} \langle f,\phi_p\rangle\, \tilde{\phi}_p = \sum_{p\in\Lambda} \langle f,\tilde{\phi}_p\rangle\, \phi_p \ . \tag{1.15}$$

Algorithms are provided to calculate these decompositions. The dual frame is also stable:

$$\forall f \in \mathbf{V} \quad , \quad B^{-1}\|f\|^2 \leqslant \sum_{m\in\Gamma} |\langle f,\tilde{\phi}_p\rangle|^2 \leqslant B^{-1}\|f\|^2 \ .$$

The frame bounds A and B are redundancy factors. If the vectors $\{\phi_p\}_{p\in\Gamma}$ are normalized and linearly independent then $A \leqslant 1 \leqslant B$. Such a dictionary is called a *Riesz basis* of $\mathbf{V}$ and the dual frame is biorthogonal:

$$\forall (p,p') \in \Lambda^2 \quad , \quad \langle \phi_p, \tilde{\phi}_{p'}\rangle = \delta[p-p'] \ .$$

Whent the basis is orthonormal then both bases are equal. Analysis and synthesis problems are then identical.

The frame theory is also be used to construct redundant dictionaries that define complete, stable and redundant signal representations, where $\mathbf{V}$ is then the whole signal space. The frame bounds measure the redundancy of such dictionaries. Chapter 5 studies the construction of windowed Fourier and wavelet frame dictionaries, by sampling their time, frequency and scaling parameters, while controlling frame bounds. In two dimensions, directional wavelet frames include wavelets sensitive to directional structures such as textures or edges in images.

To improve the sparsity of images having edges along regular geometric curves, Candès and Donoho [127] have introduced curvelet frames, with elongated waveforms having different directions, positions and scales. Images with piecewise regular edges have representations that are asymptotically more sparse by thresholding curvelet coefficients than wavelet coefficiens.

1.4.2 Ideal Dictionary Approximations

In a redundant dictionary $\mathcal{D} = \{\phi_p\}_{p\in\Gamma}$ we would like to find the best approximation support Λ with $M = |\Lambda|$ vectors, which minimize the error $\|f - f_\Lambda\|^2$. Chapter 12 proves that it is equivalent find Λ_T which minimizes the corresponding approximation Lagrangian

$$\mathcal{L}_0(T, f, \Lambda) = \|f - f_\Lambda\|^2 + T^2\,|\Lambda|\ , \tag{1.16}$$

for some multiplier T.

Compression and denoising are two applications of redundant dictionary approximations. When compressing signals by quantizing dictionary coefficients, the distortion-rate varies like the Lagrangian (1.16), with a multiplier T that depends on the quantization step. Optimizing the coder is thus equivalent to minimizing this approximation Lagrangian. For sparse representations, most of the bits are devoted to coding the geometry of the sparse approximation set Λ_T in Γ.

Estimators reducing noise from observations $X = f + W$ are also optimized by finding a best orthogonal projector over a set of dictionary vectors. The *model selection* theory of Barron, Birgé and Massart [91] proves that finding $\tilde{\Lambda}_T$ which minimizes this same Lagrangian $\mathcal{L}_0(T, X, \Lambda)$ defines an estimator whose risk is of the same order as the minimum approximation error $\|f - f_{\Lambda_T}\|^2$, up to a logarithmic factor. This is similar to the optimality result obtained for thresholding estimators in an orthonormal basis.

The bad news is that minimizing the approximation Lagrangian $\mathcal{L}_0$ is an NP-hard problem, and is therefore computationally intractable. It is thus necessary to find algorithms that are sufficiently fast, and which computes sub-optimal but "good enough" solutions.

Dictionaries of Orthonormal Bases To reduce the complexity of optimal approximations, the search can be reduced to sub-families of orthogonal dictionary vectors. In a dictionary of orthonormal bases, any family of orthogonal dictionary vectors can be complemented to form an orthogonal basis $\mathcal{B}$ included in $\mathcal{D}$. As a result the best approximation of f from orthogonal vectors in $\mathcal{B}$ is obtained by thresholding the coefficients of f in a "best basis" in $\mathcal{D}$.

For tree-dictionaries of orthonormal bases obtained by a recursive split of orthogonal vector spaces, the fast dynamic programming algorithm of Coifman and Wickerhauser [172] finds such a best basis with $O(P)$ operations, where P is the dictionary size.

Wavelet packet and local cosine bases are examples of tree-dictionaries of time-frequency orthonormal bases of size $P = N \log_2 N$. A best basis is a time-frequency tiling that best match the signal time-frequency structures.

To approximate edges that are geometrically regular, wavelet are not as efficient as curvelets, but wavelets provide more sparse representations of singularities that are not distributed along geometrically regular curves. Bandlet dictionaries, introduced by Le Pennec, Mallat and Peyré [311, 332], are dictionaries of orthonormal bases, that can adapt to the variability of the image geometric regularity. Minimax optimal asymptotic rates are derived for compression and denoising.

1.4.3 Pursuit in Dictionaries

Approximating signals only from orthogonal vectors brings rigidity which limits the ability to optimize the representation. Pursuit algorithms remove this constraint with flexible procedures that search for sparse though not necessarily optimal dictionary approximations. These approximations are computed by optimizing the choice of dictionary vectors $\{\phi_p\}_{p\in\Lambda}$.

Matching Pursuit Matching pursuit algorithms introduced by Mallat and Zhang [333] are greedy algorithms that optimize approximations by selecting one by one the dictionary vectors. The vector in $\phi_{p_0} \in \mathcal{D}$ that best approximates a signal f is

$$\phi_{p_0} = \underset{p\in\Gamma}{\operatorname{argmax}} \, |\langle f, \phi_p\rangle|$$

and the residual approximation error is

$$Rf = f - \langle f, \phi_{p_0}\rangle \, \phi_{p_0} \ .$$

A matching pursuit further approximates the residue Rf by selecting another best vector ϕ_{p_1} from the dictionary, and continues this process over the next order residues $R^m f$, which produces a signal decomposition:

$$f = \sum_{m=0}^{M-1} \langle R^m f, \phi_{p_m}\rangle \, \phi_{p_m} + R^M f \ .$$

The approximation from the M selected vectors $\{\phi_{p_m}\}_{0\leqslant m<M}$ can be refined with an orthogonal back projection on the space generated by these vectors. An orthogonal matching pursuit further improves this decomposition by orthogonalizing progressively the projection directions ϕ_{p_m}, during the decompositon. The resulting decompositions are applied to compression, denoising and pattern recognition of various types of signals, images and videos.

Basis Pursuit Approximating f with a minimum number of non-zero coefficients $a[p]$ in a dictionary $\mathcal{D}$ is equivalent to minimize the $\mathbf{l^0}$ norm $\|a\|_0$, which gives the number of non-zero coefficients. This $\mathbf{l^0}$ norm is highly non convex which explains why the resulting minimization is NP-hard. Donoho and Chen [148] thus proposed to replace the $\mathbf{l^0}$ norm by the $\mathbf{l^1}$ norm $\|a\|_1 = \sum_{p\in\Gamma} |a[p]|$ which is convex. The resulting basis pursuit algorithm computes a synthesis operator

$$f = \sum_{p\in\Gamma} a[p] \, \phi_p \quad \text{which minimizes} \quad \|a\|_1 = \sum_{p\in\Gamma} |a[p]| \ . \tag{1.17}$$

This optimal solution is calculated with a linear programming algorithm. A basis pursuit is computationally more intensive than a matching pursuit, but is a more global optimization that yields representations that can be more sparse.

In approximation, compression or denoising applications, f is recovered with an error which bounded by a precision parameter ε. The optimization (1.18) is thus relaxed by finding a synthesis such that

$$\|f - \sum_{p\in\Gamma} a[p] \, \phi_p\| \leqslant \varepsilon \quad \text{which minimizes} \quad \|a\|_1 = \sum_{p\in\Gamma} |a[p]| \ . \tag{1.18}$$

This is a convex minimization problem whose solution can be found by minimizing the corresponding $\mathbf{l^1}$ Lagrangian

$$\mathcal{L}_1(T, f, a) = \|f - \sum_{p\in\Gamma} a[p] \, \phi_p\|^2 + T \, \|a\|_1 \ ,$$

where T is a Lagrange multiplier that depends on ε. This is called an $\mathbf{l^1}$ Lagrangian pursuit in this book. A solution $\tilde{a}[p]$ is computed with iterative algorithms that are guaranteed to converge. The number of non-zero coordinates of $\tilde{a}$ typically decreases as T increases.

Incoherence for Support Recovery Matching pursuit and $\mathbf{l^1}$ Lagrangian pursuits are optimal if they recover the approximation support Λ_T which minimizes the approximation Lagrangian

$$\mathcal{L}_0(T, f, \Lambda) = \|f - f_\Lambda\|^2 + T^2 \, |\Lambda| \ ,$$

where f_Λ is the orthogonal projection of f in the space $\mathbf{V}_\Lambda$ generated by $\{\phi_p\}_{p\in\Lambda}$. This is not always true and depends on Λ_T. An *Exact Recovery Criteria* proved by Tropp [425] guarantees that pursuit algorithms do recover the optimal support Λ_T if

$$ERC(\Lambda_T) = \max_{q\notin\Lambda_T} \sum_{p\in\Lambda_T} |\langle\tilde{\phi}_p, \phi_q\rangle| < 1 \ , \tag{1.19}$$

where $\{\tilde{\phi}_p\}_{p\in\Lambda_T}$ is the biorthogonal basis of $\{\phi_p\}_{p\in\Lambda_T}$ in $\mathbf{V}_{\Lambda_T}$. This criteria implies that dictionary vectors ϕ_q outside Λ_T should have a small inner-product with vectors in Λ_T.

This recovery is stable relatively to noise perturbations if $\{\phi_p\}_{p\in\Lambda}$ has Riesz bounds that are not too far from 1. These vectors should be nearly orthogonal and hence have small inner-products. These small inner-product conditions are interpreted as a form of incoherence. A stable recovery of Λ_T is possible if vectors in Λ_T are incoherent with respect to other dictionary vectors and are incoherent between themselves. This opens the possibility of performing super-resolution with sparse representations, but it depends upon the geometric configuration of Λ_T in Γ.

1.5 Inverse Problems

Most digital measurement devices, such as cameras, microphones or medical imaging systems, can be modeled as a linear transformation of an incoming analog signal, plus a noise due to intrinsic measurement fluctuations or to electronic noises. This linear transformation can be decomposed into a stable analog to digital linear conversion followed by a discrete operator U that carries the specific transfer function of the measurement device. The resulting measured data can be written

$$Y[q] = Uf[q] + W[q]$$

where $f \in \mathbb{C}^N$ is the high resolution signal we want to recover, and $W[q]$ is the measurement noise. For a camera whose optics is out of focus, the operator U is a low-pass convolution producing a blur. For a magnetic resonance imaging system, U is a Radon transform integrating the signal along rays and the number Q of measurements is smaller than N. In such problems, U is not invertible and recovering an estimate of f is an *ill-posed* inverse problem.

Inverse problems are among the most difficult signal processing problems, with considerable applications. When data acquisition is difficult, costly or dangerous or when the signal is degraded, super-resolution is important to recover the highest possible resolution information. This applies to satellite observations, seismic exploration, medical imaging, radars, camera phones or degraded Internet videos displayed on high resolution screens. Separating mixed information sources from fewer measurements is yet another super-resolution problem in telecommunication or audio recognition.

Incoherence, sparsity and geometry play a crucial role in the solution of these ill-defined inverse problems. With a sensing matrix U with random coefficients, Candès, Romberg and Tao [132, 130] and Donoho [204] proved that super-resolution becomes stable for signals having a sufficiently sparse representation in a dictionary. This remarkable result opens the door to new compressive sensing devices and algorithms, that recover high resolution signals from few randomized linear measurements.

1.5.1 Diagonal Inverse Estimation

In an ill-posed inverse problem

$$Y = Uf + W$$

the image space $\mathbf{ImU} = \{Uh \ : \ h \in \mathbb{C}^N\}$ of U is of dimension Q smaller than the high resolution space N where f belongs. Inverse problems include two difficulties. In the image space $\mathbf{ImU}$ where U is invertible, its inverse may amplify the noise W which then needs to be reduced by an efficient denoising procedure. In the null space $\mathbf{NullU}$, all signals h are set to zero $Uh = 0$ and thus disappear in the measured data Y. Recovering the projection of f in $\mathbf{NullU}$ thus requires using some strong prior information. A super-resolution estimator recovers an estimation of f in a space of dimension larger than Q and hopefully equal to N, but this is not always possible.

Singular Value Decompositions Let $f = \sum_{m\in\Gamma} a[m]\, g_m$ be the representation of f in an orthonormal basis $\mathcal{B} = \{g_m\}_{m\in\Gamma}$. An approximation must be recovered from

$$Y = \sum_{m\in\Gamma} a[m]\, U g_m + W \ .$$

If $\mathcal{B}$ is chosen to diagonalize U^*U then U transforms a subset of Q vectors $\{g_m\}_{m\in\Gamma_Q}$ of $\mathcal{B}$ into an orthogonal basis $\{Ug_m\}_{m\in\Gamma_Q}$ of $\mathbf{ImU}$, and sets to zero all other vectors. A singular value decomposition estimates the coefficients $a[m]$ of f by projecting Y on this orthogonal basis and by renormalizing the resulting coefficients

$$\forall m \in \Gamma \quad , \quad \tilde{a}[m] = \frac{\langle Y, U g_m\rangle}{\|U g_m\|^2 + h_m^2} \ ,$$

where h_m^2 are regularization parameters. Such estimators recover non-zero coefficients in a space of dimension Q and thus brings no super-resolution. If U is a convolution operator then $\mathcal{B}$ is the Fourier basis and a singular value estimation implements a regularized inverse convolution.

Diagonal Thresholding Estimation The basis which diagonalizes U^*U rarely provides a sparse signal representation. For example, a Fourier basis which diagonalizes convolution operators does not approximate well signals including singularities.

Donoho [201] introduced more flexibility by looking for a basis $\mathcal{B}$ providing a sparse signal representation, and where a subset of Q vectors $\{g_m\}_{m\in\Gamma_Q}$ are transformed by U in a Riesz basis $\{Ug_m\}_{m\in\Gamma_Q}$ of $\mathbf{ImU}$, while the others are set to zero. With an appropriate renormalization, $\{\lambda_m\, U g_m\}_{m\in\Gamma_Q}$ has a biorthogonal basis $\{\tilde{\phi}_m\}_{m\in\Gamma_Q}$ which is normalized $\|\tilde{\phi}_m\| = 1$. The sparse coefficients of f in $\mathcal{B}$ can then be estimated with a thresholding

$$\forall m \in \Gamma_Q \quad , \quad \tilde{a}[m] = \rho_{T_m}(\lambda_m \langle Y, \tilde{\phi}_m\rangle) \quad \text{with} \quad \rho_T(x) = x\, \mathbf{1}_{|x|>T} \ ,$$

for thresholds T_m appropriately defined. For classes of signals that are sparse in $\mathcal{B}$, such thresholding estimators may yield a nearly minimax risk, but provides no super-resolution since this non-linear projector remains in a space of dimension Q. This result applies to classes of convolution operators U in wavelet or wavelet packet bases. Diagonal inverse estimators are computationally efficient and potentially optimal in cases where super-resolution is not possible.

1.5.2 Super-Resolution and Compressive Sensing

Suppose that f has a sparse representation in some dictionary $\mathcal{D}_0 = \{g_p\}_{p\in\Gamma}$ of P normalized vectors. The P vectors of the transformed dictionary $\mathcal{D} = U\mathcal{D}_0 = \{Ug_p\}_{p\in\Gamma}$ belong to the space $\mathbf{ImU}$ of dimension $Q < P$ and thus define a redundant dictionary. Vectors in the approximation support Λ of f are not restricted a priori to a particular subspace of $\mathbb{C}^N$. Super-resolution is possible if the approximation support Λ of f in $\mathcal{D}_0$ can be estimated by decomposing the noisy data Y over $\mathcal{D}$. It depends upon the properties of the approximation support Λ of f in Γ.

Geometric conditions for super-resolution Let $w_\Lambda = f - f_\Lambda$ be the approximation error of a sparse representation $f_\Lambda = \sum_{p\in\Lambda} a[p]\, g_p$ of f. The observed signal can be written

$$Y = Uf + W = \sum_{p\in\Lambda} a[p]\, U g_p + U w_\Lambda + W \ .$$

If the support Λ can be identified by finding a sparse approximation of Y in $\mathcal{D}$

$$Y_\Lambda = \sum_{p\in\Lambda} \tilde{a}[p]\, U g_p$$

then we can recover a super-resolution estimation of f

$$\tilde{F} = \sum_{p\in\Lambda} \tilde{a}[p]\, g_p \ .$$

This shows that super-resolution is possible if the approximation support Λ can be identified by decomposing Y in the redundant transformed dictionary $\mathcal{D}$. If the exact recovery criteria is satisfies $ERC(\Lambda) < 1$ and if $\{Ug_p\}_{p\in\Lambda}$ is a Riesz basis then Λ can be recovered using pursuit algorithms, with controlled error bounds.

For most operator U not all sparse approximation sets can be recovered. It is necessary to impose some further geometric conditions on Λ in Γ, which makes super-resolution difficult and often unstable. Numerical applications of $\mathbf{l^1}$ Lagrangian pursuit to tomography, super-resolution zooming and inpainting illustrate these results.

Compressive Sensing with Randomness Candès, Tao [132] and Donoho [204] proved that stable super-resolution is possible for any sufficiently sparse signal f, if U is an operator with random coefficients. Compressive sensing then becomes possible, by recovering a close approximation of $f \in \mathbb{C}^N$ from $Q \ll N$ linear measurements [126].

A recovery is stable for a sparse approximation set $|\Lambda| \leqslant M$ only if the corresponding dictionary family $\{Ug_m\}_{m\in\Lambda}$ is a Riesz basis of the space it generates. The *M-restricted isometry conditions* of Candès, Tao and Donoho [204] imposes uniform Riesz bounds for all sets $\Lambda \subset \Gamma$ with $|\Lambda| \leqslant M$:

$$\forall c \in \mathbb{C}^{|\Lambda|} \quad , \quad (1-\delta_M)\,\|c\|^2 \leqslant \|\sum_{m\in\Lambda} c[p]\,Ug_p\|^2 \leqslant (1+\delta_M)\,\|c\|^2 \; . \tag{1.20}$$

This is a strong incoherence condition on the P vectors of $\{Ug_m\}_{m\in\Gamma}$, which imposes that any subset of less than M vectors is nearly uniformly distributed on the unit sphere of $\mathbf{ImU}$. For an orthogonal basis $\mathcal{D}_0 = \{g_m\}_{m\in\Gamma}$, this is possible for $M \leqslant C\,(\log N)\,Q$, if U is a matrix with independent Gaussian random coefficients. A pursuit algorithm then provides a stable approximation of any $f \in C^N$, which is sufficiently well approximated by $M/3$ vectors in $\mathcal{D}_0$.

These results open a new compressive sensing approach to signal acquistion and representation. Instead of first discretizing linearly the signal at a high resolution N and then computing a non-linear representation over M coefficients in some dictionary, compressive sensing measures directly M randomized linear coefficients. A reconstructed signal is then recovered by a non-linear algorithm, producing an error that can be of the same order of magnitude as as the error obtained by the more classic two step approximation process, with a more economic acquisiton process. These results remain valid for several types of random matrices U. Examples of applications to single pixel camera, video super-resolution, new analog to digital converters and MRI imaging are described.

Blind Source Separation Sparsity in redundant dictionaries also provide efficient strategies to separate a family of signals $\{f_s\}_{0\leqslant s<S}$ which are linearly mixed in $K \leqslant S$ observed signals with noise:

$$Y_k[n] = \sum_{s=0}^{S-1} u_{k,s}\, f_s[n] + W_k[n] \quad \text{for} \quad 0 \leqslant n < N \quad \text{and} \quad 0 \leqslant k < K \; .$$

Most often the mixing matrix $U = \{u_{k,s}\}_{0\leqslant k<K, 0\leqslant s<S}$ is unknown. Source separation is a super-resolution problem since $S\,N$ data values must be recovered from $Q = K\,N \leqslant S\,N$ measurements. Not knowing the operator U makes it even more complicated.

If each source f_s has a sparse approximation support Λ_s in a dictionary $\mathcal{D}$, with $\sum_{s=0}^{S-1} |\Lambda_s| \ll N$, then it is likely that the sets $\{\Lambda_s\}_{0\leqslant s<s}$ are nearly disjoint. In this case, the operator U, the supports Λ_s and the sources f_s are approximated by computing sparse approximations of the observed data Y_k in $\mathcal{D}$. The distribution of these coefficients identifies the coefficients of the mixing matrix U, and the disjoint source supports. Time-frequency separation of sounds illustrate these results.

1.6 Travel Guide

Reproducible Computational Science This book covers the whole spectrum from theorems on functions of continuous variables to fast discrete algorithms and their applications. Section 1.1.2 argues that models based on continuous time functions give useful asymptotic results for understanding the behavior of discrete algorithms. Yet, a mathematical analysis alone is often unable to predict fully the behavior and suitability of algorithms for specific signals. Experiments are necessary and such experiments should be reproducible, just like experiments in other fields of sciences [117].

The reproducibility of experiments requires having the complete software and full source code for inspection, modification and application under varied parameter settings. Following this perspective, computational algorithms presented in this book are available as Matlab subroutines or in other software packages. Figures can be reproduced and the source code is available. Softwares, demonstrations and exercise solutions are available on the Web site *http://wavelet-tour.com*.

Book Road Map Some redundancy is introduced between sections to avoid imposing a linear progression through the book. The preface describes several possible programs for a sparse signal processing course.

All theorems are explained in the text and reading the proofs is not necessary to understand the results. Most theorems of the book are proved in detail, and important techniques are included. Exercises at the end of each chapter give examples of mathematical, algorithmic and numerical applications, ordered by level of difficulty from **1** to **4**, and solutions can be found in *http://wavelet-tour.com*.

The book begins with Chapters 2 and 3, which review the Fourier transform and linear discrete signal processing. They provide the necessary background for readers with no signal processing background. Important properties of linear operators, projectors and vector spaces can be found in Appendix A. Local time-frequency transforms and dictionaries are presented in Chapter 4. The wavelet and windowed Fourier transforms are introduced and compared. The measurement of instantaneous frequencies illustrates the limitations of time-frequency resolution. Dictionary stability and redundancy is introduced in Chapter 5 through the frame theory, with examples of windowed Fourier, wavelet and curvelet frames. Chapter 6 explains the relations between wavelet coefficient amplitude and local signal regularity. It is applied to the detection of singularities, edges and to the analysis of multifractals.

Wavelet bases and fast filter bank algorithms are important tools presented in Chapter 7. An overdose of orthonormal bases can strike the reader while studying the construction and properties of wavelet packets and local cosine bases in Chapter 8. It is thus important to read Chapter 9, which describes sparse approximations in bases. Signal compression and denoising applications of Chapters 10 and 11 give life to most theoretical and algorithmic results of the book. These chapters offer a practical perspective on the relevance of these linear and non-linear signal processing algorithms. Chapter 12 introduce sparse decompositions in redundant dictionaries. The resolution of inverse problems is an important area of applications, including super-resolution, compressive sensing and source separation.

II

Fourier Kingdom

The story begins in 1807 when Fourier presents a memoir to the Institut de France, where he claims that any periodic function can be represented as a series of harmonically related sinusoids. This idea had a profound impact in mathematical analysis, physics and engineering, but it took one and a half centuries to understand the convergence of Fourier series and complete the theory of Fourier integrals.

Fourier was motivated by the study of heat diffusion, which is governed by a linear differential equation. However, the Fourier transform diagonalizes all linear time-invariant operators, which are the building blocks of signal processing. It is therefore not only the starting point of our exploration but the basis of all further developments.

2.1 Linear Time-Invariant Filtering

Classical signal processing operations such as signal transmission, stationary noise removal or predictive coding are implemented with linear time-invariant operators. The time invariance of an operator L means that if the input $f(t)$ is delayed by τ, $f_\tau(t) = f(t-\tau)$, then the output is also delayed by τ:

$$g(t) = Lf(t) \quad \Rightarrow \quad g(t-\tau) = Lf_\tau(t). \tag{2.1}$$

For numerical stability, the operator L must have a weak form of continuity, which means that Lf is modified by a small amount if f is slightly modified. This weak continuity is formalized by the theory of distributions [58, 61], which guarantees that we are on a safe ground without further worrying about it.

2.1.1 Impulse Response

Linear time-invariant systems are characterized by their response to a Dirac impulse, defined in Appendix A.7. If f is continuous, its value at t is obtained by an "integration" against a Dirac located at t. Let $\delta_u(t) = \delta(t-u)$:

$$f(t) = \int_{-\infty}^{+\infty} f(u)\, \delta_u(t)\, du.$$

The continuity and linearity of L imply that

$$Lf(t) = \int_{-\infty}^{+\infty} f(u)\, L\delta_u(t)\, du.$$

Let h be the impulse response of L:

$$h(t) = L\delta(t).$$

The time-invariance proves that $L\delta_u(t) = h(t-u)$ and hence

$$Lf(t) = \int_{-\infty}^{+\infty} f(u)\, h(t-u)\, du = \int_{-\infty}^{+\infty} h(u) f(t-u)\, du = h \star f(t). \tag{2.2}$$

A time-invariant linear filter is thus equivalent to a convolution with the impulse response h. The continuity of f is not necessary. This formula remains valid for any signal f for which the convolution integral converges.

Let us recall a few useful properties of convolution products:

- Commutativity
$$f \star h(t) = h \star f(t). \tag{2.3}$$
- Differentiation
$$\frac{d}{dt}(f \star h)(t) = \frac{df}{dt} \star h(t) = f \star \frac{dh}{dt}(t). \tag{2.4}$$
- Dirac convolution
$$f \star \delta_\tau(t) = f(t-\tau). \tag{2.5}$$

Stability and Causality A filter is said to be *causal* if $Lf(t)$ does not depend on the values $f(u)$ for $u > t$. Since

$$Lf(t) = \int_{-\infty}^{+\infty} h(u)\, f(t-u)\, du,$$

this means that $h(u) = 0$ for $u < 0$. Such impulse responses are said to be *causal*.

The *stability* property guarantees that $Lf(t)$ is bounded if $f(t)$ is bounded. Since

$$|Lf(t)| \leqslant \int_{-\infty}^{+\infty} |h(u)|\, |f(t-u)|\, du \leqslant \sup_{u \in \mathbb{R}} |f(u)| \int_{-\infty}^{+\infty} |h(u)|\, du,$$

it is sufficient that $\int_{-\infty}^{+\infty} |h(u)|\, du < +\infty$. One can verify that this condition is also necessary if h is a function. We thus say that h is *stable* if it is integrable.

Example 2.1. *An amplification and delay system is defined by*

$$Lf(t) = \lambda\, f(t-\tau).$$

The impulse response of this filter is $h(t) = \lambda\, \delta(t-\tau)$.

Example 2.2. *A uniform averaging of* f *over intervals of size* T *is calculated by*

$$Lf(t) = \frac{1}{T} \int_{t-T/2}^{t+T/2} f(u)\, du.$$

This integral can be rewritten as a convolution of f *with the impulse response* $h = 1/T\, \mathbf{1}_{[-T/2,T/2]}$.

2.1.2 Transfer Functions

Complex exponentials $\mathrm{e}^{i\omega t}$ are eigenvectors of convolution operators. Indeed

$$L\mathrm{e}^{i\omega t} = \int_{-\infty}^{+\infty} h(u)\, \mathrm{e}^{i\omega(t-u)}\, du,$$

which yields

$$L\mathrm{e}^{i\omega t} = \mathrm{e}^{it\omega} \int_{-\infty}^{+\infty} h(u)\, \mathrm{e}^{-i\omega u}\, du = \hat{h}(\omega)\, \mathrm{e}^{i\omega t}.$$

The eigenvalue

$$\hat{h}(\omega) = \int_{-\infty}^{+\infty} h(u)\, \mathrm{e}^{-i\omega u}\, du$$

is the Fourier transform of h at the frequency ω. Since complex sinusoidal waves $\mathrm{e}^{i\omega t}$ are the eigenvectors of time-invariant linear systems, it is tempting to try to decompose any function f as a sum of these eigenvectors. We are then able to express Lf directly from the eigenvalues $\hat{h}(\omega)$. The Fourier analysis proves that under weak conditions on f, it is indeed possible to write it as a Fourier integral.

2.2 Fourier Integrals

To avoid convergence issues, the Fourier integral is first defined over the space $\mathbf{L^1}(\mathbb{R})$ of integrable functions [52]. It is then extended to the space $\mathbf{L^2}(\mathbb{R})$ of finite energy functions [21].

2.2.1 Fourier Transform in $\mathbf{L^1}(\mathbb{R})$

The Fourier integral

$$\hat{f}(\omega) = \int_{-\infty}^{+\infty} f(t)\, \mathrm{e}^{-i\omega t}\, dt \tag{2.6}$$

measures "how much" oscillations at the frequency ω there is in f. If $f \in \mathbf{L^1}(\mathbb{R})$ this integral does converge and

$$|\hat{f}(\omega)| \leqslant \int_{-\infty}^{+\infty} |f(t)|\, dt < +\infty. \tag{2.7}$$

The Fourier transform is thus bounded, and one can verify that it is a continuous function of ω (Exercise 2.1). If $\hat{f}$ is also integrable, the following theorem gives the inverse Fourier transform.

Theorem 2.1 (Inverse Fourier Transform)**.** *If $f \in \mathbf{L^1}(\mathbb{R})$ and $\hat{f} \in \mathbf{L^1}(\mathbb{R})$ then*

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \hat{f}(\omega)\, \mathrm{e}^{i\omega t}\, d\omega. \tag{2.8}$$

Proof. Replacing $\hat{f}(\omega)$ by its integral expression yields

$$\frac{1}{2\pi} \int_{-\infty}^{+\infty} \hat{f}(\omega)\, \exp(i\omega t)\, d\omega = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{+\infty} f(u)\, \exp[i\omega(t-u)]\, du \right) d\omega.$$

We cannot apply the Fubini Theorem A.2 directly because $f(u) \exp[i\omega(t-u)]$ is not integrable in $\mathbb{R}^2$. To avoid this technical problem, we multiply by $\exp(-\varepsilon^2\omega^2/4)$ which converges to 1 when ε goes to 0. Let us define

$$I_\varepsilon(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{+\infty} f(u)\, \exp\left(\frac{-\varepsilon^2\omega^2}{4}\right) \exp[i\omega(t-u)]\, du \right) d\omega. \tag{2.9}$$

We compute I_ε in two different ways using the Fubini theorem. The integration with respect to u gives

$$I_\varepsilon(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \hat{f}(\omega)\, \exp\left(\frac{-\varepsilon^2\omega^2}{4}\right) \exp(i\omega t)\, d\omega.$$

Since

$$\left| \hat{f}(\omega)\, \exp\left(\frac{-\varepsilon^2\omega^2}{4}\right) \exp[i\omega(t-u)] \right| \leqslant |\hat{f}(\omega)|$$

and since $\hat{f}$ is integrable, we can apply the dominated convergence Theorem A.1, which proves that

$$\lim_{\varepsilon \to 0} I_\varepsilon(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \hat{f}(\omega)\, \exp(i\omega t)\, d\omega. \tag{2.10}$$

Let us now compute the integral (2.9) differently by applying the Fubini theorem and integrating with respect to ω:

$$I_\varepsilon(t) = \int_{-\infty}^{+\infty} g_\varepsilon(t-u)\, f(u)\, du, \tag{2.11}$$

with

$$g_\varepsilon(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \exp(ix\omega)\, \exp\left(\frac{-\varepsilon^2\omega^2}{4}\right) d\omega.$$

A change of variable $\omega' = \varepsilon\omega$ shows that $g_\varepsilon(x) = \varepsilon^{-1} g_1(\varepsilon^{-1}x)$, and it is proved in (2.32) that $g_1(x) = \pi^{-1/2}\, \mathrm{e}^{-x^2}$. The Gaussian g_1 has an integral equal to 1 and a fast decay. The squeezed Gaussians g_ε have an integral that remains equal to 1, and thus they converge to a Dirac δ when ε goes to 0. By inserting (2.11) one can thus verify that

$$\lim_{\varepsilon \to 0} \int_{-\infty}^{+\infty} |I_\varepsilon(t) - f(t)|\, dt = \lim_{\varepsilon \to 0} \iint g_\varepsilon(t-u)\, |f(u) - f(t)|\, du\, dt = 0.$$

Inserting (2.10) proves (2.8). ■ ■

The inversion formula (2.8) decomposes f as a sum of sinusoidal waves $\mathrm{e}^{i\omega t}$ of amplitude $\hat{f}(\omega)$. By using this formula, we can show (Exercise 2.1) that the hypothesis $\hat{f} \in \mathbf{L^1}(\mathbb{R})$ implies that f must be continuous. The reconstruction (2.8) is therefore not proved for discontinuous functions. The extension of the Fourier transform to the space $\mathbf{L^2}(\mathbb{R})$ will address this issue.

The most important property of the Fourier transform for signal processing applications is the convolution theorem. It is another way to express the fact that sinusoidal waves $\mathrm{e}^{it\omega}$ are eigenvalues of convolution operators.

Theorem 2.2 (Convolution)**.** *Let $f \in \mathbf{L^1}(\mathbb{R})$ and $h \in \mathbf{L^1}(\mathbb{R})$. The function $g = h \star f$ is in $\mathbf{L^1}(\mathbb{R})$ and*

$$\hat{g}(\omega) = \hat{h}(\omega)\,\hat{f}(\omega). \tag{2.12}$$

Proof.

$$\hat{g}(\omega) = \int_{-\infty}^{+\infty} \exp(-it\omega)\left(\int_{-\infty}^{+\infty} f(t-u)\,h(u)\,du\right) dt.$$

Since $|f(t-u)||h(u)|$ is integrable in $\mathbb{R}^2$, we can apply the Fubini Theorem A.2, and the change of variable $(t,u) \to (v = t-u, u)$ yields

$$\begin{aligned}\hat{g}(\omega) &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \exp[-i(u+v)\omega]\,f(v)\,h(u)\,du\,dv \\ &= \left(\int_{-\infty}^{+\infty} \exp(-iv\omega)\,f(v)\,dv\right)\left(\int_{-\infty}^{+\infty} \exp(-iu\omega)\,h(u)\,du\right),\end{aligned}$$

which verifies (2.12). ■ ■

The response $Lf = g = f \star h$ of a linear time-invariant system can be calculated from its Fourier transform $\hat{g}(\omega) = \hat{f}(\omega)\,\hat{h}(\omega)$ with the inverse Fourier formula

$$g(t) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \hat{g}(\omega)\,\mathrm{e}^{i\omega t}\,d\omega, \tag{2.13}$$

which yields

$$Lf(t) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \hat{h}(\omega)\,\hat{f}(\omega)\,\mathrm{e}^{i\omega t}\,d\omega. \tag{2.14}$$

Each frequency component $\mathrm{e}^{it\omega}$ of amplitude $\hat{f}(\omega)$ is amplified or attenuated by $\hat{h}(\omega)$. Such a convolution is thus called a *frequency filtering*, and $\hat{h}$ is the *transfer function* of the filter.

The following table summarizes important properties of the Fourier transform, often used in calculations. Most of these formulas are proved with a change of variable in the Fourier integral.

Property	**Function**	**Fourier Transform**	
	$f(t)$	$\hat{f}(\omega)$	
Inverse	$\hat{f}(t)$	$2\pi\, f(-\omega)$	(2.15)
Convolution	$f_1 \star f_2(t)$	$\hat{f}_1(\omega)\,\hat{f}_2(\omega)$	(2.16)
Multiplication	$f_1(t)\,f_2(t)$	$\frac{1}{2\pi}\,\hat{f}_1 \star \hat{f}_2(\omega)$	(2.17)
Translation	$f(t-u)$	$\mathrm{e}^{-iu\omega}\,\hat{f}(\omega)$	(2.18)
Modulation	$\mathrm{e}^{i\xi t}\,f(t)$	$\hat{f}(\omega-\xi)$	(2.19)
Scaling	$f(t/s)$	$\lvert s\rvert\,\hat{f}(s\,\omega)$	(2.20)
Time derivatives	$f^{(p)}(t)$	$(i\omega)^p\,\hat{f}(\omega)$	(2.21)
Frequency derivatives	$(-it)^p\,f(t)$	$\hat{f}^{(p)}(\omega)$	(2.22)
Complex conjugate	$f^*(t)$	$\hat{f}^*(-\omega)$	(2.23)
Hermitian symmetry	$f(t) \in \mathbb{R}$	$\hat{f}(-\omega) = \hat{f}^*(\omega)$	(2.24)

2.2.2 Fourier Transform in $\mathbf{L^2}(\mathbb{R})$

The Fourier transform of the indicator function $f = \mathbf{1}_{[-1,1]}$ is

$$\hat{f}(\omega) = \int_{-1}^{1} \mathrm{e}^{-i\omega t}\, dt = \frac{2\sin\omega}{\omega}.$$

This function is not integrable because f is not continuous, but its square is integrable. The inverse Fourier transform Theorem 2.1 thus does not apply. This motivates the extension of the Fourier transform to the space $\mathbf{L^2}(\mathbb{R})$ of functions f with a finite energy $\int_{-\infty}^{+\infty} |f(t)|^2\, dt < +\infty$. By working in the Hilbert space $\mathbf{L^2}(\mathbb{R})$, we also have access to all the facilities provided by the existence of an inner product. The inner product of $f \in \mathbf{L^2}(\mathbb{R})$ and $g \in \mathbf{L^2}(\mathbb{R})$ is

$$\langle f, g\rangle = \int_{-\infty}^{+\infty} f(t)\, g^*(t)\, dt,$$

and the resulting norm in $\mathbf{L^2}(\mathbb{R})$ is

$$\|f\|^2 = \langle f, f\rangle = \int_{-\infty}^{+\infty} |f(t)|^2\, dt.$$

The following theorem proves that inner products and norms in $\mathbf{L^2}(\mathbb{R})$ are conserved by the Fourier transform up to a factor of 2π. Equations (2.25) and (2.26) are called respectively the *Parseval* and *Plancherel* formulas.

Theorem 2.3. *If f and h are in $\mathbf{L^1}(\mathbb{R}) \cap \mathbf{L^2}(\mathbb{R})$ then*

$$\int_{-\infty}^{+\infty} f(t)\, h^*(t)\, dt = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \hat{f}(\omega)\, \hat{h}^*(\omega)\, d\omega. \tag{2.25}$$

For $h = f$ it follows that

$$\int_{-\infty}^{+\infty} |f(t)|^2\, dt = \frac{1}{2\pi}\int_{-\infty}^{+\infty} |\hat{f}(\omega)|^2\, d\omega. \tag{2.26}$$

Proof. Let $g = f \star \bar{h}$ with $\bar{h}(t) = h^*(-t)$. The convolution Theorem 2.2 and property (2.23) show that $\hat{g}(\omega) = \hat{f}(\omega)\, \hat{h}^*(\omega)$. The reconstruction formula (2.8) applied to $g(0)$ yields

$$\int_{-\infty}^{+\infty} f(t)\, h^*(t)\, dt = g(0) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \hat{g}(\omega)\, d\omega = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \hat{f}(\omega)\, \hat{h}^*(\omega)\, d\omega.$$

■ ■

Density Extension in $\mathbf{L^2}(\mathbb{R})$ If $f \in \mathbf{L^2}(\mathbb{R})$ but $f \notin \mathbf{L^1}(\mathbb{R})$, its Fourier transform cannot be calculated with the Fourier integral (2.6) because $f(t)\, \mathrm{e}^{i\omega t}$ is not integrable. It is defined as a limit using the Fourier transforms of functions in $\mathbf{L^1}(\mathbb{R}) \cap \mathbf{L^2}(\mathbb{R})$.

Since $\mathbf{L^1}(\mathbb{R}) \cap \mathbf{L^2}(\mathbb{R})$ is dense in $\mathbf{L^2}(\mathbb{R})$, one can find a family $\{f_n\}_{n\in\mathbb{Z}}$ of functions in $\mathbf{L^1}(\mathbb{R}) \cap \mathbf{L^2}(\mathbb{R})$ that converges to f:

$$\lim_{n\to+\infty} \|f - f_n\| = 0.$$

Since $\{f_n\}_{n\in\mathbb{Z}}$ converges, it is a Cauchy sequence, which means that $\|f_n - f_p\|$ is arbitrarily small if n and p are large enough. Moreover, $f_n \in \mathbf{L^1}(\mathbb{R})$, so its Fourier transform $\hat{f}_n$ is well defined. The Plancherel formula (2.26) proves that $\{\hat{f}_n\}_{n\in\mathbb{Z}}$ is also a Cauchy sequence because

$$\|\hat{f}_n - \hat{f}_p\| = \sqrt{2\pi}\, \|f_n - f_p\|$$

is arbitrarily small for n and p large enough. A Hilbert space (Appendix A.2) is complete, which means that all Cauchy sequences converge to an element of the space. Hence, there exists $\hat{f} \in \mathbf{L^2}(\mathbb{R})$ such that

$$\lim_{n\to+\infty} \|\hat{f} - \hat{f}_n\| = 0.$$

By definition, $\hat{f}$ is the Fourier transform of f. This extension of the Fourier transform to $\mathbf{L^2}(\mathbb{R})$ satisfies the convolution theorem, the Parseval and Plancherel formulas, as well as all properties (2.15-2.24).

Diracs Diracs are often used in calculations; their properties are summarized in Appendix A.7. A Dirac δ associates to a function its value at $t = 0$. Since $\mathrm{e}^{i\omega t} = 1$ at $t = 0$ it seems reasonable to define its Fourier transform by

$$\hat{\delta}(\omega) = \int_{-\infty}^{+\infty} \delta(t)\, \mathrm{e}^{-i\omega t}\, dt = 1. \tag{2.27}$$

This formula is justified mathematically by the extension of the Fourier transform to tempered distributions [58, 61].

2.2.3 Examples

The following examples often appear in Fourier calculations. They also illustrate important Fourier transform properties.

- The *indicator function* $f = \mathbf{1}_{[-T,T]}$ is discontinuous at $t = \pm T$. Its Fourier transform is therefore not integrable:

  $$\hat{f}(\omega) = \int_{-T}^{T} \mathrm{e}^{-i\omega t}\, dt = \frac{2\sin(T\omega)}{\omega}. \tag{2.28}$$

- An *ideal low-pass filter* has a transfer function $\hat{\phi} = \mathbf{1}_{[-\xi,\xi]}$ that selects low frequencies over $[-\xi, \xi]$. The impulse response is calculated with the inverse Fourier integral (2.8):

  $$\phi(t) = \frac{1}{2\pi} \int_{-\xi}^{\xi} \mathrm{e}^{i\omega t}\, d\omega = \frac{\sin(\xi t)}{\pi t}. \tag{2.29}$$

- A *passive electronic circuit* implements analog filters with resistances, capacities and inductors. The input voltage $f(t)$ is related to the output voltage $g(t)$ by a differential equation with constant coefficients:

  $$\sum_{k=0}^{K} a_k\, f^{(k)}(t) = \sum_{k=0}^{M} b_k\, g^{(k)}(t). \tag{2.30}$$

  Suppose that the circuit is not charged for $t < 0$, which means that $f(t) = g(t) = 0$. The output g is a linear time-invariant function of f and can thus be written $g = f \star \phi$. Computing the Fourier transform of (2.30) and applying (2.22) proves that

  $$\hat{\phi}(\omega) = \frac{\hat{g}(\omega)}{\hat{f}(\omega)} = \frac{\sum_{k=0}^{K} a_k\, (i\omega)^k}{\sum_{k=0}^{M} b_k\, (i\omega)^k}. \tag{2.31}$$

  It is therefore a rational function of $i\omega$. An ideal low-pass transfer function $\mathbf{1}_{[-\xi,\xi]}$ thus cannot be implemented by an analog circuit. It must be approximated by a rational function. Chebyshev or Butterworth filters are often used for this purpose [12].

- A *Gaussian* $f(t) = \exp(-t^2)$ is a $\mathbf{C}^\infty$ function with a fast asymptotic decay. Its Fourier transform is also a Gaussian:

  $$\hat{f}(\omega) = \sqrt{\pi}\, \exp(-\omega^2/4). \tag{2.32}$$

  This Fourier transform is computed by showing with an integration by parts that $\hat{f}(\omega) = \int_{-\infty}^{+\infty} \exp(-t^2)\, \mathrm{e}^{-i\omega t} dt$ is differentiable and satisfies the differential equation

  $$2\, \hat{f}'(\omega) + \omega\, \hat{f}(\omega) = 0. \tag{2.33}$$

  The solution of this equation is a Gaussian $\hat{f}(\omega) = K \exp(-\omega^2/4)$, and since $\hat{f}(0) = \int_{-\infty}^{+\infty} \exp(-t^2)\, dt = \sqrt{\pi}$, we obtain (2.32).

- A Gaussian *chirp* $f(t) = \exp[-(a - ib)t^2]$ has a Fourier transform calculated with a similar differential equation:

  $$\hat{f}(\omega) = \sqrt{\frac{\pi}{a - ib}}\, \exp\left(\frac{-(a + ib)\omega^2}{4(a^2 + b^2)}\right). \tag{2.34}$$

- A translated *Dirac* $\delta_\tau(t) = \delta(t-\tau)$ has a Fourier transform calculated by evaluating $\mathrm{e}^{-i\omega t}$ at $t = \tau$:
$$\hat{\delta}_\tau(\omega) = \int_{-\infty}^{+\infty} \delta(t-\tau)\,\mathrm{e}^{-i\omega t}\,dt = \mathrm{e}^{-i\omega\tau}. \tag{2.35}$$
- The *Dirac comb* is a sum of translated Diracs
$$c(t) = \sum_{n=-\infty}^{+\infty} \delta(t-nT)$$
that is used to uniformly sample analog signals. Its Fourier transform is derived from (2.35):
$$\hat{c}(\omega) = \sum_{n=-\infty}^{+\infty} \mathrm{e}^{-inT\omega}. \tag{2.36}$$
The Poisson formula proves that it is also equal to a Dirac comb with a spacing equal to $2\pi/T$.

Theorem 2.4 (Poisson Formula)**.** *In the sense of distribution equalities (A.32),*
$$\sum_{n=-\infty}^{+\infty} \mathrm{e}^{-inT\omega} = \frac{2\pi}{T} \sum_{k=-\infty}^{+\infty} \delta\left(\omega - \frac{2\pi k}{T}\right). \tag{2.37}$$

Proof. The Fourier transform $\hat{c}$ in (2.36) is periodic with period $2\pi/T$. To verify the Poisson formula, it is therefore sufficient to prove that the restriction of $\hat{c}$ to $[-\pi/T, \pi/T]$ is equal to $2\pi/T\,\delta$. The formula (2.37) is proved in the sense of a distribution equality (A.32) by showing that for any test function $\hat{\theta}(\omega)$ with a support included in $[-\pi/T, \pi/T]$,
$$\langle \hat{c}, \hat{\theta}\rangle = \lim_{N\to+\infty} \int_{-\infty}^{+\infty} \sum_{n=-N}^{N} \exp(-inT\omega)\,\hat{\theta}(\omega)\,d\omega = \frac{2\pi}{T}\,\hat{\theta}(0).$$

The sum of the geometric series is
$$\sum_{n=-N}^{N} \exp(-inT\omega) = \frac{\sin[(N+1/2)T\omega]}{\sin[T\omega/2]}. \tag{2.38}$$

Hence
$$\langle \hat{c}, \hat{\theta}\rangle = \lim_{N\to+\infty} \frac{2\pi}{T} \int_{-\pi/T}^{\pi/T} \frac{\sin[(N+1/2)T\omega]}{\pi\omega}\,\frac{T\omega/2}{\sin[T\omega/2]}\,\hat{\theta}(\omega)\,d\omega. \tag{2.39}$$

Let
$$\hat{\psi}(\omega) = \begin{cases} \hat{\theta}(\omega)\,\frac{T\omega/2}{\sin[T\omega/2]} & \text{if } |\omega| \leqslant \pi/T \\ 0 & \text{if } |\omega| > \pi/T \end{cases}$$

and $\psi(t)$ be the inverse Fourier transform of $\hat{\psi}(\omega)$. Since $2\omega^{-1}\sin(a\omega)$ is the Fourier transform of $\mathbf{1}_{[-a,a]}(t)$, the Parseval formula (2.25) implies
$$\begin{aligned} \langle \hat{c}, \hat{\theta}\rangle &= \lim_{N\to+\infty} \frac{2\pi}{T} \int_{-\infty}^{+\infty} \frac{\sin[(N+1/2)T\omega]}{\pi\omega}\,\hat{\psi}(\omega)\,d\omega \\ &= \lim_{N\to+\infty} \frac{2\pi}{T} \int_{-(N+1/2)T}^{(N+1/2)T} \psi(t)\,dt. \end{aligned} \tag{2.40}$$

When N goes to $+\infty$ the integral converges to $\hat{\psi}(0) = \hat{\theta}(0)$. ■ ■

2.3 Properties

2.3.1 Regularity and Decay

The global regularity of a signal f depends on the decay of $|\hat{f}(\omega)|$ when the frequency ω increases. The differentiability of f is studied. If $\hat{f} \in \mathbf{L^1}(\mathbb{R})$, then the Fourier inversion formula (2.8) implies that f is continuous and bounded:

$$|f(t)| \leqslant \frac{1}{2\pi}\int_{-\infty}^{+\infty} |\mathrm{e}^{i\omega t}\hat{f}(\omega)|\,d\omega = \frac{1}{2\pi}\int_{-\infty}^{+\infty} |\hat{f}(\omega)|\,d\omega < +\infty\,. \tag{2.41}$$

The next theorem applies this property to obtain a sufficient condition that guarantees the differentiability of f at any order p.

Theorem 2.5. *A function f is bounded and p times continuously differentiable with bounded derivatives if*

$$\int_{-\infty}^{+\infty} |\hat{f}(\omega)|\,(1+|\omega|^p)\,d\omega < +\infty\,. \tag{2.42}$$

Proof. The Fourier transform of the k^{th} order derivative $f^{(k)}(t)$ is $(i\omega)^k\hat{f}(\omega)$. Applying (2.41) to this derivative proves that

$$|f^{(k)}(t)| \leqslant \int_{-\infty}^{+\infty} |\hat{f}(\omega)|\,|\omega|^k\,d\omega.$$

Condition (2.42) implies that $\int_{-\infty}^{+\infty}|\hat{f}(\omega)||\omega|^k\,d\omega < +\infty$ for any $k \leqslant p$, so $f^{(k)}(t)$ is continuous and bounded. ■ ■

This result proves that if there exist a constant K and $\varepsilon > 0$ such that

$$|\hat{f}(\omega)| \leqslant \frac{K}{1+|\omega|^{p+1+\varepsilon}}, \quad \text{then} \quad f \in \mathbf{C}^p.$$

If $\hat{f}$ has a compact support then (2.42) implies that $f \in \mathbf{C}^\infty$.

The decay of $|\hat{f}(\omega)|$ depends on the worst singular behavior of f. For example, $f = \mathbf{1}_{[-T,T]}$ is discontinuous at $t = \pm T$, so $|\hat{f}(\omega)|$ decays like $|\omega|^{-1}$. In this case, it could also be important to know that $f(t)$ is regular for $t \neq \pm T$. This information cannot be derived from the decay of $|\hat{f}(\omega)|$. To characterize local regularity of a signal f it is necessary to decompose it over waveforms that are well localized in time, as opposed to sinusoidal waves $\mathrm{e}^{i\omega t}$. Section 6.1.3 explains that wavelets are particularly well adapted to this purpose.

2.3.2 Uncertainty Principle

Can we construct a function f whose energy is well localized in time and whose Fourier transform $\hat{f}$ has an energy concentrated in a small frequency neighborhood? The Dirac $\delta(t-u)$ has a support restricted to $t = u$ but its Fourier transform $\mathrm{e}^{-iu\omega}$ has an energy uniformly spread over all frequencies. We know that $|\hat{f}(\omega)|$ decays quickly at high frequencies only if f has regular variations in time. The energy of f must therefore be spread over a relatively large domain.

To reduce the time spread of f, we can scale it by $s < 1$ while maintaining constant its total energy. If

$$f_s(t) = \frac{1}{\sqrt{s}}\,f\left(\frac{t}{s}\right) \quad \text{then} \quad \|f_s\|^2 = \|f\|^2.$$

The Fourier transform $\hat{f}_s(\omega) = \sqrt{s}\,\hat{f}(s\omega)$ is dilated by $1/s$ so we lose in frequency localization what we gained in time. Underlying is a trade-off between time and frequency localization.

Time and frequency energy concentrations are restricted by the Heisenberg uncertainty principle. This principle has a particularly important interpretation in quantum mechanics as an uncertainty as to the position and momentum of a free particle. The state of a one-dimensional particle is described by a wave function $f \in \mathbf{L^2}(\mathbb{R})$. The probability density that this particle is located at t is $\frac{1}{\|f\|^2}|f(t)|^2$. The probability density that its momentum is equal to ω is $\frac{1}{2\pi\|f\|^2}|\hat{f}(\omega)|^2$. The average location of this particle is

$$u = \frac{1}{\|f\|^2}\int_{-\infty}^{+\infty} t\,|f(t)|^2\,dt, \tag{2.43}$$

and the average momentum is

$$\xi = \frac{1}{2\pi \|f\|^2} \int_{-\infty}^{+\infty} \omega \, |\hat{f}(\omega)|^2 \, d\omega. \tag{2.44}$$

The variances around these average values are respectively

$$\sigma_t^2 = \frac{1}{\|f\|^2} \int_{-\infty}^{+\infty} (t-u)^2 \, |f(t)|^2 \, dt \tag{2.45}$$

and

$$\sigma_\omega^2 = \frac{1}{2\pi \|f\|^2} \int_{-\infty}^{+\infty} (\omega - \xi)^2 \, |\hat{f}(\omega)|^2 \, d\omega. \tag{2.46}$$

The larger σ_t, the more uncertainty there is concerning the position of the free particle; the larger σ_ω, the more uncertainty there is concerning its momentum.

Theorem 2.6 (Heisenberg Uncertainty)**.** *The temporal variance and the frequency variance of $f \in \mathbf{L^2}(\mathbb{R})$ satisfy*

$$\sigma_t^2 \, \sigma_\omega^2 \geqslant \frac{1}{4}. \tag{2.47}$$

This inequality is an equality if and only if there exist $(u, \xi, a, b) \in \mathbb{R}^2 \times \mathbb{C}^2$ such that

$$f(t) = a \, \exp[i\xi t - b(t-u)^2]. \tag{2.48}$$

Proof. The following proof due to Weyl [66] supposes that $\lim_{|t|\to+\infty} \sqrt{t} f(t) = 0$, but the theorem is valid for any $f \in \mathbf{L^2}(\mathbb{R})$. If the average time and frequency localization of f is u and ξ, then the average time and frequency location of $\exp(-i\xi t)\, f(t+u)$ is zero. It is thus sufficient to prove the theorem for $u = \xi = 0$. Observe that

$$\sigma_t^2 \, \sigma_\omega^2 = \frac{1}{2\pi \|f\|^4} \int_{-\infty}^{+\infty} |t\, f(t)|^2 \, dt \int_{-\infty}^{+\infty} |\omega \, \hat{f}(\omega)|^2 \, d\omega. \tag{2.49}$$

Since $i\omega \hat{f}(\omega)$ is the Fourier transform of $f'(t)$, the Plancherel identity (2.26) applied to $i\omega \hat{f}(\omega)$ yields

$$\sigma_t^2 \sigma_\omega^2 = \frac{1}{\|f\|^4} \int_{-\infty}^{+\infty} |t\, f(t)|^2 \, dt \int_{-\infty}^{+\infty} |f'(t)|^2 \, dt. \tag{2.50}$$

Schwarz's inequality implies

$$\begin{aligned} \sigma_t^2 \sigma_\omega^2 &\geqslant \frac{1}{\|f\|^4} \left[\int_{-\infty}^{+\infty} |t\, f'(t)\, f^*(t)| \, dt \right]^2 \\ &\geqslant \frac{1}{\|f\|^4} \left[\int_{-\infty}^{+\infty} \frac{t}{2} \, [f'(t)\, f^*(t) + f'^*(t)\, f(t)] \, dt \right]^2 \\ &\geqslant \frac{1}{4\|f\|^4} \left[\int_{-\infty}^{+\infty} t \, (|f(t)|^2)' \, dt \right]^2 . \end{aligned}$$

Since $\lim_{|t|\to+\infty} \sqrt{t}\, f(t) = 0$, an integration by parts gives

$$\sigma_t^2 \sigma_\omega^2 \geqslant \frac{1}{4\|f\|^4} \left[\int_{-\infty}^{+\infty} |f(t)|^2 \, dt \right]^2 = \frac{1}{4}. \tag{2.51}$$

To obtain an equality, Schwarz's inequality applied to (2.50) must be an equality. This implies that there exists $b \in \mathbb{C}$ such that

$$f'(t) = -2\, b\, t\, f(t). \tag{2.52}$$

Hence, there exists $a \in \mathbb{C}$ such that $f(t) = a \, \exp(-bt^2)$. The other steps of the proof are then equalities so that the lower bound is indeed reached. When $u \neq 0$ and $\xi \neq 0$ the corresponding time and frequency translations yield (2.48). ■ ■

In quantum mechanics, this theorem shows that we cannot reduce arbitrarily the uncertainty as to the position and the momentum of a free particle. In signal processing, the modulated Gaussians (2.48) that have a minimum joint time-frequency localization are called Gabor chirps. As expected, they are smooth functions with a fast time asymptotic decay.

Compact Support Despite the Heisenberg uncertainty bound, we might still be able to construct a function of compact support whose Fourier transform has a compact support. Such a function would be very useful in constructing a finite impulse response filter with a band-limited transfer function. Unfortunately, the following theorem proves that it does not exist.

Theorem 2.7. *If $f \neq 0$ has a compact support then $\hat{f}(\omega)$ cannot be zero on a whole interval. Similarly, if $\hat{f} \neq 0$ has a compact support then $f(t)$ cannot be zero on a whole interval.*

Proof. We prove only the first statement, since the second is derived from the first by applying the Fourier transform. If $\hat{f}$ has a compact support included in $[-b, b]$ then

$$f(t) = \frac{1}{2\pi} \int_{-b}^{b} \hat{f}(\omega) \, \exp(i\omega t) \, d\omega. \tag{2.53}$$

If $f(t) = 0$ for $t \in [c, d]$, by differentiating n times under the integral at $t_0 = (c+d)/2$, we obtain

$$f^{(n)}(t_0) = \frac{1}{2\pi} \int_{-b}^{b} \hat{f}(\omega) \, (i\omega)^n \, \exp(i\omega t_0) \, d\omega = 0. \tag{2.54}$$

Since

$$f(t) = \frac{1}{2\pi} \int_{-b}^{b} \hat{f}(\omega) \, \exp[i\omega(t - t_0)] \, \exp(i\omega t_0) \, d\omega, \tag{2.55}$$

developing $\exp[i\omega(t - t_0)]$ as an infinite series yields for all $t \in \mathbb{R}$

$$f(t) = \frac{1}{2\pi} \sum_{n=0}^{+\infty} \frac{[i(t - t_0)]^n}{n!} \int_{-b}^{b} \hat{f}(\omega) \, \omega^n \, \exp(i\omega t_0) \, d\omega = 0. \tag{2.56}$$

This contradicts our assumption that $f \neq 0$. ■ ■

2.3.3 Total Variation

The total variation measures the total amplitude of signal oscillations. It plays an important role in image processing, where its value depends on the length of the image level sets. We show that a low-pass filter can considerably amplify the total variation by creating Gibbs oscillations.

Variations and Oscillations If f is differentiable, its total variation is defined by

$$\|f\|_V = \int_{-\infty}^{+\infty} |f'(t)| \, dt \, . \tag{2.57}$$

If $\{x_p\}_p$ are the abscissa of the local extrema of f where $f'(x_p) = 0$, then

$$\|f\|_V = \sum_p |f(x_{p+1}) - f(x_p)| \; .$$

It thus measures the total amplitude of the oscillations of f. For example, if $f(t) = \exp(-t^2)$, then $\|f\|_V = 2$. If $f(t) = \sin(\pi t)/(\pi t)$, then f has a local extrema at $x_p \in [p, p+1]$ for any $p \in \mathbb{Z}$. Since $|f(x_{p+1}) - f(x_p)| \sim |p|^{-1}$, we derive that $\|f\|_V = +\infty$.

The total variation of non-differentiable functions can be calculated by considering the derivative in the general sense of distributions [58, 71]. This is equivalent to approximating the derivative by a finite difference on an interval h that goes to zero:

$$\|f\|_V = \lim_{h \to 0} \int_{-\infty}^{+\infty} \frac{|f(t) - f(t-h)|}{|h|} \, dt \, . \tag{2.58}$$

The total variation of discontinuous functions is thus well defined. For example, if $f = \mathbf{1}_{[a,b]}$ then (2.58) gives $\|f\|_V = 2$. We say that f has a *bounded variation* if $\|f\|_V < +\infty$.

Whether f' is the standard derivative of f or its generalized derivative in the sense of distributions, its Fourier transform is $\widehat{f'}(\omega) = i\omega \hat{f}(\omega)$. Hence

$$|\omega| \, |\hat{f}(\omega)| \leqslant \int_{-\infty}^{+\infty} |f'(t)| dt = \|f\|_V \; ,$$

which implies that

$$|\hat{f}(\omega)| \leqslant \frac{\|f\|_V}{|\omega|} \ . \tag{2.59}$$

However, $|\hat{f}(\omega)| = O(|\omega|^{-1})$ is not a sufficient condition to guarantee that f has bounded variation. For example, if $f(t) = \sin(\pi t)/(\pi t)$, then $\hat{f} = \mathbf{1}_{[-\pi,\pi]}$ satisfies $|\hat{f}(\omega)| \leqslant \pi|\omega|^{-1}$ although $\|f\|_V = +\infty$. In general, the total variation of f cannot be evaluated from $|\hat{f}(\omega)|$.

Discrete Signals Let $f_N[n] = f \star \phi_N(n/N)$ be a discrete signal obtained with an averaging filter $\phi_N(t) = \mathbf{1}_{[0,N^{-1}]}(t)$, and a uniform sampling at intervals N^{-1}. The discrete total variation is calculated by approximating the signal derivative by a finite difference over the sampling distance $h = N^{-1}$, and replacing the integral (2.58) by a Riemann sum, which gives:

$$\|f_N\|_V = \sum_n |f_N[n] - f_N[n-1]| \ . \tag{2.60}$$

If n_p are the abscissa of the local extrema of f_N, then

$$\|f_N\|_V = \sum_p |f_N[n_{p+1}] - f_N[n_p]| \ .$$

The total variation thus measures the total amplitude of the oscillations of f. In accordance with (2.58), we say that the discrete signal has a *bounded variation* if $\|f_N\|_V$ is bounded by a constant independent of the resolution N.

Gibbs Oscillations Filtering a signal with a low-pass filter can create oscillations that have an infinite total variation. Let $f_\xi = f \star \phi_\xi$ be the filtered signal obtained with an ideal low-pass filter whose transfer function is $\hat{\phi}_\xi = \mathbf{1}_{[-\xi,\xi]}$. If $f \in \mathbf{L^2}(\mathbb{R})$, then f_ξ converges to f in $\mathbf{L^2}(\mathbb{R})$ norm: $\lim_{\xi\to+\infty} \|f - f_\xi\| = 0$. Indeed, $\hat{f}_\xi = \hat{f}\,\mathbf{1}_{[-\xi,\xi]}$ and the Plancherel formula (2.26) implies that

$$\|f - f_\xi\|^2 = \frac{1}{2\pi}\int_{-\infty}^{+\infty} |\hat{f}(\omega) - \hat{f}_\xi(\omega)|^2\,d\omega = \frac{1}{2\pi}\int_{|\omega|>\xi} |\hat{f}(\omega)|^2\,d\omega,$$

which goes to zero as ξ increases. However, if f is discontinuous in t_0, then we show that f_ξ has Gibbs oscillations in the neighborhood of t_0, which prevents $\sup_{t\in\mathbb{R}} |f(t) - f_\xi(t)|$ from converging to zero as ξ increases.

Let f be a bounded variation function $\|f\|_V < +\infty$ that has an isolated discontinuity at t_0, with a left limit $f(t_0^-)$ and right limit $f(t_0^+)$. It is decomposed as a sum of f_c, which is continuous in the neighborhood of t_0, plus a Heaviside step of amplitude $f(t_0^+) - f(t_0^-)$:

$$f(t) = f_c(t) + [f(t_0^+) - f(t_0^-)]\,u(t - t_0),$$

with

$$u(t) = \begin{cases} 1 & \text{if } t \geqslant 0 \\ 0 & \text{otherwise} \end{cases} . \tag{2.61}$$

Hence

$$f_\xi(t) = f_c \star \phi_\xi(t) + [f(t_0^+) - f(t_0^-)]\,u \star \phi_\xi(t - t_0). \tag{2.62}$$

Since f_c has bounded variation and is uniformly continuous in the neighborhood of t_0, one can prove (Exercise 2.3) that $f_c \star \phi_\xi(t)$ converges uniformly to $f_c(t)$ in a neighborhood of t_0. The following theorem shows that this is not true for $u \star \phi_\xi$, which creates Gibbs oscillations.

Theorem 2.8 (Gibbs). *For any* $\xi > 0$,

$$u \star \phi_\xi(t) = \int_{-\infty}^{\xi t} \frac{\sin x}{\pi x}\,dx. \tag{2.63}$$

Proof. The impulse response of an ideal low-pass filter, calculated in (2.29), is $\phi_\xi(t) = \sin(\xi t)/(\pi t)$. Hence

$$u \star \phi_\xi(t) = \int_{-\infty}^{+\infty} u(\tau)\, \frac{\sin \xi(t-\tau)}{\pi(t-\tau)}\, d\tau = \int_{0}^{+\infty} \frac{\sin \xi(t-\tau)}{\pi(t-\tau)}\, d\tau.$$

The change of variable $x = \xi(t-\tau)$ gives (2.63). $\blacksquare$ $\blacksquare$

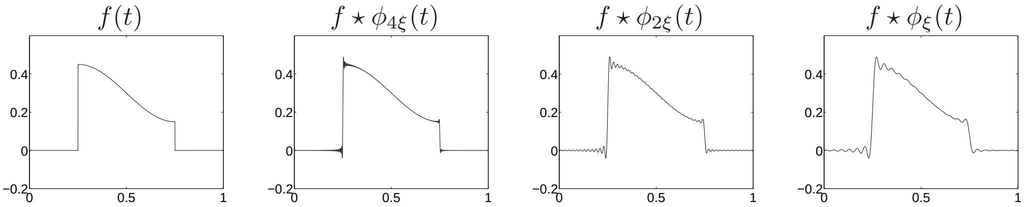


Figure 2.1: Gibbs oscillations created by low-pass filters with cut-off frequencies that decrease from left to right.

The function

$$s(\xi t) = \int_{-\infty}^{\xi t} \frac{\sin x}{\pi x}\, dx$$

is a sigmoid that increases from 0 at $t = -\infty$ to 1 at $t = +\infty$, with $s(0) = 1/2$. It has oscillations of period π/ξ, which are attenuated when the distance to 0 increases, but their total variation is infinite: $\|s\|_V = +\infty$. The maximum amplitude of the Gibbs oscillations occurs at $t = \pm\pi/\xi$, with an amplitude independent of ξ:

$$A = s(\pi) - 1 = \int_{-\infty}^{\pi} \frac{\sin x}{\pi x} dx - 1 \approx 0.045\ .$$

Inserting (2.63) in (2.62) shows that

$$f(t) - f_\xi(t) = [f(t_0^+) - f(t_0^-)]\, s(\xi(t-t_0)) + \varepsilon(\xi, t)\ , \tag{2.64}$$

where $\lim_{\xi \to +\infty} \sup_{|t-t_0|<\alpha} |\varepsilon(\xi,t)| = 0$ in some neighborhood of size $\alpha > 0$ around t_0. The sigmoid $s(\xi(t-t_0))$ centered at t_0 creates a maximum error of fixed amplitude for all ξ. This is seen in Figure 2.1, where the Gibbs oscillations have an amplitude proportional to the jump $f(t_0^+) - f(t_0^-)$ at all frequencies ξ.

Image Total Variation The total variation of an image $f(x_1, x_2)$ depends on the amplitude of its variations as well as the length of the contours along which they occur. Suppose that $f(x_1, x_2)$ is differentiable. The total variation is defined by

$$\|f\|_V = \int\int |\vec{\nabla} f(x_1, x_2)|\, dx_1\, dx_2\ , \tag{2.65}$$

where the modulus of the gradient vector is

$$|\vec{\nabla} f(x_1, x_2)| = \left(\left| \frac{\partial f(x_1, x_2)}{\partial x_1} \right|^2 + \left| \frac{\partial f(x_1, x_2)}{\partial x_2} \right|^2 \right)^{1/2} .$$

As in one dimension, the total variation is extended to discontinuous functions by taking the derivatives in the general sense of distributions. An equivalent norm is obtained by approximating the partial derivatives by finite differences:

$$\begin{aligned} |\Delta_h f(x_1, x_2)| \quad &= \quad \left(\left| \frac{f(x_1, x_2) - f(x_1 - h, x_2)}{h} \right|^2 + \right. \\ &\qquad \left. \left| \frac{f(x_1, x_2) - f(x_1, x_2 - h)}{h} \right|^2 \right)^{1/2} . \end{aligned}$$

One can verify that

$$\|f\|_V \leqslant \lim_{h\to 0} \int\int |\Delta_h f(x_1,x_2)|\, dx_1\, dx_2 \leqslant \sqrt{2}\, \|f\|_V. \tag{2.66}$$

The finite difference integral gives a larger value when $f(x_1,x_2)$ is discontinuous along a diagonal line in the (x_1,x_2) plane.

The total variation of f is related to the length of it level sets. Let us define

$$\Omega_y = \{(x_1,x_2) \in \mathbb{R}^2 \quad : \quad f(x_1,x_2) > y\} \ .$$

If f is continuous then the boundary $\partial\Omega_y$ of Ω_y is the level set of all (x_1,x_2) such that $f(x_1,x_2) = y$. Let $H^1(\partial\Omega_y)$ be the length of $\partial\Omega_y$. Formally, this length is calculated in the sense of the monodimensional Hausdorff measure. The following theorem relates the total variation of f to the length of its level sets.

Theorem 2.9 (Co-area Formula)**.** *If* $\|f\|_V < +\infty$ *then*

$$\|f\|_V = \int_{-\infty}^{+\infty} H^1(\partial\Omega_y)\, dy. \tag{2.67}$$

Proof. The proof is a highly technical result that is given in [71]. We give an intuitive explanation when f is continuously differentiable. In this case $\partial\Omega_y$ is a differentiable curve $x(y,s) \in \mathbb{R}^2$, which is parameterized by the arc-length s. Let $\vec{\tau}(x)$ be the vector tangent to this curve in the plane. The gradient $\vec{\nabla} f(x)$ is orthogonal to $\vec{\tau}(x)$. The Frenet coordinate system along $\partial\Omega_y$ is composed of $\vec{\tau}(x)$ and of the unit vector $\vec{n}(x)$ parallel to $\vec{\nabla} f(x)$. Let ds and dn be the Lebesgue measures in the direction of $\vec{\tau}$ and $\vec{n}$. We have

$$|\vec{\nabla} f(x)| = \vec{\nabla} f(x) \,.\, \vec{n} = \frac{dy}{dn} \ , \tag{2.68}$$

where dy is the differential of amplitudes across level sets. The idea of the proof is to decompose the total variation integral over the plane as an integral along the level sets and across level sets, which we write:

$$\|f\|_V = \int\int |\vec{\nabla} f(x_1,x_2)|\, dx_1\, dx_2 = \int\int_{\partial\Omega_y} |\vec{\nabla} f(x(y,s))|\, ds\, dn. \tag{2.69}$$

By using (2.68) we can get

$$\|f\|_V = \int\int_{\partial\Omega_y} ds\, dy \ .$$

But $\int_{\partial\Omega_y} ds = H^1(\partial\Omega_y)$ is the length of the level set, which justifies (2.67). ■ ■

The co-area formula gives an important geometrical interpretation of the total image variation. Images are uniformly bounded so the integral (2.67) is calculated over a finite interval and is proportional to the average length of level sets. It is finite as long as the level sets are not fractal curves. Let $f = \alpha\, \mathbf{1}_\Omega$ be proportional to the indicator function of a set $\Omega \subset \mathbb{R}^2$ which has a boundary $\partial\Omega$ of length L. The co-area formula (2.9) implies that $\|f\|_V = \alpha\, L$. In general, bounded variation images must have step edges of finite length.

Discrete Images A camera measures light intensity with photoreceptors that perform an averaging and a uniform sampling over a grid that is supposed to be uniform. For a resolution N, the sampling interval is N^{-1}. The resulting image can be written $f_N[n_1,n_2] = f \star \phi_N(n_1/N, n_2/N)$, where $\phi_N = \mathbf{1}_{[0,N^{-1}]^2}$ and f is the averaged analog image. Its total variation is defined by approximating derivatives by finite differences and the integral (2.66) by a Riemann sum:

$$\|f_N\|_V = \frac{1}{N} \sum_{n_1} \sum_{n_2} \left(\Big| f_N[n_1,n_2] - f_N[n_1-1,n_2] \Big|^2 + \Big| f_N[n_1,n_2] - f_N[n_1,n_2-1] \Big|^2 \right)^{1/2} . \tag{2.70}$$

In accordance with (2.66) we say that the image has bounded variation if $\|f_N\|_V$ is bounded by a constant independent of the resolution N. The co-area formula proves that it depends on the

(a) (b)

Figure 2.2: (a): The total variation of this image remains nearly constant when the resolution N increases. (b): Level sets $\partial\Omega_y$ obtained by sampling uniformly the amplitude variable y.

length of the level sets as the image resolution increases. The $\sqrt{2}$ upper bound factor in (2.66) comes from the fact that the length of a diagonal line can be increased by $\sqrt{2}$ if it is approximated by a zig-zag line that remains on the horizontal and vertical segments of the image sampling grid. Figure 2.2(a) shows a bounded variation image and Figure 2.2(b) displays the level sets obtained by discretizing uniformly the amplitude variable y. The total variation of this image remains nearly constant as the resolution varies.

2.4 Two-Dimensional Fourier Transform

The Fourier transform in $\mathbb{R}^n$ is a straightforward extension of the one-dimensional Fourier transform. The two-dimensional case is briefly reviewed for image processing applications. The Fourier transform of a two-dimensional integrable function $f \in \mathbf{L^1}(\mathbb{R}^2)$ is

$$\hat{f}(\omega_1,\omega_2) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x_1,x_2)\, \exp[-i(\omega_1 x_1 + \omega_2 x_2)]\, dx_1\, dx_2. \tag{2.71}$$

In polar coordinates $\exp[i(\omega_1 x + \omega_2 y)]$ can be rewritten

$$\exp[i(\omega_1 x_1 + \omega_2 x_2)] = \exp[i\xi(x_1\cos\theta + x_2\sin\theta)]$$

with $\xi = \sqrt{\omega_1^2+\omega_2^2}$. It is a plane wave that propagates in the direction of θ and oscillates at the frequency ξ. The properties of a two-dimensional Fourier transform are essentially the same as in one dimension. We summarize a few important results. We write $\omega = (\omega_1,\omega_2)$, $x = (x_1,x_2)$, $\omega\cdot x = \omega_1 x_1 + \omega_2 x_2$ and $\iint f(x_1,x_2)dx_1\,dx_2 = \iint f(x)\,dx$.

- If $f \in \mathbf{L^1}(\mathbb{R}^2)$ and $\hat{f} \in \mathbf{L^1}(\mathbb{R}^2)$ then

$$f(x) = \frac{1}{4\pi^2}\iint \hat{f}(\omega)\, \exp[i(\omega\cdot x)]\, d\omega. \tag{2.72}$$

- If $f \in \mathbf{L^1}(\mathbb{R}^2)$ and $h \in \mathbf{L^1}(\mathbb{R}^2)$ then the convolution

$$g(x) = f\star h(x) = \iint f(u)\, h(x-u)\, du$$

has a Fourier transform

$$\hat{g}(\omega) = \hat{f}(\omega)\,\hat{h}(\omega). \tag{2.73}$$

- The Parseval formula proves that

$$\iint f(x)\, g^*(x)\, dx = \frac{1}{4\pi^2}\iint \hat{f}(\omega)\,\hat{g}^*(\omega)\, d\omega\ . \tag{2.74}$$

If $f = g$, we obtain the Plancherel equality

$$\iint |f(x)|^2 \, dx = \frac{1}{4\pi^2} \iint |\hat{f}(\omega)|^2 \, d\omega \, . \tag{2.75}$$

The Fourier transform of a finite energy function thus has finite energy. With the same density based argument as in one dimension, energy equivalence makes it possible to extend the Fourier transform to any function $f \in \mathbf{L^2}(\mathbb{R}^2)$.

- If $f \in \mathbf{L^2}(\mathbb{R}^2)$ is separable, which means that

  $$f(x) = f(x_1, x_2) = g(x_1)\, h(x_2),$$

  then its Fourier transform is

  $$\hat{f}(\omega) = \hat{f}(\omega_1, \omega_2) = \hat{g}(\omega_1)\, \hat{h}(\omega_2),$$

  where $\hat{h}$ and $\hat{g}$ are the one-dimensional Fourier transforms of g and h. For example, the indicator function

  $$f(x_1, x_2) = \begin{cases} 1 & \text{if } |x_1| \leqslant T \ , \ |x_2| \leqslant T \\ 0 & \text{otherwise} \end{cases} = \mathbf{1}_{[-T,T]}(x_1) \times \mathbf{1}_{[-T,T]}(x_2)$$

  is a separable function whose Fourier transform is derived from (2.28):

  $$\hat{f}(\omega_1, \omega_2) = \frac{4 \sin(T\omega_1)\, \sin(T\omega_2)}{\omega_1\, \omega_2}.$$

- If $f(x_1, x_2)$ is rotated by θ:

  $$f_\theta(x_1, x_2) = f(x_1 \cos\theta - x_2 \sin\theta, x_1 \sin\theta + x_2 \cos\theta),$$

  then its Fourier transform is rotated by θ:

  $$\hat{f}_\theta(\omega_1, \omega_2) = \hat{f}(\omega_1 \cos\theta - \omega_2 \sin\theta, \omega_1 \sin\theta + \omega_2 \cos\theta). \tag{2.76}$$

Radon Transform A Radon transform computes integrals of $f \in \mathbf{L^2}(\mathbb{R}^2)$ along rays. It is provides a good model for some tomographic systems such as X-ray measurements in medical imaging. Inverting the Radon transform is then needed to reconstruct the 2D or 3D body from these integrals.

Let us write $\tau_\theta = (\cos\theta, \sin\theta)$. A ray $\Delta_{t,\theta}$ is a line defined by its equation

$$x \cdot \tau_\theta = x_1 \cos\theta + x_2 \sin\theta = t \ .$$

The projection p_θ of f along a parallel line of orientation θ is defined by

$$\forall\, \theta \in [0, \pi), \forall\, t \in \mathbb{R}, \quad p_\theta(t) = \int_{\Delta_{t,\theta}} f(x)\, ds = \iint f(x)\, \delta(x \cdot \tau_\theta - t)\, dx, \tag{2.77}$$

where δ is the Dirac distribution. The Radon transform maps $f(x)$ to $p_\theta(t)$ for $\theta \in [0, \pi)$. In medical imaging applications, a scanner is rotated around an object to compute the projection p_θ for many angles $\theta \in [0, \pi)$, as illustrated in Figure 2.3. The Fourier slice theorem relates the Fourier transform of p_θ to slices of the Fourier transform of f.

Theorem 2.10 (Fourier slice)**.** *The Fourier transform of projections satisfies*

$$\forall\, \theta \in [0, \pi) \ , \ \forall\, \xi \in \mathbb{R} \quad \hat{p}_\theta(\xi) = \hat{f}(\xi \cos\theta, \xi \sin\theta).$$

Proof. The Fourier transform of the projection is

$$\begin{aligned} \hat{p}_\theta(\xi) &= \int_{-\infty}^{+\infty} \left(\iint f(x)\, \delta(x \cdot \tau_\theta - t)\, dx \right) e^{-it\xi} dt \\ &= \int\!\!\int f(x) \exp\left(-i(x \cdot \tau_\theta)\xi\right) dx = \hat{f}(\xi \tau_\theta). \end{aligned}$$

■

An image f can be recovered from its projections p_θ thanks to the projection slice theorem. Indeed, the Fourier transform $\hat{f}$ is known along each ray of direction θ and f is thus obtained with the 2D inverse Fourier transform (2.71). The following back projection theorem gives an inversion formula.

Theorem 2.11 (Back projection)**.** *The image f is recovered using a one-dimensional filter $h(t)$:*

$$f(x) = \frac{1}{2\pi}\int_0^\pi p_\theta * h(x\cdot\tau_\theta)\,d\theta \quad \textit{with} \quad \hat{h}(\xi) = |\xi|\ .$$

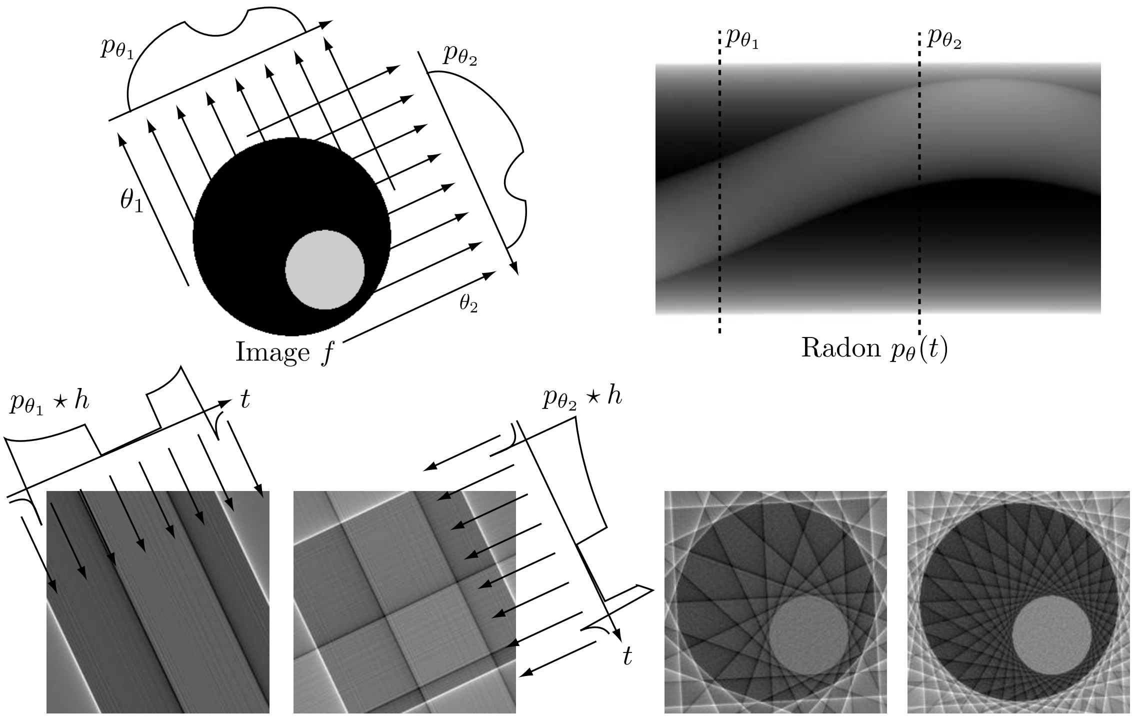
Proof. The inverse Fourier transform (2.72) in polar coordinates $(\omega_1,\omega_2) = (\xi\cos\theta, \xi\sin\theta)$, with $d\omega_1\,d\omega_2 = \xi\,d\theta\,d\xi$, can be written

$$f(x) = \frac{1}{4\pi^2}\int_0^{+\infty}\int_0^{2\pi} \hat{f}(\xi\cos\theta,\xi\sin\theta)\exp\left(i(x\cdot\tau_\theta)\xi\right)\,\xi\,d\theta\,d\xi\ .$$

Using the Fourier slice Theorem 2.10 with $p_{\theta+\pi}(t) = p_\theta(-t)$, this is rewritten as

$$f(x) = \frac{1}{2\pi}\int_0^\pi \left(\frac{1}{2\pi}\int_{-\infty}^{+\infty}|\xi|\,\hat{p}_\theta(\xi)\exp\left(i(x\cdot\tau_\theta)\xi\right)d\xi\right)d\theta\ .$$

The inner integral is the inverse Fourier transform of $\hat{p}_\theta(\xi)\,|\xi|$ evaluated at $x\cdot\tau_\theta \in \mathbb{R}$. The convolution formula (2.73) shows that it is equal to $p_\theta * h(x\cdot\tau_\theta)$. ∎



1 back projection 2 back projections 8 back projections 16 back projections

Figure 2.3: The Radon transform and its reconstruction with an increasing number of back projections.

In medical imaging applications, only a limited number of projections is available, and the Fourier transform $\hat{f}$ is thus partially known. In this case, an approximation of f can still be recovered by summing the corresponding filtered back projections $p_\theta * h(x\cdot\tau_\theta)$. Figure 2.3 describes this process, and shows the reconstruction of an image with a geometric object, using an increasing number of evenly spaced projections. Section 12.6.3 describes a non-linear super-resolution reconstruction algorithm, that recovers a more precise image by using a sparse representation.

2.5 Exercises

2.1. [1]

2.2. [2]

2.3. [2]

III

Discrete Revolution

Digital signal processing has taken over. First used in the 1950's at the service of analog signal processing to simulate analog transforms, digital algorithms have invaded most traditional fortresses, including phones, music recording, cameras, televisions, and all information processing. Analog computations performed with electronic circuits are faster than digital algorithms implemented with microprocessors, but are less precise and less flexible. Thus analog circuits are often replaced by digital chips once the computational performance of microprocessors is sufficient to operate in real time for a given application.

Whether sound recordings or images, most discrete signals are obtained by sampling an analog signal. An analog to digital conversion is a linear approximation which introduces an error which depends upon the sampling rate. Once more, the Fourier transform is unavoidable because the eigenvectors of discrete time-invariant operators are sinusoidal waves. The Fourier transform is discretized for signals of finite size and implemented with a Fast Fourier Transform algorithm.

3.1 Sampling Analog Signals

The simplest way to discretize an analog signal f is to record its sample values $\{f(ns)\}_{n\in\mathbb{Z}}$ at intervals s. An approximation of $f(t)$ at any $t \in \mathbb{R}$ may be recovered by interpolating these samples. The Shannon-Whittaker sampling theorem gives a sufficient condition on the support of the Fourier transform $\hat{f}$ to recover $f(t)$ exactly. Aliasing and approximation errors are studied when this condition is not satisfied.

Digital acquisition devices often do not satisfy the restrictive hypothesis of the Shannon-Whittaker sampling theorem. General linear analog to discrete conversion is introduced in Section 3.1.3, showing that a stable uniform discretization is a linear approximation. A digital conversion also approximates discrete coefficients with a given precision, to store them with a limited number of bits. This quantization aspect is studied in Chapter 10.

3.1.1 Shannon-Whittaker Sampling Theorem

Sampling is first studied from the more classical Shannon-Whittaker point of view, which tries to recover $f(t)$ from its samples $\{f(ns)\}_{n\in\mathbb{Z}}$. A discrete signal may be represented as a sum of Diracs. We associate to any sample $f(ns)$ a Dirac $f(ns)\delta(t-ns)$ located at $t = ns$. A uniform sampling of f thus corresponds to the weighted Dirac sum

$$f_d(t) = \sum_{n=-\infty}^{+\infty} f(ns)\,\delta(t-ns). \tag{3.1}$$

The Fourier transform of $\delta(t-ns)$ is $\mathrm{e}^{-ins\omega}$ so the Fourier transform of f_d is a Fourier series:

$$\hat{f}_d(\omega) = \sum_{n=-\infty}^{+\infty} f(ns)\,\mathrm{e}^{-ins\omega}. \tag{3.2}$$

To understand how to compute $f(t)$ from the sample values $f(ns)$ and hence f from f_d, we relate their Fourier transforms $\hat{f}$ and $\hat{f}_d$.

Theorem 3.1. *The Fourier transform of the discrete signal obtained by sampling f at intervals s is*

$$\hat{f}_d(\omega) = \frac{1}{s} \sum_{k=-\infty}^{+\infty} \hat{f}\left(\omega - \frac{2k\pi}{s}\right). \tag{3.3}$$

Proof. Since $\delta(t-ns)$ is zero outside $t = ns$,

$$f(ns)\,\delta(t-ns) = f(t)\,\delta(t-ns),$$

so we can rewrite (3.1) as multiplication with a Dirac comb:

$$f_d(t) = f(t) \sum_{n=-\infty}^{+\infty} \delta(t-ns) = f(t)\,c(t). \tag{3.4}$$

Computing the Fourier transform yields

$$\hat{f}_d(\omega) = \frac{1}{2\pi}\, \hat{f} \star \hat{c}(\omega). \tag{3.5}$$

The Poisson formula (2.4) proves that

$$\hat{c}(\omega) = \frac{2\pi}{s} \sum_{k=-\infty}^{+\infty} \delta\left(\omega - \frac{2\pi k}{s}\right). \tag{3.6}$$

Since $\hat{f} \star \delta(\omega - \xi) = \hat{f}(\omega - \xi)$, inserting (3.6) in (3.5) proves (3.3). ■ ■

Theorem 3.1 proves that sampling f at intervals s is equivalent to making its Fourier transform $2\pi/s$ periodic by summing all its translations $\hat{f}(\omega - 2k\pi/s)$. The resulting sampling theorem was first proved by Whittaker [441] in 1935 in a book on interpolation theory. Shannon rediscovered it in 1949 for applications to communication theory [390].

Theorem 3.2 (Shannon, Whittaker)**.** *If the support of $\hat{f}$ is included in $[-\pi/s, \pi/s]$ then*

$$f(t) = \sum_{n=-\infty}^{+\infty} f(ns)\,\phi_s(t-ns), \tag{3.7}$$

with

$$\phi_s(t) = \frac{\sin(\pi t/s)}{\pi t/s}. \tag{3.8}$$

Proof. If $n \neq 0$, the support of $\hat{f}(\omega - n\pi/s)$ does not intersect the support of $\hat{f}(\omega)$ because $\hat{f}(\omega) = 0$ for $|\omega| > \pi/s$. So (3.3) implies

$$\hat{f}_d(\omega) = \frac{\hat{f}(\omega)}{s} \quad \text{if} \quad |\omega| \leqslant \frac{\pi}{s}. \tag{3.9}$$

The Fourier transform of ϕ_s is $\hat{\phi}_s = s\,\mathbf{1}_{[-\pi/s,\pi/s]}$. Since the support of $\hat{f}$ is in $[-\pi/s, \pi/s]$ it results from (3.9) that $\hat{f}(\omega) = \hat{\phi}_s(\omega)\,\hat{f}_d(\omega)$. The inverse Fourier transform of this equality gives

$$\begin{aligned} f(t) = \phi_s \star f_d(t) &= \phi_s \star \sum_{n=-\infty}^{+\infty} f(ns)\,\delta(t-ns) \\ &= \sum_{n=-\infty}^{+\infty} f(ns)\,\phi_s(t-ns). \end{aligned}$$

■ ■

The sampling theorem imposes that the support of $\hat{f}$ is included in $[-\pi/s, \pi/s]$, which guarantees that f has no brutal variations between consecutive samples, and can thus be recovered with a smooth interpolation. Section 3.1.3 shows that one can impose other smoothness conditions to recover f from its samples. Figure 3.1 illustrates the different steps of a sampling and reconstruction from samples, in both the time and Fourier domains.

3.1.2 Aliasing

The sampling interval s is often imposed by computation or storage constraints and the support of $\hat{f}$ is generally not included in $[-\pi/s, \pi/s]$. In this case the interpolation formula (3.7) does not recover f. We analyze the resulting error and a filtering procedure to reduce it.

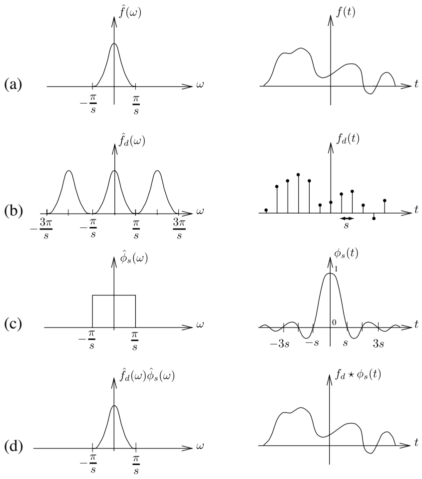


Figure 3.1: (a): Signal f and its Fourier transform $\hat{f}$. (b): A uniform sampling of f makes its Fourier transform periodic. (c): Ideal low-pass filter. (d): The filtering of (b) with (c) recovers f.

Theorem 3.1 proves that

$$\hat{f}_d(\omega) = \frac{1}{s} \sum_{k=-\infty}^{+\infty} \hat{f}\left(\omega - \frac{2k\pi}{s}\right). \tag{3.10}$$

Suppose that the support of $\hat{f}$ goes beyond $[-\pi/s, \pi/s]$. In general the support of $\hat{f}(\omega - 2k\pi/s)$ intersects $[-\pi/s, \pi/s]$ for several $k \neq 0$, as shown in Figure 3.2. This folding of high frequency components over a low frequency interval is called *aliasing*. In the presence of aliasing, the interpolated signal

$$\phi_s \star f_d(t) = \sum_{n=-\infty}^{+\infty} f(ns)\, \phi_s(t - ns)$$

has a Fourier transform

$$\hat{f}_d(\omega)\, \hat{\phi}_s(\omega) = s\, \hat{f}_d(\omega)\, \mathbf{1}_{[-\pi/s,\pi/s]}(\omega) = \mathbf{1}_{[-\pi/s,\pi/s]}(\omega) \sum_{k=-\infty}^{+\infty} \hat{f}\left(\omega - \frac{2k\pi}{s}\right) \tag{3.11}$$

which may be completely different from $\hat{f}(\omega)$ over $[-\pi/s, \pi/s]$. The signal $\phi_s \star f_d$ may not even be a good approximation of f, as shown by Figure 3.2.

Example 3.1. *Let us consider a high frequency oscillation*

$$f(t) = \cos(\omega_0 t) = \frac{\mathrm{e}^{i\omega_0 t} + \mathrm{e}^{-i\omega_0 t}}{2}.$$

Its Fourier transform is

$$\hat{f}(\omega) = \pi\Big(\delta(\omega - \omega_0) + \delta(\omega + \omega_0)\Big).$$

If $2\pi/s > \omega_0 > \pi/s$ *then (3.11) yields*

$$\begin{aligned}
&\hat{f}_d(\omega)\, \hat{\phi}_s(\omega) \\
&= \pi\, \mathbf{1}_{[-\pi/s,\pi/s]}(\omega) \sum_{k=-\infty}^{+\infty} \left(\delta\Big(\omega - \omega_0 - \frac{2k\pi}{s}\Big) + \delta\Big(\omega + \omega_0 - \frac{2k\pi}{s}\Big)\right) \\
&= \pi\Big(\delta(\omega - \frac{2\pi}{s} + \omega_0) + \delta(\omega + \frac{2\pi}{s} - \omega_0)\Big),
\end{aligned}$$

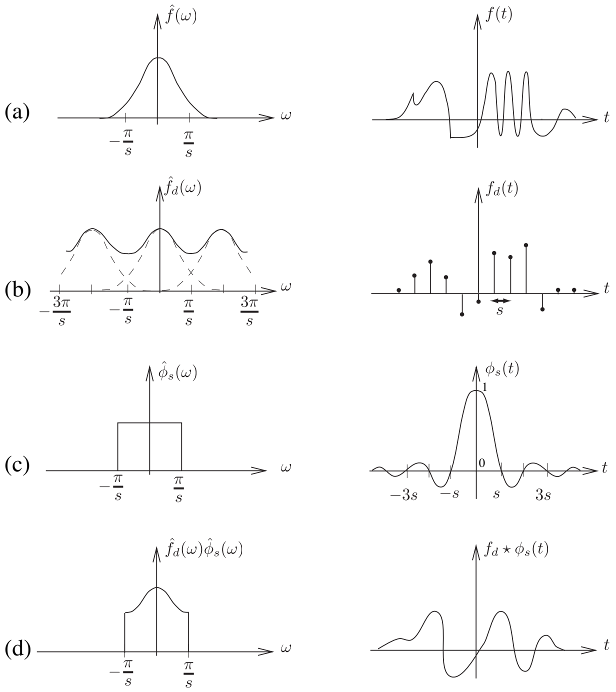


Figure 3.2: (a): Signal f and its Fourier transform $\hat{f}$. (b): Aliasing produced by an overlapping of $\hat{f}(\omega - 2k\pi/s)$ for different k, shown in dashed lines. (c): Ideal low-pass filter. (d): The filtering of (b) with (c) creates a low-frequency signal that is different from f.

so

$$f_d \star \phi_s(t) = \cos\left[\left(\frac{2\pi}{s} - \omega_0\right)t\right].$$

The aliasing reduces the high frequency ω_0 *to a lower frequency* $2\pi/s - \omega_0 \in [-\pi/s, \pi/s]$. *The same frequency folding is observed in a film that samples a fast moving object without enough images per second. A wheel turning rapidly appears as turning much more slowly in the film.*

Removal of Aliasing To apply the sampling theorem, f is approximated by the closest signal $\tilde{f}$ whose Fourier transform has a support in $[-\pi/s, \pi/s]$. The Plancherel formula (2.26) proves that

$$\begin{aligned} \|f - \tilde{f}\|^2 &= \frac{1}{2\pi}\int_{-\infty}^{+\infty} |\hat{f}(\omega) - \widehat{\tilde{f}}(\omega)|^2 \, d\omega \\ &= \frac{1}{2\pi}\int_{|\omega|>\pi/s} |\hat{f}(\omega)|^2 \, d\omega + \frac{1}{2\pi}\int_{|\omega|\leqslant\pi/s} |\hat{f}(\omega) - \widehat{\tilde{f}}(\omega)|^2 \, d\omega. \end{aligned}$$

This distance is minimum when the second integral is zero and hence

$$\widehat{\tilde{f}}(\omega) = \hat{f}(\omega)\,\mathbf{1}_{[-\pi/s,\pi/s]}(\omega) = \frac{1}{s}\,\hat{\phi}_s(\omega)\,\hat{f}(\omega). \tag{3.12}$$

It corresponds to $\tilde{f} = \frac{1}{s} f \star \phi_s$. The filtering of f by ϕ_s avoids the aliasing by removing any frequency larger than π/s. Since $\widehat{\tilde{f}}$ has a support in $[-\pi/s, \pi/s]$, the sampling theorem proves that $\tilde{f}(t)$ can be recovered from the samples $\tilde{f}(ns)$. An analog to digital converter is therefore composed of a filter that limits the frequency band to $[-\pi/s, \pi/s]$, followed by a uniform sampling at intervals s.

3.1.3 General Sampling and Linear Analog Conversions

The Shannon-Whittaker theorem is a particular example of linear discrete to analog conversion, which does not apply to all digital acquisition devices. This section describes general analog to discrete conversion and reverse discrete to analog conversion, with a general linear filtering and uniform sampling. Analog signals are approximated by linear projections on approximation spaces.

Sampling Theorems We want to recover a stable approximation of $f \in \mathbf{L^2}(\mathbb{R})$ from a filtering and uniform sampling that outputs $\{f \star \bar{\phi}_s(ns)\}_{n\in\mathbb{Z}}$, for some real filter $\bar{\phi}_s(t)$. These samples can be written as inner products in $\mathbf{L^2}(\mathbb{R})$:

$$f \star \phi_s(ns) = \int_{-\infty}^{+\infty} f(t)\, \bar{\phi}_s(ns-t)\, dt = \langle f(t), \phi_s(t-ns)\rangle \tag{3.13}$$

with $\phi_s(t) = \bar{\phi}_s(-t)$. Let $\mathbf{U}_s$ be the approximation space generated by linear combinations of the $\{\phi_s(t-ns)\}_{n\in\mathbb{Z}}$. The approximation $\tilde{f} \in \mathbf{U}_s$ which minimizes the maximum possible error $\|f - \tilde{f}\|$ is the orthognal projection of f on $\mathbf{U}_s$ (Exercice 3.3). The calculation of this orthogonal projection is stable if $\{\phi_s(t-ns)\}_{n\in\mathbb{Z}}$ is a Riesz basis of $\mathbf{U}_s$, as defined in Section 5.1.1. Following Definition 5.1, a Riesz basis is a family of linearly independent functions that yields inner product satisfying an energy equivalence. There exists $B \geqslant A > 0$ such that for any $f \in \mathbf{U}_s$

$$A\,\|f\|^2 \leqslant \sum_{n=-\infty}^{+\infty} |\langle f(t), \phi_s(t-ns)\rangle|^2 \leqslant B\,\|f\|^2. \tag{3.14}$$

The basis is orthogonal if and only if $A = B$. The following generalized sampling theorem computes the orthogonal projection on the approximation space $\mathbf{U}_s$ [429].

Theorem 3.3 (Linear sampling). *Let $\{\phi_s(t-ns)\}_{n\in\mathbb{Z}}$ be a Riesz basis of $\mathbf{U}_s$ and $\bar{\phi}_s(t) = \phi_s(-t)$. There exists a biorthogonal basis $\{\tilde{\phi}_s(t-ns)\}_{n\in\mathbb{Z}}$ of $\mathbf{U}_s$ such that*

$$\forall f \in \mathbf{L^2}(\mathbb{R}) \quad , \quad P_{\mathbf{U}_s} f(t) = \sum_{n=-\infty}^{+\infty} f \star \bar{\phi}_s(ns)\, \tilde{\phi}_s(t-ns)\ . \tag{3.15}$$

Proof. For any Riesz basis, Section 5.1.2 proves that there exists a biorthogonal basis $\{\tilde{\phi}_{s,n}(t)\}_{n\in\mathbb{Z}}$ that satisfies the biorthogonality relations

$$\forall (n,m) \in \mathbb{Z}^2 \quad , \quad \langle \phi_s(t-ns)\,,\, \tilde{\phi}_{s,m}(t-ms)\rangle = \delta[n-m]\ . \tag{3.16}$$

Since $\langle \phi_s(t-(n-m)s)\,,\, \tilde{\phi}_{s,0}(t)\rangle = \langle \phi_s(t-ns)\,,\, \tilde{\phi}_{s,0}(t-ms)\rangle = 0$ and since the dual basis is unique, necessarily $\tilde{\phi}_{s,m}(t) = \tilde{\phi}_{s,0}(t-ms)$. Section 5.1.2 proves in (5.19) that the orthogonal projection in $\mathbf{U}_s$ can be written

$$P_{\mathbf{U}_s} f(t) = \sum_{n=-\infty}^{+\infty} \langle f(t), \phi_s(t-ns)\rangle\, \tilde{\phi}_s(t-ns)$$

which proves (3.15). ■

The orthogonal projection (3.15) can be rewritten as an analog filtering of the discrete signal $f_d(t) = \sum_{n=-\infty}^{+\infty} f \star \bar{\phi}_s(ns)\, \delta(t-ns)$:

$$P_{\mathbf{U}_s} f(t) = f_d \star \tilde{\phi}_s(t)\ . \tag{3.17}$$

If $f \in \mathbf{U}_s$ then $P_{\mathbf{U}_s} f = f$ so it is exactly reconstructed by filtering the uniformly sampled discrete signal $\{f \star \bar{\phi}_s(ns)\}_{n\in\mathbb{Z}}$ with the analog filter $\tilde{\phi}(t)$. If $f \notin \mathbf{U}_s$ then (3.17) recovers the best linear approximation approximation of f in $\mathbf{U}_s$. Section 9.1 shows that the linear approximation error $\|f - P_{\mathbf{U}_s} f\|$ depends essentially on the uniform regularity of f. Given some prior information on f, optimizing the analog discretization filter ϕ_s amounts to optimize the approximation space $\mathbf{U}_s$ in order to minimize this error. The following theorem characterizes filters ϕ_s that generate a Riesz basis and computes the dual filter.

Theorem 3.4. *A filter ϕ_s generates a Riesz basis $\{\phi_s(t-ns)\}_{n\in\mathbb{Z}}$ of a space $\mathbf{U}_s$ if and only if there exists $B \geqslant A > 0$ such that*

$$\forall \omega \in [0, 2\pi/s] \quad , \quad A \leqslant \frac{1}{s} \sum_{k=-\infty}^{+\infty} |\hat{\phi}_s(\omega - \frac{2k\pi}{s})|^2 \leqslant B \ . \tag{3.18}$$

The biorthogonal basis $\{\tilde{\phi}_s(t-ns)\}_{n\in\mathbb{Z}}$ is defined by the dual filter $\tilde{\phi}_s$ which satisfies:

$$\widehat{\tilde{\phi}_s}(\omega) = \frac{s\,\hat{\phi}_s^*(\omega)}{\sum_{k=-\infty}^{+\infty} |\hat{\phi}_s(\omega - 2k\pi/s)|^2} \ . \tag{3.19}$$

Proof. Theorem 5.5 proves that $\{\phi_s(t-ns)\}_{n\in\mathbb{Z}}$ is a Riesz basis of $\mathbf{U}_s$ with Riesz bounds $B \geqslant A > 0$ if and only if it is linearly independent and

$$\forall a \in \boldsymbol{\ell}^2(\mathbb{Z}) \quad , \quad A\|a\|^2 \leqslant \|\sum_{n\in\mathbb{Z}} a[ns]\, \phi_s(t-ns)\|^2 \leqslant B\|a\|^2 \ , \tag{3.20}$$

with $\|a\|^2 = \sum_{n\in\mathbb{Z}} |a[ns]|^2$.

Let us first write these conditions in the Fourier domain. The Fourier transform of $f(t) = \sum_{n=-\infty}^{+\infty} a[ns]\, \phi_s(t-ns)$ is

$$\hat{f}(\omega) = \sum_{n=-\infty}^{+\infty} a[ns]\, \mathrm{e}^{-ins\omega}\, \hat{\phi}_s(\omega) = \hat{a}(\omega)\, \hat{\phi}_s(\omega) \tag{3.21}$$

where $\hat{a}(\omega)$ is the Fourier series $\hat{a}(\omega) = \sum_{n=-\infty}^{+\infty} a[ns]\, \mathrm{e}^{-ins\omega}$. Let us relate the norm of f and $\hat{a}$. Since $\hat{a}(\omega)$ is $2\pi/s$ periodic, inserting (3.21) in the Plancherel formula (2.26) gives

$$\begin{aligned} \|f\|^2 &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} |\hat{f}(\omega)|^2\, d\omega = \frac{1}{2\pi} \int_0^{2\pi/s} \sum_{k=-\infty}^{+\infty} |\hat{a}(\omega + 2k\pi/s)|^2\, |\hat{\phi}_s(\omega + 2k\pi/s)|^2\, d\omega \\ &= \frac{1}{2\pi} \int_0^{2\pi/s} |\hat{a}(\omega)|^2 \sum_{k=-\infty}^{+\infty} |\hat{\phi}_s(\omega + 2k\pi/s)|^2\, d\omega \ . \end{aligned} \tag{3.22}$$

Section 3.2.2 on Fourier series proves that

$$\|a\|^2 = \sum_{n=-\infty}^{+\infty} |a[ns]|^2 = \frac{s}{2\pi} \int_0^{2\pi/s} |\hat{a}(\omega)|^2\, d\omega \ . \tag{3.23}$$

As a consequence of (3.22) and (3.23), the Riesz bound inequalities (3.20) are equivalent to

$$\forall \hat{a} \in \mathbf{L^2}[0, 2\pi/s] \quad , \quad \frac{1}{2\pi} \int_0^{2\pi/s} |\hat{a}(\omega)|^2 \sum_{k=-\infty}^{+\infty} |\hat{\phi}_s(\omega + 2k\pi/s)|^2\, d\omega \leqslant \frac{B\,s}{2\pi} \int_0^{2\pi/s} |\hat{a}(\omega)|^2\, d\omega \tag{3.24}$$

and

$$\forall \hat{a} \in \mathbf{L^2}[0, 2\pi/s] \quad , \quad \frac{1}{2\pi} \int_0^{2\pi/s} |\hat{a}(\omega)|^2 \sum_{k=-\infty}^{+\infty} |\hat{\phi}_s(\omega + 2k\pi/s)|^2\, d\omega \geqslant \frac{A\,s}{2\pi} \int_0^{2\pi/s} |\hat{a}(\omega)|^2\, d\omega \ . \tag{3.25}$$

If $\hat{\phi}_s$ satisfies (3.18) then clearly (3.24) and (3.25) are valid, which proves (3.22). To prove that $\{\phi_s(t-ns)\}_{n\in\mathbb{Z}}$ is linearly independent, observe that if $f = \sum_{n\in\mathbb{Z}} a[ns]\, \phi_s(t-ns) = 0$ then (3.22) with (3.18) implies that $\hat{a} = 0$ and hence that $a = 0$. This proves that it defines a Riesz basis.

Conversely, if $\{\phi_s(ns-t)\}_{n\in\mathbb{Z}}$ is a Riesz basis. Suppose that either the upper or the lower bound of (3.18) is not satisfied for ω in a set of non-zero measure. Let $\hat{a}$ be the indicator function of this set. Then either (3.24) or (3.25) are not valid for this $\hat{a}$. This implies that the Riesz bounds (3.20) are not valid for a and hence that it is not a Riesz basis, which contradicts our hypothesis. So (3.18) is indeed valid for almost all $\omega \in [0, 2\pi/s]$.

To compute the biorthogonal basis, we are looking for $\tilde{\phi}_s \in \mathbf{U}_s$ such that $\{\tilde{\phi}_s(t-ns)\}_{n\in\mathbb{Z}}$ satisfies the biorthogonal relations (3.16). Since $\tilde{\phi}_s \in \mathbf{U}_s$ we saw in (3.21) that its Fourier transform can be written $\widehat{\tilde{\phi}_s}(\omega) = \hat{a}(\omega)\hat{\phi}_s(\omega)$ where $\hat{a}(\omega)$ is $2\pi/s$ periodic. Let us define $g(t) = \bar{\phi}_s \star \tilde{\phi}_s(t)$. Its Fourier transform is

$$\hat{g}(\omega) = \hat{\phi}_s^*(\omega)\,\hat{\tilde{\phi}}_s(\omega) = \hat{a}(\omega)|\hat{\phi}_s(\omega)|^2 \ .$$

The biorthogonal relations (3.16) are satisfied if and only if $g(ns) = 0$ if $n \neq 0$ and $g(0) = 1$. It results that $g_d(t) = \sum_{n=-\infty}^{+\infty} g(ns)\,\delta(t - ns) = \delta(t)$. Theorem 3.1 derives in (3.3) that

$$\hat{g}_d(\omega) = \frac{1}{s}\sum_{k=-\infty}^{+\infty} \hat{g}\,(\omega - 2k\pi/s) = \frac{\hat{a}(\omega)}{s}\sum_{k=-\infty}^{+\infty} |\hat{\phi}_s(\omega - 2k\pi/s)|^2 = 1.$$

It results that

$$\hat{a}(\omega) = s\left(\sum_{k=-\infty}^{+\infty} |\hat{\phi}_s(\omega - 2k\pi/s)|^2\right)^{-1}$$

which proves (3.19). ■ ■

This theorem gives a necessary and sufficient condition on the low-pass filter $\bar{\phi}_s(t) = \phi_s(-t)$ to recover a stable signal approximation from a uniform sampling at intervals s. For various sampling intervals s, the low-pass filter can be obtained by dilating a single filter $\phi_s(t) = s^{-1/2}\phi(t/s)$ and hence $\hat{\phi}_s(\omega) = s^{1/2}\hat{\phi}(s\omega)$. The necessary and sufficient Riesz basis condition (3.18) is then satisfied if and only if

$$\forall \omega \in [-\pi, \pi] \quad , \quad A \leqslant \sum_{k=-\infty}^{+\infty} |\hat{\phi}(\omega - 2k\pi)|^2 \leqslant B. \tag{3.26}$$

It results from (3.19) that the dual filter satisfies $\widehat{\tilde{\phi}_s}(\omega) = s^{1/2}\hat{\tilde{\phi}}(s\omega)$ and hence $\tilde{\phi}_s(t) = s^{-1/2}\tilde{\phi}(t/s)$. When $A = B = 1$ the Riesz basis is an orthonormal basis, which proves the following corollary.

Corollary 3.1. *The family $\{\phi_s(t - ns)\}_{n\in\mathbb{Z}}$ is an orthonormal basis of the space $\mathbf{U}_s$ it generates, with $\phi_s(t) = s^{-1/2}\phi(t/s)$, if and only if*

$$\forall \omega \in [0, 2\pi] \quad , \quad \sum_{k=-\infty}^{+\infty} |\hat{\phi}(\omega - 2k\pi)|^2 = 1 \ , \tag{3.27}$$

and the dual filter is $\tilde{\phi}_s = \phi_s$.

Shannon-Whittaker revisited The Shannon-Whittaker theorem 3.2 is defined with a sine-cardinal perfect low-pass filter ϕ_s, which we renormalize here to have a unit norm. The following theorem proves that it samples functions in an orthonormal basis.

Theorem 3.5. *If $\phi_s(t) = s^{1/2}\sin(\pi s^{-1}t)/(\pi t)$ then $\{\phi_s(t-ns)\}_{n\in\mathbb{Z}}$ is an orthonormal basis of the space $\mathbf{U}_s$ of functions whose Fourier transforms have a support included in $[-\pi/s, \pi/s]$. If $f \in \mathbf{U}_s$ then*

$$f(nT) = s^{-1/2}\, f \star \phi_s(ns) \ . \tag{3.28}$$

Proof. The filter satisfies $\phi_s(t) = s^{-1/2}\phi(t/s)$ with $\phi(t) = \sin(\pi t)/(\pi t)$. The Fourier transform $\hat{\phi}(\omega) = \mathbf{1}_{[-\pi,\pi]}(\omega)$ satisfies the condition (3.27) of Corollary 3.1, which proves that $\{\phi_s(t - ns)\}_{n\in\mathbb{Z}}$ is an orthonormal basis of a space $\mathbf{U}_s$.

Any $f(t) = \sum_{n=-\infty}^{+\infty} a[ns]\,\phi_s(t - ns) \in \mathbf{U}_s$ has a Fourier transform which can be written

$$\hat{f}(\omega) = \sum_{n=-\infty}^{+\infty} a[ns]\,\mathrm{e}^{-ins\omega}\,\hat{\phi}_s(\omega) = \hat{a}(\omega)\,s^{1/2}\,\mathbf{1}_{[-\pi/s,\pi/s]} \ , \tag{3.29}$$

which implies that $f \in \mathbf{U}_s$ if and only if f has a Fourier transform supported in $[-\pi/s, \pi/s]$.

If $f \in \mathbf{U}_s$ then decomposing it the orthonormal basis $\{\phi_s(t - ns)\}_{n\in\mathbb{Z}}$ gives

$$f(t) = P_{\mathbf{U}_s} f(t) = \sum_{n\in\mathbb{Z}} \langle f(u), \phi_s(u - ns)\rangle\, \phi_s(t - ns) \ .$$

Since $\phi_s(ps) = s^{-1/2}\delta[ps]$ and $\phi_s(-t) = \phi_s(t)$, it results that

$$f(ns) = s^{-1/2}\langle f(u), \phi_s(u-ns)\rangle = s^{-1/2}\, f \star \phi_s(ns)\ .$$

■

This theorem proves that in the particular case of the Shannon-Whittaker sampling theorem, if $f \in \mathbf{U}_s$ then the sampled low-pass filtered values $f \star \phi_s(ns)$ are proportional to the signal samples $f(ns)$. This comes from the fact that the sine-cardinal $\phi(t) = \sin(\pi t/s)/(\pi t/s)$ satisfies the interpolation property $\phi(ns) = \delta[ns]$. A generalization of such multiscale interpolations is studied in Section 7.6.

The Shannon-Whittaker sampling approximates signals by restricting their Fourier transform to a low frequency interval. It is particularly effective for smooth signals whose Fourier transform have an energy concentrated at low frequencies. It is also well adapted to sound recordings, which are well approximated by lower frequency harmonics.

For discontinuous signals such as images, a low-frequency restriction produces Gibbs oscillations studied in Section 2.3.3. The image visual quality is degraded by these oscillations, which have a total variation (2.65) that is infinite. A piecewise constant approximation has the advantage of creating no such spurious oscillations.

Block Sampler A block sampler approximates signals with piecewise constant functions. The approximation space $\mathbf{U}_s$ is the set of all functions that are constant on intervals $[ns, (n+1)s]$, for any $n \in \mathbb{Z}$. Let $\phi_s(t) = s^{-1/2}\,\mathbf{1}_{[0,s]}(t)$. The family $\{\phi_s(t-ns]\}_{n\in\mathbb{Z}}$ is an orthonormal basis of $\mathbf{U}_s$ (Exercise 3.1). If $f \notin \mathbf{U}_s$ then its orthogonal projection on $\mathbf{U}_s$ is calculated with a partial decomposition in the block orthonormal basis of $\mathbf{U}_s$

$$P_{\mathbf{U}_s} f(t) = \sum_{n=-\infty}^{+\infty} \langle f(u), \phi_s(u-ns)\rangle\, \phi_s(t-ns), \tag{3.30}$$

and each coefficient is proportional to the signal average on $[ns, (n+1)s]$

$$\langle f(u), \phi_s(u-ns)\rangle = f \star \phi_s(ns) = s^{-1/2} \int_{ns}^{(n+1)s} f(u)\, du\ .$$

This block analog to digital conversion is particularly simple to implement in analog electronic, where the integration is performed by a capacity.

In domains where f is a regular function, a piecewise constant approximation $P_{\mathbf{U}_s} f$ is not very precise and can be significantly improved. More precise approximations are obtained with approximation spaces $\mathbf{U}_s$ of higher order polynomial splines. The resulting approximations can introduce small Gibbs oscillations, but these oscillations have a finite total variation.

Spline Sampling Block samplers are generalized by a spline sampling with a space $\mathbf{U}_s$ of splines functions that are $m-1$ times continuously differentiable and equal to a polynomial of degree m on any interval $[ns, (n+1)s)$, for $n \in \mathbb{Z}$. When $m = 1$, functions in $\mathbf{U}_s$ are piecewise linear and continuous.

A Riesz basis of polynomial splines is constructed with *box splines*. A box spline ϕ of degree m is computed by convolving the box window $\mathbf{1}_{[0,1]}$ with itself $m+1$ times and centering it at 0 or 1/2. Its Fourier transform is

$$\hat{\phi}(\omega) = \left(\frac{\sin(\omega/2)}{\omega/2}\right)^{m+1} \exp\left(\frac{-i\varepsilon\omega}{2}\right). \tag{3.31}$$

If m is even then $\varepsilon = 1$ and ϕ has a support centered at $t = 1/2$. If m is odd then $\varepsilon = 0$ and $\phi(t)$ is symmetric about $t = 0$. One can verify that $\hat{\phi}(\omega)$ satisfies the sampling condition (3.26) using a closed form expression (7.19) of the resulting series. It results that for any $s > 0$, a box splines family $\{\phi_s(t-ns)\}_{n\in\mathbb{Z}}$ defines a Riesz basis of $\mathbf{U}_s$, and thus a stable sampling.

3.2 Discrete Time-Invariant Filters

3.2.1 Impulse Response and Transfer Function

Classical discrete signal processing algorithms are mostly based on time-invariant linear operators [49, 53]. The time-invariance is limited to translations on the sampling grid. To simplify notation, the sampling interval is normalized $s = 1$, and we denote $f[n]$ the sample values. A linear discrete operator L is time-invariant if an input $f[n]$ delayed by $p \in \mathbb{Z}$, $f_p[n] = f[n-p]$, produces an output also delayed by p:

$$Lf_p[n] = Lf[n-p].$$

Impulse Response We denote by $\delta[n]$ the discrete Dirac

$$\delta[n] = \begin{cases} 1 & \text{if } n = 0 \\ 0 & \text{if } n \neq 0 \end{cases} . \tag{3.32}$$

Any signal $f[n]$ can be decomposed as a sum of shifted Diracs

$$f[n] = \sum_{p=-\infty}^{+\infty} f[p]\,\delta[n-p].$$

Let $L\delta[n] = h[n]$ be the discrete *impulse response.* The linearity and time-invariance implies that

$$Lf[n] = \sum_{p=-\infty}^{+\infty} f[p]\,h[n-p] = f \star h[n]. \tag{3.33}$$

A discrete linear time-invariant operator is thus computed with a discrete convolution. If $h[n]$ has a finite support the sum (3.33) is calculated with a finite number of operations. These are called *Finite Impulse Response* (FIR) filters. Convolutions with infinite impulse response filters may also be calculated with a finite number of operations if they can be rewritten with a recursive equation (3.45).

Causality and Stability A discrete filter L is *causal* if $Lf[p]$ depends only on the values of $f[n]$ for $n \leqslant p$. The convolution formula (3.33) implies that $h[n] = 0$ if $n < 0$.

The filter is *stable* if any bounded input signal $f[n]$ produces a bounded output signal $Lf[n]$. Since

$$|Lf[n]| \leqslant \sup_{n\in\mathbb{Z}} |f[n]| \sum_{k=-\infty}^{+\infty} |h[k]|,$$

it is sufficient that $\sum_{n=-\infty}^{+\infty} |h[n]| < +\infty$, which means that $h \in \boldsymbol{\ell}^1(\mathbb{Z})$. One can verify that this sufficient condition is also necessary. The filter is thus stable if and only if $h \in \boldsymbol{\ell}^1(\mathbb{Z})$ (Exercise 3.5)

Transfer Function The Fourier transform plays a fundamental role in analyzing discrete time-invariant operators, because discrete sinusoidal waves $e_\omega[n] = \mathrm{e}^{i\omega n}$ are eigenvectors:

$$Le_\omega[n] = \sum_{p=-\infty}^{+\infty} \mathrm{e}^{i\omega(n-p)}\,h[p] = \mathrm{e}^{i\omega n} \sum_{p=-\infty}^{+\infty} h[p]\,\mathrm{e}^{-i\omega p}. \tag{3.34}$$

The eigenvalue is a Fourier series

$$\hat{h}(\omega) = \sum_{p=-\infty}^{+\infty} h[p]\,\mathrm{e}^{-i\omega p}. \tag{3.35}$$

It is the filter *transfer function.*

Example 3.2. *The uniform discrete average*

$$Lf[n] = \frac{1}{2N+1} \sum_{p=n-N}^{n+N} f[p]$$

is a time-invariant discrete filter whose impulse response is $h = (2N+1)^{-1} \mathbf{1}_{[-N,N]}$. *Its transfer function is*

$$\hat{h}(\omega) = \frac{1}{2N+1} \sum_{n=-N}^{+N} \mathrm{e}^{-in\omega} = \frac{1}{2N+1} \frac{\sin(N+1/2)\omega}{\sin \omega/2}. \tag{3.36}$$

3.2.2 Fourier Series

The properties of Fourier series are essentially the same as the properties of the Fourier transform since Fourier series are particular instances of Fourier transforms for Dirac sums. If $f(t) = \sum_{n=-\infty}^{+\infty} f[n]\, \delta(t-n)$ then $\hat{f}(\omega) = \sum_{n=-\infty}^{+\infty} f[n]\, \mathrm{e}^{-i\omega n}$.

For any $n \in \mathbb{Z}$, $\mathrm{e}^{-i\omega n}$ has period 2π, so Fourier series have period 2π. An important issue is to understand whether all functions with period 2π can be written as Fourier series. Such functions are characterized by their restriction to $[-\pi, \pi]$. We therefore consider functions $\hat{a} \in \mathbf{L^2}[-\pi, \pi]$ that are square integrable over $[-\pi, \pi]$. The space $\mathbf{L^2}[-\pi, \pi]$ is a Hilbert space with the inner product

$$\langle \hat{a}, \hat{b} \rangle = \frac{1}{2\pi} \int_{-\pi}^{\pi} \hat{a}(\omega)\, \hat{b}^*(\omega)\, d\omega \tag{3.37}$$

and the resulting norm

$$\|\hat{a}\|^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} |\hat{a}(\omega)|^2\, d\omega.$$

The following theorem proves that any function in $\mathbf{L^2}[-\pi, \pi]$ can be written as a Fourier series.

Theorem 3.6. *The family of functions* $\{\mathrm{e}^{-ik\omega}\}_{k\in\mathbb{Z}}$ *is an orthonormal basis of* $\mathbf{L^2}[-\pi, \pi]$.

Proof. The orthogonality with respect to the inner product (3.37) is established with a direct integration. To prove that $\{\exp(-ik\omega)\}_{k\in\mathbb{Z}}$ is a basis, we must show that linear expansions of these vectors are dense in $\mathbf{L^2}[-\pi, \pi]$.

We first prove that any continuously differentiable function $\hat{\phi}$ with a support included in $[-\pi, \pi]$ satisfies

$$\hat{\phi}(\omega) = \sum_{k=-\infty}^{+\infty} \langle \hat{\phi}(\xi), \exp(-ik\xi) \rangle \exp(-ik\omega)\ , \tag{3.38}$$

with a pointwise convergence for any $\omega \in [-\pi, \pi]$. Let us compute the partial sum

$$\begin{aligned} S_N(\omega) &= \sum_{k=-N}^{N} \langle \hat{\phi}(\xi), \exp(-ik\xi) \rangle \exp(-ik\omega) \\ &= \sum_{k=-N}^{N} \frac{1}{2\pi} \int_{-\pi}^{\pi} \hat{\phi}(\xi) \exp(ik\xi)\, d\xi\ \exp(-ik\omega) \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \hat{\phi}(\xi) \sum_{k=-N}^{N} \exp[ik(\xi - \omega)]\, d\xi. \end{aligned}$$

The Poisson formula (2.37) proves the distribution equality

$$\lim_{N\to+\infty} \sum_{k=-N}^{N} \exp[ik(\xi-\omega)] = 2\pi \sum_{k=-\infty}^{+\infty} \delta(\xi - \omega - 2\pi k),$$

and since the support of $\hat{\phi}$ is in $[-\pi, \pi]$ we get

$$\lim_{N\to+\infty} S_N(\omega) = \hat{\phi}(\omega).$$

Since $\hat{\phi}$ is continuously differentiable, following the steps (2.38-2.40) in the proof of the Poisson formula shows that $S_N(\omega)$ converges uniformly to $\hat{\phi}(\omega)$ on $[-\pi, \pi]$.

To prove that linear expansions of sinusoidal waves $\{\exp(-ik\omega)\}_{k\in\mathbb{Z}}$ are dense in $\mathbf{L^2}[-\pi,\pi]$, let us verify that the distance between $\hat{a} \in \mathbf{L^2}[-\pi,\pi]$ and such a linear expansion is less than ε, for any $\varepsilon > 0$. Continuously differentiable functions with a support included in $[-\pi,\pi]$ are dense in $\mathbf{L^2}[-\pi,\pi]$, hence there exists $\hat{\phi}$ such that $\|\hat{a} - \hat{\phi}\| \leqslant \varepsilon/2$. The uniform pointwise convergence proves that there exists N for which

$$\sup_{\omega\in[-\pi,\pi]} |S_N(\omega) - \hat{\phi}(\omega)| \leqslant \frac{\varepsilon}{2},$$

which implies that

$$\|S_N - \hat{\phi}\|^2 = \frac{1}{2\pi}\int_{-\pi}^{\pi} |S_N(\omega) - \hat{\phi}(\omega)|^2\, d\omega \leqslant \frac{\varepsilon^2}{4}\ .$$

It follows that $\hat{a}$ is approximated by the Fourier series S_N with an error

$$\|\hat{a} - S_N\| \leqslant \|\hat{a} - \hat{\phi}\| + \|\hat{\phi} - S_N\| \leqslant \varepsilon\ .$$

■ ■

Theorem 3.6 proves that if $f \in \boldsymbol{\ell}^2(\mathbb{Z})$, the Fourier series

$$\hat{f}(\omega) = \sum_{n=-\infty}^{+\infty} f[n]\, \mathrm{e}^{-i\omega n} \tag{3.39}$$

can be interpreted as the decomposition of $\hat{f}$ in an orthonormal basis of $\mathbf{L^2}[-\pi,\pi]$. The Fourier series coefficients can thus be written as inner products in $\mathbf{L^2}[-\pi,\pi]$:

$$f[n] = \langle \hat{f}(\omega), \mathrm{e}^{-i\omega n}\rangle = \frac{1}{2\pi}\int_{-\pi}^{\pi} \hat{f}(\omega)\, \mathrm{e}^{i\omega n}\, d\omega. \tag{3.40}$$

The energy conservation of orthonormal bases (A.10) yields a Plancherel identity:

$$\sum_{n=-\infty}^{+\infty} |f[n]|^2 = \|\hat{f}\|^2 = \frac{1}{2\pi}\int_{-\pi}^{\pi} |\hat{f}(\omega)|^2\, d\omega. \tag{3.41}$$

Pointwise Convergence The equality (3.39) is meant in the sense of mean-square convergence

$$\lim_{N\to+\infty} \left\| \hat{f}(\omega) - \sum_{k=-N}^{N} f[k]\, \mathrm{e}^{-i\omega k} \right\| = 0.$$

It does not imply a pointwise convergence at all $\omega \in \mathbb{R}$. In 1873, Dubois-Reymond constructed a periodic function $\hat{f}(\omega)$ that is continuous and whose Fourier series diverges at some points. On the other hand, if $\hat{f}(\omega)$ is continuously differentiable, then the proof of Theorem 3.6 shows that its Fourier series converges uniformly to $\hat{f}(\omega)$ on $[-\pi,\pi]$. It was only in 1966 that Carleson [141] was able to prove that if $\hat{f} \in \mathbf{L^2}[-\pi,\pi]$ then its Fourier series converges almost everywhere. The proof is very technical.

Convolutions Since $\{\mathrm{e}^{-i\omega k}\}_{k\in\mathbb{Z}}$ are eigenvectors of discrete convolution operators, we also have a discrete convolution theorem.

Theorem 3.7. *If $f \in \boldsymbol{\ell}^1(\mathbb{Z})$ and $h \in \boldsymbol{\ell}^1(\mathbb{Z})$ then $g = f \star h \in \boldsymbol{\ell}^1(\mathbb{Z})$ and*

$$\hat{g}(\omega) = \hat{f}(\omega)\, \hat{h}(\omega). \tag{3.42}$$

The proof is identical to the proof of the convolution Theorem 2.2, if we replace integrals by discrete sums. The reconstruction formula (3.40) shows that a filtered signal can be written

$$f \star h[n] = \frac{1}{2\pi}\int_{-\pi}^{\pi} \hat{h}(\omega)\hat{f}(\omega)\, \mathrm{e}^{i\omega n}\, d\omega. \tag{3.43}$$

The transfer function $\hat{h}(\omega)$ amplifies or attenuates the frequency components $\hat{f}(\omega)$ of $f[n]$.

Example 3.3. *An ideal discrete low-pass filter has a* 2π *periodic transfer function defined by* $\hat{h}(\omega) = \mathbf{1}_{[-\xi,\xi]}(\omega)$*, for* $\omega \in [-\pi, \pi]$ *and* $0 < \xi < \pi$*. Its impulse response is computed with (3.40):*

$$h[n] = \frac{1}{2\pi} \int_{-\xi}^{\xi} \mathrm{e}^{i\omega n}\, d\omega = \frac{\sin \xi n}{\pi n}. \tag{3.44}$$

It is a uniform sampling of the ideal analog low-pass filter (2.29).

Example 3.4. *A recursive filter computes* $g = Lf$ *which is solution of a recursive equation*

$$\sum_{k=0}^{K} a_k\, f[n-k] = \sum_{k=0}^{M} b_k\, g[n-k], \tag{3.45}$$

with $b_0 \neq 0$*. If* $g[n] = 0$ *and* $f[n] = 0$ *for* $n < 0$ *then* g *has a linear and time-invariant dependency upon* f*, and can thus be written* $g = f \star h$*. The transfer function is obtained by computing the Fourier transform of (3.45). The Fourier transform of* $f_k[n] = f[n-k]$ *is* $\hat{f}_k(\omega) = \hat{f}(\omega)\,\mathrm{e}^{-ik\omega}$ *so*

$$\hat{h}(\omega) = \frac{\hat{g}(\omega)}{\hat{f}(\omega)} = \frac{\sum_{k=0}^{K} a_k\, \mathrm{e}^{-ik\omega}}{\sum_{k=0}^{M} b_k\, \mathrm{e}^{-ik\omega}}.$$

It is a rational function of $\mathrm{e}^{-i\omega}$*. If* $b_k \neq 0$ *for some* $k > 0$ *then one can verify that the impulse response* h *has an infinite support. The stability of such filters is studied in Exercise* **??***. A direct calculation of the convolution sum* $g[n] = f \star h[n]$ *would require an infinite number of operations whereas (3.45) computes* $g[n]$ *with* $K + M$ *additions and multiplications from its past values.*

Window Multiplication An infinite impulse response filter h such as the ideal low-pass filter (3.44) may be approximated by a finite response filter $\tilde{h}$ by multiplying h with a window g of finite support:

$$\tilde{h}[n] = g[n]\, h[n].$$

One can verify (Exercise 3.6) that a multiplication in time is equivalent to a convolution in the frequency domain:

$$\hat{\tilde{h}}(\omega) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \hat{h}(\xi)\, \hat{g}(\omega - \xi)\, d\xi = \frac{1}{2\pi} \hat{h} \star \hat{g}(\omega). \tag{3.46}$$

Clearly $\hat{\tilde{h}} = \hat{h}$ only if $\hat{g} = 2\pi\delta$, which would imply that g has an infinite support and $g[n] = 1$. The approximation $\hat{\tilde{h}}$ is close to $\hat{h}$ only if $\hat{g}$ approximates a Dirac, which means that all its energy is concentrated at low frequencies. In time, g should therefore have smooth variations.

The rectangular window $g = \mathbf{1}_{[-N,N]}$ has a Fourier transform $\hat{g}$ computed in (3.36). It has important side lobes far away from $\omega = 0$. The resulting $\hat{\tilde{h}}$ is a poor approximation of $\hat{h}$. The Hanning window

$$g[n] = \cos^2\left(\frac{\pi n}{2N}\right) \mathbf{1}_{[-N,N]}[n]$$

is smoother and thus has a Fourier transform better concentrated at low frequencies. The spectral properties of other windows are studied in Section 4.2.2.

3.3 Finite Signals

Up to now, we have considered discrete signals $f[n]$ defined for all $n \in \mathbb{Z}$. In practice, $f[n]$ is known over a finite domain, say $0 \leqslant n < N$. Convolutions must therefore be modified to take into account the border effects at $n = 0$ and $n = N - 1$. The Fourier transform must also be redefined over finite sequences for numerical computations. The fast Fourier transform algorithm is explained as well as its application to fast convolutions.

3.3.1 Circular Convolutions

Let $\tilde{f}$ and $\tilde{h}$ be signals of N samples. To compute the convolution product

$$\tilde{f} \star \tilde{h}[n] = \sum_{p=-\infty}^{+\infty} \tilde{f}[p]\, \tilde{h}[n-p] \quad \text{for } 0 \leqslant n < N,$$

we must know $\tilde{f}[n]$ and $\tilde{h}[n]$ beyond $0 \leqslant n < N$. One approach is to extend $\tilde{f}$ and $\tilde{h}$ with a periodization over N samples, and define

$$f[n] = \tilde{f}[n \bmod N] \quad , \quad h[n] = \tilde{h}[n \bmod N].$$

The *circular convolution* of two such signals, both with period N, is defined as a sum over their period:

$$f \circledast h[n] = \sum_{p=0}^{N-1} f[p]\, h[n-p] = \sum_{p=0}^{N-1} f[n-p]\, h[p].$$

It is also a signal of period N.

The eigenvectors of a circular convolution operator

$$Lf[n] = f \circledast h[n]$$

are the discrete complex exponentials $e_k[n] = \exp(i2\pi kn/N)$. Indeed

$$Le_k[n] = \exp\left(\frac{i2\pi kn}{N}\right) \sum_{p=0}^{N-1} h[p] \exp\left(\frac{-i2\pi kp}{N}\right),$$

and the eigenvalue is the discrete Fourier transform of h:

$$\hat{h}[k] = \sum_{p=0}^{N-1} h[p] \exp\left(\frac{-i2\pi kp}{N}\right).$$

3.3.2 Discrete Fourier Transform

The space of signals of period N is an Euclidean space of dimension N and the inner product of two such signals f and g is

$$\langle f, g\rangle = \sum_{n=0}^{N-1} f[n]\, g^*[n]. \tag{3.47}$$

The following theorem proves that any signal with period N can be decomposed as a sum of discrete sinusoidal waves.

Theorem 3.8. *The family*

$$\left\{ e_k[n] = \exp\left(\frac{i2\pi kn}{N}\right) \right\}_{0 \leqslant k < N}$$

is an orthogonal basis of the space of signals of period N.

Since the space is of dimension N, any orthogonal family of N vectors is an orthogonal basis. To prove this theorem it is therefore sufficient to verify that $\{e_k\}_{0\leqslant k<N}$ is orthogonal with respect to the inner product (3.47) (Exercise 3.7). Any signal f of period N can be decomposed in this basis:

$$f = \sum_{k=0}^{N-1} \frac{\langle f, e_k\rangle}{\|e_k\|^2}\, e_k. \tag{3.48}$$

By definition, the *discrete Fourier transform* (DFT) of f is

$$\hat{f}[k] = \langle f, e_k\rangle = \sum_{n=0}^{N-1} f[n] \exp\left(\frac{-i2\pi kn}{N}\right). \tag{3.49}$$

Since $\|e_k\|^2 = N$, (3.48) gives an inverse discrete Fourier formula:

$$f[n] = \frac{1}{N} \sum_{k=0}^{N-1} \hat{f}[k] \exp\left(\frac{i2\pi kn}{N}\right). \tag{3.50}$$

The orthogonality of the basis also implies a Plancherel formula

$$\|f\|^2 = \sum_{n=0}^{N-1} |f[n]|^2 = \frac{1}{N} \sum_{k=0}^{N-1} |\hat{f}[k]|^2. \tag{3.51}$$

The discrete Fourier transform of a signal f of period N is computed from its values for $0 \leqslant n < N$. Then why is it important to consider it a periodic signal with period N rather than a finite signal of N samples? The answer lies in the interpretation of the Fourier coefficients. The discrete Fourier sum (3.50) defines a signal of period N for which the samples $f[0]$ and $f[N-1]$ are side by side. If $f[0]$ and $f[N-1]$ are very different, this produces a brutal transition in the periodic signal, creating relatively high amplitude Fourier coefficients at high frequencies. For example, Figure 3.3 shows that the "smooth" ramp $f[n] = n$ for $0 \leqslant n < N$ has sharp transitions at $n = 0$ and $n = N$ once made periodic.

Circular Convolutions Since $\{\exp(i2\pi kn/N)\}_{0 \leqslant k < N}$ are eigenvectors of circular convolutions, we derive a convolution theorem.

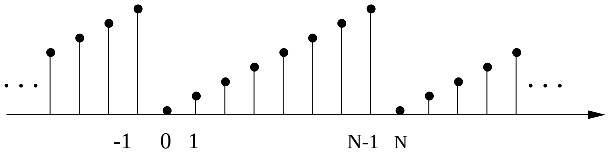


Figure 3.3: Signal $f[n] = n$ for $0 \leqslant n < N$ made periodic over N samples.

Theorem 3.9. *If f and h have period N then the discrete Fourier transform of $g = f \circledast h$ is*

$$\hat{g}[k] = \hat{f}[k]\, \hat{h}[k]. \tag{3.52}$$

The proof is similar to the proof of the two previous convolution Theorems 2.2 and 3.7. This theorem shows that a circular convolution can be interpreted as a discrete frequency filtering. It also opens the door to fast computations of convolutions using the fast Fourier transform.

3.3.3 Fast Fourier Transform

For a signal f of N points, a direct calculation of the N discrete Fourier sums

$$\hat{f}[k] = \sum_{n=0}^{N-1} f[n] \exp\left(\frac{-i2\pi kn}{N}\right) \quad , \quad \text{for } 0 \leqslant k < N\,, \tag{3.53}$$

requires N^2 complex multiplications and additions. The fast Fourier transform (FFT) algorithm reduces the numerical complexity to $O(N \log_2 N)$ by reorganizing the calculations.

When the frequency index is even, we group the terms n and $n + N/2$:

$$\hat{f}[2k] = \sum_{n=0}^{N/2-1} \Big(f[n] + f[n+N/2]\Big) \exp\left(\frac{-i2\pi kn}{N/2}\right). \tag{3.54}$$

When the frequency index is odd, the same grouping becomes

$$\hat{f}[2k+1] = \sum_{n=0}^{N/2-1} \exp\left(\frac{-i2\pi n}{N}\right) \Big(f[n] - f[n+N/2]\Big) \exp\left(\frac{-i2\pi kn}{N/2}\right). \tag{3.55}$$

Equation (3.54) proves that even frequencies are obtained by calculating the discrete Fourier transform of the $N/2$ periodic signal

$$f_e[n] = f[n] + f[n + N/2].$$

Odd frequencies are derived from (3.55) by computing the Fourier transform of the $N/2$ periodic signal

$$f_o[n] = \exp\left(\frac{-i2\pi n}{N}\right)\Big(f[n] - f[n + N/2]\Big).$$

A discrete Fourier transform of size N may thus be calculated with two discrete Fourier transforms of size $N/2$ plus $O(N)$ operations.

The inverse fast Fourier transform of $\hat{f}$ is derived from the forward fast Fourier transform of its complex conjugate $\hat{f}^*$ by observing that

$$f^*[n] = \frac{1}{N}\sum_{k=0}^{N-1} \hat{f}^*[k]\,\exp\left(\frac{-i2\pi kn}{N}\right). \tag{3.56}$$

Complexity Let $C(N)$ be the number of elementary operations needed to compute a discrete Fourier transform with the FFT. Since f is complex, the calculation of f_e and f_o requires N complex additions and $N/2$ complex multiplications. Let KN be the corresponding number of elementary operations. We have

$$C(N) = 2\,C(N/2) + K\,N. \tag{3.57}$$

Since the Fourier transform of a single point is equal to itself, $C(1) = 0$. With the change of variable $l = \log_2 N$ and the change of function $T(l) = \frac{C(N)}{N}$, we derive from (3.57) that

$$T(l) = T(l-1) + K.$$

Since $T(0) = 0$ we get $T(l) = K\,l$ and hence

$$C(N) = K\,N\,\log_2(N).$$

There exist several variations of this fast algorithm [222, 47]. The goal is to minimize the constant K. The most efficient fast discrete Fourier transform to this date is the split-radix FFT algorithm, which is slightly more complicated than the procedure just described, but which requires only $N\log_2 N$ real multiplications and $3N\log_2 N$ additions. When the input signal f is real, there are half as many parameters to compute, since $\hat{f}[-k] = \hat{f}^*[k]$. The number of multiplications and additions is thus reduced by 2.

3.3.4 Fast Convolutions

The low computational complexity of a fast Fourier transform makes it efficient to compute finite discrete convolutions by using the circular convolution Theorem 3.9. Let f and h be two signals whose samples are non-zero only for $0 \leqslant n < M$. The causal signal

$$g[n] = f \star h[n] = \sum_{k=-\infty}^{+\infty} f[k]\,h[n-k] \tag{3.58}$$

is non-zero only for $0 \leqslant n < 2M$. If h and f have M non-zero samples, calculating this convolution product with the sum (3.58) requires $M(M+1)$ multiplications and additions. When $M \geqslant 32$, the number of computations is reduced by using the fast Fourier transform [9, 47].

To use the fast Fourier transform with the circular convolution Theorem 3.9, the non-circular convolution (3.58) is written as a circular convolution. We define two signals of period $2M$:

$$a[n] = \begin{cases} f[n] & \text{if } 0 \leqslant n < M \\ 0 & \text{if } M \leqslant n < 2M \end{cases} \tag{3.59}$$

$$b[n] = \begin{cases} h[n] & \text{if } 0 \leqslant n < M \\ 0 & \text{if } M \leqslant n < 2M \end{cases}. \tag{3.60}$$

Let $c = a \circledast b$, one can verify that $c[n] = g[n]$ for $0 \leqslant n < 2M$. The $2M$ non-zero coefficients of g are thus obtained by computing $\hat{a}$ and $\hat{b}$ from a and b and then calculating the inverse discrete Fourier transform of $\hat{c} = \hat{a}\,\hat{b}$. With the fast Fourier transform algorithm, this requires a total of $O(M \log_2 M)$ additions and multiplications instead of $M(M+1)$. A single FFT or inverse FFT of a real signal of size N is calculated with $2^{-1}N \log_2 N$ multiplications, using a split-radix algorithm. The FFT convolution is thus performed with a total of $3M \log_2 M + 11M$ real multiplications. For $M \geqslant 32$ the FFT algorithm is faster than the direct convolution approach. For $M \leqslant 16$, it is faster to use a direct convolution sum.

Fast Overlap-Add Convolutions The convolution of a signal f of L non-zero samples with a smaller causal signal h of M samples is calculated with an overlap-add procedure that is faster than the previous algorithm. The signal f is decomposed with a sum of L/M blocks f_r having M non-zero samples:

$$f[n] = \sum_{r=0}^{L/M-1} f_r[n - rM] \quad \text{with} \quad f_r[n] = f[n + rM]\, \mathbf{1}_{[0,M-1]}[n]. \tag{3.61}$$

For each $0 \leqslant r < L/M$, the $2M$ non-zero samples of $g_r = f_r \star h$ are computed with the FFT based convolution algorithm, which requires $O(M \log_2 M)$ operations. These L/M convolutions are thus obtained with $O(L \log_2 M)$ operations. The block decomposition (3.61) implies that

$$f \star h[n] = \sum_{r=0}^{L/M-1} g_r[n - rM]. \tag{3.62}$$

The addition of these L/M translated signals of size $2M$ is done with $2L$ additions. The overall convolution is thus performed with $O(L \log_2 M)$ operations.

3.4 Discrete Image Processing

Two-dimensional signal processing poses many specific geometrical and topological problems that do not exist in one dimension [19, 31]. For example, a simple concept such as causality is not well defined in two dimensions. We can avoid the complexity introduced by the second dimension by extending one-dimensional algorithms with a separable approach. This not only simplifies the mathematics but also leads to fast numerical algorithms along the rows and columns of images. Appendix A.5 reviews the properties of tensor products for separable calculations.

3.4.1 Two-Dimensional Sampling Theorems

The light intensity measured by a camera is generally sampled over a rectangular array of picture elements, called *pixels*. One-dimensional sampling theorems are extended to this two-dimensional sampling array. Other two-dimensional sampling grids such as hexagonal grids are also possible, but non-rectangular sampling arrays are not often used.

Let s_1 and s_2 be the sampling intervals along the x_1 and x_2 axes of an infinite rectangular sampling grid. In the following we renormalize the axes so that $s_1 = s_2 = s$. A discrete image obtained by sampling $f(x)$ with $x = (x_1, x_2)$ can be represented as a sum of Diracs located at the grid points:

$$f_d(x) = \sum_{n \in \mathbb{Z}^2} f(sn)\, \delta(x - ns) \ .$$

The two-dimensional Fourier transform of $\delta(x - sn)$ is $\mathrm{e}^{-isn\cdot\omega}$ with $\omega = (\omega_1, \omega_2)$ and $n \cdot \omega = n_1\omega_1 + n_2\omega_2$. The Fourier transform of f_d is thus a two-dimensional Fourier series

$$\hat{f}_d(\omega) = \sum_{n \in \mathbb{Z}^2} f(sn)\, \mathrm{e}^{-isn\cdot\omega} \ . \tag{3.63}$$

It is $2\pi/s$ periodic along ω_1 and along ω_2. An extension of Theorem 3.1 relates $\hat{f}_d$ to the two-dimensional Fourier transform $\hat{f}$ of f.

Theorem 3.10. *The Fourier transform of the discrete image $f_d(x)$ is*

$$\hat{f}_d(\omega) = \frac{1}{s^2} \sum_{k \in \mathbb{Z}^2} \hat{f}(\omega - 2k\pi/s) \quad \text{with} \quad k = (k_1, k_2) \ . \tag{3.64}$$

We derive the following two-dimensional sampling theorem, which is analogous to Theorem 3.2.

Theorem 3.11. *If $\hat{f}$ has a support included in $[-\pi/s, \pi/s]^2$ then*

$$f(x) = s \sum_{n \in \mathbb{Z}^2} f(ns)\, \phi_s(x - ns) \ , \tag{3.65}$$

where

$$\phi_s(x_1, x_2) = \frac{1}{s} \frac{\sin(\pi x_1/s)}{\pi x_1/s} \frac{\sin(\pi x_2/s)}{\pi x_2/s} \ . \tag{3.66}$$

If the support of $\hat{f}$ is not included in the low-frequency rectangle $[-\pi/s, \pi/s]^2$, the interpolation formula (3.65) introduces aliasing errors. This aliasing is eliminated by prefiltering f with the ideal low-pass separable filter $\phi_s(x)$ whose Fourier transform is the indicator function of $[-\pi/s, \pi/s]^2$.

General Sampling Theorems As explained in Section 3.1.3, the Shannon-Whittaker sampling theorem is a particular case of more general linear sampling theorems with low-pass filters. The following theorem is a two-dimensional extension of Theorems 3.3 and 3.4, which characterizes these filters to obtain a stable reconstruction.

Theorem 3.12. *If there exists $B \geqslant A > 0$ such that the Fourier transform of $\phi_s \in \mathbf{L^2}(\mathbb{R}^2)$ satisfies*

$$\forall \omega \in [0, 2\pi/s]^2 \quad A \leqslant \hat{h}(\omega) = \sum_{k \in \mathbb{Z}^2} |\hat{\phi}_s(\omega - 2k\pi/s)|^2 \leqslant B \ ,$$

then $\{\phi_s(x - ns)\}_{n \in \mathbb{Z}^2}$ is a Riesz basis of a space $\mathbf{U}_s$. The Fourier transform of the dual filter $\tilde{\phi}_s$ is $\widehat{\tilde{\phi}_s}(\omega) = \hat{\phi}_s^*(\omega)/h(\omega)$, and the orthogonal projection of $f \in \mathbf{L^2}(\mathbb{R}^2)$ in $\mathbf{U}_s$ is

$$P_{\mathbf{U}_s} f(x) = \sum_{n \in \mathbb{Z}^2} f \star \bar{\phi}_s(ns)\, \tilde{\phi}_s(x - ns) \quad \text{with} \quad \bar{\phi}_s(x) = \phi_s(-x) \ . \tag{3.67}$$

This theorem gives a necessary and sufficient condition to obtain a stable linear reconstruction from samples computed with a linear filter. The proof is a direct extension of the proofs of Theorems 3.3 and 3.4. It recovers a signal approximation as an orthogonal projection, by filtering the discrete signal $f_d(x) = \sum_{n \in \mathbb{Z}^2} f \star \bar{\phi}_s(ns)\, \delta(x - ns)$

$$P_{\mathbf{U}_s} f(x) = f_d \star \tilde{\phi}_s(x) \ .$$

Like in one-dimension, the filter ϕ_s can be obtained by scaling a single filter $\phi_s(x) = s^{-1}\phi(s^{-1}x)$. The two-dimensional Shannon-Whittaker theorem is a particular example, where $\hat{\phi}_s = s\mathbf{1}_{[-\pi/s, \pi/s]^2}$, which defines an orthonormal basis of the space $\mathbf{U}_s$ of functions having a Fourier transform supported in $[-\pi/s, \pi/s]^2$.

3.4.2 Discrete Image Filtering

The properties of two-dimensional space-invariant operators are essentially the same as in one dimension. The sampling interval s is normalized to 1. A pixel value located at $n = (n_1, n_2)$ is written $f[n]$. A linear operator L is space-invariant if $Lf_p[n] = Lf[n - p]$ for any $f_p[n] = f[n - p]$, with $p = (p_1, p_2) \in \mathbb{Z}^2$. A discrete Dirac is defined by $\delta[n] = 1$ if $n = (0, 0)$ and $\delta[n] = 0$ if $n \neq (0, 0)$.

Impulse Response Since $f[n] = \sum_{p \in \mathbb{Z}^2} f[p]\, \delta[n-p]$, the linearity and time invariance implies

$$Lf[n] = \sum_{p \in \mathbb{Z}^2} f[p]\, h[n-p] = f \star h[n], \tag{3.68}$$

where $h[n]$ is the response of the impulse $h[n] = L\delta[n]$. If the impulse response is separable:

$$h[n_1, n_2] = h_1[n_1]\, h_2[n_2], \tag{3.69}$$

then the two-dimensional convolution (3.68) is computed as one-dimensional convolutions along the columns of the image followed by one-dimensional convolutions along the rows (or vice-versa):

$$f \star h[n_1, n_2] = \sum_{p_1=-\infty}^{+\infty} h_1[n_1 - p_1] \sum_{p_2=-\infty}^{+\infty} h_2[n_2 - p_2]\, f[p_1, p_2]. \tag{3.70}$$

This factorization reduces the number of operations. If h_1 and h_2 are finite impulse response filters respectively of size M_1 and M_2 then the separable calculation (3.70) requires $M_1 + M_2$ additions and multiplications per point (n_1, n_2) as opposed to $M_1\, M_2$ in a non-separable computation (3.68).

Transfer Function The Fourier transform of a discrete image f is defined by the Fourier series

$$\hat{f}(\omega) = \sum_{n \in \mathbb{Z}^2} f[n]\, \mathrm{e}^{-i\omega \cdot n} \quad \text{with} \quad \omega \cdot n = n_1 \omega_1 + n_2 \omega_2 \ . \tag{3.71}$$

The two-dimensional extension of the convolution Theorem 3.7 proves that if $g[n] = Lf[n] = f \star h[n]$ then its Fourier transform is $\hat{g}(\omega) = \hat{f}(\omega)\, \hat{h}(\omega)$, and $\hat{h}(\omega)$ is the transfer function of the filter. When a filter is separable $h[n_1, n_2] = h_1[n_1]\, h_2[n_2]$, its transfer function is also separable:

$$\hat{h}(\omega_1, \omega_2) = \hat{h}_1(\omega_1)\, \hat{h}_2(\omega_2). \tag{3.72}$$

3.4.3 Circular Convolutions and Fourier Basis

The discrete convolution of a finite image $\tilde{f}$ raises border problems. As in one dimension, these border issues are solved by extending the image, making it periodic along its rows and columns:

$$f[n_1, n_2] = \tilde{f}[n_1 \bmod N_1\, ,\, n_2 \bmod N_2] \ ,$$

where $N = N_1\, N_2$ is the image size. The resulting periodic image $f[n_1, n_2]$ is defined for all $(n_1, n_2) \in \mathbb{Z}^2$, and each of its rows and columns are periodic one-dimensional signals.

A discrete convolution is replaced by a circular convolution over the image period. If f and h have a periodicity N_1 and N_2 along (n_1, n_2) then

$$f \circledast h[n_1, n_2] = \sum_{p_1=0}^{N_1-1} \sum_{p_2=0}^{N_2-1} f[p_1, p_2]\, h[n_1 - p_1, n_2 - p_2] \ . \tag{3.73}$$

Discrete Fourier Transform The eigenvectors of circular convolutions are two-dimensional discrete sinusoidal waves:

$$e_k[n] = \mathrm{e}^{\omega_k \cdot n} \quad \text{with} \quad \omega_k = (2\pi k_1/N_1, 2\pi k_2/N_2) \quad \text{for} \quad 0 \leqslant k_1 < N_1 \ , \ 0 \leqslant k_2 < N_2 \ .$$

This family of $N = N_1\, N_2$ discrete vectors is the separable product of two one-dimensional discrete Fourier bases $\{\mathrm{e}^{i2\pi k_1 n/N_1}\}_{0 \leqslant k_1 < N_1}$ and $\{\mathrm{e}^{i2\pi k_1 n/N_2}\}_{0 \leqslant k_2 < N_2}$. Theorem A.3 thus proves that the family $\{e_k[n]\}_{0 \leqslant k_1 < N_1, 0 \leqslant k_2 < N_2}$ is an orthogonal basis of $\mathbb{C}^N = \mathbb{C}^{N_1} \otimes \mathbb{C}^{N_2}$ (Exercise 3.9). Any image $f \in \mathbb{C}^N$ can be decomposed in this orthogonal basis:

$$f[n] = \frac{1}{N} \sum_{k_1=0}^{N_1-1} \sum_{k_2=0}^{N_2-1} \hat{f}[k]\, \mathrm{e}^{i\omega_k \cdot n} \ , \tag{3.74}$$

where $\hat{f}$ is the two-dimensional discrete Fourier transform of f

$$\hat{f}[k] = \langle f, e_k \rangle = \sum_{n_1=0}^{N_1-1} \sum_{n_2=0}^{N_2-1} f[n]\, \mathrm{e}^{-i\omega_k \cdot n} \ . \tag{3.75}$$

Fast Convolutions Since $e^{i\omega_k \cdot n}$ are eigenvectors of two-dimensional circular convolutions, the discrete Fourier transform of $g = f \circledast h$ is

$$\hat{g}[k] = \hat{f}[k]\,\hat{h}[k]. \tag{3.76}$$

A direct computation of $f \circledast h$ with the summation (3.73) requires $O(N^2)$ multiplications. With the two-dimensional FFT described next, $\hat{f}[k]$ and $\hat{h}[k]$ as well as the inverse DFT of their product (3.76) are calculated with $O(N \log N)$ operations. Non-circular convolutions are computed with a fast algorithm by reducing them to circular convolutions, with the same approach as in Section 3.3.4.

Separable Basis Decomposition Let $\mathcal{B}_1 = \{e^1_{k_1}\}_{0 \leqslant k_1 < N_1}$ and $\mathcal{B}_2 = \{e^2_{k_2}\}_{0 \leqslant k_2 < N_2}$ be two orthogonal bases of $\mathbb{C}^{N_1}$ and $\mathbb{C}^{N_2}$. Suppose the calculation of decomposition coefficients of $f_1 \in \mathbb{C}^{N_1}$ in the basis $\mathcal{B}_1$ requires $C_1(N_1)$ operations and of $f_2 \in \mathbb{C}^{N_1}$ in the basis $\mathcal{B}_2$ requires $C_2(N_2)$ operations. One can verify (Exercise 3.9) that the family $\mathcal{B} = \{e_k[n] = e^1_{k_1}[n_1]\, e^2_{k_2}[n_2]\}_{0 \leqslant k_1 < N_1, 0 \leqslant k_2 < N_2}$ is an orthogonal basis of the space $\mathbb{C}^N = \mathbb{C}^{N_1} \otimes \mathbb{C}^{N_2}$ of images $f[n_1, n_2]$ of $N = N_1 N_2$ pixels. We describe a fast separable algorithm that computes the decomposition coefficients of an image f in $\mathcal{B}$ with $N_2\, C_1(N_1) + N_1\, C_2(N_2)$ operations as opposed to N^2. A fast two-dimensional FFT is derived.

Two-dimensional inner products are calculated with

$$\begin{aligned} \langle f, e^1_{k_1} e^2_{k_2} \rangle &= \sum_{n_1=0}^{N_1-1} \sum_{n_2=0}^{N_2-1} f[n_1, n_2]\, e^{1*}_{k_1}[n_1]\, e^{2*}_{k_2}[n_2] \\ &= \sum_{n_1=0}^{N_1-1} e^{1*}_{k_1}[n_1] \sum_{n_2=0}^{N_2-1} f[n_1, n_2]\, e^{2*}_{k_2}[n_2]. \end{aligned} \tag{3.77}$$

For $0 \leqslant n_1 < N_1$, we must compute

$$U f[n_1, k_2] = \sum_{n_2=0}^{N_2-1} f[n_1, n_2]\, e^{2*}_{k_2}[n_2],$$

which are the decomposition coefficients of the N_1 image rows of size N_2 in the basis $\mathcal{B}_2$. The coefficients $\{\langle f, e^1_{k_1} e^2_{k_2} \rangle\}_{0 \leqslant k_1 < N_1, 0 \leqslant k_2 < N_2}$ are calculated in (3.77) as the inner products of the columns of the transformed image $U f[n_1, k_2]$ in the basis $\mathcal{B}_1$. The overall algorithm thus requires to perform N_1 one-dimensional transforms in the basis $\mathcal{B}_2$ plus N_2 one-dimensional transforms in the basis $\mathcal{B}_1$ and hence requires $N_2\, C_1(N_1) + N_1\, C_2(N_2)$ operations.

The fast Fourier transform algorithm of Section 3.3.3 decomposes signals of size N_1 and N_2 in the discrete Fourier bases $\mathcal{B}_1 = \{e^1_{k_1}[n_1] = e^{i2\pi k_1 n_1/N_1}\}_{0 \leqslant k_1 < N_1}$ and $\mathcal{B}_2 = \{e^2_{k_2}[n_2] = e^{i2\pi k_2 n_2/N_2}\}_{0 \leqslant k_2 < N_2}$, with $C_1(N_1) = K N_1 \log_2 N_1$ and $C_2(N_2) = K N_2 \log_2 N_2$ operations. A separable implementation of a two-dimensional FFT thus requires $N_2\, C_1(N_1) + N_1\, C_2(N_2) = K N \log_2 N$ operations, with $N = N_1 N_2$. A split-radix FFT corresponds to $K = 3$.

3.5 Exercises

3.1. ^{1}

3.2. ^{1}

3.3. ^{1}

3.4. ^{1}

3.5. ^{1}

3.6. ^{1})

3.7. ^{1}

3.8. ^{1}

3.9. ^{1})

IV

Time Meets Frequency

When we listen to music, we clearly "hear" the time variation of the sound "frequencies." These localized frequency events are not pure tones but packets of close frequencies. The properties of sounds are revealed by transforms that decompose signals over elementary functions that are well concentrated in time and frequency. Windowed Fourier transforms and wavelet transforms are two important classes of local time-frequency decompositions. Measuring the time variations of "instantaneous" frequencies illustrates the limitations imposed by the Heisenberg uncertainty. These instantaneous frequencies are detected as local maxima in windowed Fourier and wavelet dictionaries, and define a signal approximation support. Audio processing algorithms are implemented by modifying the geometry of this approximation support.

There is no unique definition of time-frequency energy density. All quadratic time-frequency distributions are related through the averaging of a single quadratic form called the Wigner-Ville distribution. This framework gives another perspective on windowed Fourier and wavelet transforms.

4.1 Time-Frequency Atoms

A linear time-frequency transform correlates the signal with a dictionary of waveforms that are well concentrated in time and in frequency. These waveforms are called *time-frequency atoms*. Let us consider a general dictionary of time-frequency atoms $\mathcal{D} = \{\phi_\gamma\}_{\gamma\in\Gamma}$, where γ might be a multi-index parameter. We suppose that $\phi_\gamma \in \mathbf{L^2}(\mathbb{R})$ and that $\|\phi_\gamma\| = 1$. The corresponding linear time-frequency transform of $f \in \mathbf{L^2}(\mathbb{R})$ is defined by

$$\Phi f(\gamma) = \int_{-\infty}^{+\infty} f(t)\, \phi_\gamma^*(t)\, dt = \langle f, \phi_\gamma\rangle.$$

The Parseval formula (2.25) proves that

$$\Phi f(\gamma) = \int_{-\infty}^{+\infty} f(t)\, \phi_\gamma^*(t)\, dt = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \hat{f}(\omega)\, \hat{\phi}_\gamma^*(\omega)\, d\omega. \tag{4.1}$$

If $\phi_\gamma(t)$ is nearly zero when t is outside a neighborhood of an abscissa u, then $\langle f, \phi_\gamma\rangle$ depends only on the values of f in this neighborhood. Similarly, if $\hat{\phi}_\gamma(\omega)$ is negligible for ω far from ξ, then the right integral of (4.1) proves that $\langle f, \phi_\gamma\rangle$ reveals the properties of $\hat{f}$ in the neighborhood of ξ.

Heisenberg Boxes The slice of information provided by $\langle f, \phi_\gamma\rangle$ is represented in a time-frequency plane (t, ω) by a region whose location and width depends on the time-frequency spread of ϕ_γ. Since

$$\|\phi_\gamma\|^2 = \int_{-\infty}^{+\infty} |\phi_\gamma(t)|^2\, dt = 1,$$

we interpret $|\phi_\gamma(t)|^2$ as a probability distribution centered at

$$u_\gamma = \int_{-\infty}^{+\infty} t\,|\phi_\gamma(t)|^2\,dt. \tag{4.2}$$

The spread around u_γ is measured by the variance

$$\sigma_t^2(\gamma) = \int_{-\infty}^{+\infty} (t-u_\gamma)^2\,|\phi_\gamma(t)|^2\,dt. \tag{4.3}$$

The Plancherel formula (2.26) proves that $\int_{-\infty}^{+\infty} |\hat{\phi}_\gamma(\omega)|^2\,d\omega = 2\pi\|\phi_\gamma\|^2$. The center frequency of $\hat{\phi}_\gamma$ is therefore defined by

$$\xi_\gamma = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \omega\,|\hat{\phi}_\gamma(\omega)|^2\,d\omega, \tag{4.4}$$

and its spread around ξ_γ is

$$\sigma_\omega^2(\gamma) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} (\omega-\xi_\gamma)^2\,|\hat{\phi}_\gamma(\omega)|^2\,d\omega. \tag{4.5}$$

The time-frequency resolution of ϕ_γ is represented in the time-frequency plane (t,ω) by a Heisenberg box centered at (u_γ,ξ_γ), whose width along time is $\sigma_t(\gamma)$ and whose width along frequency is $\sigma_\omega(\gamma)$. This is illustrated by Figure 4.1. The Heisenberg uncertainty Theorem 2.6 proves that the area of the rectangle is at least 1/2:

$$\sigma_t\,\sigma_\omega \geqslant \frac{1}{2}. \tag{4.6}$$

This limits the joint resolution of ϕ_γ in time and frequency. The time-frequency plane must be manipulated carefully because a point (t_0,ω_0) is ill-defined. There is no function that is perfectly well concentrated at a point t_0 and a frequency ω_0. Only rectangles with area at least 1/2 may correspond to time-frequency atoms.

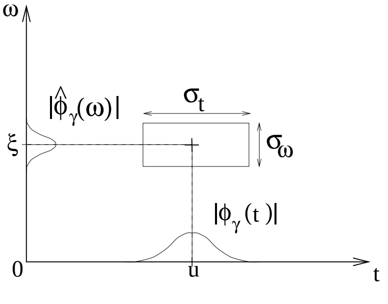


Figure 4.1: Heisenberg box representing an atom ϕ_γ.

Translation Invariant Dictionaries For pattern recognition, it can be important to construct signal representations that are translation invariant. When a pattern is translated, its numerical descriptors are then translated but not modified. Observe that for any $\phi_\gamma \in \mathcal{D}$ and any shift u

$$\langle f(t-u), \phi_\gamma(t)\rangle = \langle f(t), \phi_\gamma(t+u)\rangle\ .$$

A translation invariant representation is thus obtained if $\phi_\gamma(t+u)$ is in $\mathcal{D}$ up to a multiplicative constant. Such a dictionary is said to be translation invariant.

A translation invariant dictionary is obtained by translating a family of generators $\{\phi_\gamma\}_{\gamma\in\Gamma}$, and can be written $\mathcal{D} = \{\phi_{u,\gamma}\}_{\gamma\in\Gamma,u\in\mathbb{R}}$, with $\phi_{u,\gamma}(t) = \lambda_{u,\gamma}\,\phi_\gamma(t-u)$. The resulting time-frequency transform of f can then be written as a convolution:

$$\Phi f(u,\gamma) = \langle f, \phi_{u,\gamma}\rangle = \int_{-\infty}^{+\infty} f(t)\,\lambda_{u,\gamma}\phi_\gamma^*(t-u)\,dt = \lambda_{u,\gamma}\,f \star \tilde{\phi}_\gamma(u)$$

with $\tilde{\phi}_\gamma(t) = \phi_\gamma^*(-t)$.

Energy Density Let us suppose that $\phi_\gamma(t)$ is centered at $t = 0$ so that $\phi_{u,\gamma}(t)$ is centered at u. Let ξ_γ be the center frequency of $\hat{\phi}_\gamma(\omega)$ defined in (4.4). The time-frequency box of $\phi_{u,\gamma}$ specifies a neighborhood of (u, ξ_γ) where the energy of f is measured by

$$P_\Phi f(u,\xi_\gamma) = |\langle f, \phi_{u,\gamma}\rangle|^2 = \left| \int_{-\infty}^{+\infty} f(t)\, \phi^*_{u,\gamma}(t)\, dt \right|^2 . \tag{4.7}$$

Section 4.5.1 proves that any such energy density is an averaging of the Wigner-Ville distribution, with a kernel that depends on the atoms $\phi_{u,\gamma}$.

Example 4.1. *A windowed Fourier atom is constructed with a window g modulated by the frequency ξ and translated by u:*

$$\phi_{u,\gamma}(t) = g_{u,\xi}(t) = \mathrm{e}^{i\xi t}\, g(t-u). \tag{4.8}$$

The resulting window Fourier dictionary $\mathcal{D} = \{g_{u,\xi}(t)\}_{u,\xi\in\mathbb{R}^2}$ is translation invariant since $g_{u,\xi} = \mathrm{e}^{i\xi u}\, g_{0,\xi}(t-u)$. A windowed Fourier dictionary is also frequency shit invariant because

$$\mathrm{e}^{i\omega t}\, g_{u,\xi}(t) = g_{u,\xi+\omega}(t) \in \mathcal{D}\ .$$

This dictionary is thus particularly useful to analyze patterns that are translated in time and frequency.

A wavelet atom is a dilation by s and a translation by u of a mother wavelet ψ:

$$\phi_{u,\gamma}(t) = \psi_{u,s}(t) = \frac{1}{\sqrt{s}}\, \psi\left(\frac{t-u}{s}\right). \tag{4.9}$$

A wavelet dictionary $\mathcal{D} = \{\psi_{u,s}(t)\}_{u\in\mathbb{R}, s\in\mathbb{R}^+}$ translation invariant but also scale invariant, because scaling any wavelet produces a dilated wavelet that remains in the dictionary. A wavelet dictionary is well adapted to analyze patterns that are translated and scaled by arbitrary factors.

Wavelets and windowed Fourier atoms have their energy well localized in time, while their Fourier transform is mostly concentrated in a limited frequency band. The properties of the resulting transforms are studied in Sections 4.2 and 4.3.

4.2 Windowed Fourier Transform

In 1946, Gabor [246] introduced windowed Fourier atoms to measure the "frequency variations" of sounds. A real and symmetric window $g(t) = g(-t)$ is translated by u and modulated by the frequency ξ:

$$g_{u,\xi}(t) = \mathrm{e}^{i\xi t} g(t-u). \tag{4.10}$$

It is normalized $\|g\| = 1$ so that $\|g_{u,\xi}\| = 1$ for any $(u,\xi) \in \mathbb{R}^2$. The resulting windowed Fourier transform of $f \in \mathbf{L^2}(\mathbb{R})$ is

$$Sf(u,\xi) = \langle f, g_{u,\xi}\rangle = \int_{-\infty}^{+\infty} f(t)\, g(t-u)\, \mathrm{e}^{-i\xi t}\, dt. \tag{4.11}$$

This transform is also called the *short time Fourier transform* because the multiplication by $g(t-u)$ localizes the Fourier integral in the neighborhood of $t = u$.

As in (4.7), one can define an energy density called a *spectrogram*, denoted P_S:

$$P_S f(u,\xi) = |Sf(u,\xi)|^2 = \left| \int_{-\infty}^{+\infty} f(t)\, g(t-u)\, \mathrm{e}^{-i\xi t}\, dt \right|^2 . \tag{4.12}$$

The spectrogram measures the energy of f in a time-frequency neighborhood of (u,ξ) specified by the Heisenberg box of $g_{u,\xi}$.

Heisenberg Boxes Since g is even, $g_{u,\xi}(t) = e^{i\xi t} g(t-u)$ is centered at u. The time spread around u is independent of u and ξ:

$$\sigma_t^2 = \int_{-\infty}^{+\infty} (t-u)^2 \, |g_{u,\xi}(t)|^2 \, dt = \int_{-\infty}^{+\infty} t^2 \, |g(t)|^2 \, dt. \tag{4.13}$$

The Fourier transform $\hat{g}$ of g is real and symmetric because g is real and symmetric. The Fourier transform of $g_{u,\xi}$ is

$$\hat{g}_{u,\xi}(\omega) = \hat{g}(\omega - \xi) \, \exp[-iu(\omega - \xi)] \, . \tag{4.14}$$

It is a translation by ξ of the frequency window $\hat{g}$, so its center frequency is ξ. The frequency spread around ξ is

$$\sigma_\omega^2 = \frac{1}{2\pi} \int_{-\infty}^{+\infty} (\omega - \xi)^2 \, |\hat{g}_{u,\xi}(\omega)|^2 \, d\omega = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \omega^2 \, |\hat{g}(\omega)|^2 \, d\omega. \tag{4.15}$$

It is independent of u and ξ. Hence $g_{u,\xi}$ corresponds to a Heisenberg box of area $\sigma_t \, \sigma_\omega$ centered at (u, ξ), as illustrated by Figure 4.2. The size of this box is independent of (u, ξ), which means that a windowed Fourier transform has the same resolution across the time-frequency plane.

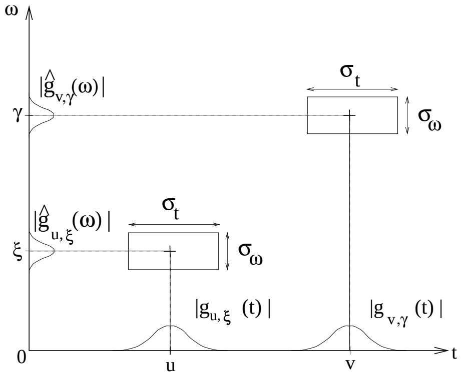


Figure 4.2: Heisenberg boxes of two windowed Fourier atoms $g_{u,\xi}$ and $g_{\nu,\gamma}$.

Example 4.2. *A sinusoidal wave* $f(t) = \exp(i\xi_0 t)$ *whose Fourier transform is a Dirac* $\hat{f}(\omega) = 2\pi\delta(\omega - \xi_0)$ *has a windowed Fourier transform*

$$Sf(u, \xi) = \hat{g}(\xi - \xi_0) \, \exp[-iu(\xi - \xi_0)] \, .$$

Its energy is spread over the frequency interval $[\xi_0 - \sigma_\omega/2, \xi_0 + \sigma_\omega/2]$.

Example 4.3. *The windowed Fourier transform of a Dirac* $f(t) = \delta(t - u_0)$ *is*

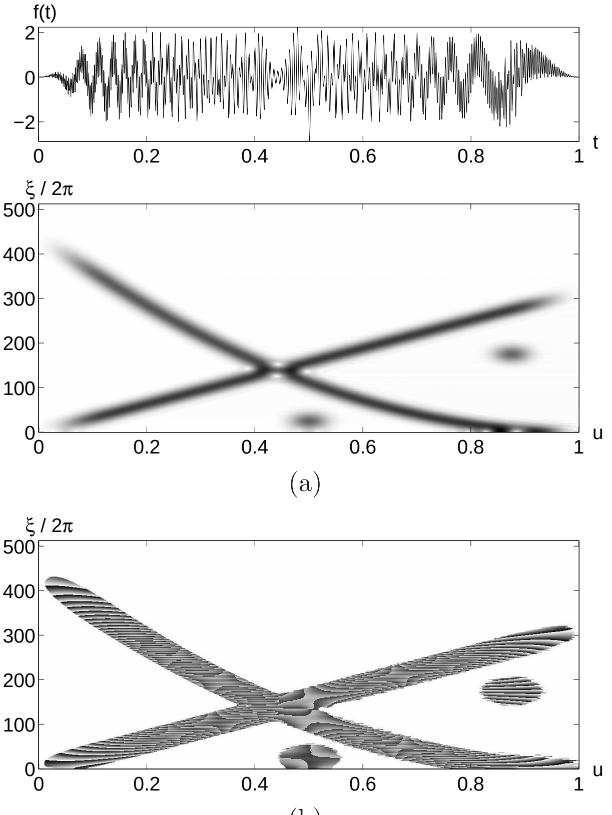
$$Sf(u, \xi) = g(u_0 - u) \, \exp(-i\xi u_0) \, .$$

Its energy is spread in the time interval $[u_0 - \sigma_t/2, u_0 + \sigma_t/2]$.

Example 4.4. *A linear chirp* $f(t) = \exp(iat^2)$ *has an "instantaneous frequency" that increases linearly in time. For a Gaussian window* $g(t) = (\pi\sigma^2)^{-1/4} \exp[-t^2/(2\sigma^2)]$, *the windowed Fourier transform of* f *is calculated using the Fourier transform (2.34) of Gaussian chirps. One can verify that its spectrogram is*

$$P_S f(u, \xi) = |Sf(u, \xi)|^2 = \left(\frac{4\pi\sigma^2}{1 + 4a^2\sigma^4} \right)^{1/2} \exp\left(-\frac{\sigma^2(\xi - 2au)^2}{1 + 4a^2\sigma^4} \right) . \tag{4.16}$$

For a fixed time u, $P_S f(u, \xi)$ *is a Gaussian that reaches its maximum at the frequency* $\xi(u) = 2au$. *Observe that if we write* $f(t) = \exp[i\phi(t)]$, *then* $\xi(u)$ *is equal to the "instantaneous frequency," defined as the derivative of the phase:* $\omega(u) = \phi'(u) = 2au$. *Section 4.4.1 explains this result.*



(a)

(b)

Figure 4.3: The signal includes a linear chirp whose frequency increases, a quadratic chirp whose frequency decreases, and two modulated Gaussian functions located at $t = 0.5$ and $t = 0.87$. (a) Spectrogram $P_S f(u, \xi)$. Dark points indicate large amplitude coefficients. (b) Complex phase of $Sf(u, \xi)$ in regions where the modulus $P_S f(u, \xi)$ is non-zero.

Example 4.5. *Figure 4.3 gives the spectrogram of a signal that includes a linear chirp, a quadratic chirp and two modulated Gaussians. The spectrogram is computed with a Gaussian window dilated by* $\sigma = 0.05$. *As expected from (4.16), the linear chirp yields large amplitude coefficients along the trajectory of its instantaneous frequency, which is a straight line. The quadratic chirp yields large coefficients along a parabola. The two modulated Gaussians produce low and high frequency blobs at* $u = 0.5$ *and* $u = 0.87$.

4.2.1 Completeness and Stability

When the time-frequency indices (u, ξ) vary across $\mathbb{R}^2$, the Heisenberg boxes of the atoms $g_{u,\xi}$ cover the whole time-frequency plane. One can thus expect that f can be recovered from its windowed Fourier transform $Sf(u, \xi)$. The following theorem gives a reconstruction formula and proves that the energy is conserved.

Theorem 4.1. *If* $f \in \mathbf{L^2}(\mathbb{R})$ *then*

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} Sf(u, \xi)\, g(t - u)\, \mathrm{e}^{i\xi t}\, d\xi\, du \tag{4.17}$$

and

$$\int_{-\infty}^{+\infty} |f(t)|^2\, dt = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} |Sf(u, \xi)|^2\, d\xi\, du. \tag{4.18}$$

Proof. The reconstruction formula (4.17) is proved first. Let us apply the Fourier Parseval formula (2.25) to the integral (4.17) with respect to the integration in u. The Fourier transform of $f_\xi(u) = Sf(u, \xi)$ with respect to u is computed by observing that

$$Sf(u, \xi) = \exp(-iu\xi) \int_{-\infty}^{+\infty} f(t)\, g(t - u)\, \exp[i\xi(u - t)]\, dt = \exp(-iu\xi)\, f \star g_\xi(u),$$

where $g_\xi(t) = g(t)\, \exp(i\xi t)$, because $g(t) = g(-t)$. Its Fourier transform is therefore

$$\hat{f}_\xi(\omega) = \hat{f}(\omega + \xi)\, \hat{g}_\xi(\omega + \xi) = \hat{f}(\omega + \xi)\, \hat{g}(\omega).$$

The Fourier transform of $g(t - u)$ with respect to u is $\hat{g}(\omega)\, \exp(-it\omega)$. Hence

$$\frac{1}{2\pi} \left(\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} Sf(u, \xi)\, g(t - u)\, \exp(i\xi t)\, du \right) d\xi \quad =$$
$$\frac{1}{2\pi} \int_{-\infty}^{+\infty} \left(\frac{1}{2\pi} \int_{-\infty}^{+\infty} \hat{f}(\omega + \xi)\, |\hat{g}(\omega)|^2\, \exp[it(\omega + \xi)]\, d\omega \right) d\xi \quad .$$

If $\hat{f} \in \mathbf{L^1}(\mathbb{R})$, we can apply the Fubini Theorem A.2 to reverse the integration order. The inverse Fourier transform proves that

$$\frac{1}{2\pi} \int_{-\infty}^{+\infty} \hat{f}(\omega + \xi)\, \exp[it(\omega + \xi)]\, d\xi = f(t).$$

Since $\frac{1}{2\pi} \int_{-\infty}^{+\infty} |\hat{g}(\omega)|^2\, d\omega = 1$, we derive (4.17). If $\hat{f} \notin \mathbf{L^1}(\mathbb{R})$, a density argument is used to verify this formula.

Let us now prove the energy conservation (4.18). Since the Fourier transform in u of $Sf(u, \xi)$ is $\hat{f}(\omega + \xi)\, \hat{g}(\omega)$, the Plancherel formula (2.26) applied to the right-hand side of (4.18) gives

$$\frac{1}{2\pi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} |Sf(u, \xi)|^2\, du\, d\xi = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \frac{1}{2\pi} \int_{-\infty}^{+\infty} |\hat{f}(\omega + \xi)\, \hat{g}(\omega)|^2\, d\omega\, d\xi.$$

The Fubini theorem applies and the Plancherel formula proves that

$$\frac{1}{2\pi} \int_{-\infty}^{+\infty} |\hat{f}(\omega + \xi)|^2\, d\xi = \|f\|^2,$$

which implies (4.18). ■ ■

The reconstruction formula (4.17) can be rewritten

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \langle f, g_{u,\xi} \rangle \, g_{u,\xi}(t) \, d\xi \, du. \tag{4.19}$$

It resembles the decomposition of a signal in an orthonormal basis but it is not, since the functions $\{g_{u,\xi}\}_{u,\xi \in \mathbb{R}^2}$ are very redundant in $\mathbf{L^2}(\mathbb{R})$. The second equality (4.18) justifies the interpretation of the spectrogram $P_S f(u,\xi) = |Sf(u,\xi)|^2$ as an energy density, since its time-frequency sum equals the signal energy.

Reproducing Kernel A windowed Fourier transform represents a one-dimensional signal $f(t)$ by a two-dimensional function $Sf(u,\xi)$. The energy conservation proves that $Sf \in \mathbf{L^2}(\mathbb{R}^2)$. Because $Sf(u,\xi)$ is redundant, it is not true that any $\Phi \in \mathbf{L^2}(\mathbb{R}^2)$ is the windowed Fourier transform of some $f \in \mathbf{L^2}(\mathbb{R})$. The next theorem gives a necessary and sufficient condition for such a function to be a windowed Fourier transform.

Theorem 4.2. *Let $\Phi \in \mathbf{L^2}(\mathbb{R}^2)$. There exists $f \in \mathbf{L^2}(\mathbb{R})$ such that $\Phi(u,\xi) = Sf(u,\xi)$ if and only if*

$$\Phi(u_0,\xi_0) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \Phi(u,\xi) \, K(u_0,u,\xi_0,\xi) \, du \, d\xi, \tag{4.20}$$

with

$$K(u_0,u,\xi_0,\xi) = \langle g_{u,\xi}, g_{u_0,\xi_0} \rangle \; . \tag{4.21}$$

Proof. Suppose that there exists f such that $\Phi(u,\xi) = Sf(u,\xi)$. Let us replace f with its reconstruction integral (4.17) in the windowed Fourier transform definition:

$$Sf(u_0,\xi_0) = \int_{-\infty}^{+\infty} \left(\frac{1}{2\pi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} Sf(u,\xi) \, g_{u,\xi}(t) \, du \, d\xi \right) g^*_{u_0,\xi_0}(t) \, dt. \tag{4.22}$$

Inverting the integral on t with the integrals on u and ξ yields (4.20). To prove that the condition (4.20) is sufficient, we define f as in the reconstruction formula (4.17):

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \Phi(u,\xi) \, g(t-u) \, \exp(i\xi t) \, d\xi \, du$$

and show that (4.20) implies that $\Phi(u,\xi) = Sf(u,\xi)$. ■ ■

Ambiguity Function The reproducing kernel $K(u_0,u,\xi_0,\xi)$ measures the time-frequency overlap of the two atoms $g_{u,\xi}$ and g_{u_0,ξ_0}. The amplitude of $K(u_0,u,\xi_0,\xi)$ decays with $u_0 - u$ and $\xi_0 - \xi$ at a rate that depends on the energy concentration of g and $\hat{g}$. Replacing $g_{u,\xi}$ and g_{u_0,ξ_0} by their expression and making the change of variable $v = t - (u+u_0)/2$ in the inner product integral (4.21) yields

$$K(u_0,u,\xi_0,\xi) = \exp\left(-\frac{i}{2} (\xi_0 - \xi)(u + u_0) \right) Ag(u_0 - u, \xi_0 - \xi) \tag{4.23}$$

where

$$Ag(\tau,\gamma) = \int_{-\infty}^{+\infty} g\left(v + \frac{\tau}{2}\right) g\left(v - \frac{\tau}{2}\right) \mathrm{e}^{-i\gamma v} \, dv \tag{4.24}$$

is called the *ambiguity function* of g. Using the Parseval formula to replace this time integral with a Fourier integral gives

$$Ag(\tau,\gamma) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \hat{g}\left(\omega + \frac{\gamma}{2}\right) \hat{g}\left(\omega - \frac{\gamma}{2}\right) \mathrm{e}^{i\tau\omega} \, d\omega. \tag{4.25}$$

The decay of the ambiguity function measures the spread of g in time and of $\hat{g}$ in frequency. For example, if g has a support included in an interval of size T, then $Ag(\tau,\omega) = 0$ for $|\tau| \geqslant T/2$. The integral (4.25) shows that the same result applies to the support of $\hat{g}$.

4.2.2 Choice of Window

The resolution in time and frequency of the windowed Fourier transform depends on the spread of the window in time and frequency. This can be measured from the decay of the ambiguity function (4.24) or more simply from the area $\sigma_t\, \sigma_\omega$ of the Heisenberg box. The uncertainty Theorem 2.6 proves that this area reaches the minimum value 1/2 if and only if g is a Gaussian. The ambiguity function $Ag(\tau, \gamma)$ is then a two-dimensional Gaussian.

Window Scale The time-frequency localization of g can be modified with a scaling. Suppose that g has a Heisenberg time and frequency width respectively equal to σ_t and σ_ω. Let $g_s(t) = s^{-1/2}\, g(t/s)$ be its dilation by s. A change of variables in the integrals (4.13) and (4.15) shows that the Heisenberg time and frequency width of g_s are respectively $s\sigma_t$ and σ_ω/s. The area of the Heisenberg box is not modified but it is dilated by s in time and compressed by s in frequency. Similarly, a change of variable in the ambiguity integral (4.24) shows that the ambiguity function is dilated in time and frequency respectively by s and $1/s$

$$Ag_s(\tau, \gamma) = Ag\left(\frac{\tau}{s}, s\gamma\right).$$

The choice of a particular scale s depends on the desired resolution trade-off between time and frequency.

Finite Support In numerical applications, g must have a compact support. Theorem 2.7 proves that its Fourier transform $\hat{g}$ necessarily has an infinite support. It is a symmetric function with a main lobe centered at $\omega = 0$, which decays to zero with oscillations. Figure 4.4 illustrates its behavior. To maximize the frequency resolution of the transform, we must concentrate the energy of $\hat{g}$ near $\omega = 0$. Three important parameters evaluate the spread of $\hat{g}$:

- The root mean-square bandwidth $\Delta\omega$, which is defined by

$$\frac{|\hat{g}(\Delta\omega/2)|^2}{|\hat{g}(0)|^2} = \frac{1}{2}.$$

- The maximum amplitude A of the first side-lobes located at $\omega = \pm\omega_0$ in Figure 4.4. It is measured in decibels:

$$A = 10\log_{10}\frac{|\hat{g}(\omega_0)|^2}{|\hat{g}(0)|^2}.$$

- The polynomial exponent p, which gives the asymptotic decay of $|\hat{g}(\omega)|$ for large frequencies:

$$|\hat{g}(\omega)| = O(\omega^{-p-1}). \tag{4.26}$$

  Table 4.1 gives the values of these three parameters for several windows g whose supports are restricted to $[-1/2, 1/2]$ [267]. Figure 4.5 shows the graph of these windows.

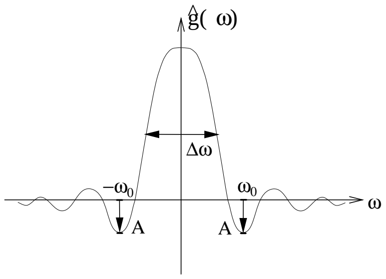


Figure 4.4: The energy spread of $\hat{g}$ is measured by its bandwidth $\Delta\omega$ and the maximum amplitude A of the first side-lobes, located at $\omega = \pm\omega_0$.

Name	$g(t)$	$\Delta\omega$	A	p
Rectangle	1	0.89	-13db	0
Hamming	$0.54 + 0.46\cos(2\pi t)$	1.36	-43db	0
Gaussian	$\exp(-18t^2)$	1.55	-55db	0
Hanning	$\cos^2(\pi t)$	1.44	-32db	2
Blackman	$0.42 + 0.5\cos(2\pi t)$ $+0.08\cos(4\pi t)$	1.68	-58db	2

Table 4.1: Frequency parameters of five windows g whose supports are restricted to $[-1/2, 1/2]$. These windows are normalized so that $g(0) = 1$ but $\|g\| \neq 1$.

To interpret these three frequency parameters, let us consider the spectrogram of a frequency tone $f(t) = \exp(i\xi_0 t)$. If $\Delta\omega$ is small, then $|Sf(u,\xi)|^2 = |\hat{g}(\xi - \xi_0)|^2$ has an energy concentrated near $\xi = \xi_0$. The side-lobes of $\hat{g}$ create "shadows" at $\xi = \xi_0 \pm \omega_0$, which can be neglected if A is also small.

If the frequency tone is embedded in a signal that has other components of much higher energy at different frequencies, the tone can still be detected if $\hat{g}(\omega - \xi)$ attenuates these components rapidly when $|\omega - \xi|$ increases. This means that $|\hat{g}(\omega)|$ has a rapid decay, and Theorem 2.5 proves that this decay depends on the regularity of g. Property (4.26) is typically satisfied by windows that are p times differentiable.

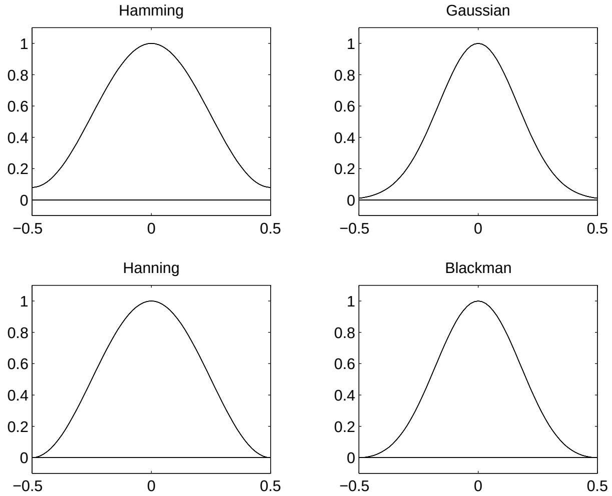


Figure 4.5: Graphs of four windows g whose support are $[-1/2, 1/2]$.

4.2.3 Discrete Windowed Fourier Transform

The discretization and fast computation of the windowed Fourier transform follow the same ideas as the discretization of the Fourier transform described in Section 3.3. We consider discrete signals of period N. The window $g[n]$ is chosen to be a symmetric discrete signal of period N with unit norm $\|g\| = 1$. Discrete windowed Fourier atoms are defined by

$$g_{m,l}[n] = g[n-m]\,\exp\left(\frac{i2\pi ln}{N}\right).$$

The discrete Fourier transform of $g_{m,l}$ is

$$\hat{g}_{m,l}[k] = \hat{g}[k-l]\,\exp\left(\frac{-i2\pi m(k-l)}{N}\right).$$

The discrete windowed Fourier transform of a signal f of period N is

$$Sf[m,l] = \langle f, g_{m,l} \rangle = \sum_{n=0}^{N-1} f[n]\, g[n-m]\, \exp\left(\frac{-i2\pi ln}{N}\right), \tag{4.27}$$

For each $0 \leqslant m < N$, $Sf[m,l]$ is calculated for $0 \leqslant l < N$ with a discrete Fourier transform of $f[n]g[n-m]$. This is performed with N FFT procedures of size N, and thus requires a total of $O(N^2 \log_2 N)$ operations. Figure 4.3 is computed with this algorithm.

Inverse Transform The following theorem discretizes the reconstruction formula and the energy conservation of Theorem 4.1.

Theorem 4.3. *If f is a signal of period N then*

$$f[n] = \frac{1}{N} \sum_{m=0}^{N-1} \sum_{l=0}^{N-1} Sf[m,l]\, g[n-m]\, \exp\left(\frac{i2\pi ln}{N}\right) \tag{4.28}$$

and

$$\sum_{n=0}^{N-1} |f[n]|^2 = \frac{1}{N} \sum_{l=0}^{N-1} \sum_{m=0}^{N-1} |Sf[m,l]|^2. \tag{4.29}$$

This theorem is proved by applying the Parseval and Plancherel formulas of the discrete Fourier transform, exactly as in the proof of Theorem 4.1 (Exercise 4.1). The energy conservation (4.29) proves that this windowed Fourier transform defines a tight frame, as explained in Chapter 5. The reconstruction formula (4.28) is rewritten

$$f[n] = \frac{1}{N} \sum_{m=0}^{N-1} g[n-m] \sum_{l=0}^{N-1} Sf[m,l]\, \exp\left(\frac{i2\pi ln}{N}\right).$$

The second sum computes for each $0 \leqslant m < N$ the inverse discrete Fourier transform of $Sf[m,l]$ with respect to l. This is calculated with N FFT procedures, requiring a total of $O(N^2 \log_2 N)$ operations.

A discrete windowed Fourier transform is an N^2 image $Sf[l,m]$ that is very redundant, since it is entirely specified by a signal f of size N. The redundancy is characterized by a discrete reproducing kernel equation, which is the discrete equivalent of (4.20) (Exercise 4.1).

4.3 Wavelet Transforms

To analyze signal structures of very different sizes, it is necessary to use time-frequency atoms with different time supports. The wavelet transform decomposes signals over dilated and translated wavelets. A wavelet is a function $\psi \in \mathbf{L^2}(\mathbb{R})$ with a zero average:

$$\int_{-\infty}^{+\infty} \psi(t)\, dt = 0. \tag{4.30}$$

It is normalized $\|\psi\| = 1$, and centered in the neighborhood of $t = 0$. A dictionary of time-frequency atoms is obtained by scaling ψ by s and translating it by u:

$$\mathcal{D} = \left\{ \psi_{u,s}(t) = \frac{1}{\sqrt{s}}\, \psi\left(\frac{t-u}{s}\right) \right\}_{u \in \mathbb{R}, s \in \mathbb{R}^+}.$$

These atoms remain normalized: $\|\psi_{u,s}\| = 1$. The wavelet transform of $f \in \mathbf{L^2}(\mathbb{R})$ at time u and scale s is

$$Wf(u,s) = \langle f, \psi_{u,s} \rangle = \int_{-\infty}^{+\infty} f(t)\, \frac{1}{\sqrt{s}}\, \psi^*\left(\frac{t-u}{s}\right) dt. \tag{4.31}$$

Linear Filtering The wavelet transform can be rewritten as a convolution product:

$$Wf(u,s) = \int_{-\infty}^{+\infty} f(t)\,\frac{1}{\sqrt{s}}\,\psi^*\left(\frac{t-u}{s}\right) dt = f \star \bar{\psi}_s(u) \tag{4.32}$$

with

$$\bar{\psi}_s(t) = \frac{1}{\sqrt{s}}\,\psi^*\left(\frac{-t}{s}\right).$$

The Fourier transform of $\bar{\psi}_s(t)$ is

$$\widehat{\bar{\psi}_s}(\omega) = \sqrt{s}\,\hat{\psi}^*(s\omega). \tag{4.33}$$

Since $\hat{\psi}(0) = \int_{-\infty}^{+\infty} \psi(t)\,dt = 0$, it appears that $\hat{\psi}$ is the transfer function of a band-pass filter. The convolution (4.32) computes the wavelet transform with dilated band-pass filters.

Analytic Versus Real Wavelets Like a windowed Fourier transform, a wavelet transform can measure the time evolution of frequency transients. This requires using a complex analytic wavelet, which can separate amplitude and phase components. The properties of this analytic wavelet transform are described in Section 4.3.2, and its application to the measurement of instantaneous frequencies is explained in Section 4.4.2. In contrast, real wavelets are often used to detect sharp signal transitions. Section 4.3.1 introduces elementary properties of real wavelets, which are developed in Chapter 6.

4.3.1 Real Wavelets

Suppose that ψ is a real wavelet. Since it has a zero average, the wavelet integral

$$Wf(u,s) = \int_{-\infty}^{+\infty} f(t)\,\frac{1}{\sqrt{s}}\,\psi^*\left(\frac{t-u}{s}\right) dt$$

measures the variation of f in a neighborhood of u, whose size is proportional to s. Section 6.1.3 proves that when the scale s goes to zero, the decay of the wavelet coefficients characterizes the regularity of f in the neighborhood of u. This has important applications for detecting transients and analyzing fractals. This section concentrates on the completeness and redundancy properties of real wavelet transforms.

Example 4.6. *Wavelets equal to the second derivative of a Gaussian are called Mexican hats. They were first used in computer vision to detect multiscale edges [446]. The normalized Mexican hat wavelet is*

$$\psi(t) = \frac{2}{\pi^{1/4}\sqrt{3\sigma}}\left(\frac{t^2}{\sigma^2} - 1\right)\exp\left(\frac{-t^2}{2\sigma^2}\right). \tag{4.34}$$

For $\sigma = 1$, Figure 4.6 plots $-\psi$ and its Fourier transform

$$\hat{\psi}(\omega) = \frac{-\sqrt{8}\,\sigma^{5/2}\,\pi^{1/4}}{\sqrt{3}}\,\omega^2\,\exp\left(\frac{-\sigma^2\omega^2}{2}\right). \tag{4.35}$$

Figure 4.7 shows the wavelet transform of a signal that is piecewise regular on the left and almost everywhere singular on the right. The maximum scale is smaller than 1 because the support of f is normalized to $[0,1]$. The minimum scale is limited by the sampling interval of the discretized signal used in numerical calculations. When the scale decreases, the wavelet transform has a rapid decay to zero in the regions where the signal is regular. The isolated singularities on the left create cones of large amplitude wavelet coefficients that converge to the locations of the singularities. This is further explained in Chapter 6.

A real wavelet transform is complete and maintains an energy conservation, as long as the wavelet satisfies a weak admissibility condition, specified by the following theorem. This theorem was first proved in 1964 by the mathematician Calderón [125], from a different point of view. Wavelets did not appear as such, but Calderón defines a wavelet transform as a convolution operator that decomposes the identity. Grossmann and Morlet [263] were not aware of Calderón's work when they proved the same formula for signal processing.

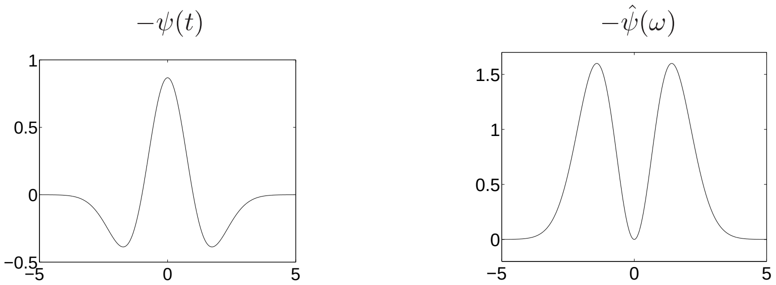


Figure 4.6: Mexican hat wavelet (4.34) for $\sigma = 1$ and its Fourier transform.

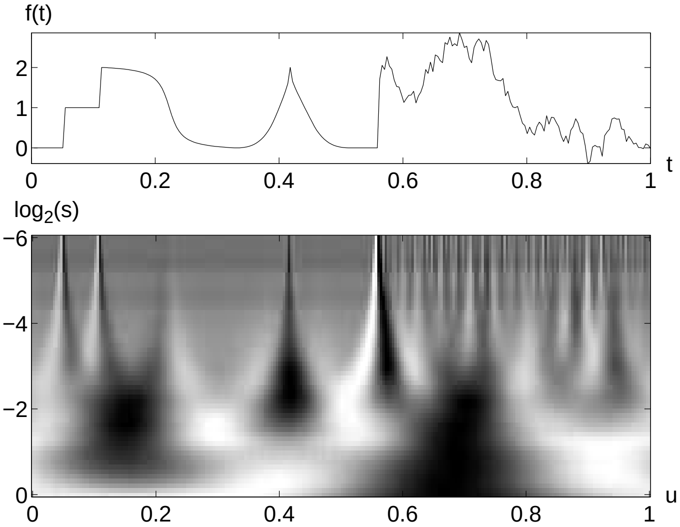


Figure 4.7: Real wavelet transform $Wf(u, s)$ computed with a Mexican hat wavelet (4.34). The vertical axis represents $\log_2 s$. Black, grey and white points correspond respectively to positive, zero and negative wavelet coefficients.

Theorem 4.4 (Calderón, Grossmann, Morlet)**.** *Let* $\psi \in \mathbf{L^2}(\mathbb{R})$ *be a real function such that*

$$C_\psi = \int_0^{+\infty} \frac{|\hat{\psi}(\omega)|^2}{\omega}\, d\omega < +\infty. \tag{4.36}$$

Any $f \in \mathbf{L^2}(\mathbb{R})$ *satisfies*

$$f(t) = \frac{1}{C_\psi} \int_0^{+\infty} \int_{-\infty}^{+\infty} Wf(u,s)\, \frac{1}{\sqrt{s}}\, \psi\left(\frac{t-u}{s}\right) du\, \frac{ds}{s^2}, \tag{4.37}$$

and

$$\int_{-\infty}^{+\infty} |f(t)|^2 dt = \frac{1}{C_\psi} \int_0^{+\infty} \int_{-\infty}^{+\infty} |Wf(u,s)|^2\, du\, \frac{ds}{s^2}. \tag{4.38}$$

Proof. The proof of (4.38) is almost identical to the proof of (4.18). Let us concentrate on the proof of (4.37). The right integral $b(t)$ of (4.37) can be rewritten as a sum of convolutions. Inserting $Wf(u,s) = f \star \bar{\psi}_s(u)$ with $\psi_s(t) = s^{-1/2}\, \psi(t/s)$ yields

$$\begin{aligned} b(t) &= \frac{1}{C_\psi} \int_0^{+\infty} Wf(.,s) \star \psi_s(t)\, \frac{ds}{s^2} \\ &= \frac{1}{C_\psi} \int_0^{+\infty} f \star \bar{\psi}_s \star \psi_s(t)\, \frac{ds}{s^2}. \end{aligned} \tag{4.39}$$

The "." indicates the variable over which the convolution is calculated. We prove that $b = f$ by showing that their Fourier transforms are equal. The Fourier transform of b is

$$\hat{b}(\omega) = \frac{1}{C_\psi} \int_0^{+\infty} \hat{f}(\omega)\, \sqrt{s}\, \hat{\psi}^*(s\omega)\, \sqrt{s}\, \hat{\psi}(s\omega)\, \frac{ds}{s^2} = \frac{\hat{f}(\omega)}{C_\psi} \int_0^{+\infty} |\hat{\psi}(s\,\omega)|^2\, \frac{ds}{s}.$$

Since ψ is real we know that $|\hat{\psi}(-\omega)|^2 = |\hat{\psi}(\omega)|^2$. The change of variable $\xi = s\omega$ thus proves that

$$\hat{b}(\omega) = \frac{1}{C_\psi} \hat{f}(\omega) \int_0^{+\infty} \frac{|\hat{\psi}(\xi)|^2}{\xi}\, d\xi = \hat{f}(\omega).$$

■ ■

The theorem hypothesis

$$C_\psi = \int_0^{+\infty} \frac{|\hat{\psi}(\omega)|^2}{\omega}\, d\omega < +\infty$$

is called the wavelet *admissibility condition.* To guarantee that this integral is finite we must ensure that $\hat{\psi}(0) = 0$, which explains why we imposed that wavelets must have a zero average. This condition is nearly sufficient. If $\hat{\psi}(0) = 0$ and $\hat{\psi}(\omega)$ is continuously differentiable then the admissibility condition is satisfied. One can verify that $\hat{\psi}(\omega)$ is continuously differentiable if ψ has a sufficient time decay

$$\int_{-\infty}^{+\infty} (1 + |t|)\, |\psi(t)|\, dt < +\infty.$$

Reproducing Kernel Like a windowed Fourier transform, a wavelet transform is a redundant representation, whose redundancy is characterized by a reproducing kernel equation. Inserting the reconstruction formula (4.37) into the definition of the wavelet transform yields

$$Wf(u_0,s_0) = \int_{-\infty}^{+\infty} \left(\frac{1}{C_\psi} \int_0^{+\infty} \int_{-\infty}^{+\infty} Wf(u,s)\, \psi_{u,s}(t)\, du \frac{ds}{s^2} \right) \psi^*_{u_0,s_0}(t)\, dt.$$

Interchanging these integrals gives

$$Wf(u_0,s_0) = \frac{1}{C_\psi} \int_{-\infty}^{+\infty} K(u, u_0, s, s_0)\, Wf(u,s)\, du\, \frac{ds}{s^2}, \tag{4.40}$$

with

$$K(u_0, u, s_0, s) = \langle \psi_{u,s}, \psi_{u_0,s_0} \rangle\ . \tag{4.41}$$

The reproducing kernel $K(u_0, u, s_0, s)$ measures the correlation of two wavelets $\psi_{u,s}$ and ψ_{u_0,s_0}. The reader can verify that any function $\Phi(u,s)$ is the wavelet transform of some $f \in \mathbf{L^2}(\mathbb{R})$ if and only if it satisfies the reproducing kernel equation (4.40).

Scaling Function When $Wf(u,s)$ is known only for $s < s_0$, to recover f we need a complement of information corresponding to $Wf(u,s)$ for $s > s_0$. This is obtained by introducing a *scaling function* ϕ that is an aggregation of wavelets at scales larger than 1. The modulus of its Fourier transform is defined by

$$|\hat{\phi}(\omega)|^2 = \int_1^{+\infty} |\hat{\psi}(s\omega)|^2 \, \frac{ds}{s} = \int_\omega^{+\infty} \frac{|\hat{\psi}(\xi)|^2}{\xi} \, d\xi, \tag{4.42}$$

and the complex phase of $\hat{\phi}(\omega)$ can be arbitrarily chosen. One can verify that $\|\phi\| = 1$ and we derive from the admissibility condition (4.36) that

$$\lim_{\omega \to 0} |\hat{\phi}(\omega)|^2 = C_\psi. \tag{4.43}$$

The scaling function can thus be interpreted as the impulse response of a low-pass filter. Let us denote

$$\phi_s(t) = \frac{1}{\sqrt{s}} \, \phi\left(\frac{t}{s}\right) \quad \text{and} \quad \bar{\phi}_s(t) = \phi_s^*(-t).$$

The low-frequency approximation of f at the scale s is

$$Lf(u,s) = \left\langle f(t), \frac{1}{\sqrt{s}} \phi\left(\frac{t-u}{s}\right) \right\rangle = f \star \bar{\phi}_s(u). \tag{4.44}$$

With a minor modification of the proof of Theorem 4.4, it can be shown that (Exercise 4.3)

$$f(t) = \frac{1}{C_\psi} \int_0^{s_0} Wf(.,s) \star \psi_s(t) \, \frac{ds}{s^2} + \frac{1}{C_\psi s_0} Lf(.,s_0) \star \phi_{s_0}(t). \tag{4.45}$$

Example 4.7. *If ψ is the second order derivative of a Gaussian whose Fourier transform is given by (4.35), then the integration (4.42) yields*

$$\hat{\phi}(\omega) = \frac{2\sigma^{3/2}\pi^{1/4}}{\sqrt{3}} \sqrt{\omega^2 + \frac{1}{\sigma^2}} \, \exp\left(-\frac{\sigma^2\omega^2}{2}\right). \tag{4.46}$$

Figure 4.8 displays ϕ and $\hat{\phi}$ for $\sigma = 1$.

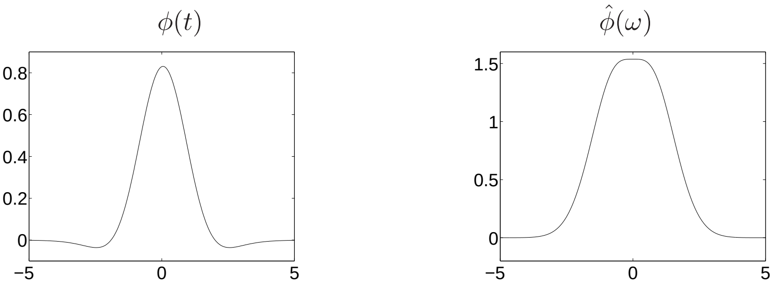


Figure 4.8: Scaling function associated to a Mexican hat wavelet and its Fourier transform calculated with (4.46).

4.3.2 Analytic Wavelets

To analyze the time evolution of frequency tones, it is necessary to use an analytic wavelet to separate the phase and amplitude information of signals. The properties of the resulting analytic wavelet transform are studied.

Analytic Signal A function $f_a \in \mathbf{L^2}(\mathbb{R})$ is said to be *analytic* if its Fourier transform is zero for negative frequencies:

$$\hat{f}_a(\omega) = 0 \ \text{ if } \omega < 0.$$

An analytic function is necessarily complex but is entirely characterized by its real part. Indeed, the Fourier transform of its real part $f = \mathrm{Re}[f_a]$ is

$$\hat{f}(\omega) = \frac{\hat{f}_a(\omega) + \hat{f}_a^*(-\omega)}{2},$$

and this relation can be inverted:

$$\hat{f}_a(\omega) = \begin{cases} 2\,\hat{f}(\omega) & \text{if } \omega \geqslant 0 \\ 0 & \text{if } \omega < 0 \end{cases}. \tag{4.47}$$

The analytic part $f_a(t)$ of a signal $f(t)$ is the inverse Fourier transform of $\hat{f}_a(\omega)$ defined by (4.47).

Discrete Analytic Part The analytic part $f_a[n]$ of a discrete signal $f[n]$ of size N is also computed by setting to zero the negative frequency components of its discrete Fourier transform. The Fourier transform values at $k = 0$ and $k = N/2$ must be carefully adjusted so that $\mathrm{Re}[f_a] = f$ (Exercise 3.4):

$$\hat{f}_a[k] = \begin{cases} \hat{f}[k] & \text{if } k = 0,\, N/2 \\ 2\,\hat{f}[k] & \text{if } 0 < k < N/2 \\ 0 & \text{if } N/2 < k < N \end{cases}. \tag{4.48}$$

We obtain $f_a[n]$ by computing the inverse discrete Fourier transform.

Example 4.8. *The Fourier transform of*

$$f(t) = a\cos(\omega_0 t + \phi) = \frac{a}{2}\Big(\exp[i(\omega_0 t + \phi)] + \exp[-i(\omega_0 t + \phi)]\Big)$$

is

$$\hat{f}(\omega) = \pi a\Big(\exp(i\phi)\,\delta(\omega - \omega_0) + \exp(-i\phi)\,\delta(\omega + \omega_0)\Big).$$

The Fourier transform of the analytic part computed with (4.47) is $\hat{f}_a(\omega) = 2\pi a\ \exp(i\phi)\,\delta(\omega - \omega_0)$ *and hence*

$$f_a(t) = a\ \exp[i(\omega_0 t + \phi)]. \tag{4.49}$$

Time-Frequency Resolution An analytic wavelet transform is calculated with an analytic wavelet ψ:

$$Wf(u,s) = \langle f, \psi_{u,s}\rangle = \int_{-\infty}^{+\infty} f(t)\,\frac{1}{\sqrt{s}}\,\psi^*\left(\frac{t-u}{s}\right)\,dt. \tag{4.50}$$

Its time-frequency resolution depends on the time-frequency spread of the wavelet atoms $\psi_{u,s}$. We suppose that ψ is centered at 0, which implies that $\psi_{u,s}$ is centered at $t = u$. With the change of variable $v = \frac{t-u}{s}$, we verify that

$$\int_{-\infty}^{+\infty} (t-u)^2\,|\psi_{u,s}(t)|^2\,dt = s^2\,\sigma_t^2, \tag{4.51}$$

with $\sigma_t^2 = \int_{-\infty}^{+\infty} t^2\,|\psi(t)|^2\,dt$. Since $\hat{\psi}(\omega)$ is zero at negative frequencies, the center frequency η of $\hat{\psi}$ is

$$\eta = \frac{1}{2\pi}\int_0^{+\infty} \omega\,|\hat{\psi}(\omega)|^2\,d\omega. \tag{4.52}$$

The Fourier transform of $\psi_{u,s}$ is a dilation of $\hat{\psi}$ by $1/s$:

$$\hat{\psi}_{u,s}(\omega) = \sqrt{s}\,\hat{\psi}(s\omega)\,\exp(-i\omega u)\,. \tag{4.53}$$

Its center frequency is therefore η/s. The energy spread of $\hat{\psi}_{u,s}$ around η/s is

$$\frac{1}{2\pi}\int_0^{+\infty}\Big(\omega-\frac{\eta}{s}\Big)^2\left|\hat{\psi}_{u,s}(\omega)\right|^2 d\omega=\frac{\sigma_\omega^2}{s^2}, \tag{4.54}$$

with

$$\sigma_\omega^2=\frac{1}{2\pi}\int_0^{+\infty}(\omega-\eta)^2\,|\hat{\psi}(\omega)|^2\,d\omega.$$

The energy spread of a wavelet time-frequency atom $\psi_{u,s}$ thus corresponds to a Heisenberg box centered at $(u,\eta/s)$, of size $s\sigma_t$ along time and σ_ω/s along frequency. The area of the rectangle remains equal to $\sigma_t\,\sigma_\omega$ at all scales but the resolution in time and frequency depends on s, as illustrated in Figure 4.9.

An analytic wavelet transform defines a local time-frequency energy density $P_W f$, which measures the energy of f in the Heisenberg box of each wavelet $\psi_{u,s}$ centered at $(u,\xi=\eta/s)$:

$$P_W f(u,\xi)=|Wf(u,s)|^2=\left|Wf\Big(u,\frac{\eta}{\xi}\Big)\right|^2. \tag{4.55}$$

This energy density is called a *scalogram.*

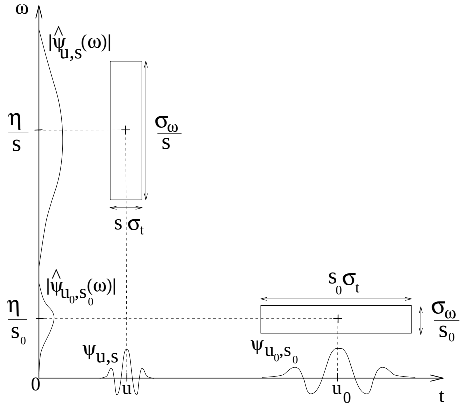


Figure 4.9: Heisenberg boxes of two wavelets. Smaller scales decrease the time spread but increase the frequency support, which is shifted towards higher frequencies.

Completeness An analytic wavelet transform of f depends only on its analytic part f_a. The following theorem derives a reconstruction formula and proves that energy is conserved for real signals.

Theorem 4.5. *For any* $f\in\mathbf{L^2}(\mathbb{R})$

$$Wf(u,s)=\frac{1}{2}\,Wf_a(u,s). \tag{4.56}$$

If $C_\psi=\int_0^{+\infty}\omega^{-1}\,|\hat{\psi}(\omega)|^2\,d\omega<+\infty$ *and* f *is real then*

$$f(t)=\frac{2}{C_\psi}\mathrm{Re}\left[\int_0^{+\infty}\int_{-\infty}^{+\infty}Wf(u,s)\,\psi_s(t-u)\,du\,\frac{ds}{s^2}\right], \tag{4.57}$$

and

$$\|f\|^2=\frac{2}{C_\psi}\int_0^{+\infty}\int_{-\infty}^{+\infty}|Wf(u,s)|^2\,du\,\frac{ds}{s^2}. \tag{4.58}$$

Proof. Let us first prove (4.56). The Fourier transform with respect to u of

$$f_s(u)=Wf(u,s)=f\star\bar{\psi}_s(u)$$

is

$$\hat{f}_s(\omega) = \hat{f}(\omega)\,\sqrt{s}\,\hat{\psi}^*(s\omega).$$

Since $\hat{\psi}(\omega) = 0$ at negative frequencies, and $\hat{f}_a(\omega) = 2\hat{f}(\omega)$ for $\omega \geqslant 0$, we derive that

$$\hat{f}_s(\omega) = \frac{1}{2}\,\hat{f}_a(\omega)\,\sqrt{s}\,\hat{\psi}^*(s\omega),$$

which is the Fourier transform of (4.56).

With the same derivations as in the proof of (4.37) one can verify that the inverse wavelet formula reconstructs the analytic part of f:

$$f_a(t) = \frac{1}{C_\psi}\int_0^{+\infty}\int_{-\infty}^{+\infty} Wf_a(u,s)\,\psi_s(t-u)\,\frac{ds}{s^2}\,du. \tag{4.59}$$

Since $f = \mathrm{Re}[f_a]$, inserting (4.56) proves (4.57).

An energy conservation for the analytic part f_a is proved as in (4.38) by applying the Plancherel formula:

$$\int_{-\infty}^{+\infty} |f_a(t)|^2\,dt = \frac{1}{C_\psi}\int_0^{+\infty}\int_{-\infty}^{+\infty} |W_a f(u,s)|^2\,du\,\frac{ds}{s^2}.$$

Since $Wf_a(u,s) = 2Wf(u,s)$ and $\|f_a\|^2 = 2\|f\|^2$, equation (4.58) follows. ■ ■

If f is real the change of variable $\xi = 1/s$ in the energy conservation (4.58) proves that

$$\|f\|^2 = \frac{2}{C_\psi}\int_0^{+\infty}\int_{-\infty}^{+\infty} P_W f(u,\xi)\,du\,d\xi.$$

It justifies the interpretation of a scalogram as a time-frequency energy density.

Wavelet Modulated Windows An analytic wavelet can be constructed with a frequency modulation of a real and symmetric window g. The Fourier transform of

$$\psi(t) = g(t)\,\exp(i\eta t) \tag{4.60}$$

is $\hat{\psi}(\omega) = \hat{g}(\omega - \eta)$. If $\hat{g}(\omega) = 0$ for $|\omega| > \eta$ then $\hat{\psi}(\omega) = 0$ for $\omega < 0$. Hence ψ is analytic, as shown in Figure 4.10. Since g is real and even, $\hat{g}$ is also real and symmetric. The center frequency of $\hat{\psi}$ is therefore η and

$$|\hat{\psi}(\eta)| = \sup_{\omega\in\mathbb{R}} |\hat{\psi}(\omega)| = \hat{g}(0). \tag{4.61}$$

A *Gabor wavelet* $\psi(t) = g(t)\,\mathrm{e}^{i\eta t}$ is obtained with a Gaussian window

$$g(t) = \frac{1}{(\sigma^2\pi)^{1/4}}\,\exp\left(\frac{-t^2}{2\sigma^2}\right). \tag{4.62}$$

The Fourier transform of this window is $\hat{g}(\omega) = (4\pi\sigma^2)^{1/4}\exp(-\sigma^2\omega^2/2)$. If $\sigma^2\eta^2 \gg 1$ then $\hat{g}(\omega) \approx 0$ for $|\omega| > \eta$. Such Gabor wavelets are thus considered to be approximately analytic.

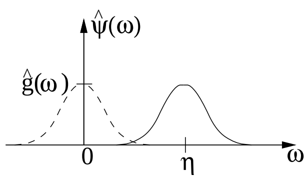


Figure 4.10: Fourier transform $\hat{\psi}(\omega)$ of a wavelet $\psi(t) = g(t)\,\exp(i\eta t)$.

Example 4.9. *The wavelet transform of* $f(t) = a\,\exp(i\omega_0 t)$ *is*

$$Wf(u,s) = a\sqrt{s}\,\hat{\psi}^*(s\omega_0)\,\exp(i\omega_0 t) = a\,\sqrt{s}\,\hat{g}(s\omega_0 - \eta)\,\exp(i\omega_0 t).$$

Observe that the normalized scalogram is maximum at $\xi = \omega_0$*:*

$$\frac{\xi}{\eta} P_W f(u,\xi) = \frac{1}{s}|Wf(u,s)|^2 = a^2\left|\hat{g}\left(\eta\left(\frac{\omega_0}{\xi} - 1\right)\right)\right|^2.$$

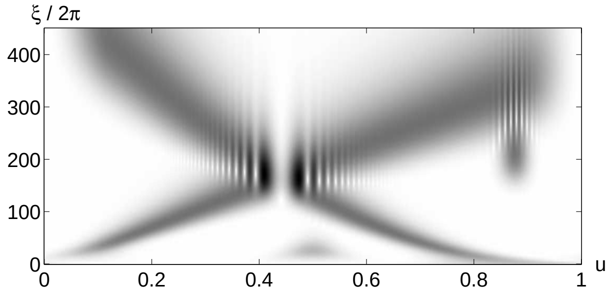
Example 4.10. *The wavelet transform of a linear chirp* $f(t) = \exp(iat^2) = \exp[i\phi(t)]$ *is computed for a Gabor wavelet whose Gaussian window is (4.62). By using the Fourier transform of Gaussian chirps (2.34) one can verify that*

$$\frac{|Wf(u,s)|^2}{s} = \left(\frac{4\pi\sigma^2}{1+4s^2a^2\sigma^4}\right)^{1/2} \exp\left(\frac{-\sigma^2}{1+4a^2s^4\sigma^4}(\eta - 2asu)^2\right).$$

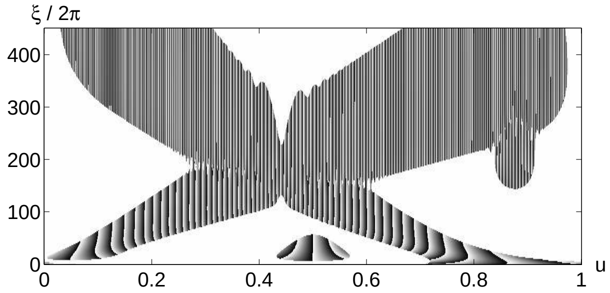
As long as $4a^2s^4\sigma^4 \ll 1$*, at a fixed time u the renormalized scalogram* $\eta^{-1}\xi P_W f(u,\xi)$ *is a Gaussian function of s that reaches its maximum at*

$$\xi(u) = \frac{\eta}{s(u)} = \phi'(u) = 2\,a\,u. \tag{4.63}$$

Section 4.4.2 explains why the amplitude is maximum at the instantaneous frequency $\phi'(u)$*.*



(a)



(b)

Figure 4.11: (a) Normalized scalogram $\eta^{-1}\xi P_W f(u,\xi)$ computed from the signal in Figure 4.3. Dark points indicate large amplitude coefficients. (b) Complex phase $\Theta_W(u,\xi)$ of $Wf(u,\eta/\xi)$, where the modulus is non-zero.

Example 4.11. *Figure 4.11 displays the normalized scalogram* $\eta^{-1}\xi P_W f(u,\xi)$*, and the complex phase* $\Theta_W(u,\xi)$ *of* $Wf(u,s)$*, for the signal f of Figure 4.3. The frequency bandwidth of wavelet atoms is proportional to* $1/s = \xi/\eta$*. The frequency resolution of the scalogram is therefore finer than the spectrogram at low frequencies but coarser than the spectrogram at higher frequencies. This explains why the wavelet transform produces interference patterns between the high frequency Gabor function at the abscissa* $t = 0.87$ *and the quadratic chirp at the same location, whereas the spectrogram in Figure 4.3 separates them well.*

4.3.3 Discrete Wavelets

Let $\bar{f}(t)$ be a continuous time signal defined over $[0, 1]$. Let $f[n]$ be a the discrete signal obtained by a low-pass filtering of $\bar{f}$ and a uniform sampling at intervals N^{-1}. Its discrete wavelet transform can only be calculated at scales $N^{-1} < s < 1$, as shown in Figure 4.7. It is calculated for $s = a^j$, with $a = 2^{1/v}$, which provides v intermediate scales in each octave $[2^j, 2^{j+1})$.

Let $\psi(t)$ be a wavelet whose support is included in $[-K/2, K/2]$. For $1 \leqslant a^j \leqslant N K^{-1}$, a discrete wavelet scaled by a^j is defined by

$$\psi_j[n] = \frac{1}{\sqrt{a^j}}\, \psi\left(\frac{n}{a^j}\right) .$$

This discrete wavelet has Ka^j non-zero values on $[-N/2, N/2]$. The scale a^j is larger than 1 otherwise the sampling interval may be larger than the wavelet support.

Fast Transform To avoid border problems, we treat $f[n]$ and the wavelets $\psi_j[n]$ as periodic signals of period N. The discrete wavelet transform can then be written as a circular convolution with $\bar{\psi}_j[n] = \psi_j^*[-n]$:

$$Wf[n, a^j] = \sum_{m=0}^{N-1} f[m]\, \psi_j^*[m-n] = f \circledast \bar{\psi}_j[n]. \tag{4.64}$$

This circular convolution is calculated with the fast Fourier transform algorithm, which requires $O(N \log_2 N)$ operations. If $a = 2^{1/v}$, there are $v \log_2(N/(2K))$ scales $a^j \in [2N^{-1}, K^{-1}]$. The total number of operations to compute the wavelet transform over all scales is therefore $O(vN(\log_2 N)^2)$ [369].

To compute the scalogram $P_W[n, \xi] = |Wf[n, \frac{\eta}{\xi}]|^2$ we calculate $Wf[n, s]$ at any scale s with a parabola interpolation. Let j be the closest integer to $\log_2 s / \log_2 a$, and $p(x)$ be the parabola such that

$$p(j-1) = Wf[n, a^{j-1}] \quad , \quad p(j) = Wf[n, a^j] \quad , \quad p(j+1) = Wf[n, a^{j+1}].$$

A second order interpolation computes

$$Wf[n, s] = p\left(\frac{\log_2 s}{\log_2 a}\right) .$$

Parabolic interpolations are used instead of linear interpolations in order to locate more precisely the ridges defined in Section 4.4.2.

Discrete Scaling Filter A wavelet transform computed up to a scale a^J is not a complete signal representation. It is necessary to add the low frequencies $Lf[n, a^J]$ corresponding to scales larger than a^J. A discrete and periodic scaling filter is computed by sampling the scaling function $\phi(t)$ defined in (4.42):

$$\phi_J[n] = \frac{1}{\sqrt{a^J}}\, \phi\left(\frac{n}{a^J}\right) \quad \text{for } n \in [-N/2, N/2].$$

Let $\bar{\phi}_J[n] = \phi_J^*[-n]$. The low frequencies are carried by

$$Lf[n, a^J] = \sum_{m=0}^{N-1} f[m]\, \phi_J^*[m-n] = f \circledast \bar{\phi}_J[n]. \tag{4.65}$$

Reconstruction An approximate inverse wavelet transform is implemented by discretizing the integral (4.45). Suppose that $a^I = 1$ is the finest scale. Since $ds/s^2 = d\log_{\mathrm{e}} s/s$ and the discrete wavelet transform is computed along an exponential scale sequence $\{a^j\}_j$ with a logarithmic increment $d\log_{\mathrm{e}} s = \log_{\mathrm{e}} a$, we obtain

$$f[n] \approx \frac{\log_{\mathrm{e}} a}{C_\psi} \sum_{j=I}^{J} \frac{1}{a^j}\, Wf[., a^j] \circledast \psi_j[n] + \frac{1}{C_\psi\, a^J}\, Lf[., a^J] \circledast \phi_J[n]. \tag{4.66}$$

The "." indicates the variable over which the convolution is calculated. These circular convolutions are calculated using the FFT, with $O(vN(\log_2 N)^2)$ operations.

Analytic wavelet transforms are often computed over real signals $f[n]$ that have no energy at low frequencies. The scaling filter component is then negligible. Theorem 4.5 shows that

$$f[n] \approx \frac{2\log_e a}{C_\psi} \operatorname{Re}\left(\sum_{j=I}^{J} \frac{1}{a^j} Wf[.,a^j] \circledast \psi_j[n] \right) . \tag{4.67}$$

The error introduced by the discretization of scales decreases when the number v of voices per octave increases. However, the approximation of continuous time convolutions with discrete convolutions also creates high frequency errors. Perfect reconstructions are obtained with a more careful design of the reconstruction filters (Exercise 4.2). Section 5.2.2 describes an exact inverse wavelet transform computed at dyadic scales $a^j = 2^j$.

4.4 Time-Frequency Geometry of Instantaneous Frequencies

When listening to music, we perceive several frequencies that change with time. In music, it is associated to the geometric perception of "movements". This notion of instantaneous frequency remains to be defined. The time variation of several instantaneous frequencies is measured with local maxima of windowed Fourier transforms and wavelet transforms. They define a geometric time-frequency support from which signal approximations are recovered. Audio processing is implemented by modifying this time-frequency support.

Analytic Instantaneous Frequency A cosine modulation

$$f(t) = a\cos(\omega_0 t + \theta_0) = a\cos\theta(t)$$

has a frequency ω_0 that is the derivative of the phase $\theta(t) = \omega_0 t + \theta_0$. To generalize this notion, real signals f are written as an amplitude $a(t)$ modulated with a time varying phase $\theta(t)$:

$$f(t) = a(t)\,\cos\theta(t) \quad \text{with} \quad a(t) \geqslant 0 . \tag{4.68}$$

The *instantaneous frequency* is defined as a positive derivative of the phase:

$$\omega(t) = \theta'(t) \geqslant 0 .$$

The derivative can be chosen to be positive by adapting the sign of $\theta(t)$. For a given $f(t)$ there are many possible choices of $a(t)$ and $\theta(t)$ satisfying (4.68), so $\omega(t)$ is not uniquely defined relative to f.

A particular decomposition (4.68) is obtained from the analytic part f_a of f, whose Fourier transform is defined in (4.47) by

$$\hat{f}_a(\omega) = \begin{cases} 2\,\hat{f}(\omega) & \text{if } \omega \geqslant 0 \\ 0 & \text{if } \omega < 0 \end{cases} . \tag{4.69}$$

This complex signal is represented by separating the modulus and the complex phase:

$$f_a(t) = a(t)\,\exp[i\theta(t)] . \tag{4.70}$$

Since $f = \operatorname{Re}[f_a]$, it follows that

$$f(t) = a(t)\,\cos\theta(t).$$

We call $a(t)$ the *analytic* amplitude of $f(t)$ and $\theta'(t)$ its *instantaneous frequency*; they are uniquely defined.

Example 4.12. *If* $f(t) = a(t)\,\cos(\omega_0 t + \theta_0)$*, then*

$$\hat{f}(\omega) = \frac{1}{2}\Big(\exp(i\theta_0)\,\hat{a}(\omega - \omega_0) + \exp(-i\theta_0)\,\hat{a}(\omega + \omega_0)\Big)\ .$$

If the variations of $a(t)$ *are slow compared to the period* $2\pi/\omega_0$*, which is achieved by requiring that the support of* $\hat{a}$ *be included in* $[-\omega_0, \omega_0]$*, then*

$$\hat{f}_a(\omega) = \hat{a}(\omega - \omega_0)\,\exp(i\theta_0)$$

so $f_a(t) = a(t)\,\exp[i(\omega_0 t + \theta_0)]$.

If a signal f is the sum of two sinusoidal waves:

$$f(t) = a\cos(\omega_1 t) + a\cos(\omega_2 t),$$

then

$$f_a(t) = a\,\exp(i\omega_1 t) + a\,\exp(i\omega_2 t) = 2a\,\cos\left(\frac{1}{2}\,(\omega_1 - \omega_2)\,t\right)\,\exp\left(\frac{i}{2}\,(\omega_1 + \omega_2)\,t\right).$$

The instantaneous frequency is $\theta'(t) = (\omega_1 + \omega_2)/2$ and the amplitude is

$$a(t) = 2a\,\left|\cos\left(\frac{1}{2}\,(\omega_1 - \omega_2)\,t\right)\right|.$$

This result is not satisfying because it does not reveal that the signal includes two sinusoidal waves of the same amplitude. It measures an average frequency value. Next sections explain how to measure the instantaneous frequencies of several spectral components by separating them with a windowed Fourier transform or a wavelet transform. We first describe two important applications of instantaneous frequencies.

Frequency Modulation In signal communications, information can be transmitted through the amplitude $a(t)$ (amplitude modulation) or the instantaneous frequency $\theta'(t)$ (frequency modulation) [57]. Frequency modulation is more robust in the presence of additive Gaussian white noise. In addition, it better resists multi-path interferences, which destroy the amplitude information. A frequency modulation sends a message $m(t)$ through a signal

$$f(t) = a\cos\theta(t) \quad \text{with} \quad \theta'(t) = \omega_0 + k\,m(t).$$

The frequency bandwidth of f is proportional to k. This constant is adjusted depending on the transmission noise and the available bandwidth. At the reception, the message $m(t)$ is restored with a frequency demodulation that computes the instantaneous frequency $\theta'(t)$ [112].

Additive Sound Models Musical sounds and voiced speech segments can be modeled with sums of sinusoidal *partials*:

$$f(t) = \sum_{k=1}^{K} f_k(t) = \sum_{k=1}^{K} a_k(t)\cos\theta_k(t)\ , \tag{4.71}$$

where a_k and θ'_k are vary slowly [374, 375]. Such decompositions are useful for pattern recognition and for modifying sound properties [308]. Sections 4.4.1 and 4.4.2 explain how to compute a_k and the instantaneous frequency θ'_k and reconstruct signals from these information.

Reducing or increasing the duration of a sound f by a factor α in time is used by radio broadcast to adjust recorded sequences to a precise time schedule. A scaling $f(\alpha t)$ transforms each $\theta_k(t)$ in $\theta_k(\alpha t)$ and hence $\theta'_k(t)$ in $\alpha\theta'_k(t)$. For sound reduction, with $\alpha > 1$ all frequencies are thus increased. To avoid modifying the values of θ'_k and a_k, a new sound is synthesized

$$f_\alpha(t) = \sum_{k=1}^{K} a_k(\alpha\,t)\,\cos\Big(\frac{1}{\alpha}\,\theta_k(\alpha\,t)\Big). \tag{4.72}$$

The partials of f_α at $t = t_0/\alpha$ and the partials of f at $t = t_0$ have the same amplitudes and the same instantaneous frequencies and the properties of these sounds are thus perceived as identical.

A frequency transposition with same duration is calculated by dividing each phase by a constant α in order to shift the sound harmonics:

$$f_\alpha(t) = \sum_{k=1}^{K} b_k(t) \cos\Big(\theta_k(t)/\alpha\Big). \tag{4.73}$$

The instantaneous frequency of each partial is now $\theta'_k(t)/\alpha$. To maintain the sound properties, the amplitudes $b_k(t)$ must be adjusted in order not to modify the global frequency envelop $F(t,\omega)$ of the harmonics:

$$a_k(t) = F\Big(t, \theta'_k(t)\Big) \quad \text{and} \quad b_k(t) = F\Big(t, \theta'_k(t)/\alpha\Big)\ . \tag{4.74}$$

Many types of sounds, musical instruments or speech, are produced by an excitation that propagates across a wave guide. Locally, $F(t,\omega)$ is the transfer function of the wave guide. In speech processing, it is called a *formant*. This transfer function is often approximated with an autoregressive filter of order M, in which case:

$$F(t,\omega) = \frac{C}{\sum_{m=0}^{M-1} c_m \, \mathrm{e}^{-im\omega}}\ . \tag{4.75}$$

The parameters c_m are identified with (4.74) from the a_k and the b_k are then derived with (4.74) and (4.75).

4.4.1 Windowed Fourier Ridges

The spectrogram $P_S f(u,\xi) = |Sf(u,\xi)|^2$ measures the energy of f in a time-frequency neighborhood of (u,ξ). The ridge algorithm computes the signal instantaneous frequencies and amplitudes from the local maxima of $P_S f(u,\xi)$. These local maxima define a geometric support in the time-frequency plane. Modifications of sound durations or frequency transpositions are computed with time or frequency dilations of the ridge support.

Time-frequency ridges were introduced by Delprat, Escudié, Guillemain, Kronland-Martinet, Tchamitchian and Torrésani [191, 62] to analyze musical sounds. Since then it has found applications for a wide range of signals [264, 62] that have time varying frequency tones.

The windowed Fourier transform is computed with a symmetric window $g(t) = g(-t)$ whose support is equal to $[-1/2, 1/2]$. The Fourier transform $\hat{g}$ is a real symmetric function. We suppose that $|\hat{g}(\omega)| \leqslant \hat{g}(0)$ for all $\omega \in \mathbb{R}$, and that $\hat{g}(0) = \int_{-1/2}^{1/2} g(t)\,dt$ is on the order of 1. Table 4.1 gives several examples of such windows. The window g is normalized so that $\|g\| = 1$. For a fixed scale s, $g_s(t) = s^{-1/2} g(t/s)$ has a support of size s and a unit norm. The corresponding windowed Fourier atoms are

$$g_{s,u,\xi}(t) = g_s(t-u)\, \mathrm{e}^{i\xi t},$$

and the resulting windowed Fourier transform is

$$Sf(u,\xi) \ = \ \langle f, g_{s,u,\xi} \rangle \ = \int_{-\infty}^{+\infty} f(t)\, g_s(t-u)\, \mathrm{e}^{-i\xi t}\, dt. \tag{4.76}$$

The following theorem relates $Sf(u,\xi)$ to the instantaneous frequency of f.

Theorem 4.6. *Let $f(t) = a(t) \cos\theta(t)$. If $\xi \geqslant 0$ then*

$$\langle f, g_{s,u,\xi} \rangle = \frac{\sqrt{s}}{2}\, a(u) \exp(i[\theta(u) - \xi u]) \Big(\hat{g}(s[\xi - \theta'(u)]) + \varepsilon(u,\xi)\Big)\ . \tag{4.77}$$

The corrective term satisfies

$$|\varepsilon(u,\xi)| \leqslant \varepsilon_{a,1} + \varepsilon_{a,2} + \varepsilon_{\theta,2} + \sup_{|\omega| \geqslant s\theta'(u)} |\hat{g}(\omega)| \tag{4.78}$$

with

$$\varepsilon_{a,1} \leqslant \frac{s\,|a'(u)|}{|a(u)|} \quad , \quad \varepsilon_{a,2} \leqslant \sup_{|t-u|\leqslant s/2} \frac{s^2\,|a''(t)|}{|a(u)|} \ , \tag{4.79}$$

and if $s\,|a'(u)|\,|a(u)|^{-1} \leqslant 1$, *then*

$$\varepsilon_{\theta,2} \leqslant \sup_{|t-u|\leqslant s/2} s^2|\theta''(t)| \ . \tag{4.80}$$

If $\xi = \theta'(u)$ *then*

$$\varepsilon_{a,1} = \frac{s\,|a'(u)|}{|a(u)|} \left|\hat{g}'\Big(2\,s\,\theta'(u)\Big)\right| \ . \tag{4.81}$$

Proof. Observe that

$$\begin{aligned} \langle f, g_{s,u,\xi}\rangle &= \int_{-\infty}^{+\infty} a(t)\,\cos\theta(t)\,g_s(t-u)\,\exp(-i\xi t)\,dt \\ &= \frac{1}{2}\int_{-\infty}^{+\infty} a(t)\,\left(\exp[i\theta(t)] + \exp[-i\theta(t)]\right)\,g_s(t-u)\,\exp[-i\xi t]\,dt \\ &= I(\theta) + I(-\theta). \end{aligned}$$

We first concentrate on

$$\begin{aligned} I(\theta) &= \frac{1}{2}\int_{-\infty}^{+\infty} a(t)\,\exp[i\theta(t)]\,g_s(t-u)\,\exp(-i\xi t)\,dt \\ &= \frac{1}{2}\int_{-\infty}^{+\infty} a(t+u)\,e^{i\theta(t+u)}\,g_s(t)\,\exp[-i\xi(t+u)]\,dt. \end{aligned}$$

This integral is computed by using second order Taylor expansions:

$$\begin{aligned} a(t+u) &= a(u) + t\,a'(u) + \frac{t^2}{2}\,\alpha(t) \ \text{ with } \ |\alpha(t)| \leqslant \sup_{h\in[u,t+u]} |a''(h)| \\ \theta(t+u) &= \theta(u) + t\,\theta'(u) + \frac{t^2}{2}\,\beta(t) \ \text{ with } \ |\beta(t)| \leqslant \sup_{h\in[u,t+u]} |\theta''(h)| \ . \end{aligned}$$

We get

$$\begin{aligned} &2 \quad \exp\Big(-i(\theta(u)-\xi u)\Big)\,I(\theta) = \\ &\int_{-\infty}^{+\infty} a(u)\,g_s(t)\,\exp\Big(-it(\xi-\theta'(u))\Big)\,\exp\Big(i\frac{t^2}{2}\beta(t)\Big)\,dt \\ &+ \int_{-\infty}^{+\infty} a'(u)\,t\,g_s(t)\,\exp\Big(-it(\xi-\theta'(u))\Big)\,\exp\Big(i\frac{t^2}{2}\beta(t)\Big)\,dt \\ &+ \frac{1}{2}\int_{-\infty}^{+\infty} \alpha(t)\,t^2\,g_s(t)\,\exp\Big(-i(t\xi+\theta(u)-\theta(t+u))\Big)\,dt \ . \end{aligned}$$

A first order Taylor expansion of $\exp(ix)$ gives

$$\exp\Big(i\frac{t^2}{2}\beta(t)\Big) = 1 + \frac{t^2}{2}\beta(t)\,\gamma(t) \quad \text{with} \quad |\gamma(t)| \leqslant 1 \ . \tag{4.82}$$

Since

$$\int_{-\infty}^{+\infty} g_s(t)\,\exp[-it(\xi-\theta'(u))]\,dt = \sqrt{s}\,\hat{g}(s[\xi-\theta'(u)])\,,$$

inserting (4.82) in the expression of $I(\theta)$ yields

$$\left|I(\theta) - \frac{\sqrt{s}}{2}a(u)\,\exp[i(\theta(u)-\xi u)]\,\hat{g}(\xi-\theta'(u))\right| \leqslant \frac{\sqrt{s}\,|a(u)|}{4}(\varepsilon^+_{a,1} + \varepsilon_{a,2} + \varepsilon_{\theta,2}) \tag{4.83}$$

with

$$\varepsilon_{a,1}^{+} = \frac{2|a'(u)|}{|a(u)|}\left|\int_{-\infty}^{+\infty} t\,\frac{1}{\sqrt{s}}\,g_s(t)\,\exp[-it(\xi-\theta'(u))]\,dt\right| , \tag{4.84}$$

$$\varepsilon_{a,2} = \int_{-\infty}^{+\infty} t^2\,|\alpha(t)|\,\frac{1}{\sqrt{s}}\,|g_s(t)|\,dt , \tag{4.85}$$

$$\varepsilon_{\theta,2} = \int_{-\infty}^{+\infty} t^2\,|\beta(t)|\,\frac{1}{\sqrt{s}}\,|g_s(t)|\,dt + \frac{|a'(u)|}{|a(u)|}\int_{-\infty}^{+\infty}|t^3|\,|\beta(t)|\,\frac{1}{\sqrt{s}}\,|g_s(t)|dt. \tag{4.86}$$

Applying (4.83) to $I(-\theta)$ gives

$$|I(-\theta)| \leqslant \frac{\sqrt{s}\,|a(u)|}{2}\,|\hat g(\xi+\theta'(u))| + \frac{\sqrt{s}\,|a(u)|}{4}(\varepsilon_{a,1}^{-}+\varepsilon_{a,2}+\varepsilon_{\theta,2}) ,$$

with

$$\varepsilon_{a,1}^{-} = \frac{2|a'(u)|}{|a(u)|}\left|\int_{-\infty}^{+\infty} t\,\frac{1}{\sqrt{s}}\,g_s(t)\,\exp[-it(\xi+\theta'(u))]\,dt\right| . \tag{4.87}$$

Since $\xi \geqslant 0$ and $\theta'(u) \geqslant 0$, we derive that

$$|\hat g(s[\xi+\theta'(u)])| \leqslant \sup_{|\omega|\geqslant s\theta'(u)} |\hat g(\omega)| ,$$

and hence

$$I(\theta)+I(-\theta) = \frac{\sqrt{s}}{2}\,a(u)\,\exp[i(\theta(u)-\xi u)]\Big(\hat g\big(s[\xi-\theta'(u)]\big)+\varepsilon(u,\xi)\Big)$$

with

$$\varepsilon(u,\xi) = \frac{\varepsilon_{a,1}^{+}+\varepsilon_{a,1}^{-}}{2} + \varepsilon_{a,2} + \varepsilon_{\theta,2} + \sup_{|\omega|\geqslant s|\theta'(u)|} |\hat g(\omega)| .$$

Let us now verify the upper bound (4.79) for $\varepsilon_{a,1} = (\varepsilon_{a,1}^{+}+\varepsilon_{a,1}^{-})/2$. Since $g_s(t) = s^{-1/2}g(t/s)$, a simple calculation shows that for $n \geqslant 0$

$$\int_{-\infty}^{+\infty} |t|^n\,\frac{1}{\sqrt{s}}\,|g_s(t)|\,dt = s^n\int_{-1/2}^{1/2} |t|^n\,|g(t)|\,dt \leqslant \frac{s^n}{2^n}\,\|g\|^2 = \frac{s^n}{2^n}. \tag{4.88}$$

Inserting this for $n=1$ in (4.84) and (4.87) gives

$$\varepsilon_{a,1} = \frac{\varepsilon_{a,1}^{+}+\varepsilon_{a,1}^{-}}{2} \leqslant \frac{s\,|a'(u)|}{|a(u)|} .$$

The upper bounds (4.79) and (4.80) of the second order terms $\varepsilon_{a,2}$ and $\varepsilon_{\theta,2}$ are obtained by observing that the remainder $\alpha(t)$ and $\beta(t)$ of the Taylor expansion of $a(t+u)$ and $\theta(t+u)$ satisfy

$$\sup_{|t|\leqslant s/2} |\alpha(t)| \leqslant \sup_{|t-u|\leqslant s/2} |a''(t)| \ , \quad \sup_{|t|\leqslant s/2} |\beta(t)| \leqslant \sup_{|t-u|\leqslant s/2} |\theta''(t)|. \tag{4.89}$$

Inserting this in (4.85) yields

$$\varepsilon_{a,2} \leqslant \sup_{|t-u|\leqslant s/2} \frac{s^2\,|a''(t)|}{|a(u)|} .$$

When $s\,|a'(u)||a(u)|^{-1} \leqslant 1$, replacing $|\beta(t)|$ by its upper bound in (4.86) gives

$$\varepsilon_{\theta,2} \leqslant \frac{1}{2}\Big(1+\frac{s\,|a'(u)|}{|a(u)|}\Big) \sup_{|t-u|\leqslant s/2} s^2|\theta''(t)| \leqslant \sup_{|t-u|\leqslant s/2} s^2|\theta''(t)|.$$

Let us finally compute ε_a when $\xi = \theta'(u)$. Since $g(t) = g(-t)$, we derive from (4.84) that

$$\varepsilon_{a,1}^{+} = \frac{2|a'(u)|}{|a(u)|}\left|\int_{-\infty}^{+\infty} t\,\frac{1}{\sqrt{s}}\,g_s(t)\,dt\right| = 0 .$$

We also derive from (2.22) that the Fourier transform of $t\frac{1}{\sqrt{s}}g_s(t)$ is $i\,s\,\hat g'(s\omega)$, so (4.87) gives

$$\varepsilon_a = \frac{1}{2}\varepsilon_{a,1}^{-} = \frac{s|a'(u)|}{|a(u)|}\,|\hat g'(2s\theta'(u))| .$$

■ ■

Delprat et al. [191] give a different proof of a similar result when $g(t)$ is a Gaussian, using a stationary phase approximation. If we can neglect the corrective term $\varepsilon(u,\xi)$ we shall see that (4.77) enables us to measure $a(u)$ and $\theta'(u)$ from $Sf(u,\xi)$. This implies that the decomposition $f(t)=a(t)\cos\theta(t)$ is uniquely defined. By reviewing the proof of Theorem 4.6, one can verify that a and θ' are the analytic amplitude and instantaneous frequencies of f.

The expressions (4.79, 4.80) show that the three corrective terms $\varepsilon_{a,1}$, $\varepsilon_{a,2}$ and $\varepsilon_{\theta,2}$ are small if $a(t)$ and $\theta'(t)$ have small relative variations over the support of the window g_s. Let $\Delta\omega$ be the bandwidth of $\hat{g}$ defined by

$$|\hat{g}(\omega)| \ll 1 \quad \text{for } |\omega| \geqslant \Delta\omega. \tag{4.90}$$

The term $\sup_{|\omega|\geqslant s|\theta'(u)|} |\hat{g}(\omega)|$ of $\varepsilon(u,\xi)$ is negligible if

$$\theta'(u) \geqslant \frac{\Delta\omega}{s} \ .$$

Ridge Points Let us suppose that $a(t)$ and $\theta'(t)$ have small variations over intervals of size s and that $\theta'(t) \geqslant \Delta\omega/s$ so that the corrective term $\varepsilon(u,\xi)$ in (4.77) can be neglected. Since $|\hat{g}(\omega)|$ is maximum at $\omega=0$, (4.77) shows that for each u the spectrogram $|Sf(u,\xi)|^2 = |\langle f, g_{s,u,\xi}\rangle|^2$ is maximum at $\xi(u)=\theta'(u)$. The corresponding time-frequency points $(u,\xi(u))$ are called *ridges*. At ridge points, (4.77) becomes

$$Sf(u,\xi) = \frac{\sqrt{s}}{2}\, a(u)\, \exp(i[\theta(u)-\xi u]) \Big(\hat{g}(0)+\varepsilon(u,\xi)\Big) \ . \tag{4.91}$$

Theorem 4.6 proves that the $\varepsilon(u,\xi)$ is smaller at a ridge point because the first order term $\varepsilon_{a,1}$ becomes negligible in (4.81). This is shown by verifying that $|\hat{g}'(2s\theta'(u))|$ is negligible when $s\theta'(u) \geqslant \Delta\omega$. At ridge points, the second order terms $\varepsilon_{a,2}$ and $\varepsilon_{\theta,2}$ are predominant in $\varepsilon(u,\xi)$.

The ridge frequency gives the instantaneous frequency $\xi(u)=\theta'(u)$ and the amplitude is calculated by

$$a(u) = \frac{2\,\big|Sf\big(u,\xi(u)\big)\big|}{\sqrt{s}\,|\hat{g}(0)|}. \tag{4.92}$$

Let $\Theta_S(u,\xi)$ be the complex phase of $Sf(u,\xi)$. If we neglect the corrective term, then (4.91) proves that ridges are also points of stationary phase:

$$\frac{\partial \Theta_S(u,\xi)}{\partial u} = \theta'(u) - \xi = 0.$$

Testing the stationarity of the phase locates the ridges more precisely.

Multiple Frequencies When the signal contains several spectral lines whose frequencies are sufficiently apart, the windowed Fourier transform separates each of these components and the ridges detect the evolution in time of each spectral component. Let us consider

$$f(t) = a_1(t)\cos\theta_1(t) + a_2(t)\cos\theta_2(t),$$

where $a_k(t)$ and $\theta'_k(t)$ have small variations over intervals of size s and $s\theta'_k(t) \geqslant \Delta\omega$. Since the windowed Fourier transform is linear, we apply (4.77) to each spectral component and neglect the corrective terms:

$$\begin{aligned} Sf(u,\xi) \ &= \ \frac{\sqrt{s}}{2}\, a_1(u)\, \hat{g}(s[\xi-\theta'_1(u)])\, \exp(i[\theta_1(u)-\xi u]) \\ &\quad + \frac{\sqrt{s}}{2}\, a_2(u)\, \hat{g}(s[\xi-\theta'_2(u)])\, \exp(i[\theta_2(u)-\xi u]) \ . \end{aligned} \tag{4.93}$$

The two spectral components are discriminated if for all u

$$\hat{g}(s|\theta'_1(u)-\theta'_2(u)|) \ll 1, \tag{4.94}$$

which means that the frequency difference is larger than the bandwidth of $\hat{g}(s\omega)$:

$$|\theta_1'(u) - \theta_2'(u)| \geqslant \frac{\Delta\omega}{s}. \tag{4.95}$$

In this case, when $\xi = \theta_1'(u)$, the second term of (4.93) can be neglected and the first term generates a ridge point from which we may recover $\theta_1'(u)$ and $a_1(u)$, using (4.92). Similarly, if $\xi = \theta_2'(u)$ the first term can be neglected and we have a second ridge point that characterizes $\theta_2'(u)$ and $a_2(u)$. The ridge points are distributed along two time-frequency lines $\xi(u) = \theta_1'(u)$ and $\xi(u) = \theta_2'(u)$. This result is valid for any number of time varying spectral components, as long as the distance between any two instantaneous frequencies satisfies (4.95). If two spectral lines are too close, they interfere, which destroys the ridge pattern.

Time-Frequency Ridge Support The number of instantaneous frequencies is typically unknown. The ridge support Λ is thus defined as the set of all (u,ξ) which are local maxima of $|Sf(u,\xi)|^2$ for u fixed and ξ varying and points of stationary phase $\partial\Theta_S(u,\xi)/\partial u \approx 0$. This support is often reduced by removing small ridge amplitudes $|Sf(u,\xi)|$ that are mostly dominated by the noise, or because smaller ridges may be "shadows" of other instantaneous frequencies created by the side-lobes of $\hat{g}(\omega)$.

Let $\{g_{s,u,\xi}\}_{(u,\xi)\in\Lambda}$ be the set of ridge atoms. For discrete signals, there is a finite number of ridge points, which thus define a frame of the space $\mathbf{V}_\Lambda$ they generate. A ridge signal approximation is computed as an orthogonal projection of f on $\mathbf{V}_\Lambda$. Section 5.1.3 shows that it is obtained with the dual frame $\{\tilde{g}_{\Lambda,u,\xi}\}_{(u,\xi)\in\Lambda}$ of $\{g_{s,u,\xi}\}_{(u,\xi)\in\Lambda}$ in $\mathbf{V}_\Lambda$:

$$f_\Lambda = \sum_{(u,\xi)\in\Lambda} Sf(u,\xi)\,\tilde{g}_{\Lambda,u,\xi} \tag{4.96}$$

The dual synthesis algorithm of Section 5.1.3 computes this orthogonal projection by inverting the symmetric operator

$$Lh = \sum_{(u,\xi)\in\Lambda} \langle h, g_{s,u,\xi}\rangle\, g_{s,u,\xi}\ . \tag{4.97}$$

The inversion requires to iterate on this operator many times. If there are few ridge points then (4.97) is efficiently computed by evaluating the inner product and the sum only for $(u,\xi)\in\Lambda$. If there many ridge points, it can be more efficient to compute the full windowed Fourier transform $Sh(u,\xi) = \langle h, g_{s,u,\xi}\rangle$ with the FFT algorithm described in Section 4.2.3, set to zero all coefficients for $(u,\xi)\notin\Lambda$, and apply the fast inverse windowed Fourier transform over all coefficients. The normalization factor N^{-1} in (4.28) must be removed (set to 1) to implement (4.97).

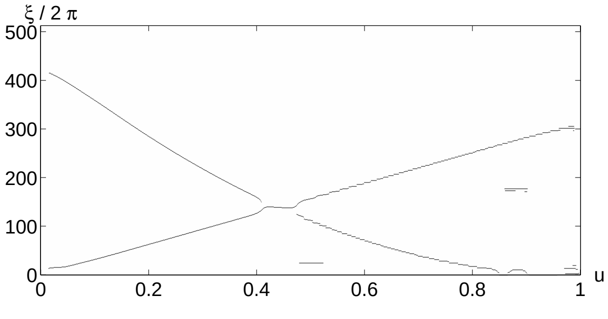


Figure 4.12: Support of larger amplitude ridges calculated from the spectrogram in Figure 4.3. These ridges give the instantaneous frequencies of the linear and quadratic chirps, and of the low and high frequency transients at $t = 0.5$ and $t = 0.87$.

Figure 4.12 displays the ridge support computed from the modulus and phase of the windowed Fourier transform shown in Figure 4.3. For $t \in [0.4, 0.5]$, the instantaneous frequencies of the linear chirp and the quadratic chirps are close and the frequency resolution of the window is not sufficient to discriminate them. As a result, the ridges detect a single average instantaneous frequency.

Time Scaling and Frequency Transpositions A reduction of sound duration by a factor α is implemented according to the deformation model (4.72), by dilating in time the ridge support Λ:

$$\Lambda_\alpha = \{(u,\xi) \ : \ (\alpha u, \xi) \in \Lambda\} \ . \tag{4.98}$$

The windowed-Fourier coefficients $c(u,\xi)$ in Λ_α are derived from the modulus and phase of ridge coefficients

$$\forall (v,\xi) \in \Lambda_\alpha \ \ , \ \ c(v,\xi) = |Sf(\alpha v, \xi)| \, \mathrm{e}^{i\Theta_S(\alpha v,\xi)/\alpha} \ . \tag{4.99}$$

The scaled signal is reconstructed from these coefficients, with the dual synthesis algorithm of Section 5.1.3, as in (4.96):

$$f_\alpha = \sum_{(v,\xi)\in\Lambda_\alpha} c(v,\xi) \, \tilde{g}_{\Lambda_\alpha,v,\xi} \ .$$

Similarly, a sound transposition is implemented according to the transposition model (4.73) by dilating in frequency the ridge support Λ:

$$\Lambda_\alpha = \{(u,\xi) \ : \ (u, \alpha\xi) \in \Lambda\} \ . \tag{4.100}$$

The transposed coefficient amplitudes $|c(u,\xi)|$ in Λ_α are calculated with (4.74). At any fixed time u_0, the ridge amplitudes at all frequencies $\{a(u_0,\xi) = |Sf(u_0,\xi)\}_{(u_0,\xi)\in\Lambda}$ are mapped to a transposed amplitudes $\{b(u_0,\eta)\}_{(u_0,\eta)\in\Lambda_\alpha}$ at frequencies $\eta = \xi/\alpha$, by computing a frequency envelop. The resulting ridge coefficients are

$$\forall (u,\eta) \in \Lambda_\alpha \ \ , \ \ c(u,\eta) = b(u,\eta) \, \mathrm{e}^{i\Theta_S(u,\alpha\eta)/\alpha} \ . \tag{4.101}$$

The transposed signal is reconstructed with the dual synthesis algorithm of Section 5.1.3:

$$f_\alpha = \sum_{(u,\eta)\in\Lambda_\alpha} c(u,\eta) \, \tilde{g}_{\Lambda_\alpha,u,\eta} \ .$$

Choice of Window The measurement of instantaneous frequencies at ridge points is valid only if the size s of the window g_s is sufficiently small so that the second order terms $\varepsilon_{a,2}$ and $\varepsilon_{\theta,2}$ in (4.79) and (4.80) are small:

$$\sup_{|t-u|\leqslant s/2} \frac{s^2\,|a_k''(t)|}{|a_k(u)|} \ll 1 \quad \text{and} \quad \sup_{|t-u|\leqslant s/2} s^2 |\theta_k''(t)| \ll 1 \ . \tag{4.102}$$

On the other hand, the frequency bandwidth $\Delta\omega/s$ must also be sufficiently small to discriminate consecutive spectral components in (4.95). The window scale s must therefore be adjusted as a trade-off between both constraints.

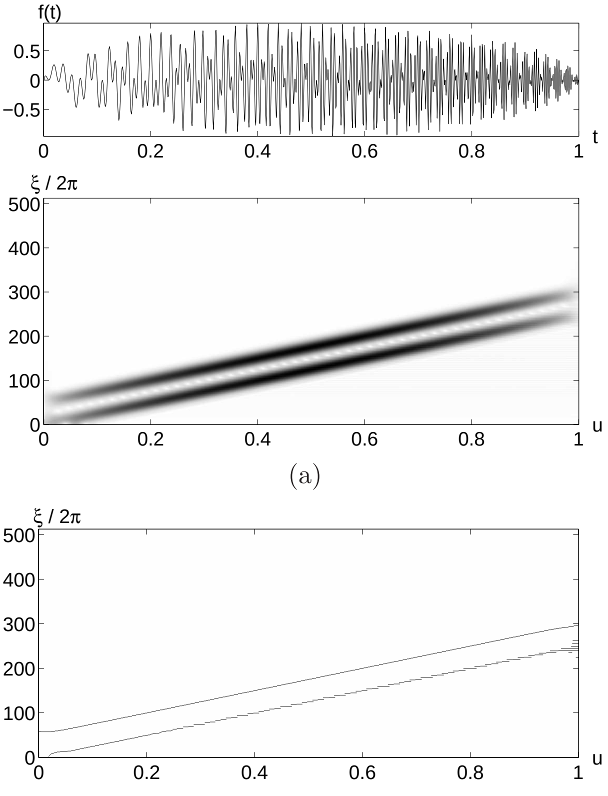
Table 4.1 gives the spectral parameters of several windows of compact support. For instantaneous frequency detection, it is particularly important to ensure that $\hat{g}$ has negligible side-lobes at $\pm\omega_0$, as illustrated by Figure 4.4. The reader can verify with (4.77) that these side-lobes "react" to an instantaneous frequency $\theta'(u)$ by creating shadow maxima of $|Sf(u,\xi)|^2$ at frequencies $\xi = \theta'(u) \pm \omega_0$. The ratio of the amplitude of these shadow maxima to the amplitude of the main local maxima at $\xi = \theta'(u)$ is $|\hat{g}(\omega_0)|^2 \, |\hat{g}(0)|^{-2}$. They can be removed by thresholding or by testing the stationarity of the phase.

Example 4.13. *The sum of two parallel linear chirps*

$$f(t) = a_1 \cos(bt^2 + ct) + a_2 \cos(bt^2) \tag{4.103}$$

has two instantaneous frequencies $\theta_1'(t) = 2bt + c$ *and* $\theta_2'(t) = 2bt$. *Figure 4.13 gives a numerical example. The window* g_s *has enough frequency resolution to discriminate both chirps if*

$$|\theta_1'(t) - \theta_2'(t)| = |c| \geqslant \frac{\Delta\omega}{s}. \tag{4.104}$$



(a)

(b)

Figure 4.13: Sum of two parallel linear chirps. (a): Spectrogram $P_S f(u, \xi) = |Sf(u, \xi)|^2$. (b): Ridge support calculated from the spectrogram.

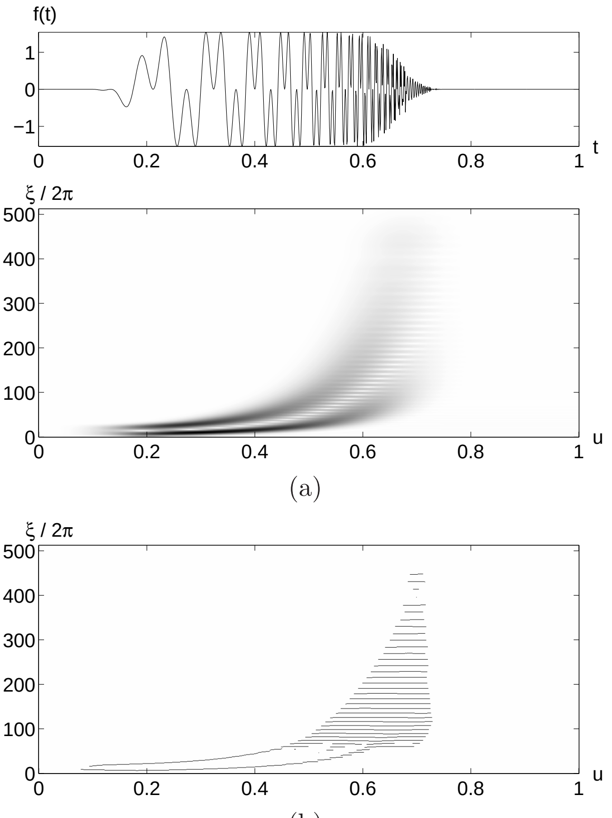
Its time support is small enough compared to their time variation if

$$s^2\,|\theta_1''(u)| = s^2\,|\theta_2''(u)| = 2\,b\,s^2 \ll 1. \tag{4.105}$$

Conditions (4.104) and (4.105) prove that there exists an appropriate window g *if and only if*

$$\frac{c}{\sqrt{b}} \gg \Delta\omega. \tag{4.106}$$

Since g *is a smooth window with a support* $[-1/2, 1/2]$*, its frequency bandwidth* $\Delta\omega$ *is on the order of* 1*. The linear chirps in Figure 4.13 satisfy (4.106). Their ridges are computed with the truncated Gaussian window of Table 4.1, with* $s = 0.5$*.*



(a)

(b)

Figure 4.14: Sum of two hyperbolic chirps. (a): Spectrogram $P_S f(u,\xi)$. (b): Ridge support calculated from the spectrogram

Example 4.14. *The hyperbolic chirp*

$$f(t) = \cos\left(\frac{\alpha}{\beta - t}\right)$$

for $0 \leqslant t < \beta$ *has an instantaneous frequency*

$$\theta'(t) = \frac{\alpha}{(\beta - t)^2},$$

which varies quickly when t *is close to* β*. The instantaneous frequency of hyperbolic chirps goes from* 0 *to* $+\infty$ *in a finite time interval. This is particularly useful for radars. These chirps are also emitted by the cruise sonars of bats [191].*

The instantaneous frequency of hyperbolic chirps cannot be estimated with a windowed Fourier transform because for any fixed window size the instantaneous frequency varies too quickly at high frequencies. When u is close enough to β then (4.102) is not satisfied because

$$s^2|\theta''(u)| = \frac{s^2\alpha}{(\beta-u)^3} > 1.$$

Figure 4.14 shows a signal that is a sum of two hyperbolic chirps:

$$f(t) = a_1 \cos\left(\frac{\alpha_1}{\beta_1 - t}\right) + a_2 \cos\left(\frac{\alpha_2}{\beta_2 - t}\right), \tag{4.107}$$

with $\beta_1 = 0.68$ and $\beta_2 = 0.72$. At the beginning of the signal, the two chirps have close instantaneous frequencies that are discriminated by the windowed Fourier ridge computed with a large size window. When getting close to β_1 and β_2, the instantaneous frequency varies too quickly relative to the window size. The resulting ridges cannot follow these instantaneous frequencies.

4.4.2 Wavelet Ridges

Windowed Fourier atoms have a fixed scale and thus cannot follow the instantaneous frequency of rapidly varying events such as hyperbolic chirps. In contrast, an analytic wavelet transform modifies the scale of its time-frequency atoms. The ridge algorithm of Delprat et al. [191] is extended to analytic wavelet transforms to accurately measure frequency tones that are rapidly changing at high frequencies.

An approximately analytic wavelet is constructed in (4.60) by multiplying a window g with a sinusoidal wave:

$$\psi(t) = g(t)\,\exp(i\eta t)\,.$$

As in the previous section, g is a symmetric window with a support equal to $[-1/2, 1/2]$, and a unit norm $\|g\| = 1$. Let $\Delta\omega$ be the bandwidth of $\hat{g}$ defined in (4.90). If $\eta > \Delta\omega$ then

$$\forall \omega < 0 \quad , \quad \hat{\psi}(\omega) = \hat{g}(\omega - \eta) \ll 1\ .$$

The wavelet ψ is not strictly analytic because its Fourier transform is not exactly equal to zero at negative frequencies.

Dilated and translated wavelets can be rewritten

$$\psi_{u,s}(t) = \frac{1}{\sqrt{s}}\,\psi\left(\frac{t-u}{s}\right) = g_{s,u,\xi}(t)\exp(-i\xi u)\,,$$

with $\xi = \eta/s$ and

$$g_{s,u,\xi}(t) = \sqrt{s}\,g\left(\frac{t-u}{s}\right)\,\exp(i\xi t).$$

The resulting wavelet transform uses time-frequency atoms similar to those of a windowed Fourier transform (4.76) but in this case the scale s varies over $\mathbb{R}^+$ while $\xi = \eta/s$:

$$Wf(u,s) = \langle f, \psi_{u,s}\rangle = \langle f, g_{s,u,\xi}\rangle \exp(i\xi u)\,.$$

Theorem 4.6 computes $\langle f, g_{s,u,\xi}\rangle$ when $f(t) = a(t)\cos\theta(t)$, which gives

$$Wf(u,s) = \frac{\sqrt{s}}{2}\,a(u)\,\exp[i\theta(u)]\Big(\hat{g}(s[\xi - \theta'(u)]) + \varepsilon(u,\xi)\Big). \tag{4.108}$$

The corrective term $\varepsilon(u,\xi)$ is negligible if $a(t)$ and $\theta'(t)$ have small variations over the support of $\psi_{u,s}$ and if $\theta'(u) \geqslant \Delta\omega/s$.

Ridge Detection The instantaneous frequency is measured from ridges defined over the wavelet transform. The normalized scalogram defined by

$$\frac{\xi}{\eta}\, P_W f(u,\xi) = \frac{|Wf(u,s)|^2}{s} \quad \text{for } \xi = \eta/s$$

is calculated with (4.108):

$$\frac{\xi}{\eta}\, P_W f(u,\xi) = \frac{1}{4}\, a^2(u) \left| \hat{g}\Big(\eta\Big[1 - \frac{\theta'(u)}{\xi}\Big]\Big) + \varepsilon(u,\xi) \right|^2 .$$

Since $|\hat{g}(\omega)|$ is maximum at $\omega = 0$, if we neglect $\varepsilon(u,\xi)$, this expression shows that the scalogram is maximum at

$$\frac{\eta}{s(u)} = \xi(u) = \theta'(u) \ . \tag{4.109}$$

The corresponding points $(u, \xi(u))$ are called *wavelet ridges*. The analytic amplitude is given by

$$a(u) = \frac{2\sqrt{\eta^{-1}\xi\, P_W f(u,\xi)}}{|\hat{g}(0)|}. \tag{4.110}$$

The complex phase of $Wf(u,s)$ in (4.108) is $\Theta_W(u,\xi) = \theta(u)$. At ridge points,

$$\frac{\partial \Theta_W(u,\xi)}{\partial u} = \theta'(u) = \xi. \tag{4.111}$$

When $\xi = \theta'(u)$, the first order term $\varepsilon_{a,1}$ calculated in (4.81) becomes negligible. The corrective term is then dominated by $\varepsilon_{a,2}$ and $\varepsilon_{\theta,2}$. To simplify their expression, we approximate the sup of a'' and θ'' in the neighborhood of u by their value at u. Since $s = \eta/\xi = \eta/\theta'(u)$, (4.79) and (4.80) imply that these second order terms become negligible if

$$\frac{\eta^2}{|\theta'(u)|^2}\, \frac{|a''(u)|}{|a(u)|} \ll 1 \quad \text{and} \quad \eta^2\, \frac{|\theta''(u)|}{|\theta'(u)|^2} \ll 1. \tag{4.112}$$

The presence of θ' in the denominator proves that a' and θ' must have slow variations if θ' is small but may vary much more quickly for large instantaneous frequencies.

Multispectral Estimation Suppose that f is a sum of two spectral components:

$$f(t) = a_1(t)\cos\theta_1(t) + a_2(t)\cos\theta_2(t).$$

As in (4.94), we verify that the second instantaneous frequency θ_2' does not interfere with the ridge of θ_1' if the dilated window has a sufficient spectral resolution at the ridge scale $s = \eta/\xi = \eta/\theta_1'(u)$:

$$\hat{g}(s|\theta_1'(u) - \theta_2'(u)|) \ll 1. \tag{4.113}$$

Since the bandwidth of $\hat{g}(\omega)$ is $\Delta\omega$, this means that

$$\frac{|\theta_1'(u) - \theta_2'(u)|}{\theta_1'(u)} \geqslant \frac{\Delta\omega}{\eta}. \tag{4.114}$$

Similarly, the first spectral component does not interfere with the second ridge located at $s = \eta/\xi = \eta/\theta_2'(u)$ if

$$\frac{|\theta_1'(u) - \theta_2'(u)|}{\theta_2'(u)} \geqslant \frac{\Delta\omega}{\eta}. \tag{4.115}$$

To separate spectral lines whose instantaneous frequencies are close, these conditions prove that the wavelet must have a small octave bandwidth $\Delta\omega/\eta$. The bandwidth $\Delta\omega$ is a fixed constant, which is on the order of 1. The frequency η is a free parameter whose value is chosen as a trade-off between the time-resolution condition (4.112) and the frequency bandwidth conditions (4.114) and (4.115).

Figure 4.15 displays the ridges computed from the normalized scalogram and the wavelet phase shown in Figure 4.11. The ridges of the high frequency transient located at $t = 0.87$ have oscillations because of the interferences with the linear chirp above. The frequency separation condition (4.114) is not satisfied. This is also the case in the time interval $[0.35, 0.55]$, where the instantaneous frequencies of the linear and quadratic chirps are too close.

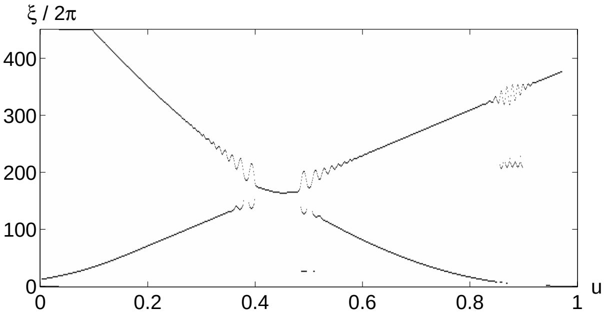
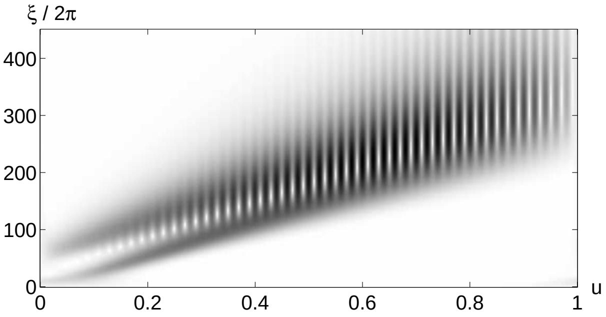
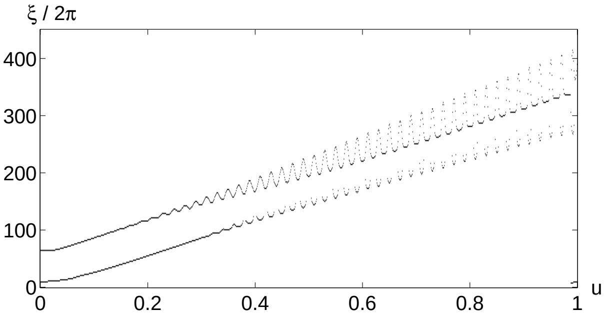


Figure 4.15: Ridge support calculated from the scalogram shown in Figure 4.11. Compare with the windowed Fourier ridges in Figure 4.12.



(a)



(b)

Figure 4.16: (a): Normalized scalogram $\eta^{-1}\xi P_W f(u,\xi)$ of two parallel linear chirps shown in Figure 4.13. (b): Wavelet ridges.

Ridge Support and Processing The wavelet ridge support Λ of f is the set of all ridge points (u, s) in the time-scale plane or $(u, \xi = \eta/s)$ in the time-frequency plane, corresponding to local maxima of $|Wf(u, s)|/s$ for a fixed u and s varying, where the complex phase $\Theta_W(u, s)$ nearly satisfies (4.111).

As in the windowed Fourier case, an orthogonal projection is computed over the space $\mathbf{V}_\Lambda$ generated by the ridge wavelets $\{\psi_{u,s}\}_{(u,s)\in\Lambda}$, by using the dual wavelet frame $\{\tilde{\psi}_{\Lambda,u,s}\}_{(u,s)\in\Lambda}$:

$$f_\Lambda = \sum_{(u,s)\in\Lambda} Wf(u, s)\, \tilde{\psi}_{\Lambda,u,s}\ , \tag{4.116}$$

It is implemented with the dual synthesis algorithm of Section 5.1.3, by inverting the symmetric operator

$$Lh = \sum_{(u,s)\in\Lambda} \langle h, \psi_{u,s}\rangle\, \psi_{u,s}\ , \tag{4.117}$$

which is performed by computing this operator many times. When there are many ridge points, instead of computing this sum only for $(u, s) \in \Lambda$, it may require less operations to compute $Wf(u, s)$ with the fast wavelet transform algorithm of Section 4.3.3. All coefficients $(u, s) \notin \Lambda$ are set to zero, and the fast inverse wavelet transform algorithm is applied. The inverse wavelet transform formula (4.66) must be modified by removing the renormalization factor a^{-j} and a^{-J} in the sum (set them to 1), to implement the operator (4.117).

Like in the windowed Fourier case, modifications of the sound durations or frequency transpositions are computed with by modifying the ridge support. A reduction of sound duration by a factor α transforms the ridge support Λ into:

$$\Lambda_\alpha = \{(u, s)\ :\ (\alpha u, s) \in \Lambda\}\ . \tag{4.118}$$

A sound transposition is implemented by modifying the scales of the time-scale ridge support Λ, which defines:

$$\Lambda_\alpha = \{(u, s)\ :\ (u, s/\alpha) \in \Lambda\}\ . \tag{4.119}$$

The wavelet coefficients over these supports are derived from the deformation model (4.72) or (4.74), similarly to (4.99) and (4.101) for the windowed Fourier transform. Processed signal are recovered from the modified wavelet coefficients and modified supports with the dual synthesis algorithm of Section 5.1.3.

Example 4.15. *The instantaneous frequencies of two linear chirps*

$$f(t) = a_1 \cos(b\,t^2 + c\,t) + a_2 \cos(b\,t^2)$$

are not well measured by wavelet ridges. Indeed

$$\frac{|\theta_2'(u) - \theta_1'(u)|}{\theta_1'(u)} = \frac{c}{b\,t}$$

converges to zero when t increases. When it is smaller than $\Delta\omega/\eta$ the two chirps interact and create interference patterns like those in Figure 4.16. The ridges follow these interferences and do not estimate properly the two instantaneous frequencies, as opposed to the windowed Fourier ridges shown in Figure 4.13.

Example 4.16. *The instantaneous frequency of a hyperbolic chirp*

$$f(t) = \cos\left(\frac{\alpha}{\beta - t}\right)$$

is $\theta'(t) = \alpha\,(1 - t)^{-2}$. Wavelet ridges can measure this instantaneous frequency if the time resolution condition (4.112) is satisfied:

$$\eta^2 \ll \frac{\theta'(t)^2}{|\theta''(t)|} = \frac{\alpha}{|t - \beta|}.$$

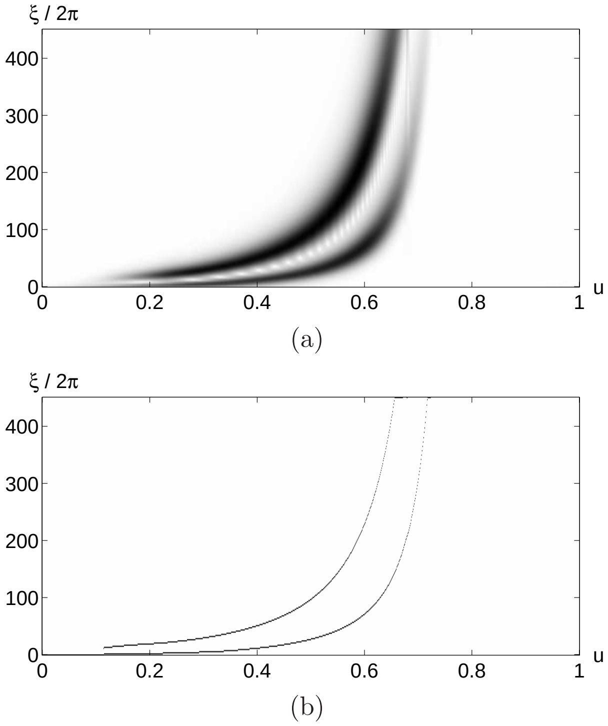


Figure 4.17: (a): Normalized scalogram $\eta^{-1}\xi P_W f(u,\xi)$ of two hyperbolic chirps shown in Figure 4.14. (b): Wavelet ridges.

This is the case if $|t-\beta|$ *is not too large.*

Figure 4.17 displays the scalogram and the ridges of two hyperbolic chirps

$$f(t) = a_1 \cos\left(\frac{\alpha_1}{\beta_1 - t}\right) + a_2 \cos\left(\frac{\alpha_2}{\beta_2 - t}\right),$$

with $\beta_1 = 0.68$ *and* $\beta_2 = 0.72$*. As opposed to the windowed Fourier ridges shown in Figure 4.14, the wavelet ridges follow the rapid time modification of both instantaneous frequencies. This is particularly useful in analyzing the returns of hyperbolic chirps emitted by radars or sonars. Several techniques have been developed to detect chirps with wavelet ridges in presence of noise [143, 416].*

Better is More Sparse The linear and hyperbolic chirp examples show that the best transform depends upon the signal time-frequency property. All examples also show that when the time-frequency transform has a resolution adapted to the signal time-frequency properties then the number of ridge points is reduced. Indeed, if signal structures do not match the dictionary time-frequency atoms then their energy is diffused over many more atoms, which produces more local maxima. Sparsity thus appears as a natural criteria to adjust the resolution of time-frequency transforms.

Section 12.3.3 studies sparse time-frequency decompositions in very redundant Gabor time-frequency dictionaries, including windowed Fourier atoms and wavelet atoms, with a computationally more intensive matching pursuit algorithm.

4.5 Quadratic Time-Frequency Energy

The wavelet and windowed Fourier transforms are computed by correlating the signal with families of time-frequency atoms. The time and frequency resolution of these transforms is thus limited by the time-frequency resolution of the corresponding atoms. Ideally, one would like to define a density of energy in a time-frequency plane, with no loss of resolution.

The Wigner-Ville distribution is a time-frequency energy density computed by correlating f with a time and frequency translation of itself. Despite its remarkable properties, the application of Wigner-Ville distributions is limited by the existence of interference terms. These interferences can be attenuated by a time-frequency averaging, but this results in a loss of resolution. It is proved that the spectrogram, the scalogram and all squared time-frequency decompositions can be written as a time-frequency averaging of the Wigner-Ville distribution, which gives a common framework to relate these transforms.

4.5.1 Wigner-Ville Distribution

To analyze time-frequency structures, in 1948 Ville [436] introduced in signal processing a quadratic form that had been studied by Wigner [443] in a 1932 article on quantum thermodynamics:

$$P_V f(u,\xi) = \int_{-\infty}^{+\infty} f\left(u+\frac{\tau}{2}\right) f^*\left(u-\frac{\tau}{2}\right) \mathrm{e}^{-i\tau\xi}\, d\tau. \tag{4.120}$$

The Wigner-Ville distribution remains real because it is the Fourier transform of $f(u+\tau/2)f^*(u-\tau/2)$, which has a Hermitian symmetry in τ. Time and frequency have a symmetrical role. This distribution can also be rewritten as a frequency integration by applying the Parseval formula:

$$P_V f(u,\xi) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \hat{f}\left(\xi+\frac{\gamma}{2}\right) \hat{f}^*\left(\xi-\frac{\gamma}{2}\right) \mathrm{e}^{i\gamma u}\, d\gamma. \tag{4.121}$$

Time-Frequency Support The Wigner-Ville transform localizes the time-frequency structures of f. If the energy of f is well concentrated in time around u_0 and in frequency around ξ_0 then $P_V f$ has its energy centered at (u_0,ξ_0), with a spread equal to the time and frequency spread of f. This property is illustrated by the following theorem, which relates the time and frequency support of $P_V f$ to the support of f and $\hat{f}$.

Theorem 4.7.
- *If the support of f is $[u_0-T/2, u_0+T/2]$, then for all ξ the support in u of $P_V f(u,\xi)$ is included in this interval.*
- *If the support of $\hat{f}$ is $[\xi_0-\Delta/2, \xi_0+\Delta/2]$, then for all u the support in ξ of $P_V f(u,\xi)$ is included in this interval.*

Proof. Let $\bar{f}(t) = f(-t)$. The Wigner-Ville distribution is rewritten

$$P_V f(u,\xi) = \int_{-\infty}^{+\infty} f\left(\frac{\tau+2u}{2}\right) \bar{f}^*\left(\frac{\tau-2u}{2}\right) \mathrm{e}^{-i\xi\tau}\, d\tau. \tag{4.122}$$

Suppose that f has a support equal to $[u_0-T/2, u_0+T/2]$. The supports of $f(\tau/2+u)$ and $\bar{f}(\tau/2-u)$ are then respectively

$$[2(u_0-u)-T, 2(u_0-u)+T] \quad\text{and}\quad [-2(u_0+u)-T, -2(u_0+u)+T].$$

The Wigner-Ville integral (4.122) shows that $P_V f(u,\xi)$ is non-zero if these two intervals overlap, which is the case only if $|u_0-u|<T$. The support of $P_V f(u,\xi)$ along u is therefore included in the support of f. If the support of $\hat{f}$ is an interval, then the same derivation based on (4.121) shows that the support of $P_V f(u,\xi)$ along ξ is included in the support of $\hat{f}$. ■ ■

Example 4.17. *Theorem 4.7 proves that the Wigner-Ville distribution does not spread the time or frequency support of Diracs or sinusoids, unlike windowed Fourier and wavelet transforms. Direct calculations yield*

$$f(t) = \delta(u-u_0) \quad\Longrightarrow\quad P_V f(u,\xi) = \delta(u-u_0)\ , \tag{4.123}$$

$$f(t) = \exp(i\xi_0 t) \quad\Longrightarrow\quad P_V f(u,\xi) = \frac{1}{2\pi}\,\delta(\xi-\xi_0)\ . \tag{4.124}$$

Example 4.18. *If f is a smooth and symmetric window then its Wigner-Ville distribution $P_V f(u,\xi)$ is concentrated in a neighborhood of $u = \xi = 0$. A Gaussian $f(t) = (\sigma^2\pi)^{-1/4}\exp(-t^2/(2\sigma^2))$ is transformed into a two-dimensional Gaussian because its Fourier transform is also a Gaussian (2.32) and one can verify that*

$$P_V f(u,\xi) = \frac{1}{\pi}\exp\left(\frac{-u^2}{\sigma^2} - \sigma^2\xi^2\right). \tag{4.125}$$

In this particular case $P_V f(u,\xi) = |f(u)|^2|\hat{f}(\xi)|^2$.

The Wigner-Ville distribution has important invariance properties. A phase shift does not modify its value:

$$f(t) = \mathrm{e}^{i\theta}\, g(t) \implies P_V f(u,\xi) = P_V g(u,\xi)\ . \tag{4.126}$$

When f is translated in time or frequency, its Wigner-Ville transform is also translated:

$$f(t) = g(t-u_0) \quad\implies\quad P_V f(u,\xi) = P_V g(u-u_0,\xi)\ , \tag{4.127}$$

$$f(t) = \exp(i\xi_0 t)g(t) \quad\implies\quad P_V f(u,\xi) = P_V g(u,\xi-\xi_0)\ . \tag{4.128}$$

If f is scaled by s and thus $\hat{f}$ is scaled by $1/s$ then the time and frequency parameters of $P_V f$ are also scaled respectively by s and $1/s$

$$f(t) = \frac{1}{\sqrt{s}}\, g\Big(\frac{t}{s}\Big) \implies P_V f(u,\xi) = P_V f\Big(\frac{u}{s}, s\xi\Big)\ . \tag{4.129}$$

Example 4.19. *If g is a smooth and symmetric window then $P_V g(u,\xi)$ has its energy concentrated in the neighborhood of $(0,0)$. The time-frequency atom*

$$f_0(t) = \frac{a}{\sqrt{s}}\exp(i\theta_0)\, g\Big(\frac{t-u_0}{s}\Big)\exp(i\xi_0 t)\quad .$$

has a Wigner-Ville distribution that is calculated with (4.126), (4.127) and (4.128):

$$P_V f_0(u,\xi) = |a|^2\, P_V g\Big(\frac{u-u_0}{s}, s(\xi-\xi_0)\Big)\quad . \tag{4.130}$$

Its energy is thus concentrated in the neighborhood of (u_0,ξ_0), on an ellipse whose axes are proportional to s in time and $1/s$ in frequency.

Instantaneous Frequency Ville's original motivation for studying time-frequency decompositions was to compute the instantaneous frequency of a signal [436]. Let f_a be the analytic part of f obtained in (4.69) by setting to zero $\hat{f}(\omega)$ for $\omega < 0$. We write $f_a(t) = a(t)\exp[i\theta(t)]$ to define the instantaneous frequency $\omega(t) = \theta'(t)$. The following theorem proves that $\theta'(t)$ is the "average" frequency computed relative to the Wigner-Ville distribution $P_V f_a$.

Theorem 4.8. *If $f_a(t) = a(t)\exp[i\theta(t)]$ then*

$$\theta'(u) = \frac{\int_{-\infty}^{+\infty}\xi\, P_V f_a(u,\xi)\, d\xi}{\int_{-\infty}^{+\infty} P_V f_a(u,\xi)\, d\xi}. \tag{4.131}$$

Proof. To prove this result, we verify that any function g satisfies

$$\iint \xi\, g\Big(u+\frac{\tau}{2}\Big)\, g^*\Big(u-\frac{\tau}{2}\Big)\exp(-i\tau\xi)\, d\xi\, d\tau = -\pi i\Big[g'(u)\, g^*(u) - g(u)\, g^{*\prime}(u)\Big]. \tag{4.132}$$

This identity is obtained by observing that the Fourier transform of $i\xi$ is the derivative of a Dirac, which gives an equality in the sense of distributions:

$$\int_{-\infty}^{+\infty}\xi\exp(-i\tau\xi)\, d\xi = -i\, 2\pi\, \delta'(\tau).$$

Since $\int_{-\infty}^{+\infty}\delta'(\tau)\,h(\tau)d\tau = -h'(0)$, inserting $h(\tau) = g(u+\tau/2)\,g^*(u-\tau/2)$ proves (4.132). If $g(u) = f_a(u) = a(u)\,\exp[i\theta(u)]$ then (4.132) gives

$$\int_{-\infty}^{+\infty}\xi\,P_V f_a(u,\xi)\,d\xi = 2\pi\,a^2(u)\,\theta'(u).$$

We will see in (4.136) that $|f_a(u)|^2 = (2\pi)^{-1}\int_{-\infty}^{+\infty}P_V f_a(u,\xi)\,d\xi$, and since $|f_a(u)|^2 = a(u)^2$ we derive (4.131). ■ ■

This theorem shows that for a fixed u the mass of $P_V f_a(u,\xi)$ is typically concentrated in the neighborhood of the instantaneous frequency $\xi = \theta'(u)$. For example, a linear chirp $f(t) = \exp(iat^2)$ is transformed into a Dirac located along the instantaneous frequency $\xi = \theta'(u) = 2au$:

$$P_V f(u,\xi) = \delta(\xi - 2au).$$

Similarly, the multiplication of f by a linear chirp $\exp(iat^2)$ makes a frequency translation of $P_V f$ by the instantaneous frequency $2au$:

$$f(t) = \exp(iat^2)\,g(t) \implies P_V f(u,\xi) = P_V g(u,\xi - 2au)\ . \tag{4.133}$$

Energy Density The Moyal [346] formula proves that the Wigner-Ville transform is unitary, which implies energy conservation properties.

Theorem 4.9 (Moyal). *For any f and g in $\mathbf{L^2}(\mathbb{R})$*

$$\left|\int_{-\infty}^{+\infty} f(t)\,g^*(t)\,dt\right|^2 = \frac{1}{2\pi}\iint P_V f(u,\xi)\,P_V g(u,\xi)\,du\,d\xi. \tag{4.134}$$

Proof. Let us compute the integral

$$\begin{aligned} I &= \iint P_V f(u,\xi)\,P_V g(u,\xi)\,du\,d\xi \\ &= \iiiint f\Big(u+\frac{\tau}{2}\Big)f^*\Big(u-\frac{\tau}{2}\Big)g\Big(u+\frac{\tau'}{2}\Big)g^*\Big(u-\frac{\tau'}{2}\Big) \\ &\qquad\qquad \exp[-i\xi(\tau+\tau')]\,d\tau\,d\tau'\,du\,d\xi. \end{aligned}$$

The Fourier transform of $h(t) = 1$ is $\hat{h}(\omega) = 2\pi\delta(\omega)$, which means that we have a distribution equality $\int\exp[-i\xi(\tau+\tau')]d\xi = 2\pi\delta(\tau+\tau')$. As a result,

$$\begin{aligned} I &= 2\pi\iiint f\Big(u+\frac{\tau}{2}\Big)f^*\Big(u-\frac{\tau}{2}\Big)g\Big(u+\frac{\tau'}{2}\Big)g^*\Big(u-\frac{\tau'}{2}\Big)\delta(\tau+\tau')\,d\tau\,d\tau'\,du \\ &= 2\pi\iint f\Big(u+\frac{\tau}{2}\Big)f^*\Big(u-\frac{\tau}{2}\Big)g\Big(u-\frac{\tau}{2}\Big)g^*\Big(u+\frac{\tau}{2}\Big)\,d\tau\,du. \end{aligned}$$

The change of variable $t = u+\tau/2$ and $t' = u-\tau/2$ yields (4.134). ■ ■

One can consider $|f(t)|^2$ and $|\hat{f}(\omega)|^2/(2\pi)$ as energy densities in time and frequency that satisfy a conservation equation:

$$\|f\|^2 = \int_{-\infty}^{+\infty}|f(t)|^2\,dt = \frac{1}{2\pi}\int_{-\infty}^{+\infty}|\hat{f}(\omega)|^2\,d\omega.$$

The following theorem shows that these time and frequency densities are recovered with marginal integrals over the Wigner-Ville distribution.

Theorem 4.10. *For any $f \in \mathbf{L^2}(\mathbb{R})$*

$$\int_{-\infty}^{+\infty} P_V f(u,\xi)\,du = |\hat{f}(\xi)|^2, \tag{4.135}$$

and

$$\frac{1}{2\pi}\int_{-\infty}^{+\infty} P_V f(u,\xi)\,d\xi = |f(u)|^2. \tag{4.136}$$

Proof. The frequency integral (4.121) proves that the one-dimensional Fourier transform of $g_\xi(u) = P_V f(u,\xi)$ with respect to u is

$$\hat{g}_\xi(\gamma) = \hat{f}\left(\xi + \frac{\gamma}{2}\right) \hat{f}^*\left(\xi - \frac{\gamma}{2}\right).$$

We derive (4.135) from the fact that is

$$\hat{g}_\xi(0) = \int_{-\infty}^{+\infty} g_\xi(u)\, du.$$

Similarly, (4.120) shows that $P_V f(u,\xi)$ is the one-dimensional Fourier transform of $f(u+\tau/2) f^*(u-\tau/2)$ with respect to τ, where ξ is the Fourier variable. Its integral in ξ thus gives the value for $\tau = 0$, which is the identity (4.136). ■ ■

This theorem suggests interpreting the Wigner-Ville distribution as a joint time-frequency energy density. However, the Wigner-Ville distribution misses one fundamental property of an energy density: positivity. Let us compute for example the Wigner-Ville distribution of $f = \mathbf{1}_{[-T,T]}$ with the integral (4.120):

$$P_V f(u,\xi) = \frac{2\sin\Big(2(T-|u|)\xi\Big)}{\xi}\, \mathbf{1}_{[-T,T]}(u).$$

It is an oscillating function that takes negative values. In fact, one can prove that translated and frequency modulated Gaussians are the only functions whose Wigner-Ville distributions remain positive. As we will see in the next section, to obtain positive energy distributions for all signals, it is necessary to average the Wigner-Ville transform and thus lose some time-frequency resolution.

4.5.2 Interferences and Positivity

At this point, the Wigner-Ville distribution may seem to be an ideal tool for analyzing the time-frequency structures of a signal. This is however not the case because of interferences created by the quadratic properties of this transform. These interferences can be removed by averaging the Wigner-Ville distribution with appropriate kernels which yield positive time-frequency densities. However, this reduces the time-frequency resolution. Spectrograms and scalograms are examples of positive quadratic distributions obtained by smoothing the Wigner-Ville distribution.

Cross Terms Let $f = f_1 + f_2$ be a composite signal. Since the Wigner-Ville distribution is a quadratic form,

$$P_V f = P_V f_1 + P_V f_2 + P_V[f_1, f_2] + P_V[f_2, f_1], \tag{4.137}$$

where $P_V[h,g]$ is the cross Wigner-Ville distribution of two signals

$$P_V[h,g](u,\xi) = \int_{-\infty}^{+\infty} h\left(u + \frac{\tau}{2}\right) g^*\left(u - \frac{\tau}{2}\right) \mathrm{e}^{-i\tau\xi}\, d\tau. \tag{4.138}$$

The interference term

$$I[f_1, f_2] = P_V[f_1, f_2] + P_V[f_2, f_1]$$

is a real function that creates non-zero values at unexpected locations of the (u,ξ) plane.

Let us consider two time-frequency atoms defined by

$$f_1(t) = a_1\, \mathrm{e}^{i\theta_1}\, g(t-u_1)\, \mathrm{e}^{i\xi_1 t} \quad \text{and} \quad f_2(t) = a_2\, \mathrm{e}^{i\theta_2}\, g(t-u_2)\, \mathrm{e}^{i\xi_2 t},$$

where g is a time window centered at $t = 0$. Their Wigner-Ville distributions computed in (4.130) are

$$P_V f_1(u,\xi) = a_1^2 P_V g(u-u_1, \xi-\xi_1) \quad \text{and} \quad P_V f_2(u,\xi) = a_2^2 P_V g(u-u_2, \xi-\xi_2).$$

Since the energy of $P_V g$ is centered at $(0,0)$, the energy of $P_V f_1$ and $P_V f_2$ is concentrated in the neighborhoods of (u_1,ξ_1) and (u_2,ξ_2) respectively. A direct calculation verifies that the interference term is

$$I[f_1, f_2](u,\xi) = 2a_1 a_2\, P_V g(u-u_0, \xi-\xi_0) \cos\Big((u-u_0)\Delta\xi - (\xi-\xi_0)\Delta u + \Delta\theta\Big)$$

with

$$u_0 = \frac{u_1+u_2}{2} \quad , \quad \xi_0 = \frac{\xi_1+\xi_2}{2}$$
$$\Delta u = u_1 - u_2 \quad , \quad \Delta\xi = \xi_1 - \xi_2$$
$$\Delta\theta = \theta_1 - \theta_2 + u_0\,\Delta\xi.$$

It is an oscillatory waveform centered at the middle point (u_0, ξ_0). This is quite counter-intuitive since f and $\hat{f}$ have very little energy in the neighborhood of u_0 and ξ_0. The frequency of the oscillations is proportional to the Euclidean distance $\sqrt{\Delta\xi^2 + \Delta u^2}$ of (u_1, ξ_1) and (u_2, ξ_2). The direction of these oscillations is perpendicular to the line that joins (u_1, ξ_1) and (u_2, ξ_2). Figure 4.18 displays the Wigner-Ville distribution of two atoms obtained with a Gaussian window g. The oscillating interference appears at the middle time-frequency point.

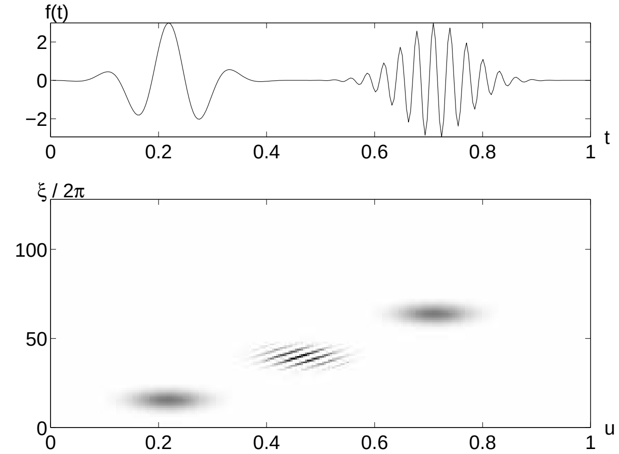


Figure 4.18: Wigner-Ville distribution $P_V f(u, \xi)$ of two Gabor atoms shown at the top. The oscillating interferences are centered at the middle time-frequency location.

This example shows that the interference $I[f_1, f_2](u, \xi)$ has some energy in regions where $|f(u)|^2 \approx 0$ and $|\hat{f}(\xi)|^2 \approx 0$. These interferences can have a complicated structure [24, 276] but they are necessarily oscillatory because the marginal integrals (4.135) and (4.136) vanish:

$$\int_{-\infty}^{+\infty} P_V f(u,\xi)d\xi = 2\pi |f(u)|^2 \quad , \quad \int_{-\infty}^{+\infty} P_V f(u,\xi)du = |\hat{f}(\xi)|^2.$$

Analytic Part Interference terms also exist in a real signal f with a single instantaneous frequency component. Let $f_a(t) = a(t)\,\exp[i\theta(t)]$ be its analytic part:

$$f = \mathrm{Re}[f_a] = \frac{1}{2}\,(f_a + f_a^*).$$

Theorem 4.8 proves that for fixed u, $P_V f_a(u, \xi)$ and $P_V f_a^*(u, \xi)$ have an energy concentrated respectively in the neighborhood of $\xi_1 = \theta'(u)$ and $\xi_2 = -\theta'(u)$. Both components create an interference term at the intermediate zero frequency $\xi_0 = (\xi_1 + \xi_2)/2 = 0$. To avoid this low frequency interference, we often compute $P_V f_a$ as opposed to $P_V f$.

Figure 4.19 displays $P_V f_a$ for a real signal f that includes a linear chirp, a quadratic chirp and two isolated time-frequency atoms. The linear and quadratic chirps are localized along narrow time frequency lines, which are spread on wider bands by the scalogram and the scalogram shown in Figure 4.3 and 4.11. However, the interference terms create complex oscillatory patterns that make it difficult to detect the existence of the two time-frequency transients at $t = 0.5$ and $t = 0.87$, which clearly appear in the spectrogram and the scalogram.

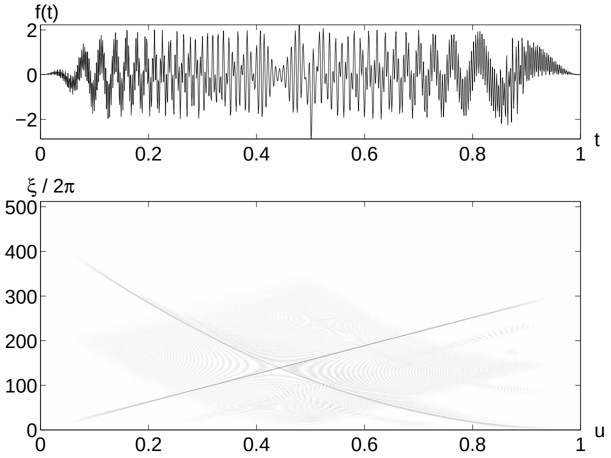


Figure 4.19: The bottom displays the Wigner-Ville distribution $P_V f_a(u,\xi)$ of the analytic part of the top signal.

Positivity Since the interference terms include positive and negative oscillations, they can be partly removed by smoothing $P_V f$ with a kernel K:

$$P_K f(u,\xi) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} P_V f(u',\xi')\, K(u,u',\xi,\xi')\, du'\, d\xi'. \tag{4.139}$$

The time-frequency resolution of this distribution depends on the spread of the kernel K in the neighborhood of (u,ξ). Since the interferences take negative values, one can guarantee that all interferences are removed by imposing that this time-frequency distribution remain positive $P_K f(u,\xi) \geqslant 0$ for all $(u,\xi) \in \mathbb{R}^2$.

The spectrogram (4.12) and scalogram (4.55) are examples of positive time-frequency energy distributions. In general, let us consider a family of time-frequency atoms $\{\phi_\gamma\}_{\gamma\in\Gamma}$. Suppose that for any (u,ξ) there exists a unique atom $\phi_{\gamma(u,\xi)}$ centered in time-frequency at (u,ξ). The resulting time-frequency energy density is

$$Pf(u,\xi) = |\langle f, \phi_{\gamma(u,\xi)}\rangle|^2.$$

The Moyal formula (4.134) proves that this energy density can be written as a time-frequency averaging of the Wigner-Ville distribution

$$Pf(u,\xi) = \frac{1}{2\pi}\iint P_V f(u',\xi')\; P_V \phi_{\gamma(u,\xi)}(u',\xi')\; du'\, d\xi'. \tag{4.140}$$

The smoothing kernel is the Wigner-Ville distribution of the atoms

$$K(u,u',\xi,\xi') = \frac{1}{2\pi}\, P_V \phi_{\gamma(u,\xi)}(u',\xi').$$

The loss of time-frequency resolution depends on the spread of the distribution $P_V \phi_{\gamma(u,\xi)}(u',\xi')$ in the neighborhood of (u,v).

Example 4.20. *A spectrogram is computed with windowed Fourier atoms*

$$\phi_{\gamma(u,\xi)}(t) = g(t-u)\, \mathrm{e}^{i\xi t}.$$

The Wigner-Ville distribution calculated in (4.130) yields

$$K(u,u',\xi,\xi') = \frac{1}{2\pi}\, P_V \phi_{\gamma(u,\xi)}(u',\xi') = \frac{1}{2\pi}\, P_V g(u'-u, \xi'-\xi). \tag{4.141}$$

For a spectrogram, the Wigner-Ville averaging (4.140) is therefore a two-dimensional convolution with $P_V g$. If g is a Gaussian window, then $P_V g$ is a two-dimensional Gaussian. This proves that *averaging $P_V f$ with a sufficiently wide Gaussian defines a positive energy density. The general class of time-frequency distributions obtained by convolving $P_V f$ with an arbitrary kernel K is studied in Section 4.5.3.*

Example 4.21. *Let ψ be an analytic wavelet whose center frequency is η. The wavelet atom $\psi_{u,s}(t) = s^{-1/2}\psi((t-u)/s)$ is centered at $(u, \xi = \eta/s)$ and the scalogram is defined by*

$$P_W f(u,\xi) = |\langle f, \psi_{u,s}\rangle|^2 \quad \text{for } \xi = \eta/s.$$

Properties (4.127,4.129) prove that the averaging kernel is

$$K(u,u',\xi,\xi') = \frac{1}{2\pi}\, P_V\psi\left(\frac{u'-u}{s}, s\xi'\right) = \frac{1}{2\pi}\, P_V\psi\left(\frac{\xi}{\eta}(u'-u), \frac{\eta}{\xi}\xi'\right).$$

Positive time-frequency distributions totally remove the interference terms but produce a loss of resolution. This is emphasized by the following theorem, due to Wigner [444].

Theorem 4.11 (Wigner)**.** *There is no positive quadratic energy distribution Pf that satisfies*

$$\int_{-\infty}^{+\infty} Pf(u,\xi)\, d\xi = 2\pi\, |f(u)|^2 \quad \text{and} \quad \int_{-\infty}^{+\infty} Pf(u,\xi)\, du = |\hat{f}(\xi)|^2. \tag{4.142}$$

Proof. Suppose that Pf is a positive quadratic distribution that satisfies these marginals. Since $Pf(u,\xi) \geqslant 0$, the integrals (4.142) imply that if the support of f is included in an interval I then $Pf(u,\xi) = 0$ for $u \notin I$. We can associate to the quadratic form Pf a bilinear distribution defined for any f and g by

$$P[f,g] = \frac{1}{4}\Big(P(f+g) - P(f-g)\Big).$$

Let f_1 and f_2 be two non-zero signals whose supports are two intervals I_1 and I_2 that do not intersect, so that $f_1\, f_2 = 0$. Let $f = a\, f_1 + b\, f_2$:

$$Pf = |a|^2\, Pf_1 + ab^*\, P[f_1,f_2] + a^* b\, P[f_2,f_1] + |b|^2\, Pf_2.$$

Since I_1 does not intersect I_2, $Pf_1(u,\xi) = 0$ for $u \in I_2$. Remember that $Pf(u,\xi) \geqslant 0$ for all a and b so necessarily $P[f_1,f_2](u,\xi) = P[f_2,f_1](u,\xi) = 0$ for $u \in I_2$. Similarly we prove that these cross terms are zero for $u \in I_1$ and hence

$$Pf(u,\xi) = |a|^2\, Pf_1(u,\xi) + |b|^2\, Pf_2(u,\xi).$$

Integrating this equation and inserting (4.142) yields

$$|\hat{f}(\xi)|^2 = |a|^2\, |\hat{f}_1(\xi)|^2 + |b|^2\, |\hat{f}_2(\xi)|^2.$$

Since $\hat{f}(\xi) = a\, \hat{f}_1(\xi) + b\, \hat{f}_2(\xi)$ it follows that $\hat{f}_1(\xi)\, \hat{f}_2^*(\xi) = 0$. But this is not possible because f_1 and f_2 have a compact support in time and Theorem 2.7 proves that $\hat{f}_1$ and $\hat{f}_2$ are $\mathbf{C}^\infty$ functions that cannot vanish on a whole interval. We thus conclude that one cannot construct a positive quadratic distribution Pf that satisfies the marginals (4.142). ■ ■

4.5.3 Cohen's Class

While attenuating the interference terms with a smoothing kernel K, we may want to retain certain important properties. Cohen [166] introduced a general class of quadratic time-frequency distributions that satisfy the time translation and frequency modulation invariance properties (4.127) and (4.128). If a signal is translated in time or frequency, its energy distribution is just translated by the corresponding amount. This was the beginning of a systematic study of quadratic time-frequency distributions obtained as a weighted average of a Wigner-Ville distribution [6, 24, 167, 275].

Section 2.1 proves that linear translation invariant operators are convolution products. The translation invariance properties (4.127,4.128) are thus equivalent to imposing that the smoothing kernel in (4.139) be a convolution kernel

$$K(u,u',\xi,\xi') = K(u-u', \xi-\xi'), \tag{4.143}$$

and hence

$$P_K f(u,\xi) = P_V f \star K(u,\xi) = \iint K(u-u',\xi-\xi')\, P_V f(u',\xi')\, du'\, d\xi'. \tag{4.144}$$

The spectrogram is an example of Cohen's class distribution, whose kernel in (4.141) is the Wigner-Ville distribution of the window

$$K(u,\xi) = \frac{1}{2\pi}\, P_V g(u,\xi) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} g\left(u+\frac{\tau}{2}\right) g\left(u-\frac{\tau}{2}\right) \mathrm{e}^{-i\tau\xi}\, d\tau. \tag{4.145}$$

Ambiguity Function The properties of the convolution (4.144) are more easily studied by calculating the two-dimensional Fourier transform of $P_V f(u,\xi)$ with respect to u and ξ. We denote by $Af(\tau,\gamma)$ this Fourier transform

$$Af(\tau,\gamma) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} P_V f(u,\xi)\, \exp[-i(u\gamma+\xi\tau)]\, du\, d\xi.$$

Note that the Fourier variables τ and γ are inverted with respect to the usual Fourier notation. Since the one-dimensional Fourier transform of $P_V f(u,\xi)$ with respect to u is $\hat{f}(\xi+\gamma/2)\, \hat{f}^*(\xi-\gamma/2)$, applying the one-dimensional Fourier transform with respect to ξ gives

$$Af(\tau,\gamma) = \int_{-\infty}^{+\infty} \hat{f}\left(\xi+\frac{\gamma}{2}\right) \hat{f}^*\left(\xi-\frac{\gamma}{2}\right) \mathrm{e}^{-i\tau\xi}\, d\xi. \tag{4.146}$$

The Parseval formula yields

$$Af(\tau,\gamma) = \int_{-\infty}^{+\infty} f\left(u+\frac{\tau}{2}\right) f^*\left(u-\frac{\tau}{2}\right) \mathrm{e}^{-i\gamma u}\, du. \tag{4.147}$$

We recognize the *ambiguity function* encountered in (4.24) when studying the time-frequency resolution of a windowed Fourier transform. It measures the energy concentration of f in time and in frequency.

Kernel Properties The Fourier transform of $K(u,\xi)$ is

$$\hat{K}(\tau,\gamma) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} K(u,\xi)\, \exp[-i(u\gamma+\xi\tau)]\, du\, d\xi.$$

As in the definition of the ambiguity function (4.146), the Fourier parameters τ and γ of $\hat{K}$ are inverted. The following theorem gives necessary and sufficient conditions to ensure that P_K satisfies marginal energy properties like those of the Wigner-Ville distribution. The Wigner Theorem 4.11 proves that in this case $P_K f(u,\xi)$ takes negative values.

Theorem 4.12. *For all* $f \in \mathbf{L^2}(\mathbb{R})$

$$\int_{-\infty}^{+\infty} P_K f(u,\xi)\, d\xi = 2\pi\, |f(u)|^2 \quad , \quad \int_{-\infty}^{+\infty} P_K f(u,\xi)\, du = |\hat{f}(\xi)|^2, \tag{4.148}$$

if and only if

$$\forall(\tau,\gamma)\in\mathbb{R}^2 \quad , \quad \hat{K}(\tau,0) = \hat{K}(0,\gamma) = 1. \tag{4.149}$$

Proof. Let $A_K f(\tau,\gamma)$ be the two-dimensional Fourier transform of $P_K f(u,\xi)$. The Fourier integral at $(0,\gamma)$ gives

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} P_K f(u,\xi)\, \mathrm{e}^{-iu\gamma}\, d\xi\, du = A_K f(0,\gamma). \tag{4.150}$$

Since the ambiguity function $Af(\tau,\gamma)$ is the Fourier transform of $P_V f(u,\xi)$, the two-dimensional convolution (4.144) gives

$$A_K(\tau,\gamma) = Af(\tau,\gamma)\, \hat{K}(\tau,\gamma). \tag{4.151}$$

The Fourier transform of $2\pi|f(u)|^2$ is $\hat f \star \overline{\hat f}(\gamma)$, with $\overline{\hat f}(\gamma) = \hat f^*(-\gamma)$. The relation (4.150) shows that (4.148) is satisfied if and only if

$$A_K f(0,\gamma) = Af(0,\gamma)\,\hat K(0,\gamma) = \hat f \star \overline{\hat f}(\gamma). \tag{4.152}$$

Since $P_V f$ satisfies the marginal property (4.135), we similarly prove that

$$Af(0,\gamma) = \hat f \star \overline{\hat f}(\gamma).$$

Requiring that (4.152) be valid for any $\hat f(\gamma)$, is equivalent to requiring that $\hat K(0,\gamma) = 1$ for all $\gamma \in \mathbb{R}$. The same derivation applied to the other marginal integration yields $\hat K(\xi, 0) = 1$. ■ ■

In addition to requiring time-frequency translation invariance, it may be useful to guarantee that P_K satisfies the same scaling property as a Wigner-Ville distribution:

$$g(t) = \frac{1}{\sqrt{s}} f\left(\frac{t}{s}\right) \quad \Longrightarrow \quad P_K g(u,\xi) = P_K f\left(\frac{u}{s}, s\xi\right) .$$

Such a distribution P_K is *affine* invariant. One can verify (Exercise 4.5) that affine invariance is equivalent to imposing that

$$\forall s \in \mathbb{R}^+ \quad , \quad K\left(s\,u, \frac{\xi}{s}\right) = K(u,\xi), \tag{4.153}$$

and hence

$$K(u,\xi) = K(u\,\xi, 1) = \beta(u\,\xi).$$

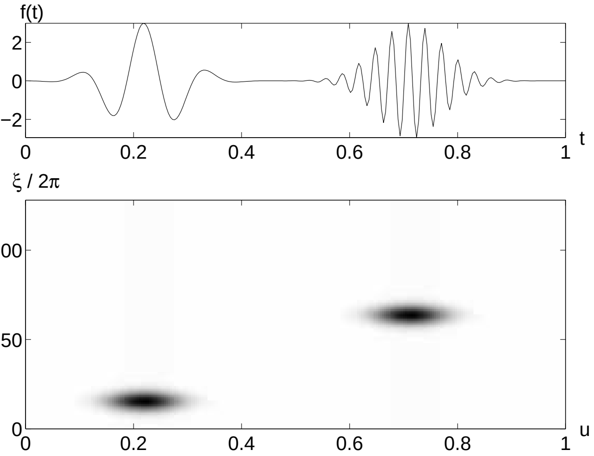


Figure 4.20: Choi-William distribution $P_K f(u,\xi)$ of the two Gabor atoms shown at the top. The interference term that appears in the Wigner-Ville distribution of Figure 4.18 has nearly disappeared.

Example 4.22. *The Rihaczek distribution is an affine invariant distribution whose convolution kernel is*

$$\hat K(\tau,\gamma) = \exp\left(\frac{i\,\tau\,\gamma}{2}\right). \tag{4.154}$$

A direct calculation shows that

$$P_K f(u,\xi) = f(u)\,\hat f^*(\xi)\,\exp(-iu\xi). \tag{4.155}$$

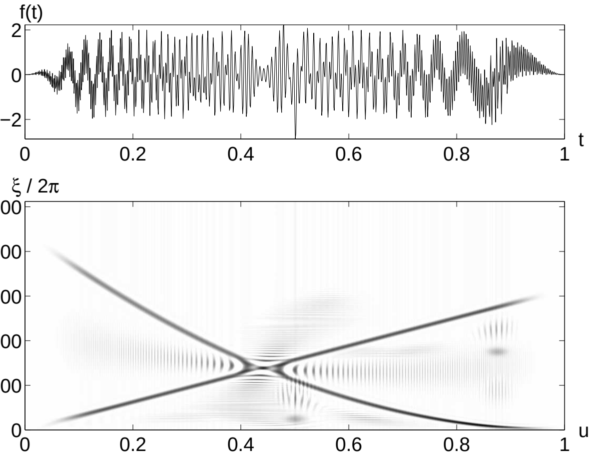


Figure 4.21: Choi-William distribution $P_K f_a(u,\xi)$ of the analytic part of the signal shown at the top. The interferences remain visible.

Example 4.23. *The kernel of the Choi-William distribution is [151]*

$$\hat{K}(\tau,\gamma) = \exp(-\sigma^2\, \tau^2\, \gamma^2)\ . \tag{4.156}$$

It is symmetric and thus corresponds to a real function $K(u,\xi)$*. This distribution satisfies the marginal conditions (4.149). Since* $\lim_{\sigma\to 0}\hat{K}(\tau,\gamma) = 1$*, when* σ *is small the Choi-William distribution is close to a Wigner-Ville distribution. Increasing* σ *attenuates the interference terms, but spreads* $K(u,\xi)$*, which reduces the time-frequency resolution of the distribution.*

Figure 4.20 shows that the interference terms of two modulated Gaussians nearly disappear when the Wigner-Ville distribution of Figure 4.18 is averaged by a Choi-William kernel having a sufficiently large σ*. Figure 4.21 gives the Choi-William distribution of the analytic signal whose Wigner-Ville distribution is in Figure 4.19. The energy of the linear and quadratic chirps are spread over wider time-frequency bands but the interference terms are attenuated, although not totally removed. It remains difficult to isolate the two modulated Gaussians at* $t = 0.5$ *and* $t = 0.87$*, which clearly appear in the spectrogram of Figure 4.3.*

4.5.4 Discrete Wigner-Ville Computations

The Wigner integral (4.120) is the Fourier transform of $f(u+\tau/2)f^*(u-\tau/2)$:

$$P_V f(u,\xi) = \int_{-\infty}^{+\infty} f\left(u+\frac{\tau}{2}\right)\, f^*\left(u-\frac{\tau}{2}\right)\, \mathrm{e}^{-i\tau\xi}\, d\tau. \tag{4.157}$$

For a discrete signal $f[n]$ defined over $0 \leqslant n < N$, the integral is replaced by a discrete sum:

$$P_V f[n,k] = \sum_{p=-N}^{N-1} f\left[n+\frac{p}{2}\right]\, f^*\left[n-\frac{p}{2}\right]\, \exp\left(\frac{-i2\pi kp}{N}\right). \tag{4.158}$$

When p is odd, this calculation requires knowing the value of f at half integers. These values are computed by interpolating f, with an addition of zeroes to its Fourier transform. This is necessary to avoid the aliasing produced by the discretization of the Wigner-Ville integral [155].

The interpolation $\tilde{f}$ of f is a signal of size $2N$ whose discrete Fourier transform $\widehat{\tilde{f}}$ is defined from the discrete Fourier transform $\hat{f}$ of f by

$$\widehat{\tilde{f}}[k] = \begin{cases} 2\hat{f}[k] & \text{if } 0 \leqslant k < N/2 \\ 0 & \text{if } N/2 < k < 3N/2 \\ 2\hat{f}[k-N] & \text{if } 3N/2 < k < 2N \\ \hat{f}[N/2] & \text{if } k = N/2\,,\ 3N/2 \end{cases}.$$

Computing the inverse discrete Fourier transform shows that $\tilde{f}[2n] = f[n]$ for $n \in [0, N-1]$. When $n \notin [0, 2N-1]$, we set $\tilde{f}[n] = 0$. The Wigner summation (4.158) is calculated from $\tilde{f}$:

$$\begin{aligned} P_V f[n,k] &= \sum_{p=-N}^{N-1} \tilde{f}[2n+p]\, \tilde{f}^*[2n-p] \exp\left(\frac{-i2\pi kp}{N}\right) \\ &= \sum_{p=0}^{2N-1} \tilde{f}[2n+p-N]\, \tilde{f}^*[2n-p+N] \exp\left(\frac{-i2\pi(2k)p}{2N}\right). \end{aligned}$$

For $0 \leqslant n < N$ fixed, $P_V f[n,k]$ is the discrete Fourier transform of size $2N$ of $g[p] = \tilde{f}[2n+p-N]\, \tilde{f}^*[2n-p+N]$ at the frequency $2k$. The discrete Wigner-Ville distribution is thus calculated with N FFT procedures of size $2N$, which requires $O(N^2 \log N)$ operations. To compute the Wigner-Ville distribution of the analytic part f_a of f, we use (4.48).

Cohen's Class A Cohen's class distribution is calculated with a circular convolution of the discrete Wigner-Ville distribution with a kernel $K[p,q]$:

$$P_K[n,k] = P_V \circledast K[n,k]. \tag{4.159}$$

Its two-dimensional discrete Fourier transform is therefore

$$A_K[p,q] = Af[p,q]\, \hat{K}[p,q]. \tag{4.160}$$

The signal $Af[p,q]$ is the discrete ambiguity function, calculated with a two-dimensional FFT of the discrete Wigner-Ville distribution $P_V f[n,k]$. As in the case of continuous time, we have inverted the index p and q of the usual two-dimensional Fourier transform. The Cohen's class distribution (4.159) is obtained by calculating the inverse Fourier transform of (4.160). This also requires a total of $O(N^2 \log N)$ operations.

4.6 Exercises

4.1. ^{1}).

4.2. ^{1}

4.3. ^{2}

4.4. ^{1}

4.5. ^{2}

V

Frames

A signal representation may provide "analysis" coefficients which are inner products with a family of vectors, or "synthesis" coefficients that compute an approximation by recombining these vectors. Frames are families of vectors where these representations are stable, and computed with a dual frame. Frames are potentially redundant and thus more general than bases, with a redundancy measured by frame bounds. They provide the flexibility needed to build signal representations with unstructured families of vectors.

Complete and stable wavelet and windowed Fourier transforms are constructed with frames of wavelets and windowed Fourier atoms. In two dimensions, frames of directional wavelets and curvelets are introduced to analyze and process multiscale image structures.

5.1 Frames and Riesz Bases

5.1.1 Stable Analysis and Synthesis Operators

The frame theory was originally developed by Duffin and Schaeffer [220] to reconstruct band-limited signals from irregularly spaced samples. They established general conditions to recover a vector f in a Hilbert space $\mathbf{H}$ from its inner products with a family of vectors $\{\phi_n\}_{n\in\Gamma}$. The index set Γ might be finite or infinite. The following frame definition gives an energy equivalence to invert the operator Φ defined by

$$\forall n \in \Gamma \quad , \quad \Phi f[n] = \langle f, \phi_n \rangle. \tag{5.1}$$

Definition 5.1 (Frame and Riesz Basis)**.** *The sequence $\{\phi_n\}_{n\in\Gamma}$ is a frame of $\mathbf{H}$ if there exist two constants $B \geqslant A > 0$ such that*

$$\forall f \in \mathbf{H} \quad , \quad A\,\|f\|^2 \leqslant \sum_{n\in\Gamma} |\langle f, \phi_n\rangle|^2 \leqslant B\,\|f\|^2. \tag{5.2}$$

When $A = B$ the frame is said to be tight. If the $\{\phi_n\}_{n\in\Gamma}$ are linearly independant then the frame is not redundant and is called a Riesz basis.

If the frame condition is satisfied then Φ is called a frame analysis operator. Section 5.1.2 proves that (5.2) is a necessary and sufficient condition guaranteeing that Φ is invertible on its image space, with a bounded inverse. A frame thus defines a complete and stable signal representation, which may also be redundant.

Frame Synthesis Let us consider the space of finite energy coefficients

$$\boldsymbol{\ell}^2(\Gamma) = \{a \ : \ \|a\|^2 = \sum_{n\in\Gamma} |a[n]|^2 < +\infty\} \ .$$

The adjoint Φ^* of Φ is defined over $\boldsymbol{\ell}^2(\Gamma)$ and satisfies for any $f \in \mathbf{H}$ and $a \in \boldsymbol{\ell}^2(\Gamma)$

$$\langle \Phi^* a, f\rangle \;=\; \langle a, \Phi f\rangle \;=\; \sum_{n\in\Gamma} a[n]\,\langle f, \phi_n\rangle^*.$$

It is therefore the synthesis operator

$$\Phi^* a = \sum_{n\in\Gamma} a[n]\,\phi_n \;. \tag{5.3}$$

The frame condition (5.2) can be rewritten

$$\forall f \in \mathbf{H} \quad , \quad A\,\|f\|^2 \leqslant \|\Phi f\|^2 = \langle \Phi^*\Phi f, f\rangle \leqslant B\,\|f\|^2, \tag{5.4}$$

with

$$\Phi^*\Phi f = \sum_{m\in\Gamma} \langle f, \phi_n\rangle\,\phi_n \;.$$

It results that A and B are the infimum and supremum values of the spectrum of the symmetric operator $\Phi^*\Phi$, which correspond to the smallest and largest eigenvalues in finite dimension. The following theorem derives that the frame synthesis operator is also stable.

Theorem 5.1. *The family* $\{\phi_n\}_{n\in\Gamma}$ *is a frame with bounds* $0 < A \leqslant B$ *if and only if*

$$\forall a \in \boldsymbol{\ell}^2(\Gamma) \quad , \quad A\,\|a\|^2 \leqslant \|\sum_{n\in\Gamma} a[n]\,\phi_n\|^2 \leqslant B\,\|a\|^2 \,. \tag{5.5}$$

Proof. Since $\Phi^* a = \sum_{n\in\Gamma} a[n]\,\phi_n$, it results that

$$\|\sum_{n\in\Gamma} a[n]\,\phi_n\|^2 = \langle \Phi\Phi^* a, a\rangle \;.$$

The operator Φ is a frame if and only if its spectrum is bound by A and B. The inequality (5.5) states that the the spectrum of $\Phi\Phi^*$ is also bounded by A and B. Both statements are proved to be equivalent by verifying that sup and inf of the spectrum of $\Phi^*\Phi$ is equal to the sup and inf of the spectrum of $\Phi^*\Phi$.

In finite dimension, if λ is an eigenvalue of $\Phi^*\Phi$ with eigenvector f then λ is also an eigenvalue of $\Phi^*\Phi$ with eigenvector Φf. Indeed, $\Phi^*\Phi f = \lambda f$ so $\Phi\Phi^*(\Phi f) = \lambda \Phi f$ and $\Phi f \neq 0$ because the left frame inequality (5.2) implies that $\|\Phi f\|^2 \leqslant A\,\|f\|^2$. It results that the maximum and minimum eigenvectors of $\Phi^*\Phi$ and $\Phi\Phi^*$ are identical.

In a Hilbert space of infinite dimension, we prove that the sup and inf of the spectrum of both operators remain identical by growing the space dimension, and computing the limit of the largest and smallest eigenvalues when the space dimension tends to infinity. ■

This theorem proves that linear combination of frame vectors define a stable signal representation. Section 5.1.2 proves that synthesis coefficients are computed with a dual frame. The operator $\Phi\Phi^*$ is the *Gram* matrix, whose coefficients are $\{\langle \phi_n, \phi_p\rangle\}_{(m,p)\in\boldsymbol{\ell}^2(\Gamma)}$:

$$\Phi\Phi^* a[p] = \sum_{m\in\Gamma} a[m]\,\langle \phi_n, \phi_p\rangle \,. \tag{5.6}$$

Redundancy When the frame vectors are normalized $\|\phi_n\| = 1$, the following theorem shows that the frame redundancy is measured by the frame bounds A and B.

Theorem 5.2. *In a space of finite dimension* N*, a frame of* $P \geqslant N$ *normalized vectors has frame bounds* A *and* B *which satisfy*

$$A \leqslant \frac{P}{N} \leqslant B. \tag{5.7}$$

For a tight frame $A = B = P/N$*.*

Proof. It results from (5.4) that all eigenvalues of $\Phi^*\Phi$ are between A and B. The trace of $\Phi^*\Phi$ thus satisfies

$$A\,N \leqslant \mathrm{tr}(\Phi^*\Phi) \leqslant B\,N\ .$$

But since the trace is not modified by commuting matrices (Exercise 5.2), and $\|\phi_n\| = 1$

$$A\,N \leqslant \mathrm{tr}(\Phi^*\Phi) = \mathrm{tr}(\Phi\Phi^*) = \sum_{n=1}^{P} |\langle \phi_n, \phi_n\rangle|^2 = P \leqslant B\,N\ ,$$

which implies (5.7). ■

If $\{\phi_n\}_{n\in\Gamma}$ is a normalized Riesz basis and is therefore linearly independent then (5.7) proves that $A \leqslant 1 \leqslant B$. This result remains valid in infinite dimension. Inserting $f = \phi_n$ in the frame inequality (5.2) proves that the frame is orthonormal if and only if $B = 1$ in which case $A = 1$.

Example 5.1. *Let $\{g_1, g_2\}$ be an orthonormal basis of an $N = 2$ two-dimensional plane $\mathbf{H}$. The $P = 3$ normalized vectors*

$$\phi_1 = g_1\ \ ,\ \ \phi_2 = -\frac{g_1}{2} + \frac{\sqrt{3}}{2}\,g_2\ \ ,\ \ \phi_3 = -\frac{g_1}{2} - \frac{\sqrt{3}}{2}\,g_2$$

have equal angles of $2\pi/3$ between each other. For any $f \in \mathbf{H}$

$$\sum_{n=1}^{3} |\langle f, \phi_n\rangle|^2 = \frac{3}{2}\,\|f\|^2.$$

These three vectors thus define a tight frame with $A = B = 3/2$.

Example 5.2. *For any $0 \leqslant k < K$, suppose that $\{\phi_{k,n}\}_{n\in\Gamma}$ is an orthonormal basis of $\mathbf{H}$. The union of these K orthonormal bases $\{\phi_{k,n}\}_{n\in\Gamma, 0\leqslant k<K}$ is a tight frame with $A = B = K$. Indeed, the energy conservation in an orthonormal basis implies that for any $f \in \mathbf{H}$,*

$$\sum_{n\in\mathbb{Z}} |\langle f, \phi_{k,n}\rangle|^2 = \|f\|^2,$$

hence

$$\sum_{k=0}^{K-1}\sum_{n\in\mathbb{Z}} |\langle f, \phi_{k,n}\rangle|^2 = K\,\|f\|^2.$$

One can verify (Exercice 5.5) that a finite set of N vectors $\{\phi_n\}_{1\leqslant n\leqslant N}$ is always a frame of the space $\mathbf{V}$ generated by linear combinations of these vectors. When N increases, the frame bounds A and B may go respectively to 0 and $+\infty$. This illustrates the fact that in infinite dimensional spaces, a family of vectors may be complete and not yield a stable signal representation.

Irregular Sampling Let $\mathbf{U}_s$ be the space of $\mathbf{L^2}(\mathbb{R})$ functions whose Fourier transforms have a support included in $[-\pi/s, \pi/s]$. For a uniform sampling, $t_n = ns$, Theorem 3.5 proves that if $\phi_s(t) = s^{1/2}\sin(\pi s^{-1}t)/(\pi t)$ then $\{\phi_s(t{-}ns)\}_{n\in\mathbb{Z}}$ is an orthonormal basis of $\mathbf{U}_s$. The reconstruction of f from its samples is then given by the sampling Theorem 3.2.

The irregular sampling conditions of Duffin and Schaeffer [220] for constructing a frame were later refined by several researchers [97, 457, 77]. Grochenig proved [260] that if $\lim_{n\to+\infty} t_n = +\infty$ and $\lim_{n\to-\infty} t_n = -\infty$, and if the maximum sampling distance δ satisfies

$$\delta = \sup_{n\in\mathbb{Z}} |t_{n+1} - t_n| < s\ , \tag{5.8}$$

then

$$\{\lambda_n\,\phi_s(t - t_n)\}_{n\in\mathbb{Z}}\ \ \text{with}\ \ \lambda_n = \sqrt{\frac{t_{n+1} - t_{n-1}}{2s}}$$

is a frame with frame bounds $A \geqslant (1 - \delta/s)^2$ and $B \leqslant (1 + \delta/s)^2$. The amplitude factor λ_n compensates for the increase of sample density relatively to s. The reconstruction of f requires inverting the frame operator $\Phi f[n] = \langle f(u), \lambda_n, \phi_s(u - t_n)\rangle$.

5.1.2 Dual Frame and Pseudo Inverse

The reconstruction of f from its frame coefficients $\Phi f[n]$ is calculated with a pseudo inverse also called Moore-Penrose pseudo-inverse. This pseudo inverse is a bounded operator that is implements a dual frame reconstruction. For Riesz bases, this dual frame is a biorthogonal basis.

For any operator U, we denote by **ImU** the image space of all Uf and by **NullU** the null space of all h such that $Uh = 0$.

Theorem 5.3. *If $\{\phi_n\}_{n\in\Gamma}$ is a frame but not a Riesz basis then Φ admits an infinite number of left inverses.*

Proof. We know that $\mathbf{Null}\Phi^* = (\mathbf{Im}\Phi)^\perp$ is the orthogonal complement of $\mathbf{Im}\Phi$ in $\ell^2(\Gamma)$ (Exercice 5.3). If Φ is a frame and not a Riesz basis then $\{\phi_n\}_{n\in\Gamma}$ is linearly dependent so there exists $a \in \mathbf{Null}\Phi^* = (\mathbf{Im}\Phi)^\perp$ with $a \neq 0$.

A frame operator Φ is injective (one to one). Indeed, the frame inequality (5.2) guarantees that $\Phi f = 0$ implies $f = 0$. Its restriction to $\mathbf{Im}\Phi$ is thus invertible, which means that Φ admits a left inverse. There is an infinite number of left inverses since the restriction of a left inverse to $(\mathbf{Im}\Phi)^\perp \neq \{0\}$ may be any arbitrary linear operator. ■ ■

The more redundant the frame $\{\phi_n\}_{n\in\Gamma}$, the larger the orthogonal complement $(\mathbf{Im}\Phi)^\perp$ of $\mathbf{Im}\Phi$ in $\ell^2(\Gamma)$. The pseudo inverse, that we write Φ^+, is defined as the left inverse that is zero on $(\mathbf{Im}\Phi)^\perp$:

$$\forall f \in \mathbf{H} \quad , \quad \Phi^+\Phi f = f \quad \text{and} \quad \forall a \in (\mathbf{Im}\Phi)^\perp \quad , \quad \Phi^+ a = 0.$$

The following theorem computes this pseudo-inverse.

Theorem 5.4 (Pseudo inverse)**.** *If Φ is a frame operator then $\Phi^*\Phi$ is invertible and the pseudo inverse satisfies*

$$\Phi^+ = (\Phi^*\Phi)^{-1}\Phi^*. \tag{5.9}$$

Proof. The frame condition in (5.4) is rewritten

$$\forall f \in \mathbf{H} \quad , \quad A\,\|f\|^2 \leqslant \langle \Phi^*\Phi f, f\rangle \leqslant B\,\|f\|^2.$$

It result that $\Phi^*\Phi$ is an injective self-adoint opeator: $\Phi^*\Phi f = 0$ if and only if $f = 0$. It is therefore invertible. For all $f \in \mathbf{H}$

$$\Phi^+\Phi f = (\Phi^*\Phi)^{-1}\Phi^*\Phi f = f \ ,$$

so Φ^+ is a left inverse. Since $(\mathbf{Im}\Phi)^\perp = \mathbf{Null}\Phi^*$ it results that $\Phi^+ a = 0$ for any $a \in (\mathbf{Im}\Phi)^\perp = \mathbf{Null}\Phi^*$. Since this left inverse vanishes on $(\mathbf{Im}\Phi)^\perp$, it is the pseudo-inverse. ■ ■

Dual Frame The pseudo inverse of a frame operator implements a reconstruction with a dual frame, which is specified by the following theorem.

Theorem 5.5. *Let $\{\phi_n\}_{n\in\Gamma}$ be a frame with bounds $0 < A \leqslant B$. The dual operator defined by*

$$\forall n \in \Gamma \quad , \quad \tilde{\Phi} f[n] = \langle f, \tilde{\phi}_n\rangle \quad \text{with} \quad \tilde{\phi}_n = (\Phi^*\Phi)^{-1}\phi_n \tag{5.10}$$

satisfies $\tilde{\Phi}^* = \Phi^+$ and hence

$$f = \sum_{n\in\Gamma} \langle f, \phi_n\rangle\, \tilde{\phi}_n = \sum_{n\in\Gamma} \langle f, \tilde{\phi}_n\rangle\, \phi_n. \tag{5.11}$$

It defines a dual frame

$$\forall f \in \mathbf{H} \quad , \quad \frac{1}{B}\,\|f\|^2 \leqslant \sum_{n\in\Gamma} |\langle f, \tilde{\phi}_n\rangle|^2 \leqslant \frac{1}{A}\,\|f\|^2\,. \tag{5.12}$$

If the frame is tight (i.e., $A = B$), then $\tilde{\phi}_n = A^{-1}\,\phi_n$.

Proof. The dual operator can be written $\tilde{\Phi} = \Phi(\Phi^*\Phi)^{-1}$. Indeed,

$$\tilde{\Phi} f[n] = \langle f, \tilde{\phi}_n \rangle = \langle f, (\Phi^*\Phi)^{-1}\phi_n \rangle = \langle (\Phi^*\Phi)^{-1} f, \phi_n \rangle = \Phi(\Phi^*\Phi)^{-1} f.$$

We thus derive from (5.9) that its adjoint is the pseudo-inverse of Φ:

$$\tilde{\Phi}^* = (\Phi^*\Phi)^{-1}\Phi^* = \Phi^+$$

It results that $\Phi^+\,\Phi = \tilde{\Phi}^*\Phi = \mathrm{Id}$ and hence that $\Phi^*\tilde{\Phi} = \mathrm{Id}$, which proves (5.11).

Let us now prove the frame bounds (5.12). Frame conditions are rewritten in (5.4):

$$\forall f \in \mathbf{H} \quad , \quad A\,\|f\|^2 \leqslant \langle \Phi^*\Phi f, f \rangle \leqslant B\,\|f\|^2. \tag{5.13}$$

The following lemma applied to $L = \Phi^*\Phi$ proves that

$$\forall f \in \mathbf{H} \quad , \quad B^{-1}\,\|f\|^2 \leqslant \langle (\Phi^*\Phi)^{-1} f, f \rangle \leqslant A^{-1}\,\|f\|^2. \tag{5.14}$$

Since for any $f \in \mathbf{H}$

$$\|\tilde{\Phi} f\|^2 = \langle \Phi(\Phi^*\Phi)^{-1} f, \Phi(\Phi^*\Phi)^{-1} f \rangle = \langle f, (\Phi^*\Phi)^{-1} f \rangle \ ,$$

the dual frame bounds (5.12) are derived from (5.14).

If $A = B$ then $\langle \Phi^*\Phi f, f \rangle = A\,\|f\|^2$. The spectrum of $\Phi^*\Phi$ is thus reduced to A and hence $\Phi^*\Phi = A\,\mathrm{Id}$. As a result $\tilde{\phi}_n = (\Phi^*\Phi)^{-1}\phi_n = A^{-1}\phi_n$.

Lemma 5.1. *If L is a self-adjoint operator such that there exist $B \geqslant A > 0$ satisfying*

$$\forall f \in \mathbf{H} \quad , \quad A\,\|f\|^2 \,\leqslant\, \langle Lf, f \rangle \,\leqslant B\,\|f\|^2 \tag{5.15}$$

then L is invertible and

$$\forall f \in \mathbf{H} \quad , \quad \frac{1}{B}\,\|f\|^2 \,\leqslant\, \langle L^{-1} f, f \rangle \,\leqslant\, \frac{1}{A}\,\|f\|^2. \tag{5.16}$$

In finite dimensions, since L is self-adjoint we know that it is diagonalized in an orthonormal basis. The inequality (5.15) proves that its eigenvalues are between A and B. It is therefore invertible with eigenvalues between B^{-1} and A^{-1}, which proves (5.16). In a Hilbert space of infinite dimension, we prove that same result on the sup and inf of the spectrum by growing the space dimension, and computing the limit of the largest and smallest eigenvalues when the space dimension tends to infinity. ■ ■

This theorem proves that f is reconstructed from frame coefficients $\Phi f[n] = \langle f, \phi_n \rangle$ with the dual frame $\{\tilde{\phi}_n\}_{n\in\Gamma}$. The synthesis coefficients of f in $\{\phi_n\}_{n\in\Gamma}$ are the dual frame coefficients $\tilde{\Phi} f[n] = \langle f, \phi_n \rangle$. If the frame is tight then both decompositions are identical

$$f = \frac{1}{A} \sum_{n\in\Gamma} \langle f, \phi_n \rangle\, \phi_n. \tag{5.17}$$

Biorthogonal Bases A Riesz basis is a frame of vectors that are linearly independent, which implies that $\mathbf{Im}\Phi = \boldsymbol{\ell}^2(\Gamma)$, then its dual frame is also linearly independent. Inserting $f = \phi_p$ in (5.11) yields

$$\phi_p = \sum_{n\in\Gamma} \langle \phi_p, \tilde{\phi}_n \rangle\, \phi_n,$$

and the linear independence implies that

$$\langle \phi_p, \tilde{\phi}_n \rangle = \delta[p-n].$$

Dual Riesz bases are thus biorthogonal families of vectors. If the basis is normalized (i.e., $\|\phi_n\| = 1$), then

$$A \leqslant 1 \leqslant B. \tag{5.18}$$

This is proved by inserting $f = \phi_p$ in the frame inequality (5.12):

$$\frac{1}{B}\,\|\phi_p\|^2 \leqslant \sum_{n\in\Gamma} |\langle \phi_p, \tilde{\phi}_n \rangle|^2 = 1 \leqslant \frac{1}{A}\,\|\phi_p\|^2.$$

5.1.3 Dual Frame Analysis and Synthesis Computations

Suppose that $\{\phi_n\}_{n\in\Lambda}$ is a frame of a subspace $\mathbf{V}$ of the whole signal space. The best linear approximation of f in $\mathbf{V}$ is the orthogonal projection of f in $\mathbf{V}$. The following theorem shows that this orthogonal projection is computed with the dual frame. Two iterative numerical algorithms are described to implement such computations.

Theorem 5.6. *Let $\{\phi_n\}_{n\in\Lambda}$ be a frame of $\mathbf{V}$ and $\{\tilde\phi_n\}_{n\in\Lambda}$ it's dual frame in $\mathbf{V}$. The orthogonal projection of $f \in \mathbf{H}$ in $\mathbf{V}$ is*

$$P_{\mathbf{V}} f = \sum_{n\in\Lambda} \langle f, \phi_n\rangle \, \tilde\phi_n = \sum_{n\in\Lambda} \langle f, \tilde\phi_n\rangle \, \phi_n \ . \tag{5.19}$$

Proof. Since both frames are dual in $\mathbf{V}$, if $f \in \mathbf{V}$ then (5.11) proves that the operator $P_{\mathbf{V}}$ defined in (5.19) satisfies $P_{\mathbf{V}} f = f$. To prove that it is an orthogonal projection it is sufficient to verify that for all $f \in \mathbf{H}$ $\langle f - P_{\mathbf{V}} f, \phi_p\rangle = 0$ for all $p \in \Lambda$. Indeed,

$$\langle f - P_{\mathbf{V}} f, \phi_p\rangle = \langle f, \phi_p\rangle - \sum_{n\in\Lambda} \langle f, \phi_n\rangle \, \langle \tilde\phi_n, \phi_p\rangle = 0$$

because the dual frame property implies that $\sum_{n\in\Lambda} \langle \tilde\phi_n, \phi_p\rangle \, \phi_n = \phi_p$. ■

If Λ is finite then $\{\phi_n\}_{n\in\Lambda}$ is necessarily a frame of the space $\mathbf{V}$ it generates, and (5.19) reconstructs the best linear approximation of f in $\mathbf{V}$. This result is particularly important for approximating signals from a finite set of vectors.

Dual Analysis and Synthesis Algorithms When constructing sparse signal representations, we are facing two types of signal reconstruction problems. In a *dual synthesis* problem, the orthogonal projection is computed from the frame coefficients $\{\Phi f[n] = \langle f, \phi_n\rangle\}_{n\in\Lambda}$ with the dual frame synthesis operator:

$$P_{\mathbf{V}} f = \tilde\Phi^* \, \Phi f = \sum_{n\in\Lambda} \langle f, \phi_n\rangle \, \tilde\phi_n \ . \tag{5.20}$$

In a *dual analysis* problem, the orthogonal projection is computed from the frame vectors $\{\phi_n\}_{n\in\Lambda}$ with the dual frame analysis operator $\tilde\Phi f[n] = \langle f, \tilde\phi_n\rangle$:

$$P_{\mathbf{V}} f = \Phi^* \, \tilde\Phi f = \sum_{n\in\Lambda} \langle f, \tilde\phi_n\rangle \, \phi_n \ . \tag{5.21}$$

We describe efficient numerical algorithms to solve both problems by inverting a symmetric operator. We write Φ_Λ the restriction of Φ to $\mathbf{V}$. The operator $\Phi^*\Phi_\Lambda$ is invertible on $\mathbf{V}$ and we write $(\Phi^*\Phi_\Lambda)^{-1}$ its inverse.

If the frame $\{\phi_n\}_{n\in\Lambda}$ does not depend upon the signal f then the dual frame vectors are precomputed with (5.10):

$$\forall n \in \Lambda \ \ , \ \ \tilde\phi_n = (\Phi^*\Phi_\Lambda)^{-1}\phi_n,$$

and the dual synthesis or analysis problems are solved directly with (5.20) or (5.21).

In many applications, the frame vectors $\{\phi_n\}_{n\in\Lambda}$ depend on the signal f, in which case the dual frame vectors $\tilde\phi_n$ cannot be computed in advance, and it is highy inefficient to compute them. This is the case when the $\{\phi_n\}_{n\in\Lambda}$ are selected in a larger dictionary, to build a sparse signal representation. For example, the time-frequency ridge vectors in Sections 4.4.1 and 4.4.2, are selected from the local maxima of f in highly redundant windowed Fourier or wavelet dictionaries.

In a dual synthesis problem, Φf is known and we must compute

$$P_{\mathbf{V}} f = \tilde\Phi \Phi f = (\Phi^*\Phi_\Lambda)^{-1} \Phi^* \Phi f \ .$$

The dual synthesis algorithm computes

$$y = \Phi^*\Phi f = \sum_{n\in\Lambda} \langle f, \phi_n\rangle \, \phi_n \in \mathbf{V}$$

and derives $P_{\mathbf{V}} f = L^{-1} y = z$ by applying the inverse of the symmetric operator $L = \Phi^* \Phi_\Lambda$ to y, with

$$\forall h \in \mathbf{V} \quad , \quad Lh = \sum_{n \in \Lambda} \langle h, \phi_n \rangle \, \phi_n \ . \tag{5.22}$$

The eigenvalues of L are between A and B.

In a dual analysis problem, the dual coefficient $a[n] = \tilde{\Phi} f[n]$ must be calculated as well as

$$P_{\mathbf{V}} f = \Phi^* a = \sum_{n \in \Lambda} a[n] \, \phi_n \ . \tag{5.23}$$

Since $\Phi P_{\mathbf{V}} f = \Phi f$, we have $\Phi \Phi^* a = \Phi f$, and since $\Phi \Phi^*$ is invertible

$$a = (\Phi \Phi^*)^{-1} \Phi \, f \ .$$

The dual analysis algorithm thus computes $y = \Phi f = \{\langle f, \phi_n \rangle\}_{n \in \Lambda}$ and derives the dual coefficients $a = L^{-1} y = z$ by applying the inverse of the Gram operator $L = \Phi \Phi^*$ to y, with

$$Lh[n] = \sum_{p \in \Lambda} h[n] \, \langle \phi_n, \phi_p \rangle \ . \tag{5.24}$$

The eigenvalues of L are also between A and B. The orthogonal projection of f is recovered with (5.23).

The key computational step of a dual analysis or a dual synthesis problem is to compute $z = L^{-1} y$, where L is a symmetric operator whose eigenvalues are between A and B. Theorems 5.7 and 5.8 describe two iterative algorithms with exponential convergence. The *Richardson iteration procedure* is simpler but requires knowing the frame bounds A and B. *Conjugate gradient* iterations converge more quickly when B/A is large, and do not require knowing the values of A and B.

Richardson Inversion of Symmetric Operators

Theorem 5.7. *To compute $z = L^{-1} y$, let z_0 be an initial value and $\gamma > 0$ be a relaxation parameter. For any $n > 0$, define*

$$z_n = z_{n-1} + \gamma \, (y - L z_{n-1}). \tag{5.25}$$

If

$$\delta = \max \{ |1 - \gamma A|, |1 - \gamma B| \} < 1, \tag{5.26}$$

then

$$\|z - z_n\| \leqslant \delta^n \, \|z - z_0\|, \tag{5.27}$$

and hence $\lim\limits_{n \to +\infty} z_n = z$.

Proof. The induction equation (5.25) can be rewritten

$$z - z_n = z - z_{n-1} - \gamma \, L(z - z_{n-1}).$$

Let

$$R = Id - \gamma \, L,$$
$$z - z_n = R(z - z_{n-1}) = R^n (z - z_0). \tag{5.28}$$

Since the eigenvalues of L are between A and B

$$A \, \|z\|^2 \leqslant \langle Lz, z \rangle \leqslant B \, \|z\|^2.$$

This implies that $R = I - \gamma L$ satisfies

$$|\langle Rz, z \rangle| \leqslant \delta \, \|z\|^2,$$

where δ is given by (5.26). Since R is symmetric, this inequality proves that $\|R\| \leqslant \delta$. We thus derive (5.27) from (5.28). The error $\|z - z_n\|$ clearly converges to zero if $\delta < 1$. ■ ■

The convergence is guaranteed for all initial values z_0. If an estimation z_0 of the solution z is known then this estimation can be chosen, otherwise z_0 is often set to 0. For frame inversion, the Richardson iteration algorithm is sometimes called the *frame algorithm* [17]. The convergence rate is maximized when δ is minimum:

$$\delta = \frac{B-A}{B+A} = \frac{1-A/B}{1+A/B},$$

which corresponds to the relaxation parameter

$$\gamma = \frac{2}{A+B}. \tag{5.29}$$

The algorithm converges quickly if A/B is close to 1. If A/B is small then

$$\delta \approx 1 - 2\,\frac{A}{B}. \tag{5.30}$$

The inequality (5.27) proves that we obtain an error smaller than ε for a number n of iterations, which satisfies:

$$\frac{\|z-z_n\|}{\|z-z_0\|} \leqslant \delta^n = \varepsilon.$$

Inserting (5.30) gives

$$n \approx \frac{\log_{\mathrm{e}} \varepsilon}{\log_{\mathrm{e}}(1-2A/B)} \approx \frac{-B}{2A}\,\log_{\mathrm{e}} \varepsilon. \tag{5.31}$$

The number of iterations thus increases proportionally to the frame bound ratio B/A.

The exact values of A and B are often not known, and A is generally more difficult to compute. The upper frame bound is $B = \|\Phi\,\Phi^*\|_S = \|\Phi^*\,\Phi\|_S$. If we choose

$$\gamma < 2\,\|\Phi\,\Phi^*\|_S^{-1} \tag{5.32}$$

then (5.26) shows that the algorithm is guaranteed to converge, but the convergence rate depends on A. Since $0 < A \leqslant B$, the optimal relaxation parameter γ in (5.29) is in the range: $\|\Phi\,\Phi^*\|_S^{-1} \leqslant \gamma < 2\,\|\Phi\,\Phi^*\|_S^{-1}$.

Conjugate-Gradient Inversion The conjugate gradient algorithm computes $z = L^{-1}y$ with a gradient descent along orthogonal directions with respect to the norm induced by the symmetric operator L:

$$\|z\|_L^2 = \|Lz\|^2. \tag{5.33}$$

This L norm is used to estimate the error. Grochenig's [262] implementation of the conjugate gradient algorithm is given by the following theorem.

Theorem 5.8 (Conjugate gradient)**.** *To compute $z = L^{-1}y$, we initialize*

$$z_0 = 0 \quad , \quad r_0 = p_0 = y \quad , \quad p_{-1} = 0. \tag{5.34}$$

For any $n \geqslant 0$, we define by induction

$$\lambda_n = \frac{\langle r_n, p_n\rangle}{\langle p_n, Lp_n\rangle} \tag{5.35}$$

$$z_{n+1} = z_n + \lambda_n\, p_n \tag{5.36}$$

$$r_{n+1} = r_n - \lambda_n\, Lp_n \tag{5.37}$$

$$p_{n+1} = Lp_n - \frac{\langle Lp_n, Lp_n\rangle}{\langle p_n, Lp_n\rangle}\, p_n - \frac{\langle Lp_n, Lp_{n-1}\rangle}{\langle p_{n-1}, Lp_{n-1}\rangle}\, p_{n-1}. \tag{5.38}$$

If $\sigma = \frac{\sqrt{B}-\sqrt{A}}{\sqrt{B}+\sqrt{A}}$ then

$$\|z-z_n\|_L \leqslant \frac{2\sigma^n}{1+\sigma^{2n}}\,\|z\|_L, \tag{5.39}$$

and hence $\lim\limits_{n\to+\infty} z_n = z$.

Proof. We give the main steps of the proof as outlined by Grochenig [262].

Step 1: Let $\mathbf{U}_n$ be the subspace generated by $\{L^j z\}_{1 \leqslant j \leqslant n}$. By induction on n, we derive from (5.38) that $p_j \in \mathbf{U}_n$, for $j < n$.

Step 2: We prove by induction that $\{p_j\}_{0 \leqslant j < n}$ is an orthogonal basis of $\mathbf{U}_n$ with respect to the inner product $\langle z, h \rangle_L = \langle z, Lh \rangle$. Assuming that $\langle p_n, Lp_j \rangle = 0$, for $j \leqslant n-1$, it can be shown that $\langle p_{n+1}, Lp_j \rangle = 0$, for $j \leqslant n$.

Step 3: We verify that z_n is the orthogonal projection of z onto $\mathbf{U}_n$ with respect to $\langle ., . \rangle_L$ which means that

$$\forall g \in \mathbf{U}_n \quad , \quad \|z - g\|_L \geqslant \|z - z_n\|_L .$$

Since $z_n \in \mathbf{U}_n$, this requires proving that $\langle z - z_n, p_j \rangle_L = 0$, for $j < n$.

Step 4: We compute the orthogonal projection of z in embedded spaces $\mathbf{U}_n$ of dimension n, and one can verify that $\lim_{n \to +\infty} \|z - z_n\|_L = 0$. The exponential convergence (5.39) is proved in [262]. ■

■

As opposed to the Richardson alorithm, the initial value z_0 must be set to 0. As in the Richardson iteration algorithm, the convergence is slower when A/B is small. In this case

$$\sigma = \frac{1 - \sqrt{A/B}}{1 + \sqrt{A/B}} \approx 1 - 2\sqrt{\frac{A}{B}}.$$

The upper bound (5.39) proves that we obtain a relative error

$$\frac{\|z - z_n\|_L}{\|z\|_L} \leqslant \varepsilon$$

for a number of iterations

$$n \approx \frac{\log_e \frac{\varepsilon}{2}}{\log_e \sigma} \approx \frac{-\sqrt{B}}{2\sqrt{A}} \log_e \frac{\varepsilon}{2}.$$

Comparing this result with (5.31) shows that when A/B is small, the conjugate gradient algorithm needs many fewer iterations than the Richardson iteration algorithm to compute $z = L^{-1} y$ at a fixed precision.

5.1.4 Frame Projector and Reproducing Kernel

Frame redundancy is useful in reducing noise added to the frame coefficients. The vector computed with noisy frame coefficients is projected on the image of Φ to reduce the amplitude of the noise. This technique is used for high precision analog to digital conversion based on oversampling. The following theorem specifies the orthogonal projector on $\mathbf{Im}\Phi$.

Theorem 5.9. *Let $\{\phi_n\}_{n \in \Gamma}$ be a frame of $\mathbf{H}$. The orthogonal projection from $\boldsymbol{\ell}^2(\Gamma)$ onto $\mathbf{Im}\Phi$ is*

$$Pa[n] = \Phi\Phi^+ a[n] = \sum_{p \in \Gamma} a[p] \, \langle \tilde{\phi}_p, \phi_n \rangle \; . \tag{5.40}$$

Coefficients $a \in \boldsymbol{\ell}^2(\Gamma)$ are frame coefficients $a \in \mathbf{Im}\Phi$ if and only if they satisfy the reproducing kernel equation

$$a[n] = \sum_{p \in \Gamma} a[p] \, \langle \tilde{\phi}_p, \phi_n \rangle. \tag{5.41}$$

Proof. If $a \in \mathbf{Im}\Phi$ then $a = \Phi f$ and

$$Pa = \Phi\Phi^+\Phi f = \Phi f = a.$$

If $a \in (\mathbf{Im}\Phi)^\perp$ then $Pa = 0$ because $\Phi^+ a = 0$. This proves that P is an orthogonal projector on $\mathbf{Im}\Phi$. Since $\Phi f[n] = \langle f, \phi_n \rangle$ and $\Phi^+ a = \sum_{p \in \Gamma} a[p] \, \tilde{\phi}_p$, we derive (5.40).

A vector $a \in \boldsymbol{\ell}^2(\Gamma)$ belongs to $\mathbf{Im}\Phi$ if and only if $a = Pa$, which proves (5.41). ■ ■

The reproducing kernel equation (5.41) expresses the redundancy of frame coefficients. If the frame is not redundant and is a Riesz basis then $\langle \tilde{\phi}_p, \phi_n \rangle = 0$, so this equation vanishes.

Noise Reduction Suppose that each frame coefficient $\Phi f[n]$ is contaminated by an additive noise $W[n]$, which is a random variable. Applying the projector P gives

$$P(\Phi f + W) = \Phi f + PW,$$

with

$$PW[n] = \sum_{p \in \Gamma} W[p] \, \langle \tilde{\phi}_p, \phi_n \rangle.$$

Since P is an orthogonal projector, $\|PW\| \leqslant \|W\|$. This projector removes the component of W that is in $(\mathbf{Im}\Phi)^{\perp}$. Increasing the redundancy of the frame reduces the size of $\mathbf{Im}\Phi$ and thus increases $(\mathbf{Im}\Phi)^{\perp}$, so a larger portion of the noise is removed. If W is a white noise, its energy is uniformly distributed in the space $\boldsymbol{\ell}^2(\Gamma)$. The following theorem proves that its energy is reduced by at least A if the frame vectors are normalized.

Theorem 5.10. *Suppose that $\|\phi_n\| = C$, for all $n \in \Gamma$. If W is a zero-mean white noise of variance $\mathsf{E}\{|W[n]|^2\} = \sigma^2$, then*

$$\mathsf{E}\{|PW[n]|^2\} \leqslant \frac{\sigma^2 \, C^2}{A}. \tag{5.42}$$

If the frame is tight then this inequality is an equality.

Proof. Let us compute

$$\mathsf{E}\{|PW[n]|^2\} = \mathsf{E}\left\{ \left(\sum_{p \in \Gamma} W[p] \, \langle \tilde{\phi}_p, \phi_n \rangle \right) \left(\sum_{l \in \Gamma} W^*[l] \, \langle \tilde{\phi}_l, \phi_n \rangle^* \right) \right\}.$$

Since W is white,

$$\mathsf{E}\{W[p] \, W^*[l]\} = \sigma^2 \, \delta[p-l],$$

and therefore

$$\mathsf{E}\{|PW[n]|^2\} = \sigma^2 \sum_{p \in \Gamma} |\langle \tilde{\phi}_p, \phi_n \rangle|^2 \leqslant \frac{\sigma^2 \, \|\phi_n\|^2}{A} = \frac{\sigma^2 \, C^2}{A}.$$

The last inequality is an equality if the frame is tight. ■ ■

Oversampling This noise reduction strategy is used by high precision analog to digital converters. After a low-pass filter, a band-limited analog signal $f(t)$ is uniformly sampled and quantized. In hardware, it is often easier to increase the sampling rate rather than the quantization precision. Increasing the sampling rate introduces a redundancy between the sample values of the band-limited signal. These samples can thus be interpreted as frame coefficients. For a wide range of signals, it has been shown that the quantization error is nearly a white noise [254]. It can thus be significantly reduced by a frame projector, which in this case is a low-pass convolution operator (Exercise **??**).

The noise can be further reduced if it is not white and if its energy is better concentrated in $(\mathbf{Im}\Phi)^{\perp}$. This can be done by transforming the quantization noise into a noise whose energy is mostly concentrated at high frequencies. Sigma-Delta modulators produce such quantization noises by integrating the signal before its quantization [85]. To compensate for the integration, the quantized signal is differentiated. This differentiation increases the energy of the quantized noise at high frequencies and reduces its energy at low frequencies [417].

5.1.5 Translation Invariant Frames

To construct translation invariant signal representations, Section 4.1 introduces translation invariant dictionaries obtained by translating a family of generators $\{\phi_n\}_{n \in \Gamma}$. In multiple dimensions for $\phi_n \in \mathbf{L^2}(\mathbb{R}^d)$, the resulting dictionary can be written $\mathcal{D} = \{\phi_{u,n}(x)\}_{n \in \Gamma, u \in \mathbb{R}^d}$, with $\phi_{u,n}(x) = \lambda_{u,n} \, \phi_n(x-u)$. In a translation invariant wavelet dictionary, the generators are obtained by dilating a wavelet $\psi(t)$ with scales s_n: $\phi_n(t) = s_n^{-1/2} \psi(x/s_n)$. In a window Fourier dictionary, the generators are obtained by modulating a window $g(x)$ at frequencies ξ_n: $\phi_n(x) = \mathrm{e}^{i \xi_n x} g(x)$.

The decomposition coefficients of f in $\mathcal{D}$ are convolution products

$$\Phi f(u,n) = \langle f, \phi_{u,n}\rangle = \lambda_{u,n}\, f \star \bar{\phi}_n(u) \quad \text{with} \quad \bar{\phi}_n(x) = \phi_n^*(-x)\ . \tag{5.43}$$

Suppose that Γ is a countable set. The overall index set $\mathbb{R}^d \times \Gamma$ is not countable so the dictionary $\mathcal{D}$ can not strictly speaking be considered as a frame. However, if we consider the overall energy of dictionary coefficients, calculated with a sum and a multidimensional integral

$$\sum_{n\in\Gamma} \|\Phi f(u,n)\|^2 = \sum_{n\in\Gamma} \int |\Phi f(u,n)|\, du,$$

and if there exist two constants $A > 0$ and $B > 0$ such that for all $f \in \mathbf{L^2}(\mathbb{R})$,

$$A\,\|f\|^2 \leqslant \sum_{n\in\Gamma} \|\Phi f(u,n)\|^2 \leqslant B\,\|f\|^2 \tag{5.44}$$

then the frame theory results of the previous section apply. With an abuse of language, such translation invariant dictionaries will thus also be called frames. The following theorem proves that the frame condition (5.44) is equivalent to a condition on the Fourier transform $\hat{\phi}_n(\omega)$ of the generators.

Theorem 5.11. *If there exist two constants $B \geqslant A > 0$ such that for almost all ω in $\mathbb{R}^d$*

$$A \leqslant \sum_{n\in\Gamma} |\hat{\phi}_n(\omega)|^2 \leqslant B, \tag{5.45}$$

then the frame inequality (5.44) is valid for all $f \in \mathbf{L^2}(\mathbb{R}^d)$. Any $\{\tilde{\phi}_n\}_{n\in\Gamma}$ which satisfies for almost all ω in $\mathbb{R}^d$

$$\sum_{n\in\Gamma} \hat{\phi}_n^*(\omega)\, \widehat{\tilde{\phi}}_n(\omega) = 1, \tag{5.46}$$

defines a left inverse

$$f(t) = \sum_{n\in\Gamma} \Phi f(.,n) \star \tilde{\phi}_n(t). \tag{5.47}$$

The pseudo-invese (dual frame) is implemented by

$$\widehat{\tilde{\phi}_n}(\omega) = \frac{\hat{\phi}_n(\omega)}{\sum_{n\in\Gamma} |\hat{\phi}_n(\omega)|^2}. \tag{5.48}$$

Proof. The frame condition (5.44) means that $\Phi^*\Phi$ has a spectrum bounded by A and B. It results from (5.43) that

$$\Phi^*\Phi f(x) = f \star (\sum_{n\in\Gamma} \phi_n \star \bar{\phi}_n)(x)\ . \tag{5.49}$$

The spectrum of this convolution operator is given by the Fourier transform of $\sum_{n\in\Gamma} \phi_n \star \bar{\phi}_n(x)$, which is $\sum_{n\in\Gamma} |\hat{\phi}_n(\omega)|^2$. The frame inequality (5.44) is thus equivalent to condition (5.45).

Equation (5.47) is proved by taking the Fourier transform on both sides and inserting (5.46).

Theorem 5.5 proves that the dual frame vectors implementing the pseudo-inverse are $\tilde{\phi}_{n,u} = (\Phi^*\Phi)^{-1}\phi_{n,u}$. Since $\Phi^*\Phi$ is the convolution operator (5.49), its inverse is calculated by inverting its tranfer function, which yields (5.48). ■ ■

For wavelet or windowed Fourier translation invariant dictionaries, the theorem condition (5.45) becomes a condition on the Fourier transform of the wavelet $\hat{\psi}(\omega)$ or on the Fourier transform of the window $\hat{g}(\omega)$. As explained in Section 5.3 and 5.4, more conditions are needed to obtain a frame by discretizing the translation parameter u.

Discrete Translation Invariant Frames For finite dimensional signals $f[n] \in \mathbb{C}^N$ a circular translation invariant frame is obtained with a periodic shift modulo N of a finite number of generators $\{\phi_m[n]\}_{0 \leqslant m < M}$:

$$\mathcal{D} = \{\phi_{m,p}[n] = \phi_m[(n-p) \mathrm{mod} N]\}_{0 \leqslant m < M, 0 \leqslant p < N} .$$

Such translation invariant frames appear in Section 11.2.3 to define translation invariant thresholding estimators for noise removal. Similarly to Theorem 5.11, the following theorem gives a necessary and sufficient condition on the discrete Fourier transform $\hat\phi_m[k] = \sum_{n=0}^{N-1} \phi_m[n] \mathrm{e}^{-i2\pi k/N}$ of the generators $\phi_m[n]$ to obtain a frame.

Theorem 5.12. *A circular translation invariant dictionary $\mathcal{D} = \{\phi_{m,p}[n]\}_{0 \leqslant m < M, 0 \leqslant n < N}$ is a frame with frame bounds $0 < A \leqslant B$ if and only if*

$$\forall 0 \leqslant k < N \quad A \leqslant \sum_{m=0}^{M-1} |\hat\phi_m[k]|^2 \leqslant B. \tag{5.50}$$

The proof proceeds essentially like the proof of Theorem 5.11, and is left in Exercise 5.8.

5.2 Translation Invariant Dyadic Wavelet Transform

The continuous wavelet transform of Section 4.3 decomposes one-dimensional signals $f \in \mathbf{L^2}(\mathbb{R})$ over a dictionary of translated and dilated wavelets

$$\psi_{u,s}(t) = \frac{1}{\sqrt{s}} \, \psi\left(\frac{t-u}{s}\right) .$$

Translation invariant wavelet dictionaries are constructed by sampling the scale parameter s along an exponential sequence $\{\nu^j\}_{j\in\mathbb{Z}}$, while keeping all translation parameters u. We choose $\nu = 2$ to simplifiy computer implementations:

$$\mathcal{D} = \left\{\psi_{u,2^j}(t) = \frac{1}{\sqrt{2^j}} \, \psi\left(\frac{t-u}{2^j}\right)\right\}_{u\in\mathbb{R}, j\in\mathbb{Z}} .$$

The resulting dyadic wavelet transform of $f \in \mathbf{L^2}(\mathbb{R})$ is defined by

$$Wf(u,2^j) = \langle f, \psi_{u,2^j} \rangle = \int_{-\infty}^{+\infty} f(t) \, \frac{1}{\sqrt{2^j}} \, \psi\left(\frac{t-u}{2^j}\right) dt = f \star \bar\psi_{2^j}(u), \tag{5.51}$$

with

$$\bar\psi_{2^j}(t) = \psi_{2^j}(-t) = \frac{1}{2^j} \, \psi\left(\frac{-t}{2^j}\right).$$

Translation invariant dyadic wavelet transforms are used in pattern recognition applications and for denoising with translation invariant wavelet thresholding estimators, as explained in Section 11.3.1. Fast computations with filter banks are presented in the next two sections.

Theorem 5.11 on translation invariant dictionaries can be applied to the multiscale wavelet generators $\phi_n(t) = 2^{-j/2} \, \psi_{2^j}(t)$. Since $\hat\phi_n(\omega) = \hat\psi(2^j\omega)$, the Fourier condition (5.45) means that there exist two constants $A > 0$ and $B > 0$ such that

$$\forall \omega \in \mathbb{R} - \{0\} \quad , \quad A \leqslant \sum_{j=-\infty}^{+\infty} |\hat\psi(2^j\omega)|^2 \leqslant B, \tag{5.52}$$

and since $\Phi f(u,n) = 2^{-j/2} Wf(u,n)$, Theorem 5.11 proves the frame inequality

$$A \, \|f\|^2 \leqslant \sum_{j=-\infty}^{+\infty} \frac{1}{2^j} \, \|Wf(u,2^j)\|^2 \leqslant B \, \|f\|^2. \tag{5.53}$$

This shows that if the frequency axis is completely covered by dilated dyadic wavelets, as illustrated by Figure 5.1, then it a dyadic wavelet transform defines a complete and stable representation.

Moreover, if $\tilde{\psi}$ satisfies

$$\forall \omega \in \mathbb{R} - \{0\} \quad , \quad \sum_{j=-\infty}^{+\infty} \hat{\psi}^*(2^j\omega)\, \widehat{\tilde{\psi}}(2^j\omega) = 1, \tag{5.54}$$

then (5.47) applied to $\tilde{\phi}_n(t) = 2^{-j}\tilde{\psi}(2^{-j}t)$ proves that

$$f(t) = \sum_{j=-\infty}^{+\infty} \frac{1}{2^j} Wf(.,2^j) \star \tilde{\psi}_{2^j}(t). \tag{5.55}$$

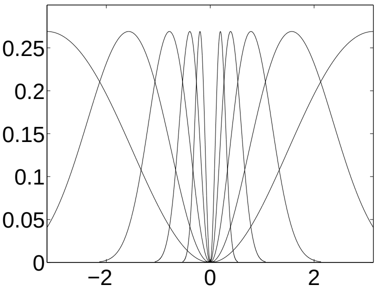


Figure 5.1: Scaled Fourier transforms $|\hat{\psi}(2^j\omega)|^2$ computed with (5.66), for $1 \leqslant j \leqslant 5$ and $\omega \in [-\pi, \pi]$.

Figure 5.2 gives a dyadic wavelet transform computed over 5 scales with the quadratic spline wavelet shown in Figure 5.3.

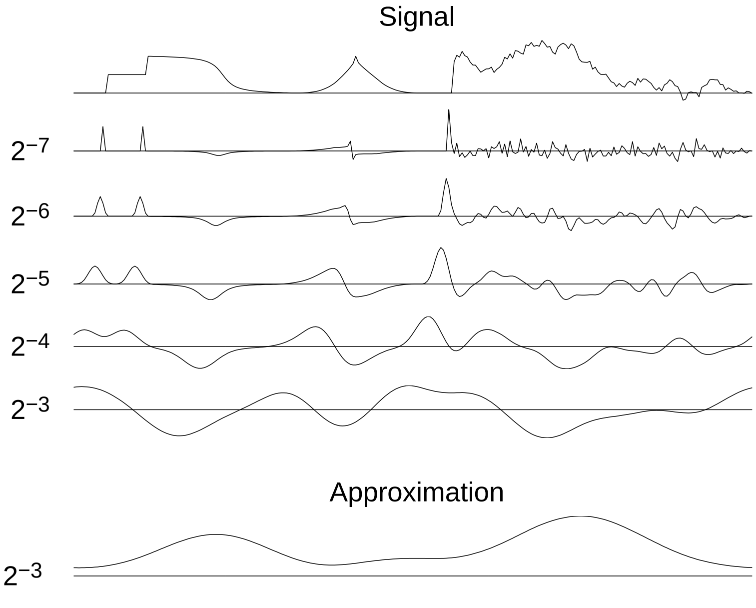


Figure 5.2: Dyadic wavelet transform $Wf(u, 2^j)$ computed at scales $2^{-7} \leqslant 2^j \leqslant 2^{-3}$ with the filter bank algorithm of Section 5.2.2, for a signal defined over $[0, 1]$. The bottom curve carries the lower frequencies corresponding to scales larger than 2^{-3}.

5.2.1 Dyadic Wavelet Design

A discrete dyadic wavelet transform can be computed with a fast filter bank algorithm if the wavelet is appropriately designed. The synthesis of these dyadic wavelets is similar to the construction of biorthogonal wavelet bases, explained in Section 7.4. All technical issues related to the convergence of infinite cascades of filters are avoided in this section. Reading Chapter 7 first is necessary for understanding the main results.

Let h and g be a pair of finite impulse response filters. Suppose that h is a low-pass filter whose transfer function satisfies $\hat{h}(0) = \sqrt{2}$. As in the case of orthogonal and biorthogonal wavelet bases, we construct a scaling function whose Fourier transform is

$$\hat{\phi}(\omega) = \prod_{p=1}^{+\infty} \frac{\hat{h}(2^{-p}\omega)}{\sqrt{2}} = \frac{1}{\sqrt{2}}\,\hat{h}\left(\frac{\omega}{2}\right)\hat{\phi}\left(\frac{\omega}{2}\right). \tag{5.56}$$

We suppose here that this Fourier transform is a finite energy function so that $\phi \in \mathbf{L^2}(\mathbb{R})$. The corresponding wavelet ψ has a Fourier transform defined by

$$\hat{\psi}(\omega) = \frac{1}{\sqrt{2}}\,\hat{g}\left(\frac{\omega}{2}\right)\hat{\phi}\left(\frac{\omega}{2}\right). \tag{5.57}$$

Theorem 7.5 proves that both ϕ and ψ have a compact support because h and g have a finite number of non-zero coefficients. The number of vanishing moments of ψ is equal to the number of zeroes of $\hat{\psi}(\omega)$ at $\omega = 0$. Since $\hat{\phi}(0) = 1$, (5.57) implies that it is also equal to the number of zeros of $\hat{g}(\omega)$ at $\omega = 0$.

Reconstructing Wavelets Reconstructing wavelets that satisfy (5.46) are calculated with a pair of finite impulse response dual filters $\tilde{h}$ and $\tilde{g}$. We suppose that the following Fourier transform has a finite energy:

$$\hat{\tilde{\phi}}(\omega) = \prod_{p=1}^{+\infty} \frac{\hat{\tilde{h}}(2^{-p}\omega)}{\sqrt{2}} = \frac{1}{\sqrt{2}}\,\hat{\tilde{h}}\left(\frac{\omega}{2}\right)\hat{\tilde{\phi}}\left(\frac{\omega}{2}\right). \tag{5.58}$$

Let us define

$$\hat{\tilde{\psi}}(\omega) = \frac{1}{\sqrt{2}}\,\hat{\tilde{g}}\left(\frac{\omega}{2}\right)\hat{\tilde{\phi}}\left(\frac{\omega}{2}\right). \tag{5.59}$$

The following theorem gives a sufficient condition to guarantee that $\hat{\tilde{\psi}}$ is the Fourier transform of a reconstruction wavelet.

Theorem 5.13. *If the filters satisfy*

$$\forall \omega \in [-\pi, \pi] \quad , \quad \hat{\tilde{h}}(\omega)\,\hat{h}^*(\omega) + \hat{\tilde{g}}(\omega)\,\hat{g}^*(\omega) = 2 \tag{5.60}$$

then

$$\forall \omega \in \mathbb{R} - \{0\} \quad , \quad \sum_{j=-\infty}^{+\infty} \hat{\psi}^*(2^j\omega)\,\hat{\tilde{\psi}}(2^j\omega) = 1. \tag{5.61}$$

Proof. The Fourier transform expressions (5.57) and (5.59) prove that

$$\hat{\tilde{\psi}}(\omega)\,\hat{\psi}^*(\omega) = \frac{1}{2}\,\hat{\tilde{g}}\left(\frac{\omega}{2}\right)\,\hat{g}^*\left(\frac{\omega}{2}\right)\,\hat{\tilde{\phi}}\left(\frac{\omega}{2}\right)\,\hat{\phi}^*\left(\frac{\omega}{2}\right).$$

Equation (5.60) implies

$$\begin{aligned} \hat{\tilde{\psi}}(\omega)\,\hat{\psi}^*(\omega) &= \frac{1}{2}\left[2 - \hat{\tilde{h}}\left(\frac{\omega}{2}\right)\,\hat{h}^*\left(\frac{\omega}{2}\right)\right]\hat{\tilde{\phi}}\left(\frac{\omega}{2}\right)\,\hat{\phi}^*\left(\frac{\omega}{2}\right) \\ &= \hat{\tilde{\phi}}\left(\frac{\omega}{2}\right)\,\hat{\phi}^*\left(\frac{\omega}{2}\right) - \hat{\tilde{\phi}}(\omega)\,\hat{\phi}^*(\omega). \end{aligned}$$

Hence

$$\sum_{j=-l}^{k} \hat{\tilde{\psi}}(2^j\omega)\,\hat{\psi}^*(2^j\omega) = \hat{\phi}^*(2^{-l}\omega)\,\hat{\tilde{\phi}}(2^{-l}\omega) - \hat{\phi}^*(2^k\omega)\,\hat{\tilde{\phi}}(2^k\omega).$$

Since $\hat{g}(0) = 0$, (5.60) implies $\hat{\tilde{h}}(0)\,\hat{h}^*(0) = 2$. We also impose that $\hat{h}(0) = \sqrt{2}$ so one can derive from (5.56) and (5.58) that $\hat{\tilde{\phi}}(0) = \hat{\phi}^*(0) = 1$. Since ϕ and $\tilde{\phi}$ belong to $\mathbf{L^1}(\mathbb{R})$, $\hat{\phi}$ and $\hat{\tilde{\phi}}$ are continuous, and the Riemann-Lebesgue lemma (Exercise 2.2) proves that $|\hat{\phi}(\omega)|$ and $|\hat{\tilde{\phi}}(\omega)|$ decrease to zero when ω goes to ∞. For $\omega \neq 0$, letting k and l go to $+\infty$ yields (5.61). ■ ■

Observe that (5.60) is the same as the unit gain condition (7.117) for biorthogonal wavelets. The aliasing cancellation condition (7.116) of biorthogonal wavelets is not required because the wavelet transform is not sampled in time.

Finite Impulse Response Solution Let us shift h and g to obtain causal filters. The resulting transfer functions $\hat{h}(\omega)$ and $\hat{g}(\omega)$ are polynomials in $\mathrm{e}^{-i\omega}$. We suppose that these polynomials have no common zeros. The Bezout Theorem 7.8 on polynomials proves that if $P(z)$ and $Q(z)$ are two polynomials of degree n and l, with no common zeros, then there exists a unique pair of polynomials $\tilde{P}(z)$ and $\tilde{Q}(z)$ of degree $l-1$ and $n-1$ such that

$$P(z)\,\tilde{P}(z) + Q(z)\,\tilde{Q}(z) = 1. \tag{5.62}$$

This guarantees the existence of $\hat{\tilde{h}}(\omega)$ and $\hat{\tilde{g}}(\omega)$ that are polynomials in $\mathrm{e}^{-i\omega}$ and satisfy (5.60). These are the Fourier transforms of the finite impulse response filters $\tilde{h}$ and $\tilde{g}$. One must however be careful because the resulting scaling function $\hat{\tilde{\phi}}$ in (5.58) does not necessarily have a finite energy.

Spline Dyadic Wavelets A *box spline* of degree m is a translation of $m+1$ convolutions of $\mathbf{1}_{[0,1]}$ with itself. It is centered at $t = 1/2$ if m is even and at $t = 0$ if m is odd. Its Fourier transform is

$$\hat{\phi}(\omega) = \left(\frac{\sin(\omega/2)}{\omega/2}\right)^{m+1} \exp\left(\frac{-i\varepsilon\omega}{2}\right) \quad \text{with } \varepsilon = \begin{cases} 1 & \text{if } m \text{ is even} \\ 0 & \text{if } m \text{ is odd} \end{cases}, \tag{5.63}$$

so

$$\hat{h}(\omega) = \sqrt{2}\,\frac{\hat{\phi}(2\omega)}{\hat{\phi}(\omega)} = \sqrt{2}\left(\cos\frac{\omega}{2}\right)^{m+1} \exp\left(\frac{-i\varepsilon\omega}{2}\right). \tag{5.64}$$

We construct a wavelet that has one vanishing moment by choosing $\hat{g}(\omega) = O(\omega)$ in the neighborhood of $\omega = 0$. For example

$$\hat{g}(\omega) = -i\,\sqrt{2}\,\sin\frac{\omega}{2}\,\exp\left(\frac{-i\varepsilon\omega}{2}\right). \tag{5.65}$$

The Fourier transform of the resulting wavelet is

$$\hat{\psi}(\omega) = \frac{1}{\sqrt{2}}\,\hat{g}\left(\frac{\omega}{2}\right)\hat{\phi}\left(\frac{\omega}{2}\right) = \frac{-i\omega}{4}\left(\frac{\sin(\omega/4)}{\omega/4}\right)^{m+2} \exp\left(\frac{-i\omega(1+\varepsilon)}{4}\right). \tag{5.66}$$

It is the first derivative of a box spline of degree $m+1$ centered at $t = (1+\varepsilon)/4$. For $m = 2$, Figure 5.3 shows the resulting quadratic splines ϕ and ψ. The dyadic admissibility condition (5.45) is verified numerically for $A = 0.505$ and $B = 0.522$.

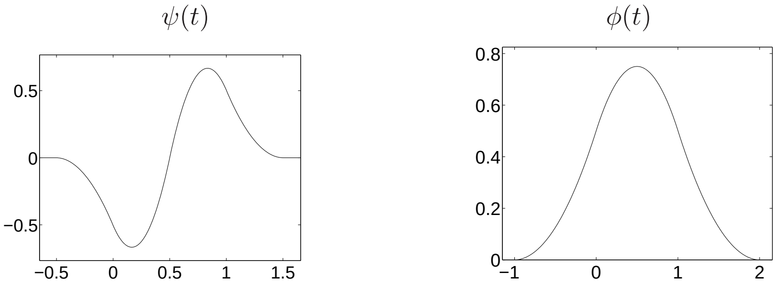


Figure 5.3: Quadratic spline wavelet and scaling function.

To design dual scaling functions $\tilde{\phi}$ and wavelets $\tilde{\psi}$ which are splines, we choose $\hat{\tilde{h}} = \hat{h}$. As a consequence, $\phi = \tilde{\phi}$ and the reconstruction condition (5.60) implies that

$$\hat{\tilde{g}}(\omega) = \frac{2 - |\hat{h}(\omega)|^2}{\hat{g}^*(\omega)} = -i\,\sqrt{2}\,\exp\left(\frac{-i\omega}{2}\right)\sin\frac{\omega}{2}\sum_{n=0}^{m}\left(\cos\frac{\omega}{2}\right)^{2n}. \tag{5.67}$$

Table 5.1 gives the corresponding filters for $m = 2$.

n	$h[n]/\sqrt{2}$	$\tilde{h}[n]/\sqrt{2}$	$g[n]/\sqrt{2}$	$\tilde{g}[n]/\sqrt{2}$
-2				-0.03125
-1	0.125	0.125		-0.21875
0	0.375	0.375	-0.5	-0.6875
1	0.375	0.375	0.5	0.6875
2	0.125	0.125		0.21875
3				0.03125

Table 5.1: Coefficients of the filters computed from their transfer functions (5.64, 5.65, 5.67) for $m = 2$. These filters generate the quadratic spline scaling functions and wavelets shown in Figure 5.3.

5.2.2 "Algorithme à Trous"

Suppose that the scaling functions and wavelets ϕ, ψ, $\tilde{\phi}$ and $\tilde{\psi}$ are designed with the filters h, g, $\tilde{h}$ and $\tilde{g}$. A fast dyadic wavelet transform is calculated with a filter bank algorithm called in French the *algorithme à trous*, introduced by Holschneider, Kronland-Martinet, Morlet and Tchamitchian [277]. It is similar to a fast biorthogonal wavelet transform, without subsampling [394, 334].

Fast Dyadic Transform The samples $a_0[n]$ of the input discrete signal are written as a low-pass filtering with ϕ of an analog signal f, in the neighborhood of $t = n$:

$$a_0[n] = f \star \bar{\phi}(n) = \langle f(t), \phi(t-n) \rangle = \int_{-\infty}^{+\infty} f(t)\, \phi(t-n)\, dt.$$

This is further justified in Section 7.3.1. For any $j \geqslant 0$, we denote

$$a_j[n] = \langle f(t), \phi_{2^j}(t-n) \rangle \quad \text{with} \quad \phi_{2^j}(t) = \frac{1}{\sqrt{2^j}} \phi\Big(\frac{t}{2^j}\Big) .$$

The dyadic wavelet coefficients are computed for $j > 0$ over the integer grid

$$d_j[n] = Wf(n, 2^j) = \langle f(t), \psi_{2^j}(t-n) \rangle.$$

For any filter $x[n]$, we denote by $x_j[n]$ the filters obtained by inserting $2^j - 1$ zeros between each sample of $x[n]$. Its Fourier transform is $\hat{x}(2^j \omega)$. Inserting zeros in the filters creates holes (*trous* in French). Let $\bar{x}_j[n] = x_j[-n]$. The next theorem gives convolution formulas that are cascaded to compute a dyadic wavelet transform and its inverse.

Theorem 5.14. *For any $j \geqslant 0$,*

$$a_{j+1}[n] = a_j \star \bar{h}_j[n] \quad , \qquad d_{j+1}[n] = a_j \star \bar{g}_j[n] \quad , \tag{5.68}$$

and

$$a_j[n] = \frac{1}{2} \Big(a_{j+1} \star \tilde{h}_j[n] + d_{j+1} \star \tilde{g}_j[n] \Big) . \tag{5.69}$$

Proof. Proof of (5.68). Since

$$a_{j+1}[n] = f \star \bar{\phi}_{2^{j+1}}(n) \quad \text{and} \quad d_{j+1}[n] = f \star \bar{\psi}_{2^{j+1}}(n),$$

we verify with (3.3) that their Fourier transforms are respectively

$$\hat{a}_{j+1}(\omega) = \sum_{k=-\infty}^{+\infty} \hat{f}(\omega + 2k\pi)\, \hat{\phi}^*_{2^{j+1}}(\omega + 2k\pi)$$

and

$$\hat{d}_{j+1}(\omega) = \sum_{k=-\infty}^{+\infty} \hat{f}(\omega + 2k\pi)\, \hat{\psi}^*_{2^{j+1}}(\omega + 2k\pi).$$

The properties (5.58) and (5.59) imply that

$$\hat{\phi}_{2^{j+1}}(\omega) = \sqrt{2^{j+1}}\,\hat{\phi}(2^{j+1}\omega) = \hat{h}(2^j\omega)\,\sqrt{2^j}\,\hat{\phi}(2^j\omega),$$

$$\hat{\psi}_{2^{j+1}}(\omega) = \sqrt{2^{j+1}}\,\hat{\psi}(2^{j+1}\omega) = \hat{g}(2^j\omega)\,\sqrt{2^j}\,\hat{\phi}(2^j\omega).$$

Since $j \geqslant 0$, both $\hat{h}(2^j\omega)$ and $\hat{g}(2^j\omega)$ are 2π periodic, so

$$\hat{a}_{j+1}(\omega) = \hat{h}^*(2^j\omega)\,\hat{a}_j(\omega) \quad\text{and}\quad \hat{d}_{j+1}(\omega) = \hat{g}^*(2^j\omega)\,\hat{a}_j(\omega). \tag{5.70}$$

These two equations are the Fourier transforms of (5.68).

Proof of (5.69). Equations (5.70) imply

$$\begin{aligned}\hat{a}_{j+1}(\omega)\,\hat{\tilde{h}}(2^j\omega) + \hat{d}_{j+1}(\omega)\,\hat{\tilde{g}}(2^j\omega) \quad &= \\ \hat{a}_j(\omega)\,\hat{h}^*(2^j\omega)\,\hat{\tilde{h}}(2^j\omega) \quad &+ \quad \hat{a}_j(\omega)\,\hat{g}^*(2^j\omega)\,\hat{\tilde{g}}(2^j\omega).\end{aligned}$$

Inserting the reconstruction condition (5.60) proves that

$$\hat{a}_{j+1}(\omega)\,\hat{\tilde{h}}(2^j\omega) + \hat{d}_{j+1}(\omega)\,\hat{\tilde{g}}(2^j\omega) = 2\,\hat{a}_j(\omega),$$

which is the Fourier transform of (5.69). ■ ■

The dyadic wavelet representation of a_0 is defined as the set of wavelet coefficients up to a scale 2^J plus the remaining low-frequency information a_J:

$$\Big[\{d_j\}_{1\leqslant j\leqslant J}\ ,\ a_J\Big]. \tag{5.71}$$

It is computed from a_0 by cascading the convolutions (5.68) for $0 \leqslant j < J$, as illustrated in Figure 5.4(a). The dyadic wavelet transform of Figure 5.2 is calculated with this filter bank algorithm. The original signal a_0 is recovered from its wavelet representation (5.71) by iterating (5.69) for $J > j \geqslant 0$, as illustrated in Figure 5.4(b).

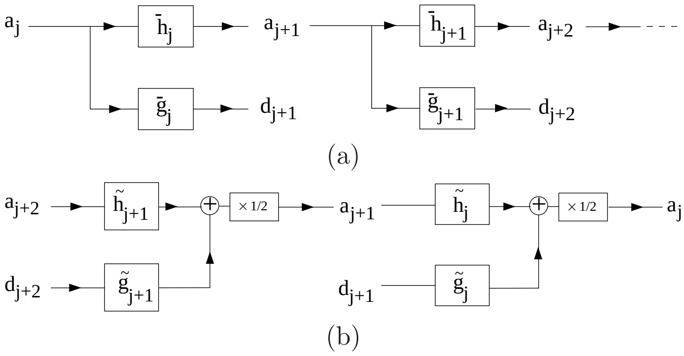


Figure 5.4: (a): The dyadic wavelet coefficients are computed by cascading convolutions with dilated filters $\bar{h}_j$ and $\bar{g}_j$. (b): The original signal is reconstructed through convolutions with $\tilde{h}_j$ and $\tilde{g}_j$. A multiplication by 1/2 is necessary to recover the next finer scale signal a_j.

If the input signal $a_0[n]$ has a finite size of N samples, the convolutions (5.68) are replaced by circular convolutions. The maximum scale 2^J is then limited to N, and for $J = \log_2 N$ one can verify that $a_J[n]$ is constant and equal to $N^{-1/2}\sum_{n=0}^{N-1} a_0[n]$. Suppose that h and g have respectively K_h and K_g non-zero samples. The "dilated" filters h_j and g_j have the same number of non-zero coefficients. The number of multiplications needed to compute a_{j+1} and d_{j+1} from a_j or the reverse is thus equal to $(K_h + K_g)N$. For $J = \log_2 N$, the dyadic wavelet representation (5.71) and its inverse are thus calculated with $(K_h + K_g)N \log_2 N$ multiplications and additions.

5.3 Subsampled Wavelet Frames

Wavelet frames are constructed by sampling the scale parameter but also the translation parameter of a wavelet dictionary. A real continuous wavelet transform of $f \in \mathbf{L^2}(\mathbb{R})$ is defined in Section 4.3 by

$$Wf(u,s) = \langle f, \psi_{u,s} \rangle \quad \text{with} \quad \psi_{u,s}(t) = \frac{1}{\sqrt{s}}\, \psi\left(\frac{t-u}{s}\right)$$

where ψ is a real wavelet. Imposing $\|\psi\| = 1$ implies that $\|\psi_{u,s}\| = 1$.

Intuitively, to construct a frame we need to cover the time-frequency plane with the Heisenberg boxes of the corresponding discrete wavelet family. A wavelet $\psi_{u,s}$ has an energy in time that is centered at u over a domain proportional to s. Over positive frequencies, its Fourier transform $\hat\psi_{u,s}$ has a support centered at a frequency η/s, with a spread proportional to $1/s$. To obtain a full cover, we sample s along an exponential sequence $\{a^j\}_{j\in\mathbb{Z}}$, with a sufficiently small dilation step $a > 1$. The time translation u is sampled uniformly at intervals proportional to the scale a^j, as illustrated in Figure 5.5. Let us denote

$$\psi_{j,n}(t) = \frac{1}{\sqrt{a^j}}\, \psi\left(\frac{t - n u_0 a^j}{a^j}\right).$$

In the following, we give without proofs some necessary conditions and sufficient conditions on ψ, a and u_0 so that $\{\psi_{j,n}\}_{(j,n)\in\mathbb{Z}^2}$ is a frame of $\mathbf{L^2}(\mathbb{R})$.

Necessary Conditions We suppose that ψ is real, normalized, and satisfies the admissibility condition of Theorem 4.4:

$$C_\psi = \int_0^{+\infty} \frac{|\hat\psi(\omega)|^2}{\omega}\, d\omega < +\infty. \tag{5.72}$$

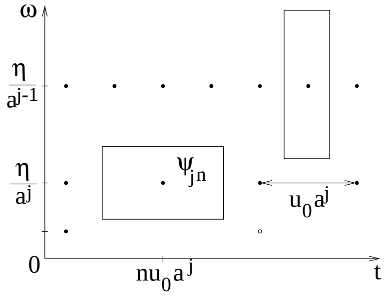


Figure 5.5: The Heisenberg box of a wavelet $\psi_{j,n}$ scaled by $s = a^j$ has a time and frequency width proportional respectively to a^j and a^{-j}. The time-frequency plane is covered by these boxes if u_0 and a are sufficiently small.

Theorem 5.15 (Daubechies). *If $\{\psi_{j,n}\}_{(j,n)\in\mathbb{Z}^2}$ is a frame of $\mathbf{L^2}(\mathbb{R})$ then the frame bounds satisfy*

$$A \leqslant \frac{C_\psi}{u_0 \log_e a} \leqslant B, \tag{5.73}$$

$$\forall \omega \in \mathbb{R} - \{0\} \quad , \quad A \leqslant \frac{1}{u_0} \sum_{j=-\infty}^{+\infty} |\hat\psi(a^j \omega)|^2 \leqslant B. \tag{5.74}$$

This theoroem is proved in [153, 17]. Condition (5.74) is equivalent to the frame condition (5.52) for a translation invariant dyadic wavelet transform, for which the parameter u is not sampled. It requires that the Fourier axis is covered by wavelets dilated by $\{a^j\}_{j\in\mathbb{Z}}$. The inequality (5.73), which relates the sampling density $u_0 \log_e a$ to the frame bounds, is proved in [17]. It shows that the frame is an orthonormal basis if and only if

$$A = B = \frac{C_\psi}{u_0 \log_e a} = 1.$$

Chapter 7 constructs wavelet orthonormal bases of $\mathbf{L^2}(\mathbb{R})$ with regular wavelets of compact support.

Sufficient Conditions The following theorem proved by Daubechies [17] provides a lower and upper bound for the frame bounds A and B, depending on ψ, u_0 and a.

Theorem 5.16 (Daubechies)**.** *Let us define*

$$\theta(\xi) = \sup_{1 \leqslant |\omega| \leqslant a} \sum_{j=-\infty}^{+\infty} |\hat{\psi}(a^j \omega)| \, |\hat{\psi}(a^j \omega + \xi)| \tag{5.75}$$

and

$$\Delta = \sum_{\substack{k=-\infty \\ k \neq 0}}^{+\infty} \left[\theta\left(\frac{2\pi k}{u_0}\right) \theta\left(\frac{-2\pi k}{u_0}\right) \right]^{1/2}.$$

If u_0 and a are such that

$$A_0 = \frac{1}{u_0} \left(\inf_{1 \leqslant |\omega| \leqslant a} \sum_{j=-\infty}^{+\infty} |\hat{\psi}(a^j \omega)|^2 - \Delta \right) > 0, \tag{5.76}$$

and

$$B_0 = \frac{1}{u_0} \left(\sup_{1 \leqslant |\omega| \leqslant a} \sum_{j=-\infty}^{+\infty} |\hat{\psi}(a^j \omega)|^2 + \Delta \right) < +\infty, \tag{5.77}$$

then $\{\psi_{j,n}\}_{(j,n) \in \mathbb{Z}^2}$ is a frame of $\mathbf{L^2}(\mathbb{R})$. The constants A_0 and B_0 are respectively lower and upper bounds of the frame bounds A and B.

The sufficient conditions (5.76) and (5.77) are similar to the necessary condition (5.74). If Δ is small relative to $\inf_{1 \leqslant |\omega| \leqslant a} \sum_{j=-\infty}^{+\infty} |\hat{\psi}(a^j \omega)|^2$ then A_0 and B_0 are close to the optimal frame bounds A and B. For a fixed dilation step a, the value of Δ decreases when the time sampling interval u_0 decreases.

Dual Frame Theorem 5.5 gives a general formula for computing the dual wavelet frame vectors

$$\tilde{\psi}_{j,n} = (\Phi^* \Phi)^{-1} \psi_{j,n}. \tag{5.78}$$

One could reasonably hope that the dual functions $\tilde{\psi}_{j,n}$ would be obtained by scaling and translating a dual wavelet $\tilde{\psi}$. The sad reality is that this is generally not true. In general the operator $\Phi^* \Phi$ does not commute with dilations by a^j, so $(\Phi^* \Phi)^{-1}$ does not commute with these dilations either. On the other hand, one can prove that $(\Phi^* \Phi)^{-1}$ commutes with translations by $n a^j u_0$, which means that

$$\tilde{\psi}_{j,n}(t) = \tilde{\psi}_{j,0}(t - n a^j u_0). \tag{5.79}$$

The dual frame $\{\tilde{\psi}_{j,n}\}_{(j,n) \in \mathbb{Z}^2}$ is thus obtained by calculating each elementary function $\tilde{\psi}_{j,0}$ with (5.78), and translating them with (5.79). The situation is much simpler for tight frames, where the dual frame is equal to the original wavelet frame.

Mexican Hat Wavelet The normalized second derivative of a Gaussian is

$$\psi(t) = \frac{2}{\sqrt{3}} \pi^{-1/4} (t^2 - 1) \exp\left(\frac{-t^2}{2}\right). \tag{5.80}$$

Its Fourier transform is

$$\hat{\psi}(\omega) = -\frac{\sqrt{8}\, \pi^{1/4}\, \omega^2}{\sqrt{3}} \exp\left(\frac{-\omega^2}{2}\right).$$

The graph of these functions is shown in Figure 4.6.

The dilation step a is generally set to be $a = 2^{1/v}$ where v is the number of intermediate scales (voices) for each octave. Table 5.2 gives the estimated frame bounds A_0 and B_0 computed by Daubechies [17] with the formula of Theorem 5.16. For $v \geqslant 2$ voices per octave, the frame is nearly

a	u_0	A_0	B_0	B_0/A_0
2	0.25	13.091	14.183	1.083
2	0.5	6.546	7.092	1.083
2	1.0	3.223	3.596	1.116
2	1.5	0.325	4.221	12.986
$2^{\frac{1}{2}}$	0.25	27.273	27.278	1.0002
$2^{\frac{1}{2}}$	0.5	13.673	13.639	1.0002
$2^{\frac{1}{2}}$	1.0	6.768	6.870	1.015
$2^{\frac{1}{2}}$	1.75	0.517	7.276	14.061
$2^{\frac{1}{4}}$	0.25	54.552	54.552	1.0000
$2^{\frac{1}{4}}$	0.5	27.276	27.276	1.0000
$2^{\frac{1}{4}}$	1.0	13.586	13.690	1.007
$2^{\frac{1}{4}}$	1.75	2.928	12.659	4.324

Table 5.2: Estimated frame bounds for the Mexican hat wavelet computed with Theorem 5.16 [17].

tight when $u_0 \leqslant 0.5$, in which case the dual frame can be approximated by the original wavelet frame. As expected from (5.73), when $A \approx B$

$$A \approx B \approx \frac{C_\psi}{u_0 \log_{\mathrm{e}} a} = \frac{v}{u_0} C_\psi \log_2 \mathrm{e}.$$

The frame bounds increase proportionally to v/u_0. For $a = 2$, we see that A_0 decreases brutally from $u_0 = 1$ to $u_0 = 1.5$. For $u_0 = 1.75$ the wavelet family is not a frame anymore. For $a = 2^{1/2}$, the same transition appears for a larger u_0.

5.4 Windowed Fourier Frames

Frame theory gives conditions for discretizing the windowed Fourier transform while retaining a complete and stable representation. The windowed Fourier transform of $f \in \mathbf{L^2}(\mathbb{R})$ is defined in Section 4.2 by

$$Sf(u,\xi) = \langle f, g_{u,\xi} \rangle,$$

with

$$g_{u,\xi}(t) = g(t-u)\, \mathrm{e}^{i\xi t}.$$

Setting $\|g\| = 1$ implies that $\|g_{u,\xi}\| = 1$. A discrete windowed Fourier transform representation

$$\{Sf(u_n,\xi_k) = \langle f, g_{u_n,\xi_k} \rangle\}_{(n,k)\in\mathbb{Z}^2}$$

is complete and stable if $\{g_{u,n,\xi_k}\}_{(n,k)\in\mathbb{Z}^2}$ is a frame of $\mathbf{L^2}(\mathbb{R})$.

Intuitively, one can expect that the discrete windowed Fourier transform is complete if the Heisenberg boxes of all atoms $\{g_{u,n,\xi_k}\}_{(n,k)\in\mathbb{Z}^2}$ fully cover the time-frequency plane. Section 4.2 shows that the Heisenberg box of g_{u_n,ξ_k} is centered in the time-frequency plane at (u_n, ξ_k). Its size is independent of u_n and ξ_k. It depends on the time-frequency spread of the window g. A complete cover of the plane is thus obtained by translating these boxes over a uniform rectangular grid, as illustrated in Figure 5.6. The time and frequency parameters (u,ξ) are discretized over a rectangular grid with time and frequency intervals of size u_0 and ξ_0. Let us denote

$$g_{n,k}(t) = g(t - nu_0)\, \exp(ik\xi_0 t).$$

The sampling intervals (u_0, ξ_0) must be adjusted to the time-frequency spread of g.

Window Scaling Suppose that $\{g_{n,k}\}_{(n,k)\in\mathbb{Z}^2}$ is a frame of $\mathbf{L^2}(\mathbb{R})$ with frame bounds A and B. Let us dilate the window $g_s(t) = s^{-1/2} g(t/s)$. It increases by s the time width of the Heisenberg box of g and reduces by s its frequency width. We thus obtain the same cover of the time-frequency plane by increasing u_0 by s and reducing ξ_0 by s. Let

$$g_{s,n,k}(t) = g_s(t - nsu_0)\,\exp\left(ik\frac{\xi_0}{s}t\right). \tag{5.81}$$

We prove that $\{g_{s,n,k}\}_{(n,k)\in\mathbb{Z}^2}$ satisfies the same frame inequalities as $\{g_{n,k}\}_{(n,k)\in\mathbb{Z}^2}$, with the same frame bounds A and B, by a change of variable $t' = ts$ in the inner product integrals.

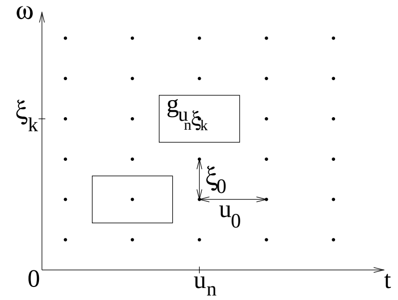


Figure 5.6: A windowed Fourier frame is obtained by covering the time-frequency plane with a regular grid of windowed Fourier atoms, translated by $u_n = n\,u_0$ in time and by $\xi_k = k\,\xi_0$ in frequency.

Tight Frames Tight frames are easier to manipulate numerically since the dual frame is equal to the original frame. Daubechies, Grossmann and Meyer [184] give sufficient conditions for building a window of compact support that generates a tight frame.

Theorem 5.17 (Daubechies, Grossmann, Meyer)**.** *Let g be a window whose support is included in $[-\pi/\xi_0, \pi/\xi_0]$. If*

$$\forall t \in \mathbb{R} \quad , \quad \frac{2\pi}{\xi_0} \sum_{n=-\infty}^{+\infty} |g(t - nu_0)|^2 = A > 0 \tag{5.82}$$

then $\{g_{n,k}(t) = g(t - nu_0)\,\mathrm{e}^{ik\xi_0 t}\}_{(n,k)\in\mathbb{Z}^2}$ is a tight frame $\mathbf{L^2}(\mathbb{R})$ with a frame bound equal to A.

Proof. The function $g(t - nu_0)\,f(t)$ has a support in $[nu_0 - \pi/\xi_0, nu_0 + \pi/\xi_0]$. Since $\{\mathrm{e}^{ik\xi_0 t}\}_{k\in\mathbb{Z}}$ is an orthogonal basis of this space we have

$$\int_{-\infty}^{+\infty} |g(t - nu_0)|^2\,|f(t)|^2 dt = \int_{nu_0-\pi/\xi_0}^{nu_0+\pi/\xi_0} |g(t - nu_0)|^2\,|f(t)|^2 dt = \frac{\xi_0}{2\pi} \sum_{k=-\infty}^{+\infty} |\langle g(u - nu_0)\,f(u), \mathrm{e}^{ik\xi_0 u}\rangle|^2.$$

Since $g_{n,k}(t) = g(t - nu_0)\,\mathrm{e}^{ik\xi_0 t}$, we get

$$\int_{-\infty}^{+\infty} |g(t - nu_0)|^2\,|f(t)|^2 dt = \frac{\xi_0}{2\pi} \sum_{k=-\infty}^{+\infty} |\langle f, g_{n,k}\rangle|^2.$$

Summing over n and inserting (5.82) proves that $A\,\|f\|^2 = \sum_{k,n=-\infty}^{+\infty} |\langle f, g_{n,k}\rangle|^2$ and hence that $\{g_{n,k}\}_{(n,k)\in\mathbb{Z}^2}$ is a tight frame of $\mathbf{L^2}(\mathbb{R})$. ■

Since g has a support in $[-\pi/\xi_0, \pi/\xi_0]$ the condition (5.82) implies that

$$\frac{2\pi}{u_0\,\xi_0} \geqslant 1$$

so that there is no whole between consecutive windows $g(t-nu_0)$ and $g(t-(n+1)u_0)$. If we impose that $1 \leqslant 2\pi/(u_0\xi_0) \leqslant 2$ then only consecutive windows have supports that overlap. The square root of a Hanning window

$$g(t)=\sqrt{\frac{\xi_0}{\pi}}\cos\left(\frac{\xi_0 t}{2}\right)\mathbf{1}_{[-\pi/\xi_0,\pi/\xi_0]}(t)$$

is positive normalized window that satisfies (5.82) with $u_0=\pi/\xi_0$ and a redundancy factor $A=2$. The design of other windows is studied in Section 8.4.2 for local cosine bases.

Discrete Window Fourier Tight Frames To construct a windowed Fourier tight frame of $\mathbb{C}^N$, the Fourier basis $\{e^{ik\xi_0 t}\}_{k\in\mathbb{Z}}$ of $\mathbf{L^2}[-\pi/\xi_0,\pi/\xi_0]$ is replaced by the discrete Fourier basis $\{e^{i2\pi kn/K}\}_{0\leqslant k<K}$ of $\mathbb{C}^K$. The following theorem is a discrete equivalent of Theorem 5.17.

Theorem 5.18. *Let $g[n]$ be an N periodic discrete window whose support restricted to $[-N/2,N/2]$ is included in $[-K/2,K/2-1]$. If M divides N and*

$$\forall 0\leqslant n<N\ ,\quad K\sum_{m=0}^{N/M-1}|g[n-mM]|^2=A>0 \tag{5.83}$$

then $\{g_{m,k}[n]=g[n-mM]\,e^{i2\pi kn/K}\}_{0\leqslant k<K,0\leqslant m<N/M}$ is a tight frame $\mathbb{C}^N$ with a frame bound equal to A.

The proof of this theorem follows the same steps as the proof of Theorem 5.18. It is left in Exercise 5.4. There are N/M translated windows and hence NK/M windowed Fourier coefficients. For a fixed window position indexed by m, the discrete windowed Fourier coefficients are the discrete Fourier coefficients of the windowed signal

$$Sf[m,k]=\langle f,g_{m,k}\rangle=\sum_{n=K/2}^{K/2-1}f[n]\,g[n-mM]\,e^{-i2\pi kn/K}\quad\text{for}\quad 0\leqslant k<K.$$

They are computed with $O(K\log_2 K)$ operations with an FFT. Over all windows, this requires a total of $O(NK/M\log_2 K)$ operations. We generally choose $1<K/M\leqslant 2$ so that only consecutive windows overlap. The square root of a Hanning window $g[n]=\sqrt{2/K}\cos(\pi n/K)$ satisfies (5.83) for $M=K/2$ and a redundancy factor $A=2$. Figure 5.7 shows the log spectrogram $\log|Sf[m,k]|^2$ of the windowed Fourier frame coefficients computed with a square root Hanning window for a musical recording.

Necessary Frame Conditions For general windowed Fourier frames of $\mathbf{L^2}(\mathbb{R}^2)$, Daubechies [17] proved several necessary conditions on g, u_0 and ξ_0 to guarantee that $\{g_{n,k}\}_{(n,k)\in\mathbb{Z}^2}$ is a frame of $\mathbf{L^2}(\mathbb{R})$. We do not reproduce the proofs, but summarize the main results.

Theorem 5.19 (Daubechies)**.** *The windowed Fourier family $\{g_{n,k}\}_{(n,k)\in\mathbb{Z}^2}$ is a frame only if*

$$\frac{2\pi}{u_0\,\xi_0}\geqslant 1. \tag{5.84}$$

The frame bounds A and B necessarily satisfy

$$A\leqslant\frac{2\pi}{u_0\,\xi_0}\leqslant B, \tag{5.85}$$

$$\forall t\in\mathbb{R}\ ,\quad A\leqslant\frac{2\pi}{\xi_0}\sum_{n=-\infty}^{+\infty}|g(t-nu_0)|^2\leqslant B, \tag{5.86}$$

$$\forall\omega\in\mathbb{R}\ ,\quad A\leqslant\frac{1}{u_0}\sum_{k=-\infty}^{+\infty}|\hat g(\omega-k\xi_0)|^2\leqslant B. \tag{5.87}$$

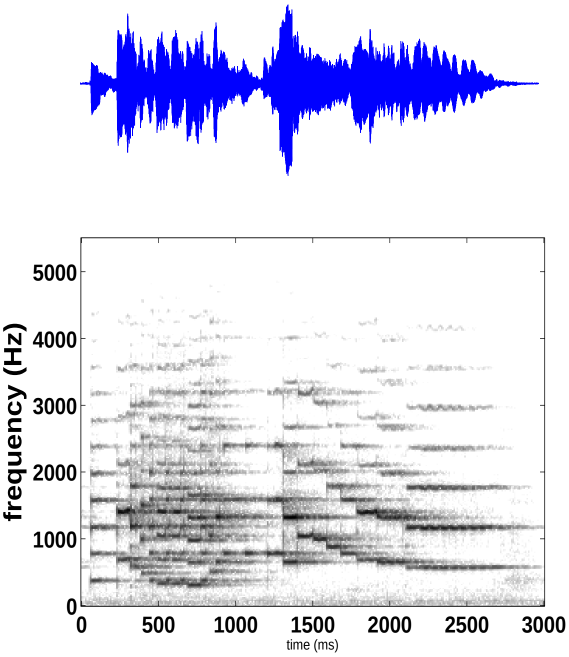


Figure 5.7: Top: Musical recording. Bottom: Log spectrogram $\log|Sf[m,k]|^2$ computed with a square root Hanning window.

The ratio $2\pi/(u_0\xi_0)$ measures the density of windowed Fourier atoms in the time-frequency plane. The first condition (5.84) ensures that this density is greater than 1 because the covering ability of each atom is limited. The inequalities (5.86) and (5.87) are proved in full generality by Chui and Shi [153]. They show that the uniform time translations of g must completely cover the time axis, and the frequency translations of its Fourier transform $\hat{g}$ must similarly cover the frequency axis.

Since all windowed Fourier vectors are normalized, the frame is an orthogonal basis only if $A = B = 1$. The frame bound condition (5.85) shows that this is possible only at the critical sampling density $u_0\xi_0 = 2\pi$. The Balian-Low Theorem [89] proves that g is then either non-smooth or has a slow time decay.

Theorem 5.20 (Balian-Low)**.** *If* $\{g_{n,k}\}_{(n,k)\in\mathbb{Z}^2}$ *is a windowed Fourier frame with* $u_0\xi_0 = 2\pi$*, then*

$$\int_{-\infty}^{+\infty} t^2\,|g(t)|^2\,dt = +\infty \quad \text{or} \quad \int_{-\infty}^{+\infty} \omega^2\,|\hat{g}(\omega)|^2\,d\omega = +\infty. \tag{5.88}$$

This theorem proves that we cannot construct an orthogonal windowed Fourier basis with a differentiable window g of compact support. On the other hand, one can verify that the discontinuous rectangular window

$$g = \frac{1}{\sqrt{u_0}}\,\mathbf{1}_{[-u_0/2,u_0/2]}$$

yields an orthogonal windowed Fourier basis for $u_0\xi_0 = 2\pi$. This basis is rarely used because of the bad frequency localization of $\hat{g}$.

Sufficient Conditions The following theorem proved by Daubechies [182] gives sufficient conditions on u_0, ξ_0 and g for constructing a windowed Fourier frame.

Theorem 5.21 (Daubechies)**.** *Let us define*

$$\theta(u) = \sup_{0\leqslant t\leqslant u_0} \sum_{n=-\infty}^{+\infty} |g(t-nu_0)|\,|g(t-nu_0+u)| \tag{5.89}$$

and

$$\Delta = \sum_{\substack{k=-\infty \\ k\neq 0}}^{+\infty} \left[\theta\left(\frac{2\pi k}{\xi_0}\right) \theta\left(\frac{-2\pi k}{\xi_0}\right) \right]^{1/2}. \tag{5.90}$$

If u_0 *and* ξ_0 *satisfy*

$$A_0 = \frac{2\pi}{\xi_0} \left(\int_{0\leqslant t\leqslant u_0} \sum_{n=-\infty}^{+\infty} |g(t-nu_0)|^2 - \Delta \right) > 0 \tag{5.91}$$

and

$$B_0 = \frac{2\pi}{\xi_0} \left(\sup_{0\leqslant t\leqslant u_0} \sum_{n=-\infty}^{+\infty} |g(t-nu_0)|^2 + \Delta \right) < +\infty, \tag{5.92}$$

then $\{g_{n,k}\}_{(n,k)\in\mathbb{Z}^2}$ *is a frame. The constants* A_0 *and* B_0 *are respectively lower bounds and upper bounds of the frame bounds* A *and* B.

Observe that the only difference between the sufficient conditions (5.91, 5.92) and the necessary condition (5.86) is the addition and subtraction of Δ. If Δ is small compared to $\inf_{0\leqslant t\leqslant u_0} \sum_{n=-\infty}^{+\infty} |g(t-nu_0)|^2$ then A_0 and B_0 are close to the optimal frame bounds A and B.

Dual Frame Theorem 5.5 proves that the dual windowed frame vectors are

$$\tilde{g}_{n,k} = (\Phi^*\Phi)^{-1} g_{n,k}. \tag{5.93}$$

The following theorem shows that this dual frame is also a windowed Fourier frame, which means that its vectors are time and frequency translations of a new window $\tilde{g}$.

Theorem 5.22. *Dual windowed Fourier vectors can be rewritten*

$$\tilde{g}_{n,k}(t) = \tilde{g}(t-nu_0)\,\exp(ik\xi_0 t)$$

where $\tilde{g}$ *is the dual window*

$$\tilde{g} = (\Phi^*\Phi)^{-1} g. \tag{5.94}$$

Proof. This result is proved by showing first that $\Phi^*\Phi$ commutes with time and frequency translations proportional to u_0 and ξ_0. If $\phi \in \mathbf{L^2}(\mathbb{R})$ and $\phi_{m,l}(t) = \phi(t-mu_0)\exp(il\xi_0 t)$ we verify that

$$\Phi^*\Phi\phi_{m,l}(t) = \exp(il\xi_0 t)\,\Phi^*\Phi h(t-mu_0).$$

Indeed

$$\Phi^*\Phi\phi_{m,l} = \sum_{(n,k)\in\mathbb{Z}^2} \langle \phi_{m,l}, g_{n,k}\rangle\, g_{n,k}$$

and a change of variable yields

$$\langle \phi_{m,l}, g_{n,k}\rangle = \langle \phi, g_{n-m,k-l}\rangle.$$

Consequently

$$\begin{aligned} \Phi^*\Phi\phi_{m,l}(t) &= \sum_{(n,k)\in\mathbb{Z}^2} \langle \phi, g_{n-m,k-l}\rangle\, \exp(il\xi_0 t)\, g_{n-m,k-l}(t-mu_0) \\ &= \exp(il\xi_0 t)\,\Phi^*\Phi\phi(t-mu_0). \end{aligned}$$

Since $\Phi^*\Phi$ commutes with these translations and frequency modulations we verify that $(\Phi^*\Phi)^{-1}$ necessarily commutes with the same group operations. Hence

$$\tilde{g}_{n,k}(t) = (\Phi^*\Phi)^{-1} g_{n,k} = \exp(ik\xi_0)\,(\Phi^*\Phi)^{-1} g_{0,0}(t-nu_0) = \exp(ik\xi_0)\,\tilde{g}(t-nu_0).$$

■ ■

$u_0\xi_0$	A_0	B_0	B_0/A_0
$\pi/2$	3.9	4.1	1.05
$3\pi/4$	2.5	2.8	1.1
π	1.6	2.4	1.5
$4\pi/3$	0.58	2.1	3.6
1.9π	0.09	2.0	22

Table 5.3: Frame bounds estimated with Theorem 5.21 for the Gaussian window (5.95) and $u_0 = \xi_0$.

Gaussian Window The Gaussian window

$$g(t) = \pi^{-1/4} \exp\left(\frac{-t^2}{2}\right) \tag{5.95}$$

has a Fourier transform $\hat{g}$ that is a Gaussian with the same variance. The time and frequency spreads of this window are identical. We therefore choose equal sampling intervals in time and frequency: $u_0 = \xi_0$. For the same product $u_0\xi_0$ other choices would degrade the frame bounds. If g is dilated by s then the time and frequency sampling intervals must become su_0 and ξ_0/s.

If the time-frequency sampling density is above the critical value: $2\pi/(u_0\xi_0) > 1$, then Daubechies [182] proves that $\{g_{n,k}\}_{(n,k)\in\mathbb{Z}^2}$ is a frame. When $u_0\xi_0$ tends to 2π, the frame bound A tends to 0. For $u_0\xi_0 = 2\pi$, the family $\{g_{n,k}\}_{(n,k)\in\mathbb{Z}^2}$ is complete in $\mathbf{L^2}(\mathbb{R})$, which means that any $f \in \mathbf{L^2}(\mathbb{R})$ is entirely characterized by the inner products $\{\langle f, g_{n,k}\rangle\}_{(n,k)\in\mathbb{Z}^2}$. However, the Balian-Low Theorem 5.20 proves that it cannot be a frame and one can indeed verify that $A = 0$ [182]. This means that the reconstruction of f from these inner products is unstable.

Table 5.3 gives the estimated frame bounds A_0 and B_0 calculated with Theorem 5.21, for different values of $u_0 = \xi_0$. For $u_0\xi_0 = \pi/2$, which corresponds to time and frequency sampling intervals that are half the critical sampling rate, the frame is nearly tight. As expected, $A \approx B \approx 4$, which verifies that the redundancy factor is 4 (2 in time and 2 in frequency). Since the frame is almost tight, the dual frame is approximately equal to the original frame, which means that $\tilde{g} \approx g$. When $u_0\xi_0$ increases we see that A decreases to zero and $\tilde{g}$ deviates more and more from a Gaussian. In the limit $u_0\xi_0 = 2\pi$, the dual window $\tilde{g}$ is a discontinuous function that does not belong to $\mathbf{L^2}(\mathbb{R})$. These results can be extended to discrete windowed Fourier transforms computed with a discretized Gaussian window [458].

5.5 Multiscale Directional Frames for Images

To reveal geometric image properties, wavelet frames are constructed with mother wavelets having a direction selectivity, providing information on the direction of sharp transitions such as edges and textures. Directional wavelet frames described in Section 5.5.1.

Wavelet frames yield high amplitude coefficients in the neighborhood of edges, and can not take advantage of their geometric regularity to improve the sparsity of the representation. Curvelet frames, described in Section 5.5.2, are constructed with elongated waveforms that follow directional image structures and improve the representation sparsity.

5.5.1 Directional Wavelet Frames

A directional wavelet transform decomposes images over directional wavelets that are translated, rotated and dilated at dyadic scales. Such transforms appear in many image processing applications and physiological models. Applications to texture discrimination are discussed.

A directional wavelet $\psi^\alpha(x)$ with $x = (x_1, x_2) \in \mathbb{R}^2$ of angle α is a wavelet having p directional vanishing moments along any one-dimensional line of direction $\alpha + \pi/2$ in the plane:

$$\forall \rho \in \mathbb{R} \ , \quad \int \psi^\alpha(\rho\cos\alpha - u\sin\alpha, \rho\sin\alpha + u\cos\alpha)\, u^k\, du = 0 \ \text{ for } \ 0 \leqslant k < p \ , \tag{5.96}$$

but which does not have directional vanishing moments along the direction α. Such a wavelet oscillates in the direction of $\alpha + \pi/2$ but not in the direction α. It is orthogonal to any two-dimensional polynomial of degree strictly smaller than p (Exercise 5.1).

The set Θ of chosen directions are typically uniform in $[0, \pi]$: $\Theta = \{\alpha = k\pi/K \text{ for } 0 \leqslant k < K\}$. Dilating these directional wavelets by factors 2^j and translating them by any $u \in \mathbb{R}$ yields a translation invariant directional wavelet family

$$\{\psi^{\alpha}_{2^j}(x-u)\}_{u\in\mathbb{R}^2, j\in\mathbb{Z}, \alpha\in\Theta} \quad \text{with} \quad \psi^{\alpha}_{2^j}(x) = 2^{-j}\psi^{\alpha}(2^{-j}x) \; . \tag{5.97}$$

Directional wavelets may be derived by rotating a single mother wavelet $\psi(x_1, x_2)$ having vanishing moments in the horizontal direction, with a rotation operator R_α of angle α in $\mathbb{R}^2$.

A dyadic directional wavelet transform of f computes inner product with each wavelet:

$$Wf(u, 2^j, \alpha) = \langle f, \, \psi^{\alpha}_{2^j,u}\rangle \qquad \text{where} \qquad \psi^{\alpha}_{2^j,u}(x) = \psi^{\alpha}_{2^j}(x-u).$$

This dyadic wavelet transform can also be written as convolutions with directional wavelets:

$$Wf(u, 2^j, \alpha) = f \star \bar{\psi}^{\alpha}_{2^j}(u) \quad \text{where} \quad \bar{\psi}^{\alpha}_{2^j}(x) = \psi^{\alpha}_{2^j}(-x).$$

A wavelet $\psi^{\alpha}_{2^j}(x-u)$ has a support dilated by 2^j, located in the neighborhood of u and oscillates in the direction of $\alpha + \pi/2$. If $f(x)$ is constant over the support of $\psi^{\alpha}_{2^j,u}$ along lines of direction $\alpha + \pi/2$ then $\langle f, \psi^{\alpha}_{2^j,u}\rangle = 0$ because of its directional vanishing moments. In particular, this coefficient vanishes in the neighborhood of an edge whose tangent has a direction $\alpha + \pi/2$ or any structure that is geometrically regular in the direction α. If the edge angle deviates from $\alpha + \pi/2$ then it produces large amplitude coefficients, with a maximum typically when the edge has a direction α. The amplitude of wavelet coefficients thus depends upon the local orientation of the image structures. In Figure 5.9, the central texture has a vertical direction (regular vertically) and thus creates small amplitude Gabor wavelet coefficients for $\alpha = 0$, and large amplitude coefficients for $\alpha = \pi/2$. Similarly, the amplitude of steerable wavelet coefficients in Figure 5.10 depends upon the local orientation of edges, and vanishes when the edge direction is aligned with the vanishing moment direction.

Theorem 5.11 proves that the translation invariant wavelet family is a frame if there exist $B \geqslant A > 0$ such that the generators $\phi_n(x) = 2^{-j}\psi^{\alpha}_{2^j}(x)$ have Fourier transforms $\hat{\phi}_n(\omega) = \hat{\psi}^{\alpha}(\omega)$ which satisfy

$$\forall \omega = (\omega_1, \omega_2) \in \mathbb{R}^2 - \{(0,0)\} \quad , \quad A \leqslant \sum_{\alpha\in\Theta} \sum_{j=-\infty}^{+\infty} |\hat{\psi}^{\alpha}(2^j\omega)|^2 \leqslant B. \tag{5.98}$$

It results from Theorem 5.11 that there exists a dual family of reconstructing wavelets $\{\tilde{\psi}^{\alpha}\}_{\alpha\in\Theta}$ whose Fourier transforms satisfy

$$\sum_{j=-\infty}^{+\infty} \sum_{\alpha\in\Theta} \widehat{\tilde{\psi}^{\alpha}}(2^j\omega)\, \widehat{\psi}^{\alpha*}(2^j\omega) = 1 \, , \tag{5.99}$$

which yields

$$f(x) = \sum_{j=-\infty}^{+\infty} \frac{1}{2^{2j}} \sum_{\alpha\in\Theta} Wf(\cdot, 2^j, \alpha) \star \tilde{\psi}^{\alpha}_{2^j}(x) \, . \tag{5.100}$$

Examples of directional wavelets obtained by rotating a single mother wavelet are constructed with Gabor functions and steerable derivatives.

Gabor Wavelets In the cat's visual cortex, Hubel and Wiesel [279] discovered a class of cells, called simple cells, whose responses depend on the frequency and direction of the visual stimuli. Numerous physiological experiments [363] have shown that these cells can be modeled as linear filters, whose impulse responses have been measured at different locations of the visual cortex.

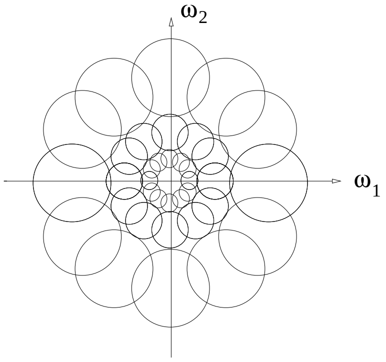


Figure 5.8: Each circle represents the frequency domain in a direction $\alpha+\pi/2$ where the amplitude of a Gabor wavelet Fourier transform $|\hat{\psi}^{\alpha}_{2^j}(\omega)|$ is large. It is proportional to 2^{-j} and its position rotates with α.

Daugmann [187] showed that these impulse responses can be approximated by *Gabor wavelets*, obtained with a Gaussian window $g(x_1,x_2)=(2\pi)^{-1}\mathrm{e}^{-(x_1^2+x_2^2)/2}$ multiplied by a sinusoidal wave:

$$\psi^{\alpha}(x_1,x_2)=g(x_1,x_2)\,\exp[-i\eta(-x_1\sin\alpha+x_2\cos\alpha)]. \tag{5.101}$$

These findings suggest the existence of some sort of wavelet transform in the visual cortex, combined with subsequent non-linearities [365]. The "physiological" wavelets have a frequency resolution on the order of 1–1.5 octaves, and are thus similar to dyadic wavelets.

The Fourier transform of $g(x_1,x_2)$ is $\hat{g}(\omega_1,\omega_2)=\mathrm{e}^{-(\omega_1^2+\omega_2^2)/2}$. It results from (5.101) that

$$\hat{\psi}^{\alpha}_{2^j}(\omega_1,\omega_2)=\sqrt{2^j}\,\hat{g}(2^j\omega_1+\eta\sin\alpha,2^j\omega_2-\eta\cos\alpha).$$

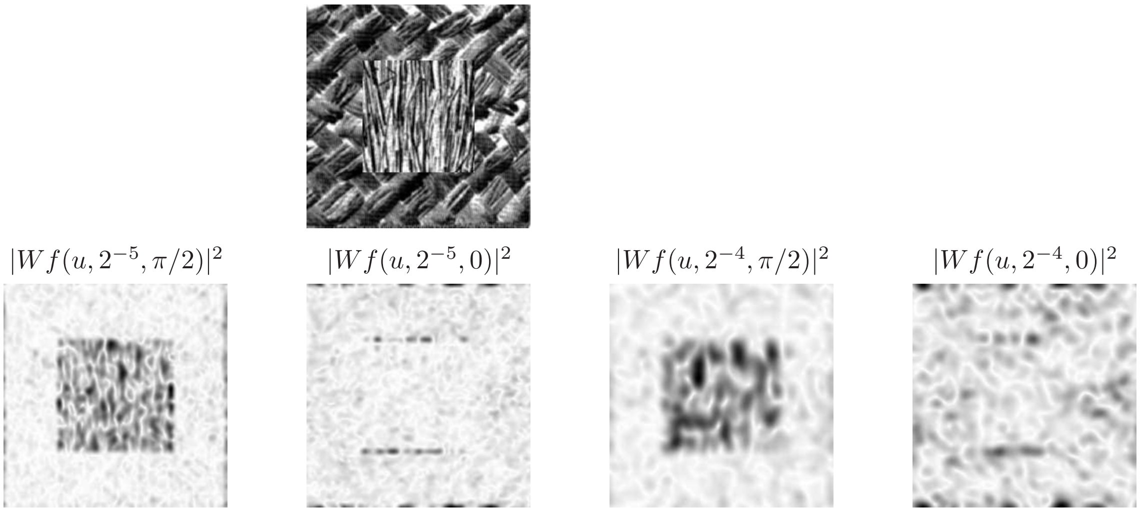
In the Fourier plane, the energy of this Gabor wavelet is mostly concentrated around $(-2^{-j}\eta\sin\alpha,2^{-j}\eta\cos\alpha)$, in a neighborhood proportional to 2^{-j}.

The direction are chosen to be uniform $\alpha=l\pi/K$ for $-K<l\leqslant K$. Figure 5.8 shows a cover of the frequency plane by dyadic Gabor wavelets (5.101) with $K=6$. If $K\geqslant 4$ and η is of the order of 1 then (5.98) is satisfied with stable bounds. Since images $f(x)$ are real, $\hat{f}(-\omega)=\hat{f}^*(\omega)$, and f can thus be reconstructed by covering only half of the frequency plane, with $-K<l\leqslant 0$. This is a two-dimensional equivalent of the one-dimensional analytic wavelet transform, studied in Section 4.3.2, with wavelets having a Fourier transform support restricted to positive frequencies. For texture analysis, Gabor wavelets provide information on the local image frequencies.

Texture Discrimination Despite many attempts, there are no appropriate mathematical models for "homogeneous image textures." The notion of texture homogeneity is still defined with respect to our visual perception. A texture is said to be homogeneous if it is preattentively perceived as being homogeneous by a human observer.

The texton theory of Julesz [293] was a first important step in understanding the different parameters that influence the perception of textures. The direction of texture elements and their frequency content seem to be important clues for discrimination. This motivated early researchers to study the repartition of texture energy in the Fourier domain [88]. For segmentation purposes, it is however necessary to localize texture measurements over neighborhoods of varying sizes. The Fourier transform was thus replaced by localized energy measurements at the output of filter banks that compute a wavelet transform [288, 307, 364, 428]. Besides the algorithmic efficiency of this approach, this model is partly supported by physiological studies of the visual cortex.

Since $Wf(u,2^j,\alpha)=f\star\bar{\psi}^{\alpha}_{2^j}(u)$, Gabor wavelet coefficients measure the energy of f in a spatial neighborhood of u of size 2^j, and in a frequency neighborhood of $(-2^{-j}\eta\sin\alpha,2^{-j}\eta\cos\alpha)$ of size 2^{-j}, where is located the support of $\hat{\psi}^{\alpha}_{2^j}(\omega)$ illustrated in Figure 5.8. Varying the scale 2^j and the angle α modifies the frequency channel [111]. The wavelet transform energy $|Wf(u,2^j,\alpha)|^2$ is large when the angle α and scale 2^j match the direction and scale of high energy texture components in the neighborhood of u. The amplitude of $|Wf(u,2^j,\alpha)|^2$ can thus be used to discriminate textures.



$|Wf(u,2^{-5},\pi/2)|^2$ $|Wf(u,2^{-5},0)|^2$ $|Wf(u,2^{-4},\pi/2)|^2$ $|Wf(u,2^{-4},0)|^2$

Figure 5.9: Directional Gabor wavelet transform $|Wf(u,2^j,\alpha)|^2$ of a texture patch, at the scales $2^j = 2^{-4}, 2^{-5}$, along two directions $\alpha = 0, \pi/2$. The darker a pixel, the larger the wavelet coefficient amplitude.

Figure 5.9 shows the dyadic wavelet transform of two textures, computed along horizontal and vertical directions, at the scales 2^{-4} and 2^{-5} (the image support is normalized to $[0,1]^2$). The central texture is regular vertically and has more energy along horizontal high frequencies than the peripheric texture. These two textures are therefore discriminated by the wavelet of angle $\alpha = \pi/2$ whereas the other wavelet with $\alpha = 0$ produces similar responses for both textures.

For segmentation, one must design an algorithm that aggregates the wavelet responses at all scales and directions in order to find the boundaries of homogeneous textured regions. Both clustering procedures and detection of sharp transitions over wavelet energy measurements have been used to segment the image [288, 364, 428]. These algorithms work well experimentally but rely on ad hoc parameter settings.

Steerable Wavelets Steerable directional wavelets along any angle α can be written as a linear expansion of few mother wavelets [402]. For example, a steerable wavelet in the direction α can be defined as the partial derivative of order p of a window $\theta(x)$ in the direction of the vector $\vec{n} = (-\sin\alpha, \cos\alpha)$:

$$\psi^\alpha(x) = \frac{\partial^p \theta(x)}{\partial \vec{n}^p} = \Big(-\sin\alpha \,\frac{\partial}{\partial x_1} + \cos\alpha \,\frac{\partial}{\partial x_2}\Big)^p \theta(x). \tag{5.102}$$

Let R_α be the planar rotation by an angle α. If the window is invariant under rotations, $\theta(x) = \theta(R_\alpha x)$, then these wavelets are generated by the rotation of a single mother wavelet: $\psi^\alpha(x) = \psi(R_\alpha x)$ with $\psi = \partial^p\theta / \partial x_2^p$.

Furthermore, the expansion of the derivatives in (5.102) proves that each ψ^α can be expressed as a linear combination of $p+1$ partial derivatives

$$\psi^\alpha(x) = \sum_{i=0}^{p} a_i(\alpha)\, \rho^i(x), \quad \text{where} \quad a_i(\alpha) = \begin{pmatrix} p \\ i \end{pmatrix} (-\sin\alpha)^i \,(\cos\alpha)^{p-i}, \tag{5.103}$$

with

$$\forall\, 0 \leqslant i \leqslant p, \quad \rho^i(x) = \frac{\partial^p \theta(x)}{\partial x_1^i \, \partial x_2^{p-i}}.$$

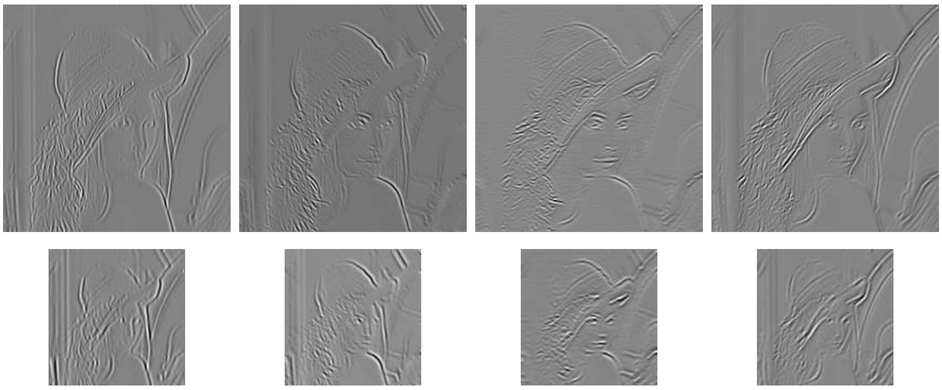
The waveforms $\rho^i(x)$ can also be considered as wavelets functions with vanishing moments. It results from (5.103), that the directional wavelet transform at any angle α can be calculated from

$p+1$ convolutions of f with the ρ^i dilated:

$$Wf(u,2^j,\alpha) = \sum_{i=0}^{p} a_i(\alpha)(f \star \bar{\rho}^i_{2^j})(u) \quad \text{for} \quad \bar{\rho}^i_{2^j}(x) = 2^{-j}\rho^i(-2^{-j}x) \ .$$

Exercise 5.6 gives conditions on θ so that for a set Θ of $p+1$ angles $\alpha = k\pi/(p+1)$ with $0 \leqslant k < p$ the resulting oriented wavelets ψ^α define a a family of dyadic wavelets that satisfy (5.98). Section 6.3 uses such directional wavelets, with $p = 1$, to detect multiscale edges in images.

Image f



$\alpha = \pi/2$ $\alpha = 3\pi/4$ $\alpha = 0$ $\alpha = \pi/4$

Figure 5.10: Decomposition of an image in a frame of steerable directional wavelets [402] along 4 directions, for $\alpha = 0, \pi/4, \pi/2, 3\pi/4$ at two consecutive scales 2^j and 2^j+1. Black, grey and white pixels correspond to wavelet coefficients of negative, zero and positive values.

Discretization of the Translation A translation invariant wavelet transforms $Wf(u, 2^j, \alpha)$ for all scales 2^j and angles α requires a large amount of memory. To reduce computation and memory storage, the translation parameter is discretized. Like in one the dimensional case a frame is obtained by sampling uniformly the translation parameter u with intervals $u_0 2^j n$ with $n = (n_1, n_2) \in \mathbb{Z}^2$, proportional to the scale 2^j. The discretized wavelet derived from the translated invariant wavelet family (5.97) is:

$$\{\psi^\alpha_{2^j}(x - 2^j u_0 n)\}_{n\in\mathbb{Z}^2, j\in\mathbb{Z}, \alpha\in\Theta} \quad \text{with} \quad \psi^\alpha_{2^j}(x) = 2^{-j}\psi^\alpha(2^{-j}x) \ . \tag{5.104}$$

Necessary and sufficient conditions similar to Theorems 5.15 and 5.16 can be established to guarantee that such a wavelet family defines a frame of $\mathbf{L^2}(\mathbb{R}^2)$.

To decompose images in such wavelet frames with a fast filter filter bank, directional wavelets can be synthesized as a product of discrete filters. The steerable pyramid of Simoncelli, Freeman, Adelson and Heeger [402] decomposes images in such a directional wavelet frame, with a cascade of convolutions with low-pass filters and directional band-pass filters. The filter coefficients are optimized to yield wavelets that satisfy approximately the steerability condition (5.103) and produce a tight frame. The sampling interval is $u_0 = 1/2$.

Figure 5.10 shows an example of decomposition on such a steerable pyramid with $K = 4$ directions. For discrete images of N pixels, the finest scale is $2^j = 2N^{-1}$. Since $u_0 = 1/2$, wavelet coefficients at the finest scale define an image of N pixels for each direction. The wavelet image size then decrease as the scale 2^j increases. The total number of wavelet coefficients is $4KN/3$ and the tight frame factor is $4K/3$ [402]. Steerable wavelet frames are used to remove noise with wavelet thresholding estimators [366] and for texture analysis and synthesis [399].

Chapter 9 explains that sparse image representation can be obtained by keeping large amplitude coefficients above a threshold. Large wavelet coefficient amplitude appear where the image has a sharp transition, when the wavelet oscillates in a direction approximately perpendicular to the direction of the edge. However, even when directions are not perfectly aligned, wavelet coefficient amplitude remain non-negligible in the neighborhood of edges. The number of large wavelet coefficients is thus typically proportional to the length of edges in images. Reducing the number of large coefficients requires using waveforms that are more sensitive to direction properties, as shown in the next section.

5.5.2 Curvelet Frames

Curvelet frames were introduced by Candès and Donoho [127] to construct sparse representation for images including edges that are geometrically regular. Similarly to directional wavelets, curvelet frames are obtained by rotating, dilating and translating elementary waveforms. However, curvelets have a highly elongated support obtained with a parabolic scaling using different scaling factors along the curvelet width and length. These anisotropic waveforms have a much better direction sensitivity than directional wavelets. Section 9.3.2 studies applications to sparse approximations of geometrically regular images.

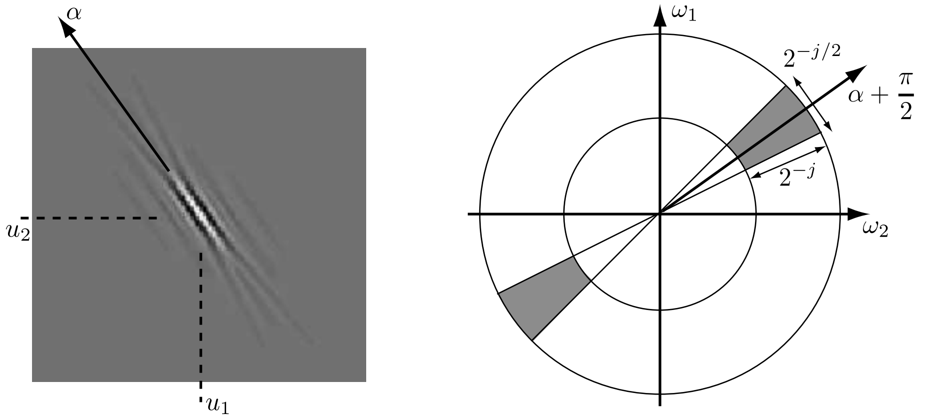


Figure 5.11: Left: Example of curvelet $c^{\alpha}_{2^j,u}(x)$. Right: the frequency support of $\hat{c}^{\alpha}_{2^j,u}(\omega)$ is a wedge obtained as a product of a radial window with an angular window.

Dyadic Curvelet Transform A curvelet is function $c(x)$ having vanishing moments along the horizontal direction like a wavelet. However, as opposed to wavelets, dilated curvelets are obtained with a parabolic scaling law that produces highly elongated waveforms at fine scales:

$$c_{2^j}(x_1, x_2) \approx 2^{-3j/4}\, c(2^{-j/2} x_1, 2^{-j} x_2). \tag{5.105}$$

They have a *width* proportional to their *length*^{2}. Dilated curvelets are then rotated $c^{\alpha}_{2^j} = c_{2^j}(R_\alpha x)$, where R_α is the planar rotation of angle α, and translated like wavelets: $c^{\alpha}_{2^j,u} = c^{\alpha}_{2^j}(x-u)$. The resulting translation invariant dyadic curvelet transform of $f \in \mathbf{L^2}(\mathbb{R}^2)$ is defined by

$$Cf(u, 2^j, \alpha) = \langle f, c^{\alpha}_{2^j,u} \rangle = f \star \bar{c}^{\alpha}_{2^j}(u) \quad \text{with} \quad \bar{c}^{\alpha}_{2^j}(x) = c^{\alpha}_{2^j}(-x)\ .$$

To obtain a tight frame, the Fourier transform of a curvelet at a scale 2^j is defined by

$$\hat{c}_{2^j}(\omega) = 2^{3j/4} \hat{\psi}(2^j r) \hat{\phi}\left(\frac{2\theta}{2^{\lfloor j/2 \rfloor}\pi}\right), \quad \text{with} \quad \omega = r(\cos\theta\,, \sin\theta), \tag{5.106}$$

where $\hat{\psi}$ is the Fourier transform of a one-dimensional wavelet and $\hat{\phi}$ is a one-dimensional angular window which localize the frequency support of $\hat{c}_{2^j}$ in a polar parabolic wedge illustrated in Figure 5.11. The wavelet $\hat{\psi}$ is chosen to have a compact support in $[1/2, 2]$ and satisfies the dyadic frequency covering:

$$\forall\, r \in \mathbb{R}^* \quad , \quad \sum_{j=-\infty}^{+\infty} |\hat{\psi}(2^j r)|^2 = 1. \tag{5.107}$$

One may for example choose a Meyer wavelet as defined in (7.82). The angular window $\hat{\phi}$ is chosen to be supported in $[-1, 1]$ and satisfies

$$\forall u, \qquad \sum_{k=-\infty}^{+\infty} |\hat{\phi}(u-k)|^2 = 1. \tag{5.108}$$

As a result of these two properties, one can verify that for uniformly distributed angles

$$\Theta_j = \{\alpha = k\pi 2^{\lfloor j/2 \rfloor - 1} \;\text{ for }\; 0 \leqslant k < 2^{-\lfloor j/2 \rfloor + 2}\}$$

curvelets cover the frequency plane

$$\forall \omega \in \mathbb{R}^2 - \{0\} \quad , \quad \sum_{j \in \mathbb{Z}} \sum_{\alpha \in \Theta_j} 2^{-3j/2} |\hat{c}^{\alpha}_{2^j}(\omega)|^2 = 1. \tag{5.109}$$

Real valued curvelets are obtained with a symmetric version of (5.106): $\hat{c}_{2^j}(\omega) + \hat{c}_{2^j}(-\omega)$. Applying Theorem 5.11 proves that translation invariant dyadic curvelet dictionary $\{c^{\alpha}_{2^j,u}\}_{\alpha \in \Theta_j, j \in \mathbb{Z}, u \in \mathbb{R}^2}$ is a dyadic translation invariant tight frame that defines a complete and stable signal representation [135].

Theorem 5.23 (Candès, Donoho). *For any* $f \in \mathbf{L^2}(\mathbb{R}^2)$

$$\|f\|^2 = \sum_{j \in \mathbb{Z}} 2^{-3j/2} \sum_{\alpha \in \Theta_j} \|Cf(\cdot, 2^j, \alpha)\|^2,$$

and

$$f(x) = \sum_{j \in \mathbb{Z}} 2^{-3j/2} \sum_{\alpha \in \Theta_j} Cf(\cdot, 2^j, \alpha) * c^{\alpha}_{2^j}(x).$$

Curvelet Properties Since $\hat{c}_{2^j}(\omega)$ is a smooth function with a support included in a rectangle of size proportional to $2^{-j/2} \times 2^{-j}$, the spatial curvelet $c_{2^j}(x)$ is a regular function with a fast decay outside a rectangle of size $2^{j/2} \times 2^j$. The rotated and translated curvelet $c^{\alpha}_{2^j,u}$ is supported around the point u in an elongated rectangle along the direction α whose shape has a parabolic ratio *width=length*^{2}, as shown by Figure 5.11.

Since the Fourier transform $\hat{c}_{2^j}(\omega_1, \omega_2)$ is zero in the neighborhood of the vertical axis $\omega_1 = 0$, $c_{2^j}(x_1, x_2)$ has an infinite number of vanishing moments in the horizontal direction

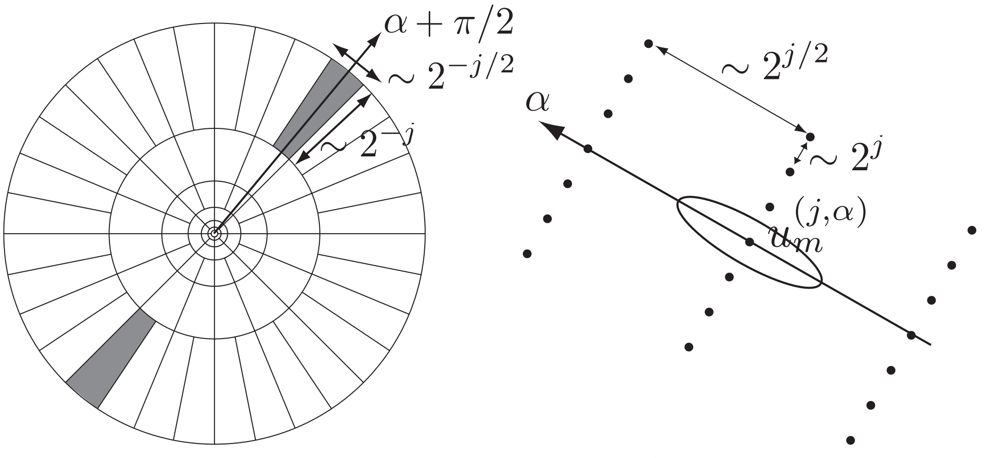
$$\forall\, \omega_1, \quad \frac{\partial^q \hat{c}_j}{\partial^q \omega_1}(0, \omega_1) = 0 \quad \Longrightarrow \quad \forall\, q \geqslant 0, \quad \forall\, x_2, \quad \int c_{2^j}(x_1, x_2) x_1^q \, dx_1 = 0\;.$$

A rotated curvelet $c^{\alpha}_{2^j,u}$ has vanishing moments in the direction $\alpha + \pi/2$ and thus oscillates in the direction $\alpha + \pi/2$ whereas its support is elongated in the direction α.

Discretization of Translation Curvelet tight frames are constructed by sampling the translation parameter u [127]. These tight frames provide sparse representations of signals including regular geometric structures.

The curvelet sampling grid depends on the scale 2^j and on the angle α. Sampling intervals are proportional to the curvelet width 2^j in the direction $\alpha + \pi/2$ and to its length $2^{j/2}$ in the direction α:

$$\forall m = (m_1, m_2) \in \mathbb{Z}^2 \quad , \quad u_m^{(j,\alpha)} = R_\alpha(2^{j/2} m_1, 2^j m_2) \tag{5.110}$$



Frequency tiling Spatial sampling

Figure 5.12: Left: Curvelet polar tiling of the frequency plane with with parabolic wedges. Right: curvelet sampling grid $u_m^{(j,\alpha)}$ at a scale 2^j and direction α.

Figure 5.12 illustrates this sampling grid. The resulting dictionary of translated curvelets is

$$\left\{ c_{j,m}^{\alpha}(x) = c_{2^j}^{\alpha}(x - u_m^{(j,\alpha)}) \right\}_{j\in\mathbb{Z},\alpha\in\Theta_j,m\in\mathbb{Z}^2} .$$

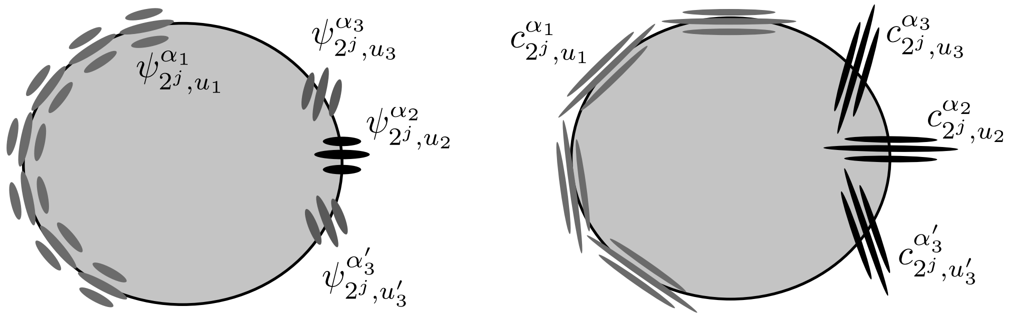
The following theorem proves that this curvelet family is a tight frame of $\mathbf{L^2}(\mathbb{R}^2)$. The proof is not give here and can be found in [135].

Theorem 5.24 (Candès, Donoho). *For any* $f \in \mathbf{L^2}(\mathbb{R}^2)$

$$\|f\|^2 = \sum_{j\in\mathbb{Z}} \sum_{\alpha\in\Theta_j} \sum_{m\in\mathbb{Z}^2} |\langle f,\, c_{j,m}^{\alpha}\rangle|^2 \tag{5.111}$$

and

$$f(x) = \sum_{j\in\mathbb{Z}} \sum_{\alpha\in\Theta_j} \sum_{m\in\mathbb{Z}^2} \langle f,\, c_{j,m}^{\alpha}\rangle c_{j,m}^{\alpha}(x) .$$



Directional wavelets Curvelets

Figure 5.13: A regular edge creates more large amplitude wavelet coefficients than curvelet coefficients. Curvelets coefficients have a large amplitude when their support is aligned with the edge direction, but there are few such curvelets.

Wavelet versus Curvelet Coefficients In the neighborhood of an edge whose tangent has a direction θ, large amplitude coefficients are created by curvelets and wavelets of direction $\alpha = \theta$, which have their vanishing moment in the direction $\theta + \pi/2$. These curvelets have a support elongated in the edge direction θ, as illustrated in Figure 5.13. In this direction, the sampling grid of a curvelet frame has an interval $2^{j/2}$ which is much larger than the sampling interval 2^j of a wavelet frame. An edge is thus covered by fewer curvelets than wavelets having a direction equal to the edge direction. If the angle α of the curvelet deviates from θ, then curvelet coefficients decay quickly because of the narrow frequency localizaton of curvelets. This gives a high directional selectivity to curvelets.

Eventhough wavelet coefficients vanish when $\alpha = \theta + \pi/2$ they have a smaller directional selectivity then curvelets and wavelet coefficients amplitudes decay more slowly as α deviates from θ. Indeed, the frequency support of wavelets is much wider and their spatial support is nearly isotropic. As a consequence, an edge produces large amplitude wavelet coefficients in several directions. This is illustrated by Figure 5.10, where Lena's shoulder edge creates large coefficients in 3 directions, and small coefficients only when the wavelet and the edge directions are aligned.

As a result, edges and image structures having some directional regularity create fewer large amplitude curvelet coefficients than wavelet coefficients. Section 9.3.3 derives a theorem which proves that for images having regular edges, curvelet tight frames are asymptotically more efficient than wavelet bases, to build sparse representations.

Fast Curvelet Decomposition Algorithm To compute the curvelet transform of a discrete image $f[n_1, n_2]$ uniformly sampled over N pixels, one must take into account the discretization grid, which imposes constraints on curvelet angles. The fast curvelet transform [133] replaces the polar tiling of the Fourier domain by a recto-polar tiling illustrated on the right of Figure 5.14. The directions α are uniformly discretized so that the slopes of the wedges containing the support of the curvelets are uniformly distributed in each of the North, South, West and East Fourier quadrants. Each wedge is the support of the 2D discrete Fourier transform $\hat{c}_j^\alpha[k_1, k_2]$ of a discrete curvelet $c_j^\alpha[n_1, n_2]$. The curvelet translation parameters are not chosen according to (5.110), but remain on a subgrid of the original image sampling grid. At a scale 2^j, there is one sampling grid $(2^{\rfloor j/2 \lfloor} m_1, 2^j m_2)$ for curvelets in the East and West quadrants. In these directions, curvelet coefficients are

$$\langle f[n_1, n_2], c_j^\alpha[n_1 - 2^{\lfloor j/2 \rfloor} m_1, n_2 - 2^j m_2] \rangle = f \star \bar{c}_j^\alpha[2^{\lfloor j/2 \rfloor} m_1, 2^j m_2] \tag{5.112}$$

with $\bar{c}_j^\alpha[n_1, n_2] = c_j^\alpha[-n_1, -n_2]$. For curvelets in the North and South quadrants the translation grid is $(2^j m_1, 2^{\lfloor j/2 \rfloor} m_2)$, which correspond to curvelet coefficients

$$\langle f[n_1, n_2], c_j^\alpha[n_1 - 2^j m_1, n_2 - 2^{\lfloor j/2 \rfloor} m_2] \rangle = f \star \bar{c}_j^\alpha[2^j m_1, 2^{\lfloor j/2 \rfloor} m_2] \ . \tag{5.113}$$

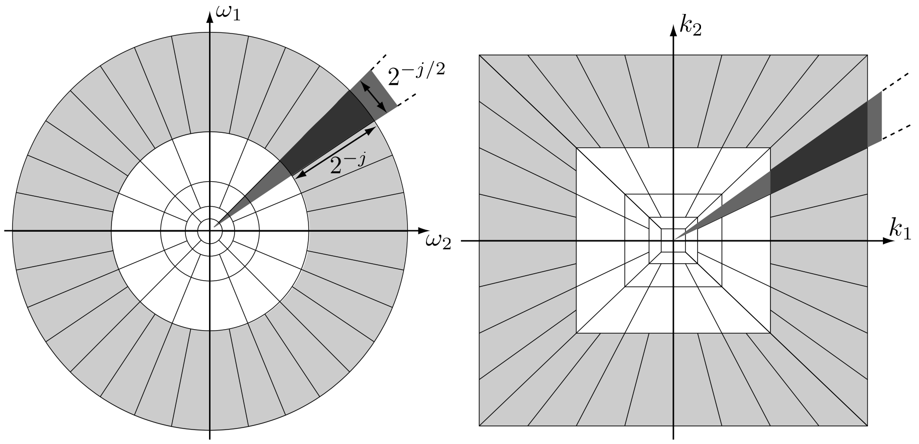
The discrete curvelet transform computes the curvelet filtering and sampling with an two-dimensional FFT. The 2D discrete Fourier transforms of $f[n]$ and $\bar{c}_j^\alpha[n]$ are $\hat{c}_j^\alpha[-k]$ and $\hat{f}[k]$. The algorithm proceeds as follow.

- Computation of the 2D discrete Fourier transform $\hat{f}[k]$ of $f[n]$.
- For each j and the corresponding $2^{-\lfloor j/2 \rfloor + 2}$ angles α, calculation of $\hat{f}[k]\, \hat{c}_j^\alpha[-k]$.
- Computation of the inverse Fourier transform of $\hat{f}[k]\, \hat{c}_j^\alpha[-k]$ on the smallest possible warped frequency rectangle including the wedge support of $\hat{c}_j^\alpha[-k]$.

The critical step is the last inverse Fourier transform. A computationally more expensive would compute $f \star \bar{c}_j^\alpha[n]$ for all $n = (n_1, n_2)$ and then subsample this convolution along the grids (5.112) and (5.113). Instead, the subsampled curvelet coefficients are calculated directly by restricting the FFT to a bounding box which contains the support of $\hat{c}_j^\alpha[-k]$. A horizontal or vertical warping maps this bounding box to an elongated rectangular frequency box on which the inverse FFT is calculated. One can verify that the resulting coefficients corresponds to the subsampled curvelet coefficients. The overall complexity of the algorithm is then $O(N \log_2(N))$, as detailed in [133]. The tight frame redundancy bound obtained with this discrete algorithm is $A \approx 5$.

An orthogonal curvelet type transform has been developed by Do and Vetterli [199]. The resulting contourlets are not redundant but do not have the appropriate time and frequency localization needed to obtain asymptotic approximation results similar to curvelets.

1



Frequency plane tiling Discrete frequency tiling

Figure 5.14: Left: curvelet frequency tiling. The dark gray area is a wedge obtained as the product of a radial window and an angular window. Right: Discrete curvelet frequency tiling. The radial and angular windows define trapezoidal wedges as shown in dark gray.

5.6 Exercises

5.1. [1]
5.2. [1]
5.3. [1]
5.4. [1]
5.5. [1]
5.6.
5.7.
5.8.

VI

Wavelet Zoom

A wavelet transform can focus on localized signal structures with a zooming procedure that progressively reduces the scale parameter. Singularities and irregular structures often carry essential information in a signal. For example, discontinuities in images may correspond to occlusion contours of objects in a scene. The wavelet transform amplitude across scales is related to the local signal regularity and Lipschitz exponents. Singularities and edges are detected from wavelet transform local maxima at multiple scales. These maxima define a geometric scale-space support, from which signal and image approximations are recovered.

Non-isolated singularities appear in highly irregular signals such as multifractals. The wavelet transform takes advantage of multifractal self-similarities to compute the distribution of their singularities. This singularity spectrum characterizes multifractal properties. Throughout this chapter, wavelets are real functions.

6.1 Lipschitz Regularity

To characterize singular structures, it is necessary to precisely quantify the local regularity of a signal $f(t)$. Lipschitz exponents provide uniform regularity measurements over time intervals, but also at any point v. If f has a singularity at v, which means that it is not differentiable at v, then the Lipschitz exponent at v characterizes this singular behavior.

The next section relates the uniform Lipschitz regularity of f over $\mathbb{R}$ to the asymptotic decay of the amplitude of its Fourier transform. This global regularity measurement is useless in analyzing the signal properties at particular locations. Section 6.1.3 studies zooming procedures that measure local Lipschitz exponents from the decay of the wavelet transform amplitude at fine scales.

6.1.1 Lipschitz Definition and Fourier Analysis

The Taylor formula relates the differentiability of a signal to local polynomial approximations. Suppose that f is m times differentiable in $[v-h, v+h]$. Let p_v be the Taylor polynomial in the neighborhood of v:

$$p_v(t) = \sum_{k=0}^{m-1} \frac{f^{(k)}(v)}{k!} (t-v)^k. \tag{6.1}$$

The Taylor formula proves that the approximation error

$$\varepsilon_v(t) = f(t) - p_v(t)$$

satisfies

$$\forall t \in [v-h, v+h] \quad , \quad |\varepsilon_v(t)| \leqslant \frac{|t-v|^m}{m!} \sup_{u\in[v-h,v+h]} |f^m(u)|. \tag{6.2}$$

The m^{th} order differentiability of f in the neighborhood of v yields an upper bound on the error $\varepsilon_v(t)$ when t tends to v. The Lipschitz regularity refines this upper bound with non-integer exponents. Lipschitz exponents are also called *Hölder* exponents in the mathematical literature.

Definition 6.1 (Lipschitz).
- *A function f is pointwise Lipschitz $\alpha \geqslant 0$ at v, if there exist $K > 0$, and a polynomial p_v of degree $m = \lfloor \alpha \rfloor$ such that*

$$\forall t \in \mathbb{R} \quad , \quad |f(t) - p_v(t)| \leqslant K\,|t - v|^{\alpha}. \tag{6.3}$$

- *A function f is uniformly Lipschitz α over $[a, b]$ if it satisfies (6.3) for all $v \in [a, b]$, with a constant K that is independent of v.*
- *The Lipschitz regularity of f at v or over $[a, b]$ is the sup of the α such that f is Lipschitz α.*

At each v the polynomial $p_v(t)$ is uniquely defined. If f is $m = \lfloor \alpha \rfloor$ times continuously differentiable in a neighborhood of v, then p_v is the Taylor expansion of f at v. Pointwise Lipschitz exponents may vary arbitrarily from abscissa to abscissa. One can construct multifractal functions with non-isolated singularities, where f has a different Lipschitz regularity at each point. In contrast, uniform Lipschitz exponents provide a more global measurement of regularity, which applies to a whole interval. If f is uniformly Lipschitz $\alpha > m$ in the neighborhood of v then one can verify that f is necessarily m times continuously differentiable in this neighborhood.

If $0 \leqslant \alpha < 1$ then $p_v(t) = f(v)$ and the Lipschitz condition (6.3) becomes

$$\forall t \in \mathbb{R} \quad , \quad |f(t) - f(v)| \leqslant K\,|t - v|^{\alpha}.$$

A function that is bounded but discontinuous at v is Lipschitz 0 at v. If the Lipschitz regularity is $\alpha < 1$ at v, then f is not differentiable at v and α characterizes the singularity type.

Fourier Condition The uniform Lipschitz regularity of f over $\mathbb{R}$ is related to the asymptotic decay of its Fourier transform. The following theorem can be interpreted as a generalization of Theorem 2.5.

Theorem 6.1. *A function f is bounded and uniformly Lipschitz α over $\mathbb{R}$ if*

$$\int_{-\infty}^{+\infty} |\hat{f}(\omega)|\,(1 + |\omega|^{\alpha})\,d\omega < +\infty. \tag{6.4}$$

Proof. To prove that f is bounded, we use the inverse Fourier integral (2.8) and (6.4) which shows that

$$|f(t)| \leqslant \int_{-\infty}^{+\infty} |\hat{f}(\omega)|\,d\omega < +\infty.$$

Let us now verify the Lipschitz condition (6.3) when $0 \leqslant \alpha \leqslant 1$. In this case $p_v(t) = f(v)$ and the uniform Lipschitz regularity means that there exists $K > 0$ such that for all $(t, v) \in \mathbb{R}^2$

$$\frac{|f(t) - f(v)|}{|t - v|^{\alpha}} \leqslant K.$$

Since

$$\begin{aligned} f(t) &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} \hat{f}(\omega)\,\exp(i\omega t)\,d\omega, \\ \frac{|f(t) - f(v)|}{|t - v|^{\alpha}} &\leqslant \frac{1}{2\pi} \int_{-\infty}^{+\infty} |\hat{f}(\omega)|\,\frac{|\exp(i\omega t) - \exp(i\omega v)|}{|t - v|^{\alpha}}\,d\omega. \end{aligned} \tag{6.5}$$

For $|t - v|^{-1} \leqslant |\omega|$,

$$\frac{|\exp(i\omega t) - \exp(i\omega v)|}{|t - v|^{\alpha}} \leqslant \frac{2}{|t - v|^{\alpha}} \leqslant 2\,|\omega|^{\alpha}.$$

For $|t - v|^{-1} \geqslant |\omega|$,

$$\frac{|\exp(i\omega t) - \exp(i\omega v)|}{|t - v|^{\alpha}} \leqslant \frac{|\omega|\,|t - v|}{|t - v|^{\alpha}} \leqslant |\omega|^{\alpha}.$$

Cutting the integral (6.5) in two for $|\omega| < |t-v|^{-1}$ and $|\omega| \geqslant |t-v|^{-1}$ yields

$$\frac{|f(t)-f(v)|}{|t-v|^{\alpha}} \leqslant \frac{1}{2\pi} \int_{-\infty}^{+\infty} 2\,|\hat{f}(\omega)|\,|\omega|^{\alpha}\,d\omega = K.$$

If (6.4) is satisfied, then $K < +\infty$ so f is uniformly Lipschitz α.

Let us extend this result for $m = \lfloor \alpha \rfloor > 0$. We proved in (2.42) that (6.4) implies that f is m times continuously differentiable. One can verify that f is uniformly Lipschitz α over $\mathbb{R}$ if and only if $f^{(m)}$ is uniformly Lipschitz $\alpha - m$ over $\mathbb{R}$. The Fourier transform of $f^{(m)}$ is $(i\omega)^m \hat{f}(\omega)$. Since $0 \leqslant \alpha - m < 1$, we can use our previous result which proves that $f^{(m)}$ is uniformly Lipschitz $\alpha - m$, and hence that f is uniformly Lipschitz α. ■ ■

The Fourier transform is a powerful tool for measuring the minimum global regularity of functions. However, it is not possible to analyze the regularity of f at a particular point v from the decay of $|\hat{f}(\omega)|$ at high frequencies ω. In contrast, since wavelets are well localized in time, the wavelet transform gives Lipschitz regularity over intervals *and* at points.

6.1.2 Wavelet Vanishing Moments

To measure the local regularity of a signal, it is not so important to use a wavelet with a narrow frequency support, but vanishing moments are crucial. If the wavelet has n vanishing moments then we show that the wavelet transform can be interpreted as a multiscale differential operator of order n. This yields a first relation between the differentiability of f and its wavelet transform decay at fine scales.

Polynomial Suppression The Lipschitz property (6.3) approximates f with a polynomial p_v in the neighborhood of v:

$$f(t) = p_v(t) + \varepsilon_v(t) \quad \text{with} \quad |\varepsilon_v(t)| \leqslant K\,|t-v|^{\alpha}\ . \tag{6.6}$$

A wavelet transform estimates the exponent α by ignoring the polynomial p_v. For this purpose, we use a wavelet that has $n > \alpha$ *vanishing moments*:

$$\int_{-\infty}^{+\infty} t^k\,\psi(t)\,dt = 0 \quad \text{for} \quad 0 \leqslant k < n\ .$$

A wavelet with n vanishing moments is orthogonal to polynomials of degree $n-1$. Since $\alpha < n$, the polynomial p_v has degree at most $n-1$. With the change of variable $t' = (t-u)/s$ we verify that

$$Wp_v(u,s) = \int_{-\infty}^{+\infty} p_v(t)\,\frac{1}{\sqrt{s}}\,\psi\left(\frac{t-u}{s}\right)\,dt = 0. \tag{6.7}$$

Since $f = p_v + \varepsilon_v$,

$$Wf(u,s) = W\varepsilon_v(u,s). \tag{6.8}$$

Section 6.1.3 explains how to measure α from $|Wf(u,s)|$ when u is in the neighborhood of v.

Multiscale Differential Operator The following theorem proves that a wavelet with n vanishing moments can be written as the n^{th} order derivative of a function θ; the resulting wavelet transform is a multiscale differential operator. We suppose that ψ has a fast decay which means that for any decay exponent $m \in \mathbb{N}$ there exists C_m such that

$$\forall t \in \mathbb{R}\ ,\ |\psi(t)| \leqslant \frac{C_m}{1+|t|^m}\ . \tag{6.9}$$

Theorem 6.2. *A wavelet ψ with a fast decay has n vanishing moments if and only if there exists θ with a fast decay such that*

$$\psi(t) = (-1)^n\,\frac{d^n\theta(t)}{dt^n}. \tag{6.10}$$

As a consequence

$$Wf(u,s) = s^n \frac{d^n}{du^n}(f \star \bar{\theta}_s)(u) \ , \tag{6.11}$$

with $\bar{\theta}_s(t) = s^{-1/2}\theta(-t/s)$. *Moreover,* ψ *has no more than* n *vanishing moments if and only if* $\int_{-\infty}^{+\infty} \theta(t)\, dt \neq 0$.

Proof. The fast decay of ψ implies that $\hat{\psi}$ is $\mathbf{C}^\infty$. This is proved by setting $f = \hat{\psi}$ in Theorem 2.5. The integral of a function is equal to its Fourier transform evaluated at $\omega = 0$. The derivative property (2.22) implies that for any $k < n$

$$\int_{-\infty}^{+\infty} t^k\, \psi(t)\, dt = (i)^k \hat{\psi}^{(k)}(0) = 0. \tag{6.12}$$

We can therefore make the factorization

$$\hat{\psi}(\omega) = (-i\omega)^n\, \hat{\theta}(\omega), \tag{6.13}$$

and $\hat{\theta}(\omega)$ is bounded. The fast decay of θ is proved with an induction on n. For $n = 1$,

$$\theta(t) = \int_{-\infty}^{t} \psi(u)\, du = \int_{t}^{+\infty} \psi(u)\, du,$$

and the fast decay of θ is derived from (6.9). We then similarly verify that increasing by 1 the order of integration up to n maintains the fast decay of θ.

Conversely, $|\hat{\theta}(\omega)| \leqslant \int_{-\infty}^{+\infty} |\theta(t)|\, dt < +\infty$, because θ has a fast decay. The Fourier transform of (6.10) yields (6.13) which implies that $\hat{\psi}^{(k)}(0) = 0$ for $k < n$. It follows from (6.12) that ψ has n vanishing moments.

To test whether ψ has more than n vanishing moments, we compute with (6.13)

$$\int_{-\infty}^{+\infty} t^n\, \psi(t)\, dt = (i)^n\, \hat{\psi}^{(n)}(0) = (-i)^n\, n!\, \hat{\theta}(0).$$

Clearly, ψ has no more than n vanishing moments if and only if $\hat{\theta}(0) = \int_{-\infty}^{+\infty} \theta(t)\, dt \neq 0$.

The wavelet transform (4.32) can be written

$$Wf(u,s) = f \star \bar{\psi}_s(u) \ \text{ with } \ \bar{\psi}_s(t) = \frac{1}{\sqrt{s}}\, \psi\left(\frac{-t}{s}\right) \ . \tag{6.14}$$

We derive from (6.10) that $\bar{\psi}_s(t) = s^n\, \frac{d^n \bar{\theta}_s(t)}{dt^n}$. Commuting the convolution and differentiation operators yields

$$Wf(u,s) = s^n\, f \star \frac{d^n \bar{\theta}_s}{dt^n}(u) = s^n\, \frac{d^n}{du^n}(f \star \bar{\theta}_s)(u).$$

■ ■

If $K = \int_{-\infty}^{+\infty} \theta(t)\, dt \neq 0$ then the convolution $f \star \bar{\theta}_s(t)$ can be interpreted as a weighted average of f with a kernel dilated by s. So (6.11) proves that $Wf(u,s)$ is an n^{th} order derivative of an averaging of f over a domain proportional to s. Figure 6.1 shows a wavelet transform calculated with $\psi = -\theta'$, where θ is a Gaussian. The resulting $Wf(u,s)$ is the derivative of f averaged in the neighborhood of u with a Gaussian kernel dilated by s.

Since θ has a fast decay, one can verify that

$$\lim_{s \to 0} \frac{1}{\sqrt{s}}\, \bar{\theta}_s = K\, \delta \ ,$$

in the sense of the weak convergence (A.30). This means that for any ϕ that is continuous at u,

$$\lim_{s \to 0} \phi \star \frac{1}{\sqrt{s}}\, \bar{\theta}_s(u) = K\, \phi(u).$$

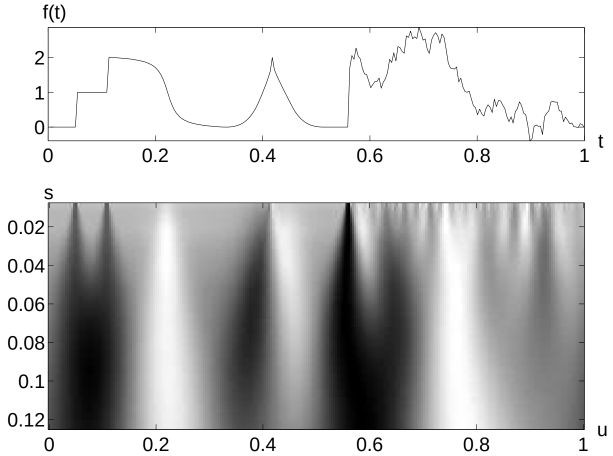


Figure 6.1: Wavelet transform $Wf(u,s)$ calculated with $\psi = -\theta'$ where θ is a Gaussian, for the signal f shown above. The position parameter u and the scale s vary respectively along the horizontal and vertical axes. Black, grey and white points correspond respectively to positive, zero and negative wavelet coefficients. Singularities create large amplitude coefficients in their cone of influence.

If f is n times continuously differentiable in the neighborhood of u then (6.11) implies that

$$\lim_{s \to 0} \frac{Wf(u,s)}{s^{n+1/2}} = \lim_{s \to 0} f^{(n)} \star \frac{1}{\sqrt{s}}\, \bar{\theta}_s(u) = K\, f^{(n)}(u) \ . \tag{6.15}$$

In particular, if f is $\mathbf{C}^n$ with a bounded n^{th} order derivative then $|Wf(u,s)| = O(s^{n+1/2})$. This is a first relation between the decay of $|Wf(u,s)|$ when s decreases and the uniform regularity of f. Finer relations are studied in the next section.

6.1.3 Regularity Measurements with Wavelets

The decay of the wavelet transform amplitude across scales is related to the uniform and pointwise Lipschitz regularity of the signal. Measuring this asymptotic decay is equivalent to zooming into signal structures with a scale that goes to zero. We suppose that the wavelet ψ has n vanishing moments and is $\mathbf{C}^n$ with derivatives that have a fast decay. This means that for any $0 \leqslant k \leqslant n$ and $m \in \mathbb{N}$ there exists C_m such that

$$\forall t \in \mathbb{R} \ , \ |\psi^{(k)}(t)| \leqslant \frac{C_m}{1+|t|^m} \ . \tag{6.16}$$

The following theorem relates the uniform Lipschitz regularity of f on an interval to the amplitude of its wavelet transform at fine scales.

Theorem 6.3. *If $f \in \mathbf{L^2}(\mathbb{R})$ is uniformly Lipschitz $\alpha \leqslant n$ over $[a,b]$, then there exists $A > 0$ such that*

$$\forall (u,s) \in [a,b] \times \mathbb{R}^+ \quad , \quad |Wf(u,s)| \leqslant A\, s^{\alpha+1/2}. \tag{6.17}$$

Conversely, suppose that f is bounded and that $Wf(u,s)$ satisfies (6.17) for an $\alpha < n$ that is not an integer. Then f is uniformly Lipschitz α on $[a+\varepsilon, b-\varepsilon]$, for any $\varepsilon > 0$.

Proof. This theorem is proved with minor modifications in the proof of Theorem 6.4. Since f is Lipschitz α at any $v \in [a,b]$, Theorem 6.4 shows in (6.20) that

$$\forall (u,s) \in \mathbb{R} \times \mathbb{R}^+ \quad , \quad |Wf(u,s)| \leqslant A\, s^{\alpha+1/2} \left(1 + \left|\frac{u-v}{s}\right|^{\alpha}\right) \ .$$

For $u \in [a,b]$, we can choose $v = u$, which implies that $|Wf(u,s)| \leqslant A\, s^{\alpha+1/2}$. We verify from the proof of (6.20) that the constant A does not depend on v because the Lipschitz regularity is uniform over $[a,b]$.

To prove that f is uniformly Lipschitz α over $[a+\varepsilon, b-\varepsilon]$ we must verify that there exists K such that for all $v \in [a+\varepsilon, b-\varepsilon]$ we can find a polynomial p_v of degree $\lfloor \alpha \rfloor$ such that

$$\forall t \in \mathbb{R} \quad , \quad |f(t) - p_v(t)| \leqslant K\, |t-v|^{\alpha} \ . \tag{6.18}$$

When $t \notin [a+\varepsilon/2, b-\varepsilon/2]$ then $|t-v| \geqslant \varepsilon/2$ and since f is bounded, (6.18) is verified with a constant K that depends on ε. For $t \in [a+\varepsilon/2, b-\varepsilon/2]$, the proof follows the same derivations as the proof of pointwise Lipschitz regularity from (6.21) in Theorem 6.4. The upper bounds (6.26) and (6.27) are replaced by

$$\forall t \in [a+\varepsilon/2, b-\varepsilon/2] \quad , \quad |\Delta_j^{(k)}(t)| \leqslant K\, 2^{(\alpha-k)j} \quad \text{for } 0 \leqslant k \leqslant \lfloor \alpha \rfloor + 1 \ . \tag{6.19}$$

This inequality is verified by computing an upper bound integral similar to (6.25) but which is divided in two, for $u \in [a,b]$ and $u \notin [a,b]$. When $u \in [a,b]$, the condition (6.21) is replaced by $|Wf(u,s)| \leqslant A\, s^{\alpha+1/2}$ in (6.25). When $u \notin [a,b]$, we just use the fact that $|Wf(u,s)| \leqslant \|f\|\, \|\psi\|$ and derive (6.19) from the fast decay of $|\psi^{(k)}(t)|$, by observing that $|t-u| \geqslant \varepsilon/2$ for $t \in [a+\varepsilon/2, b-\varepsilon/2]$. The constant K depends on A and ε but not on v. The proof then proceeds like the proof of Theorem 6.4, and since the resulting constant K in (6.29) does not depend on v, the Lipschitz regularity is uniform over $[a-\varepsilon, b+\varepsilon]$. ■ ■

The inequality (6.17) is really a condition on the asymptotic decay of $|Wf(u,s)|$ when s goes to zero. At large scales it does not introduce any constraint since the Cauchy-Schwarz inequality guarantees that the wavelet transform is bounded:

$$|Wf(u,s)| = |\langle f, \psi_{u,s} \rangle| \leqslant \|f\|\, \|\psi\|.$$

When the scale s decreases, $Wf(u,s)$ measures fine scale variations in the neighborhood of u. Theorem 6.3 proves that $|Wf(u,s)|$ decays like $s^{\alpha+1/2}$ over intervals where f is uniformly Lipschitz α.

Observe that the upper bound (6.17) is similar to the sufficient Fourier condition of Theorem 6.1, which supposes that $|\hat{f}(\omega)|$ decays faster than $\omega^{-\alpha}$. The wavelet scale s plays the role of a "localized" inverse frequency ω^{-1}. As opposed to the Fourier transform Theorem 6.1, the wavelet transform gives a Lipschitz regularity condition that is localized over any finite interval and it provides a necessary condition which is nearly sufficient. When $[a,b] = \mathbb{R}$ then (6.17) is a necessary and sufficient condition for f to be uniformly Lipschitz α on $\mathbb{R}$.

If ψ has exactly n vanishing moments then the wavelet transform decay gives no information concerning the Lipschitz regularity of f for $\alpha > n$. If f is uniformly Lipschitz $\alpha > n$ then it is $\mathbf{C}^n$ and (6.15) proves that $\lim_{s \to 0} s^{-n-1/2}\, Wf(u,s) = K\, f^{(n)}(u)$ with $K \neq 0$. This proves that $|Wf(u,s)| \sim s^{n+1/2}$ at fine scales despite the higher regularity of f.

If the Lipschitz exponent α is an integer then (6.17) is not sufficient in order to prove that f is uniformly Lipschitz α. When $[a,b] = \mathbb{R}$, if $\alpha = 1$ and ψ has two vanishing moments, then the class of functions that satisfy (6.17) is called the *Zygmund class* [42]. It is slightly larger than the set of functions that are uniformly Lipschitz 1. For example, $f(t) = t \log_e t$ belongs to the Zygmund class although it is not Lipschitz 1 at $t = 0$.

Pointwise Lipschitz Regularity The study of pointwise Lipschitz exponents with the wavelet transform is a delicate and beautiful topic which finds its mathematical roots in the characterization of Sobolev spaces by Littlewood and Paley in the 1930's. Characterizing the regularity of f at a point v can be difficult because f may have very different types of singularities that are aggregated in the neighborhood of v. In 1984, Bony [110] introduced the "two-microlocalization" theory which refines the Littlewood-Paley approach to provide pointwise characterization of singularities, which he used to study the solution of hyperbolic partial differential equations. These technical results became much simpler through the work of Jaffard [285] who proved that the two-microlocalization properties are equivalent to specific decay conditions on the wavelet transform amplitude. The following theorem gives a necessary condition and a sufficient condition on the wavelet transform for estimating the Lipschitz regularity of f at a point v. Remember that the wavelet ψ has n vanishing moments and n derivatives having a fast decay.

Theorem 6.4 (Jaffard). *If $f \in \mathbf{L^2}(\mathbb{R})$ is Lipschitz $\alpha \leqslant n$ at v, then there exists A such that*

$$\forall (u,s) \in \mathbb{R} \times \mathbb{R}^+ \quad , \quad |Wf(u,s)| \leqslant A\, s^{\alpha+1/2} \left(1 + \left|\frac{u-v}{s}\right|^{\alpha}\right) \ . \tag{6.20}$$

Conversely, if $\alpha < n$ is not an integer and there exist A and $\alpha' < \alpha$ such that

$$\forall (u,s) \in \mathbb{R} \times \mathbb{R}^+ \quad , \quad |Wf(u,s)| \leqslant A\, s^{\alpha+1/2} \left(1 + \left|\frac{u-v}{s}\right|^{\alpha'}\right) \tag{6.21}$$

then f is Lipschitz α at v.

Proof. The necessary condition is relatively simple to prove but the sufficient condition is much more difficult.

• *Proof of (6.20)* Since f is Lipschitz α at v, there exists a polynomial p_v of degree $\lfloor \alpha \rfloor < n$ and K such that $|f(t) - p_v(t)| \leqslant K|t-v|^{\alpha}$. Since ψ has n vanishing moments, we saw in (6.7) that $Wp_v(u,s) = 0$ and hence

$$\begin{aligned} |Wf(u,s)| &= \left| \int_{-\infty}^{+\infty} \big(f(t) - p_v(t)\big) \frac{1}{\sqrt{s}}\, \psi\Big(\frac{t-u}{s}\Big)\, dt \right| \\ &\leqslant \int_{-\infty}^{+\infty} K\, |t-v|^{\alpha}\, \frac{1}{\sqrt{s}} \left|\psi\Big(\frac{t-u}{s}\Big)\right| dt. \end{aligned}$$

The change of variable $x = (t-u)/s$ gives

$$|Wf(u,s)| \leqslant \sqrt{s} \int_{-\infty}^{+\infty} K\, |sx + u - v|^{\alpha}\, |\psi(x)|\, dx.$$

Since $|a+b|^{\alpha} \leqslant 2^{\alpha}\,(|a|^{\alpha} + |b|^{\alpha})$,

$$|Wf(u,s)| \leqslant K\, 2^{\alpha} \sqrt{s} \left(s^{\alpha} \int_{-\infty}^{+\infty} |x|^{\alpha}\, |\psi(x)|\, dx + |u-v|^{\alpha} \int_{-\infty}^{+\infty} |\psi(x)|\, dx \right)$$

which proves (6.20).

• *Proof of (6.21)* The wavelet reconstruction formula (4.37) proves that f can be decomposed in a Littlewood-Paley type sum

$$f(t) = \sum_{j=-\infty}^{+\infty} \Delta_j(t) \tag{6.22}$$

with

$$\Delta_j(t) = \frac{1}{C_\psi} \int_{-\infty}^{+\infty} \int_{2^j}^{2^{j+1}} Wf(u,s)\, \frac{1}{\sqrt{s}}\, \psi\left(\frac{t-u}{s}\right) \frac{ds}{s^2}\, du \ . \tag{6.23}$$

Let $\Delta_j^{(k)}$ be its k^{th} order derivative. To prove that f is Lipschitz α at v we shall approximate f with a polynomial that generalizes the Taylor polynomial

$$p_v(t) = \sum_{k=0}^{\lfloor \alpha \rfloor} \left(\sum_{j=-\infty}^{+\infty} \Delta_j^{(k)}(v) \right) \frac{(t-v)^k}{k!} \ . \tag{6.24}$$

If f is n times differentiable at v then p_v corresponds to the Taylor polynomial but this is not necessarily true. We shall first prove that $\sum_{j=-\infty}^{+\infty} \Delta_j^{(k)}(v)$ is finite by getting upper bounds on $|\Delta_j^{(k)}(t)|$. These sums may be thought of as a generalization of pointwise derivatives.

To simplify the notation, we denote by K a generic constant which may change value from one line to the next but that does not depend on j and t. The hypothesis (6.21) and the asymptotic decay condition (6.16) imply that

$$\begin{aligned} |\Delta_j(t)| &= \frac{1}{C_\psi} \int_{-\infty}^{+\infty} \int_{2^j}^{2^{j+1}} A\, s^{\alpha} \left(1 + \left|\frac{u-v}{s}\right|^{\alpha'}\right) \frac{C_m}{1 + |(t-u)/s|^m}\, \frac{ds}{s^2}\, du \\ &\leqslant K \int_{-\infty}^{+\infty} 2^{\alpha j} \left(1 + \left|\frac{u-v}{2^j}\right|^{\alpha'}\right) \frac{1}{1 + |(t-u)/2^j|^m}\, \frac{du}{2^j} \end{aligned} \tag{6.25}$$

Since $|u-v|^{\alpha'} \leqslant 2^{\alpha'}(|u-t|^{\alpha'}+|t-v|^{\alpha'})$, the change of variable $u'=2^{-j}(u-t)$ yields

$$|\Delta_j(t)| \leqslant K\,2^{\alpha j}\int_{-\infty}^{+\infty}\frac{1+|u'|^{\alpha'}+\left|(v-t)/2^j\right|^{\alpha'}}{1+|u'|^m}\,du'.$$

Choosing $m=\alpha'+2$ yields

$$|\Delta_j(t)| \leqslant K\,2^{\alpha j}\left(1+\left|\frac{v-t}{2^j}\right|^{\alpha'}\right). \tag{6.26}$$

The same derivations applied to the derivatives of $\Delta_j(t)$ yield

$$\forall k \leqslant \lfloor\alpha\rfloor+1 \quad , \quad |\Delta_j^{(k)}(t)| \leqslant K\,2^{(\alpha-k)j}\left(1+\left|\frac{v-t}{2^j}\right|^{\alpha'}\right). \tag{6.27}$$

At $t=v$ it follows that

$$\forall k \leqslant \lfloor\alpha\rfloor \quad , \quad |\Delta_j^{(k)}(v)| \leqslant K\,2^{(\alpha-k)j}. \tag{6.28}$$

This guarantees a fast decay of $|\Delta_j^{(k)}(v)|$ when 2^j goes to zero, because α is not an integer so $\alpha > \lfloor\alpha\rfloor$. At large scales 2^j, since $|Wf(u,s)| \leqslant \|f\|\,\|\psi\|$ with the change of variable $u'=(t-u)/s$ in (6.23) we have

$$|\Delta_j^{(k)}(v)| \leqslant \frac{\|f\|\,\|\psi\|}{C_\psi}\int_{-\infty}^{+\infty}|\psi^{(k)}(u')|\,du'\int_{2^j}^{2^{j+1}}\frac{ds}{s^{3/2+k}}$$

and hence $|\Delta_j^{(k)}(v)| \leqslant K 2^{-(k+1/2)j}$. Together with (6.28) this proves that the polynomial p_v defined in (6.24) has coefficients that are finite.

With the Littlewood-Paley decomposition (6.22) we compute

$$|f(t)-p_v(t)| = \left|\sum_{j=-\infty}^{+\infty}\left(\Delta_j(t)-\sum_{k=0}^{\lfloor\alpha\rfloor}\Delta_j^{(k)}(v)\,\frac{(t-v)^k}{k!}\right)\right|.$$

The sum over scales is divided in two at 2^J such that $2^J \geqslant |t-v| \geqslant 2^{J-1}$. For $j \geqslant J$, we can use the classical Taylor theorem to bound the Taylor expansion of Δ_j:

$$\begin{aligned} I &= \sum_{j=J}^{+\infty}\left|\Delta_j(t)-\sum_{k=0}^{\lfloor\alpha\rfloor}\Delta_j^{(k)}(v)\frac{(t-v)^k}{k!}\right| \\ &\leqslant \sum_{j=J}^{+\infty}\frac{(t-v)^{\lfloor\alpha\rfloor+1}}{(\lfloor\alpha\rfloor+1)!}\sup_{h\in[t,v]}|\Delta_j^{\lfloor\alpha\rfloor+1}(h)|\,. \end{aligned}$$

Inserting (6.27) yields

$$I \leqslant K\,|t-v|^{\lfloor\alpha\rfloor+1}\sum_{j=J}^{+\infty}2^{-j(\lfloor\alpha\rfloor+1-\alpha)}\left|\frac{v-t}{2^j}\right|^{\alpha'}$$

and since $2^J \geqslant |t-v| \geqslant 2^{J-1}$ we get $I \leqslant K\,|v-t|^{\alpha}$.

Let us now consider the case $j<J$

$$\begin{aligned} II &= \sum_{j=-\infty}^{J-1}\left|\Delta_j(t)-\sum_{k=0}^{\lfloor\alpha\rfloor}\Delta_j^{(k)}(v)\,\frac{(t-v)^k}{k!}\right| \\ &\leqslant K\sum_{j=-\infty}^{J-1}\left(2^{\alpha j}\left(1+\left|\frac{v-t}{2^j}\right|^{\alpha'}\right)+\sum_{k=0}^{\lfloor\alpha\rfloor}\frac{(t-v)^k}{k!}\,2^{j(\alpha-k)}\right) \\ &\leqslant K\left(2^{\alpha J}+2^{(\alpha-\alpha')J}\,|t-v|^{\alpha'}+\sum_{k=0}^{\lfloor\alpha\rfloor}\frac{(t-v)^k}{k!}\,2^{J(\alpha-k)}\right) \end{aligned}$$

and since $2^J \geqslant |t-v| \geqslant 2^{J-1}$ we get $II \leqslant K\,|v-t|^{\alpha}$. As a result

$$|f(t)-p_v(t)| \leqslant I+II \leqslant K\,|v-t|^{\alpha} \tag{6.29}$$

which proves that f is Lipschitz α at v. ■

Cone of Influence To interpret more easily the necessary condition (6.20) and the sufficient condition (6.21), we shall suppose that ψ has a compact support equal to $[-C, C]$. The *cone of influence* of v in the scale-space plane is the set of points (u, s) such that v is included in the support of $\psi_{u,s}(t) = s^{-1/2}\,\psi((t-u)/s)$. Since the support of $\psi((t-u)/s)$ is equal to $[u - Cs, u + Cs]$, the cone of influence of v is defined by

$$|u - v| \leqslant Cs. \tag{6.30}$$

It is illustrated in Figure 6.2. If u is in the cone of influence of v then $Wf(u, s) = \langle f, \psi_{u,s}\rangle$ depends on the value of f in the neighborhood of v. Since $|u - v|/s \leqslant C$, the conditions (6.20,6.21) can be written

$$|Wf(u, s)| \leqslant A'\, s^{\alpha+1/2}$$

which is identical to the uniform Lipschitz condition (6.17) given by Theorem 6.3. In Figure 6.1, the high amplitude wavelet coefficients are in the cone of influence of each singularity.

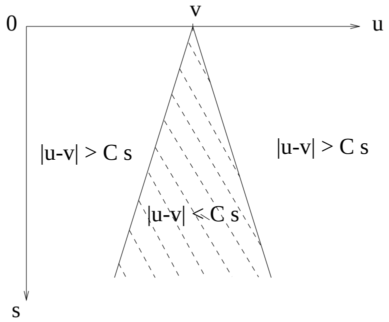


Figure 6.2: The cone of influence of an abscissa v consists of the scale-space points (u, s) for which the support of $\psi_{u,s}$ intersects $t = v$.

Oscillating Singularities It may seem surprising that (6.20) and (6.21) also impose a condition on the wavelet transform outside the cone of influence of v. Indeed, this corresponds to wavelets whose support does not intersect v. For $|u - v| > Cs$ we get

$$|Wf(u, s)| \leqslant A'\, s^{\alpha-\alpha'+1/2}\, |u - v|^{\alpha}. \tag{6.31}$$

We shall see that it is indeed necessary to impose this decay when u tends to v in order to control the oscillations of f that might generate singularities.

Let us consider the generic example of a highly oscillatory function

$$f(t) = \sin\frac{1}{t}\ ,$$

which is discontinuous at $v = 0$ because of the acceleration of its oscillations. Since ψ is a smooth $\mathbf{C}^n$ function, if it is centered close to zero then the rapid oscillations of $\sin t^{-1}$ produce a correlation integral $\langle\,\sin t^{-1}, \psi_{u,s}\rangle$ that is very small. With an integration by parts, one can verify that if (u, s) is in the cone of influence of $v = 0$, then $|Wf(u, s)| \leqslant A\, s^{2+1/2}$. This looks as if f is Lipschitz 2 at 0. However, Figure 6.3 shows high energy wavelet coefficients outside the cone of influence of $v = 0$, which are responsible for the discontinuity. To guarantee that f is Lipschitz α, the amplitude of such coefficients is controlled by the upper bound (6.31).

To explain why the high frequency oscillations appear outside the cone of influence of v, we use the results of Section 4.4.2 on the estimation of instantaneous frequencies with wavelet ridges. The instantaneous frequency of $\sin t^{-1} = \sin\theta(t)$ is $|\theta'(t)| = t^{-2}$. Let ψ^a be the analytic part of ψ, defined in (4.47). The corresponding complex analytic wavelet transform is $W^a f(u, s) = \langle f, \psi^a_{u,s}\rangle$. It was proved in (4.109) that for a fixed time u, the maximum of $s^{-1/2}|W^a f(u, s)|$ is located at the scale

$$s(u) = \frac{\eta}{\theta'(u)} = \eta\, u^2,$$

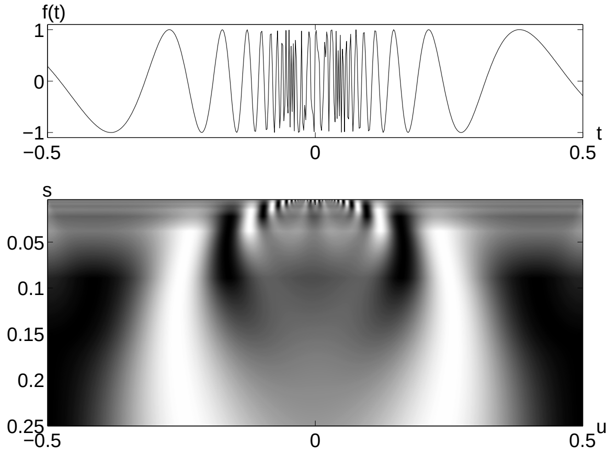


Figure 6.3: Wavelet transform of $f(t) = \sin(a\,t^{-1})$ calculated with $\psi = -\theta'$ where θ is a Gaussian. High amplitude coefficients are along a parabola outside the cone of influence of $t = 0$.

where η is the center frequency of $\hat{\psi}^a(\omega)$. When u varies, the set of points $(u, s(u))$ define a *ridge* that is a parabola located outside the cone of influence of $v = 0$ in the plane (u, s). Since $\psi = \mathrm{Re}[\psi^a]$, the real wavelet transform is

$$Wf(u,s) = \mathrm{Re}[W^a f(u,s)].$$

The high amplitude values of $Wf(u,s)$ are thus located along the same parabola ridge curve in the scale-space plane, which clearly appears in Figure 6.3. Real wavelet coefficients $Wf(u,s)$ change sign along the ridge because of the variations of the complex phase of $W^a f(u,s)$.

The example of $f(t) = \sin t^{-1}$ can be extended to general oscillating singularities [30]. A function f has an oscillating singularity at v if there exist $\alpha \geqslant 0$ and $\beta > 0$ such that for t in a neighborhood of v

$$f(t) \sim |t-v|^\alpha \, g\left(\frac{1}{|t-v|^\beta}\right) ,$$

where $g(t)$ is a $\mathbf{C}^\infty$ oscillating function whose primitives at any order are bounded. The function $g(t) = \sin t^{-1}$ is a typical example. The oscillations have an instantaneous frequency $\theta'(t)$ that increases to infinity faster than $|t|^{-1}$ when t goes to v. High energy wavelet coefficients are located along the ridge $s(u) = \eta/\theta'(u)$, and this curve is necessarily outside the cone of influence $|u - v| \leqslant Cs$.

6.2 Wavelet Transform Modulus Maxima

Theorems 6.3 and 6.4 prove that the local Lipschitz regularity of f at v depends on the decay at fine scales of $|Wf(u,s)|$ in the neighborhood of v. Measuring this decay directly in the time-scale plane (u,s) is not necessary. The decay of $|Wf(u,s)|$ can indeed be controlled from its local maxima values. Section 6.2.1 studies the detection and characterization of singularities from wavelet local maxima. Signal approximations are recovered in Section 6.2.2, from the scale-space support of these local maxima at dyadic scales.

6.2.1 Detection of Singularities

Singularities are detected by finding the abscissa where the wavelet modulus maxima converge at fine scales. A wavelet *modulus maximum* is defined as a point (u_0, s_0) such that $|Wf(u, s_0)|$ is locally maximum at $u = u_0$. This implies that

$$\frac{\partial Wf(u_0,s_0)}{\partial u} = 0.$$

This local maximum should be a strict local maximum in either the right or the left neighborhood of u_0, to avoid having any local maxima when $|Wf(u,s_0)|$ is constant. We call *maxima line* any connected curve $s(u)$ in the scale-space plane (u,s) along which all points are modulus maxima. Figure 6.5(b) shows the wavelet modulus maxima of a signal.

To better understand the properties of these maxima, the wavelet transform is written as a multiscale differential operator. Theorem 6.2 proves that if ψ has exactly n vanishing moments and a compact support, then there exists θ of compact support such that $\psi = (-1)^n \theta^{(n)}$ with $\int_{-\infty}^{+\infty} \theta(t)\, dt \neq 0$. The wavelet transform is rewritten in (6.11) as a multiscale differential operator

$$Wf(u,s) = s^n \, \frac{d^n}{du^n}(f \star \bar{\theta}_s)(u). \tag{6.32}$$

If the wavelet has only one vanishing moment, wavelet modulus maxima are the maxima of the first order derivative of f smoothed by $\bar{\theta}_s$, as illustrated by Figure 6.4. These multiscale modulus maxima are used to locate discontinuities, and edges in images. If the wavelet has two vanishing moments, the modulus maxima correspond to high curvatures. The following theorem proves that if $Wf(u,s)$ has no modulus maxima at fine scales, then f is locally regular.

Theorem 6.5 (Hwang, Mallat). *Suppose that ψ is $\mathbf{C}^n$ with a compact support, and $\psi = (-1)^n \, \theta^{(n)}$ with $\int_{-\infty}^{+\infty} \theta(t)dt \neq 0$. Let $f \in \mathbf{L^1}[a,b]$. If there exists $s_0 > 0$ such that $|Wf(u,s)|$ has no local maximum for $u \in [a,b]$ and $s < s_0$, then f is uniformly Lipschitz n on $[a+\varepsilon, b-\varepsilon]$, for any $\varepsilon > 0$.*

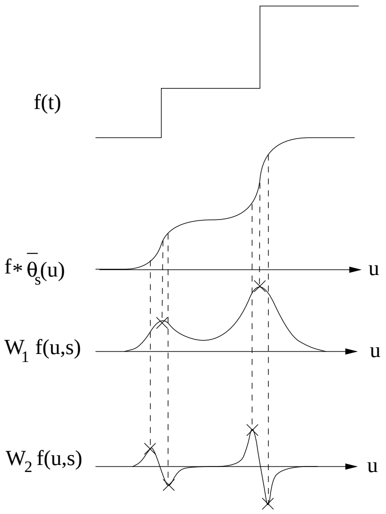


Figure 6.4: The convolution $f \star \bar{\theta}_s(u)$ averages f over a domain proportional to s. If $\psi = -\theta'$ then $W_1 f(u,s) = s\frac{d}{du}(f \star \bar{\theta}_s)(u)$ has modulus maxima at sharp variation points of $f \star \bar{\theta}_s(u)$. If $\psi = \theta''$ then the modulus maxima of $W_2 f(u,s) = s^2 \frac{d^2}{du^2}(f \star \bar{\theta}_s)(u)$ correspond to locally maximum curvatures.

This theorem is proved in [331]. It implies that f can be singular (not Lipschitz 1) at a point v only if there is a sequence of wavelet maxima points $(u_p, s_p)_{p\in\mathbb{N}}$ that converges towards v at fine scales:

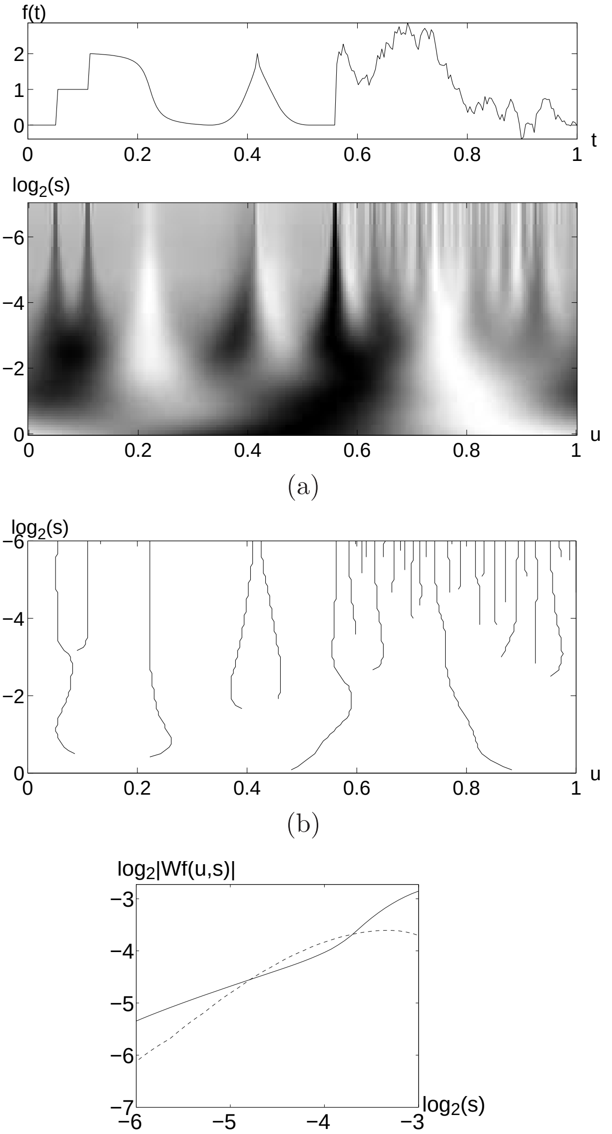
$$\lim_{p\to+\infty} u_p = v \quad \text{and} \quad \lim_{p\to+\infty} s_p = 0 \ .$$

These modulus maxima points may or may not be along the same maxima line. This result guarantees that all singularities are detected by following the wavelet transform modulus maxima at fine scales. Figure 6.5 gives an example where all singularities are located by following the maxima lines.

Maxima Propagation For all $\psi = (-1)^n \, \theta^{(n)}$, we are not guaranteed that a modulus maxima located at (u_0, s_0) belongs to a maxima line that propagates towards finer scales. When s decreases,

$Wf(u,s)$ may have no more maxima in the neighborhood of $u = u_0$. The following theorem proves that this is never the case if θ is a Gaussian. The wavelet transform $Wf(u,s)$ can then be written as the solution of the heat diffusion equation, where s is proportional to the diffusion time. The maximum principle applied to the heat diffusion equation proves that maxima may not disappear when s decreases. Applications of the heat diffusion equation to the analysis of multiscale averaging have been studied by several computer vision researchers [282, 299, 454].

Theorem 6.6 (Hummel, Poggio, Yuille)**.** *Let* $\psi = (-1)^n \theta^{(n)}$ *where* θ *is a Gaussian. For any* $f \in \mathbf{L^2}(\mathbb{R})$*, the modulus maxima of* $Wf(u,s)$ *belong to connected curves that are never interrupted when the scale decreases.*



(a)

(b)

(c)

Figure 6.5: (a): Wavelet transform $Wf(u,s)$. The horizontal and vertical axes give respectively u and $\log_2 s$. (b): Modulus maxima of $Wf(u,s)$. (c): The full line gives the decay of $\log_2 |Wf(u,s)|$ as a function of $\log_2 s$ along the maxima line that converges to the abscissa $t = 0.05$. The dashed line gives $\log_2 |Wf(u,s)|$ along the left maxima line that converges to $t = 0.42$.

Proof. To simplify the proof, we suppose that θ is a normalized Gaussian $\theta(t) = 2^{-1}\pi^{-1/2}\,\exp(-t^2/4)$

whose Fourier transform is $\hat{\theta}(\omega) = \exp(-\omega^2)$. Theorem 6.2 proves that

$$Wf(u,s) = s^n\, f^{(n)} \star \theta_s(u), \tag{6.33}$$

where the n^{th} derivative $f^{(n)}$ is defined in the sense of distributions. Let τ be the diffusion time. The solution of

$$\frac{\partial g(u,\tau)}{\partial \tau} = \frac{\partial^2 g(u,\tau)}{\partial u^2} \tag{6.34}$$

with initial condition $g(u,0) = g_0(u)$ is obtained by computing the Fourier transform with respect to u of (6.34):

$$\frac{\partial \hat{g}(\omega,u)}{\partial \tau} = -\omega^2\, \hat{g}(\omega,\tau).$$

It follows that $\hat{g}(\omega,\tau) = \hat{g}_0(\omega)\, \exp(-\tau\omega^2)$ and hence

$$g(u,\tau) = \frac{1}{\sqrt{\tau}}\, g_0 \star \theta_{\sqrt{\tau}}(u).$$

For $\tau = s^2$, setting $g_0 = f^{(n)}$ and inserting (6.33) yields $Wf(u,s) = s^{n+1}\, g(u,s^2)$. The wavelet transform is thus proportional to a heat diffusion with initial condition $f^{(n)}$.

The maximum principle for the parabolic heat equation [33] proves that a global maximum of $|g(u,s^2)|$ for $(u,s) \in [a,b] \times [s_0,s_1]$ is necessarily either on the boundary $u = a,b$ or at $s = s_0$. A modulus maxima of $Wf(u,s)$ at (u_1,s_1) is a local maxima of $|g(u,s^2)|$ for a fixed s and u varying. Suppose that a line of modulus maxima is interrupted at (u_1,s_1), with $s_1 > 0$. One can then verify that there exists $\varepsilon > 0$ such that a global maximum of $|g(u,s^2)|$ over $[u_1-\varepsilon, u_1+\varepsilon] \times [s_1-\varepsilon, s_1]$ is at (u_1,s_1). This contradicts the maximum principle, and thus proves that all modulus maxima propagate towards finer scales. ■ ■

Derivatives of Gaussians are most often used to guarantee that all maxima lines propagate up to the finest scales. Chaining together maxima into maxima lines is also a procedure for removing spurious modulus maxima created by numerical errors in regions where the wavelet transform is close to zero.

Isolated Singularities A wavelet transform may have a sequence of local maxima that converge to an abscissa v even though f is perfectly regular at v. This is the case of the maxima line of Figure 6.5 that converges to the abscissa $v = 0.23$. To detect singularities it is therefore not sufficient to follow the wavelet modulus maxima across scales. The Lipschitz regularity is calculated from the decay of the modulus maxima amplitude.

Let us suppose that for $s < s_0$ all modulus maxima that converge to v are included in a cone

$$|u - v| \leqslant Cs. \tag{6.35}$$

This means that f does not have oscillations that accelerate in the neighborhood of v. The potential singularity at v is necessarily isolated. Indeed, we can derive from Theorem 6.5 that the absence of maxima outside the cone of influence implies that f is uniformly Lipschitz n in the neighborhood of any $t \neq v$ with $t \in (v - Cs_0, v + Cs_0)$. The decay of $|Wf(u,s)|$ in the neighborhood of v is controlled by the decay of the modulus maxima included in the cone $|u - v| \leqslant Cs$. Theorem 6.3 implies that f is uniformly Lipschitz α in the neighborhood of v if and only if there exists $A > 0$ such that each modulus maximum (u,s) in the cone (6.35) satisfies

$$|Wf(u,s)| \leqslant A\, s^{\alpha+1/2}, \tag{6.36}$$

which is equivalent to

$$\log_2 |Wf(u,s)| \leqslant \log_2 A + \left(\alpha + \frac{1}{2}\right) \log_2 s. \tag{6.37}$$

The Lipschitz regularity at v is thus the maximum slope of $\log_2 |Wf(u,s)|$ as a function of $\log_2 s$ along the maxima lines converging to v.

In numerical calculations, the finest scale of the wavelet transform is limited by the resolution of the discrete data. From a sampling at intervals N^{-1}, Section 4.3.3 computes the discrete wavelet transform at scales $s \geqslant \lambda N^{-1}$, where λ is large enough to avoid sampling coarsely the wavelets at the finest scale. The Lipschitz regularity α of a singularity is then estimated by measuring the decay slope of $\log_2 |Wf(u,s)|$ as a function of $\log_2 s$ for $2^J \geqslant s \geqslant \lambda N^{-1}$. The largest scale 2^J should be smaller than the distance between two consecutive singularities to avoid having other singularities influence the value of $Wf(u,s)$. The sampling interval N^{-1} must therefore be small enough to measure α accurately. The signal in Figure 6.5(a) is defined by $N = 256$ samples. Figure 6.5(c) shows the decay of $\log_2 |Wf(u,s)|$ along the maxima line converging to $t = 0.05$. It has slope $\alpha + 1/2 \approx 1/2$ for $2^{-4} \geqslant s \geqslant 2^{-6}$. As expected, $\alpha = 0$ because the signal is discontinuous at $t = 0.05$. Along the second maxima line converging to $t = 0.42$ the slope is $\alpha + 1/2 \approx 1$, which indicates that the singularity is Lipschitz $1/2$.

When f is a function whose singularities are not isolated, finite resolution measurements are not sufficient to distinguish individual singularities. Section 6.4 describes a global approach that computes the singularity spectrum of multifractals by taking advantage of their self-similarity.

Smoothed Singularities The signal may have important variations that are infinitely continuously differentiable. For example, at the border of a shadow the grey level of an image varies quickly but is not discontinuous because of the diffraction effect. The smoothness of these transitions is modeled as a diffusion with a Gaussian kernel whose variance is measured from the decay of wavelet modulus maxima.

In the neighborhood of a sharp transition at v, we suppose that

$$f(t) = f_0 \star g_\sigma(t), \tag{6.38}$$

where g_σ is a Gaussian of variance σ^2:

$$g_\sigma(t) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-t^2}{2\sigma^2}\right). \tag{6.39}$$

If f_0 has a Lipschitz α singularity at v that is isolated and non-oscillating, it is uniformly Lipschitz α in the neighborhood of v. For wavelets that are derivatives of Gaussians, the following theorem [334] relates the decay of the wavelet transform to σ and α.

Theorem 6.7. *Let $\psi = (-1)^n \theta^{(n)}$ with $\theta(t) = \lambda \exp(-t^2/(2\beta^2))$. If $f = f_0 \star g_\sigma$ and f_0 is uniformly Lipschitz α on $[v-h, v+h]$ then there exists A such that*

$$\forall (u,s) \in [v-h, v+h] \times \mathbb{R}^+ \quad , \quad |Wf(u,s)| \leqslant A\, s^{\alpha+1/2} \left(1 + \frac{\sigma^2}{\beta^2 s^2}\right)^{-(n-\alpha)/2}. \tag{6.40}$$

Proof. The wavelet transform can be written

$$Wf(u,s) = s^n \frac{d^n}{du^n}(f \star \bar{\theta}_s)(u) = s^n \frac{d^n}{du^n}(f_0 \star g_\sigma \star \bar{\theta}_s)(u). \tag{6.41}$$

Since θ is a Gaussian, one can verify with a Fourier transform calculation that

$$\bar{\theta}_s \star g_\sigma(t) = \sqrt{\frac{s}{s_0}}\, \bar{\theta}_{s_0}(t) \quad \text{with} \quad s_0 = \sqrt{s^2 + \frac{\sigma^2}{\beta^2}}. \tag{6.42}$$

Inserting this result in (6.41) yields

$$Wf(u,s) = s^n \sqrt{\frac{s}{s_0}} \frac{d^n}{du^n}(f_0 \star \bar{\theta}_{s_0})(u) = \left(\frac{s}{s_0}\right)^{n+1/2} Wf_0(u, s_0). \tag{6.43}$$

Since f_0 is uniformly Lipschitz α on $[v-h, v+h]$, Theorem 6.3 proves that there exists $A > 0$ such that

$$\forall (u,s) \in [v-h, v+h] \times \mathbb{R}^+ \quad , \quad |Wf_0(u,s)| \leqslant A\, s^{\alpha+1/2}. \tag{6.44}$$

Inserting this in (6.43) gives

$$|Wf(u,s)| \leqslant A \left(\frac{s}{s_0}\right)^{n+1/2} s_0^{\alpha+1/2}, \tag{6.45}$$

from which we derive (6.40) by inserting the expression (6.42) of s_0. ■ ■

This theorem explains how the wavelet transform decay relates to the amount of diffusion of a singularity. At large scales $s \gg \sigma/\beta$, the Gaussian averaging is not "felt" by the wavelet transform which decays like $s^{\alpha+1/2}$. For $s \leqslant \sigma/\beta$, the variation of f at v is not sharp relative to s because of the Gaussian averaging. At these fine scales, the wavelet transform decays like $s^{n+1/2}$ because f is $\mathbf{C}^\infty$.

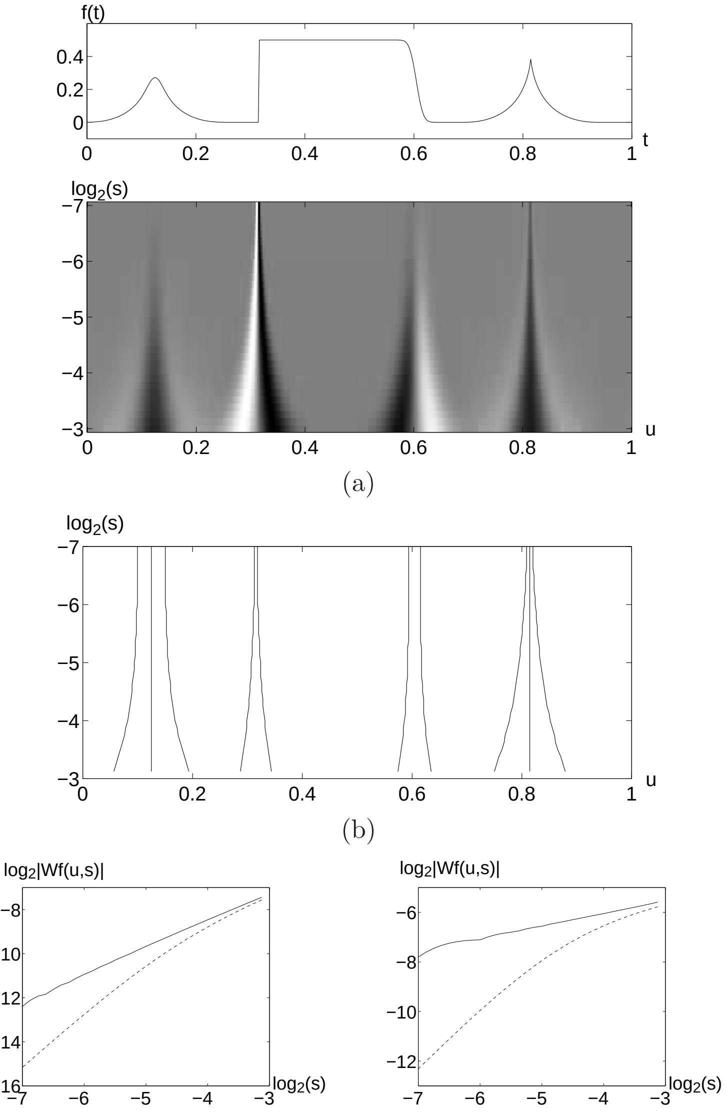


Figure 6.6: (a): Wavelet transform $Wf(u,s)$. (b): Modulus maxima of a wavelet transform computed $\psi = \theta''$, where θ is a Gaussian with variance $\beta = 1$. (c): Decay of $\log_2 |Wf(u,s)|$ along maxima curves. In the left figure, the solid and dotted lines correspond respectively to the maxima curves converging to $t = 0.81$ and $t = 0.12$. In the right figure, they correspond respectively to the curves converging to $t = 0.38$ and $t = 0.55$. The diffusion at $t = 0.12$ and $t = 0.55$ modifies the decay for $s \leqslant \sigma = 2^{-5}$.

The parameters K, α, and σ are numerically estimated from the decay of the modulus maxima along the maxima curves that converge towards v. The variance β^2 depends on the choice of wavelet and is known in advance. A regression is performed to approximate

$$\log_2 |Wf(u,s)| \approx \log_2(K) + \Big(\alpha + \frac{1}{2}\Big) \log_2 s - \frac{n-\alpha}{2} \log_2\Big(1 + \frac{\sigma^2}{\beta^2 s^2}\Big).$$

Figure 6.6 gives the wavelet modulus maxima computed with a wavelet that is a second derivative of a Gaussian. The decay of $\log_2 |Wf(u,s)|$ as a function of $\log_2 s$ is given along several maxima lines corresponding to smoothed and non-smoothed singularities. The wavelet is normalized so that $\beta = 1$ and the diffusion scale is $\sigma = 2^{-5}$.

6.2.2 Dyadic Maxima Representation

Wavelet transform maxima carry the properties of sharp signal transitions and singularities. By recovering a signal approximation from these maxima, signal singularities can be modified or removed by processing the wavelet modulus maxima.

For fast numerical computations, the detection of wavelet transform maxima is limited to dyadic scales $\{2^j\}_{j\in\mathbb{Z}}$. Suppose that ψ is a dyadic wavelet, which means that there exist $A > 0$ and B such that

$$\forall \omega \in \mathbb{R} - \{0\} \quad , \quad A \leqslant \sum_{j=-\infty}^{+\infty} |\hat{\psi}(2^j \omega)|^2 \leqslant B. \tag{6.46}$$

As a consequence of Theorem 5.11 on translation invariant frames, it is proved in Section 5.2 that the resulting translation invariant dyadic wavelet transform $\{Wf(u, 2^j)\}_{j\in\mathbb{Z}}$ is complete and stable. This dyadic wavelet transform has the same properties as a continuous wavelet transform $Wf(u,s)$. All theorems of Sections 6.1.3 and 6.2 remain valid if we restrict s to the dyadic scales $\{2^j\}_{j\in\mathbb{Z}}$. Singularities create sequences of maxima that converge towards the corresponding location at fine scales, and the Lipschitz regularity is calculated from the decay of the maxima amplitude.

Scale-space Maxima Support Mallat and Zhong [334] introduced a dyadic wavelet maxima representation, with a scale-space approximation support Λ of modulus maxima $(u, 2^j)$ of Wf.

Wavelet maxima can be interpreted as points of 0 or π phase for an appropriate complex wavelet transform. Let ψ' be the derivative of ψ and $\psi'_{u,2^j}(t) = 2^{-j/2}\psi'(2^{-j}(t-u))$. If Wf has a local extremum at u_0 then

$$\frac{\partial Wf(u_0, 2^j)}{\partial u} = -2^{-j} \, \langle f, \psi'_{2^j,u_0} \rangle = 0 \ .$$

Let us introduce a complex wavelet $\psi^c(t) = \psi(t) + i\psi'(t)$. If $(u,s) \in \Lambda$ then the resulting complex wavelet transform value is

$$W^c f(u, 2^j) = \langle f, \psi^c_{2^j,u} \rangle = \langle f, \psi_{2^j,u} \rangle + i \, \langle f, \psi'_{2^j,u} \rangle = Wf(u,s) \tag{6.47}$$

because $\langle f, \psi'_{2^j,u} \rangle = 0$. The complex wavelet value $W^c f(u,s)$ has a phase equal to 0 or π depending upon the sign of $Wf(u,s)$, and a modulus $|W^c f(u,s)| = |Wf(u,s)|$.

Figure 6.7(c) gives an example computed with the quadratic spline dyadic wavelet in Figure 5.3. This adaptive sampling of u produces a translation-invariant representation, which is important for pattern recognition. When f is translated by τ each $Wf(2^j, u)$ is translated by τ so the maxima support is translated by τ, as illustrated by Figure 6.8. This is not the case for wavelet frame coefficients, where the translation parameter u is sampled with an interval proportional to the scale a^j, as explained in Section 5.3.

Approximations from Wavelet Maxima Mallat and Zhong [334] recover signal approximations from their wavelet maxima with an alternate projection algorithm, but several other algorithms have been proposed [142, 177, 261]. In the following we concentrate on orthogonal projection approximations on the space generated by wavelets in the scale-space maxima support. Numerical experiments show that with dyadic wavelets of compact support, recover signal approximations with a relative mean-square error that is typically of the order of 10^{-2}. For general dyadic wavelets, Meyer [43] and Berman [101] proved that exact reconstruction is not possible. They found families of continuous or discrete signals whose dyadic wavelet transforms have the same modulus maxima. However, signals with same wavelet maxima differ from each other by small amplitude errors introducing no oscillation, which explains the success of numerical reconstructions [334]. If the signal has a band-limited Fourier transform and if $\hat{\psi}$ has a compact support, then Kicey and Lennard [298] proved that wavelet modulus maxima define a complete and stable signal representation.

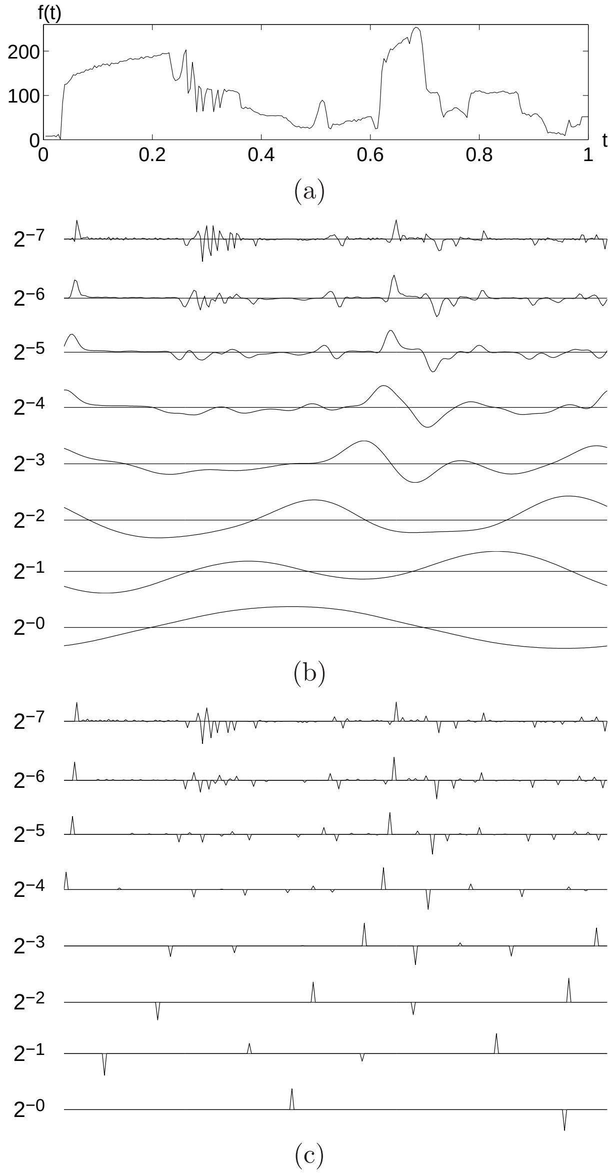


Figure 6.7: (a): Intensity variation along one row of the Lena image. (b): Dyadic wavelet transform computed at all scales $2N^{-1} \leqslant 2^j \leqslant 1$, with the quadratic spline wavelet $\psi = -\theta'$ shown in Figure 5.3. (c): Modulus maxima of the dyadic wavelet transform.

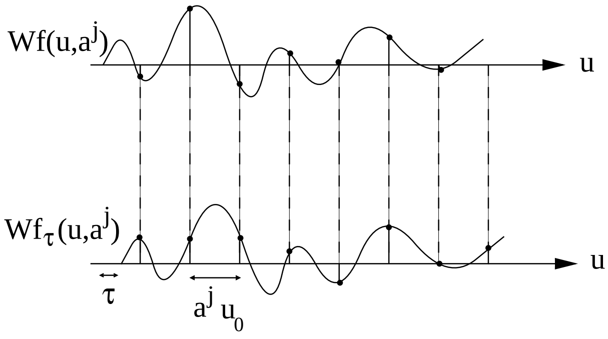


Figure 6.8: If $f_\tau(t) = f(t-\tau)$ then $Wf_\tau(u,a^j) = Wf(u-\tau,a^j)$. Uniformly sampling $Wf_\tau(u,a^j)$ and $Wf(u,a^j)$ at $u = na^j u_0$ may yield very different values if $\tau \neq ku_0a^j$.

As a result of (6.47), the wavelet modulus maxima specifies the complex wavelet inner products $\{\langle f, \psi^c_{u,2^j}\rangle\}_{(u,2^j)\in\Lambda}$. A modulus maxima approximation can thus be computed as an orthogonal projection of f on the space generated by the complex wavelets $\{\psi^c_{u,2^j}\}_{(u,2^j)\in\Lambda}$. To reduce computations, the explicit extrema condition $\langle \tilde{f}, \psi'_{u,2^j}\rangle = 0$ is often removed, because it is indirectly almost obtained by the calculating the orthogonal projection over the space $\mathbf{V}_\Lambda$ generated by the real maxima wavelets $\{\psi_{u,2^j}\}_{(u,2^j)\in\Lambda}$. Section 5.1.3 shows that this orthogonal projection is obtained from the dual frame $\{\tilde{\psi}_{\Lambda,u,2^j}\}_{(u,2^j)\in\Lambda}$ of $\{\psi_{u,2^j}\}_{(u,2^j)\in\Lambda}$ in $\mathbf{V}_\Lambda$:

$$f_\Lambda = \sum_{(u,2^j)\in\Lambda} \langle f, \psi_{u,2^j}\rangle\, \tilde{\psi}_{u,2^j} \ . \tag{6.48}$$

The dual synthesis algorithm of Section 5.1.3 computes this orthogonal projection by inverting the symmetric operator

$$Ly = \sum_{(u,2^j)\in\Lambda} \langle y, \psi_{u,2^j}\rangle\, \psi_{u,2^j} \ , \tag{6.49}$$

with a conjugate gradient algorithm.

Example 6.1. *Figure 6.9(b) shows the approximation f_Λ, recovered with 10 conjugate gradient iterations, from the wavelet maxima in Figure 6.7(c). This reconstruction is calculated with real instead of complex wavelets. After 20 iterations, the reconstruction error is $\|f - \tilde{f}\|/\|f\| = 2.5\, 10^{-2}$. Figure 6.9(c) shows the signal reconstructed from the 50% of wavelet maxima that have the largest amplitude. Sharp signal transitions corresponding to large wavelet maxima have not been affected, but small texture variations disappear because the corresponding maxima are removed. The resulting signal is piecewise regular.*

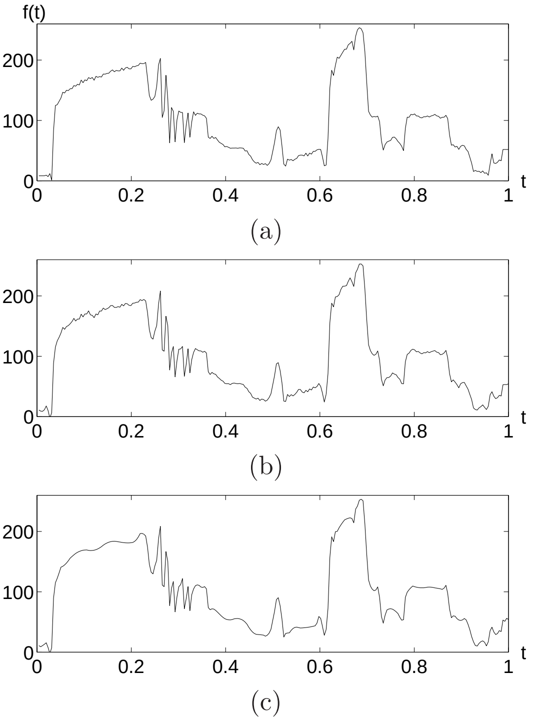


Figure 6.9: (a): Original signal f. (b): Signal approximation f_Λ recovered from the dyadic wavelet maxima shown in Figure 6.7(c). (c): Approximation recovered from the 50% largest maxima.

Fast Discrete Calculations The conjugate gradient inversion of the operator (6.49) requires to iterate on this operator many times. If there many local maxima, it is more efficient to compute

$Wy(u,2^j) = \langle y, \psi_{u,2^j} \rangle$ for all $(u,2^j)$, with the "algorithme à trous" of Section 5.2.2. For a signal of size N, it cascades convolutions with two filters $h[n]$ and $g[n]$, up to a maximum scale $J = \log_2 N$, with $O(N \log_2 N)$ operations. All non-maxima coefficients for $(u,2^j) \notin \Lambda$ are then set to zero. The reconstruction of Ly is computed by modifying the filter bank reconstruction given by Theorem 5.14, which also requires $O(N \log_2 N)$ operations. The decomposition and reconstruction wavelets are the same in (6.49), so the reconstruction filters are $\tilde{h}[n] = h[n]$ and $\tilde{g}[n] = g[n]$. The factor 1/2 in (5.69) is also removed because the reconstruction wavelets in (6.49) are not attenuated by 2^{-j} as in a non-sampled wavelet reconstruction (5.47). For $J = \log_2 N$, we initialize $\tilde{a}_J[n] = C/\sqrt{N}$ where C is average signal value and for $\log_2 N > j \geqslant 0$ we compute

$$\tilde{a}_j[n] = \tilde{a}_{j+1} \star h_j[n] + \tilde{d}_{j+1} \star g_j[n]. \tag{6.50}$$

One can verify that $Ly[n] = \tilde{a}_0[n]$ with the same derivations as in the proof of Theorem 5.14.

The signal approximations shown in Figure 6.9 are computed with the filters of Table 5.1. About 10 iterations of conjugate gradient are usually sufficient to recover an approximation with $\|f_\Lambda - f\|/\|f\|$ of the order of 10^{-2}, if all wavelet maxima are kept.

6.3 Multiscale Edge Detection

Image edges are often important for pattern recognition. This is well illustrated by our visual ability to recognize an object from a drawing that gives a rough outline of contours. But, what is an edge? It could be defined as points where the image intensity has sharp transitions. A closer look shows that this definition is often not satisfactory. Image textures do have sharp intensity variations that are often not considered as edges. When looking at a brick wall, we may decide that the edges are the contours of the wall whereas the bricks define a texture. Alternatively, we may include the contours of each brick in the set of edges and consider the irregular surface of each brick as a texture. The discrimination of edges versus textures depends on the scale of analysis. This has motivated computer vision researchers to detect sharp image variations at different scales [40, 376]. The next section describes the multiscale Canny edge detector [138]. It is equivalent to detecting modulus maxima in a two-dimensional dyadic wavelet transform [334]. The scale-space support of these modulus maxima thus correspond to multiscale edges. The Lipschitz regularity of edge points is derived from the decay of wavelet modulus maxima across scales. Image approximations are recovered with an orthogonal projection on the wavelets of the modulus maxima support, with no visual degradation. Image processing algorithms can thus be implemented on multiscale edges.

6.3.1 Wavelet Maxima for Images

Canny Edge Detection The Canny algorithm detects points of sharp variation in an image $f(x_1,x_2)$ by calculating the modulus of its gradient vector

$$\vec{\nabla} f = \left(\frac{\partial f}{\partial x_1} \, , \, \frac{\partial f}{\partial x_2} \right) \, . \tag{6.51}$$

The partial derivative of f in the direction of a unit vector $\vec{n} = (\cos\alpha, \sin\alpha)$ in the $x = (x_1,x_2)$ plane is calculated as an inner product with the gradient vector

$$\frac{\partial f}{\partial \vec{n}} = \vec{\nabla} f \, . \, \vec{n} = \frac{\partial f}{\partial x_1} \cos\alpha + \frac{\partial f}{\partial x_2} \sin\alpha \, .$$

The absolute value of this partial derivative is maximum if $\vec{n}$ is colinear to $\vec{\nabla} f$. This shows that $\vec{\nabla} f(x)$ is parallel to the direction of maximum change of the surface $f(x)$. A point $y \in \mathbb{R}^2$ is defined as an edge if $|\vec{\nabla} f(x)|$ is locally maximum at $x = y$ when $x = y + \lambda \vec{\nabla} f(y)$ for $|\lambda|$ small enough. This means that the partial derivatives of f reach a local maximum at $x = y$, when x varies in a one-dimensional neighborhood of y along the direction of maximum change of f at y. These edge points are inflection points of f.

Multiscale Edge Detection A multiscale version of this edge detector is implemented by smoothing the surface with a convolution kernel $\theta(x)$ that is dilated. This is computed with two wavelets that are the partial derivatives of θ:

$$\psi^1 = -\frac{\partial \theta}{\partial x_1} \quad \text{and} \quad \psi^2 = -\frac{\partial \theta}{\partial x_2}. \tag{6.52}$$

The scale varies along the dyadic sequence $\{2^j\}_{j \in \mathbb{Z}}$ to limit computations and storage. For $1 \leqslant k \leqslant 2$, we denote for $x = (x_1, x_2)$

$$\psi^k_{2^j}(x_1, x_2) = \frac{1}{2^j}\, \psi^k\left(\frac{x_1}{2^j}, \frac{x_2}{2^j}\right) \quad \text{and} \quad \bar{\psi}^k_{2^j}(x) = \psi^k_{2^j}(-x).$$

In the two directions indexed by $1 \leqslant k \leqslant 2$, the dyadic wavelet transform of $f \in \mathbf{L^2}(\mathbb{R}^2)$ at $u = (u_1, u_2)$ is

$$W^k f(u, 2^j) = \langle f(x), \psi^k_{2^j}(x - u)\rangle = f \star \bar{\psi}^k_{2^j}(u) \ . \tag{6.53}$$

Section 5.5 gives necessary and sufficient conditions for obtaining a complete and stable representation.

Let us denote $\theta_{2^j}(x) = 2^{-j}\,\theta(2^{-j}x)$ and $\bar{\theta}_{2^j}(x) = \theta_{2^j}(-x)$. The two scaled wavelets can be rewritten

$$\bar{\psi}^1_{2^j} = 2^j\, \frac{\partial \bar{\theta}_{2^j}}{\partial x_1} \quad \text{and} \quad \bar{\psi}^2_{2^j} = 2^j\, \frac{\partial \bar{\theta}_{2^j}}{\partial x_2} \ .$$

We thus derive from (6.53) that the wavelet transform components are proportional to the coordinates of the gradient vector of f smoothed by $\bar{\theta}_{2^j}$:

$$\begin{pmatrix} W^1 f(u, 2^j) \\ W^2 f(u, 2^j) \end{pmatrix} = 2^j \begin{pmatrix} \frac{\partial}{\partial u_1}(f \star \bar{\theta}_{2^j})(u) \\ \frac{\partial}{\partial u_2}(f \star \bar{\theta}_{2^j})(u) \end{pmatrix} = 2^j\, \vec{\nabla}(f \star \bar{\theta}_{2^j})(u) \ . \tag{6.54}$$

The modulus of this gradient vector is proportional to the wavelet transform modulus

$$Mf(u, 2^j) = \sqrt{|W^1 f(u, 2^j)|^2 + |W^2 f(u, 2^j)|^2}. \tag{6.55}$$

Let $Af(u, 2^j)$ be the angle of the wavelet transform vector (6.54) in the plane (x_1, x_2)

$$Af(u, 2^j) = \begin{cases} \alpha(u) & \text{if } W^1 f(u, 2^j) \geqslant 0 \\ \pi + \alpha(u) & \text{if } W^1 f(u, 2^j) < 0 \end{cases} \tag{6.56}$$

with

$$\alpha(u) = \tan^{-1}\left(\frac{W^2 f(u, 2^j)}{W^1 f(u, 2^j)}\right) \ .$$

The unit vector $\vec{n}_j(u) = (\cos Af(u, 2^j), \sin Af(u, 2^j))$ is colinear to $\vec{\nabla}(f \star \bar{\theta}_{2^j})(u)$. An edge point at the scale 2^j is a point v such that $Mf(u, 2^j)$ is locally maximum at $u = v$ when $u = v + \lambda\, \vec{n}_j(v)$ for $|\lambda|$ small enough. These points are also called wavelet transform *modulus maxima*. The smoothed image $f \star \bar{\theta}_{2^j}$ has an inflection point at a modulus maximum location. Figure 6.10 gives an example where the wavelet modulus maxima are located along the contour of a circle.

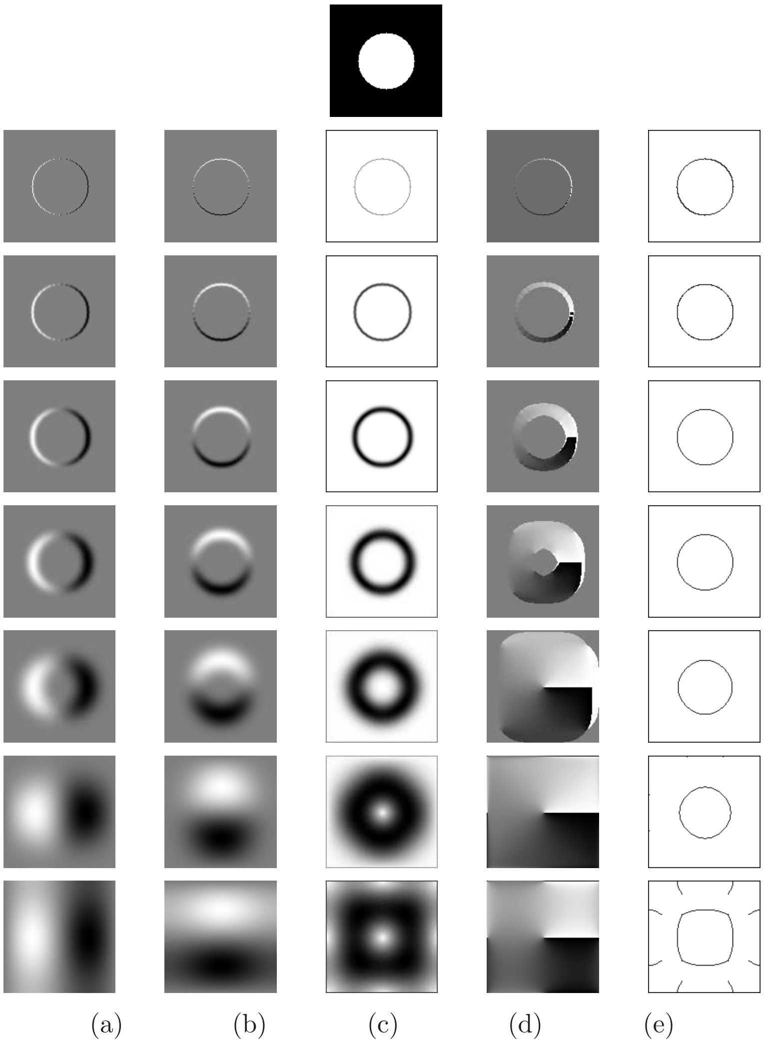
Maxima curves Edge points are distributed along curves that often correspond to the boundary of important structures. Individual wavelet modulus maxima are chained together to form a maxima curve that follows an edge. At any location, the tangent of the edge curve is approximated by computing the tangent of a level set. This tangent direction is used to chain wavelet maxima that are along the same edge curve.

The level sets of $g(x)$ are the curves $x(s)$ in the (x_1, x_2) plane where $g(x(s))$ is constant. The parameter s is the arc-length of the level set. Let $\vec{\tau} = (\tau_1, \tau_2)$ be the direction of the tangent of $x(s)$. Since $g(x(s))$ is constant when s varies,

$$\frac{\partial g(x(s))}{\partial s} = \frac{\partial g}{\partial x_1}\, \tau_1 + \frac{\partial g}{\partial x_2}\, \tau_2 = \vec{\nabla} g \, . \, \vec{\tau} = 0 \ .$$

So $\vec{\nabla} g(x)$ is perpendicular to the direction $\vec{\tau}$ of the tangent of the level set that goes through x.

This level set property applied to $g = f \star \bar{\theta}_{2^j}$ proves that at a maximum point v the vector $\vec{n}_j(v)$ of angle $Af(v, 2^j)$ is perpendicular to the level set of $f \star \bar{\theta}_{2^j}$ going through v. If the intensity profile remains constant along an edge, then the inflection points (maxima points) are along a level set. The tangent of the maxima curve is therefore perpendicular to $\vec{n}_j(v)$. The intensity profile of an edge may not be constant but its variations are often negligible over a neighborhood of size 2^j for a sufficiently small scale 2^j, unless we are near a corner. The tangent of the maxima curve is then nearly perpendicular to $\vec{n}_j(v)$. In discrete calculations, maxima curves are thus recovered by chaining together any two wavelet maxima at v and $v + \vec{n}$, which are neighbors over the image sampling grid and such that $\vec{n}$ is nearly perpendicular to $\vec{n}_j(v)$.



(a) (b) (c) (d) (e)

Figure 6.10: The top image has $N = 128^2$ pixels. (a): Wavelet transform in the horizontal direction, with a scale 2^j that increases from top to bottom: $\{W^1 f(u, 2^j)\}_{-6 \leqslant j \leqslant 0}$. Black, grey and white pixels correspond respectively to negative, zero and positive values. (b): Vertical direction: $\{W^2 f(u, 2^j)\}_{-6 \leqslant j \leqslant 0}$. (c): Wavelet transform modulus $\{Mf(u, 2^j)\}_{-6 \leqslant j \leqslant 0}$. White and black pixels correspond respectively to zero and large amplitude coefficients. (d): Angles $\{Af(u, 2^j)\}_{-6 \leqslant j \leqslant 0}$ at points where the modulus is non-zero. (e): The wavelet modulus maxima support is shown in black.

Example 6.2. *The dyadic wavelet transform of the image in Figure 6.10 yields modulus images* $Mf(v, 2^j)$ *whose maxima are along the boundary of a disk. This circular edge is also a level set of the image. The vector* $\vec{n}_j(v)$ *of angle* $Af(v, 2^j)$ *is thus perpendicular to the edge at the maxima locations.*

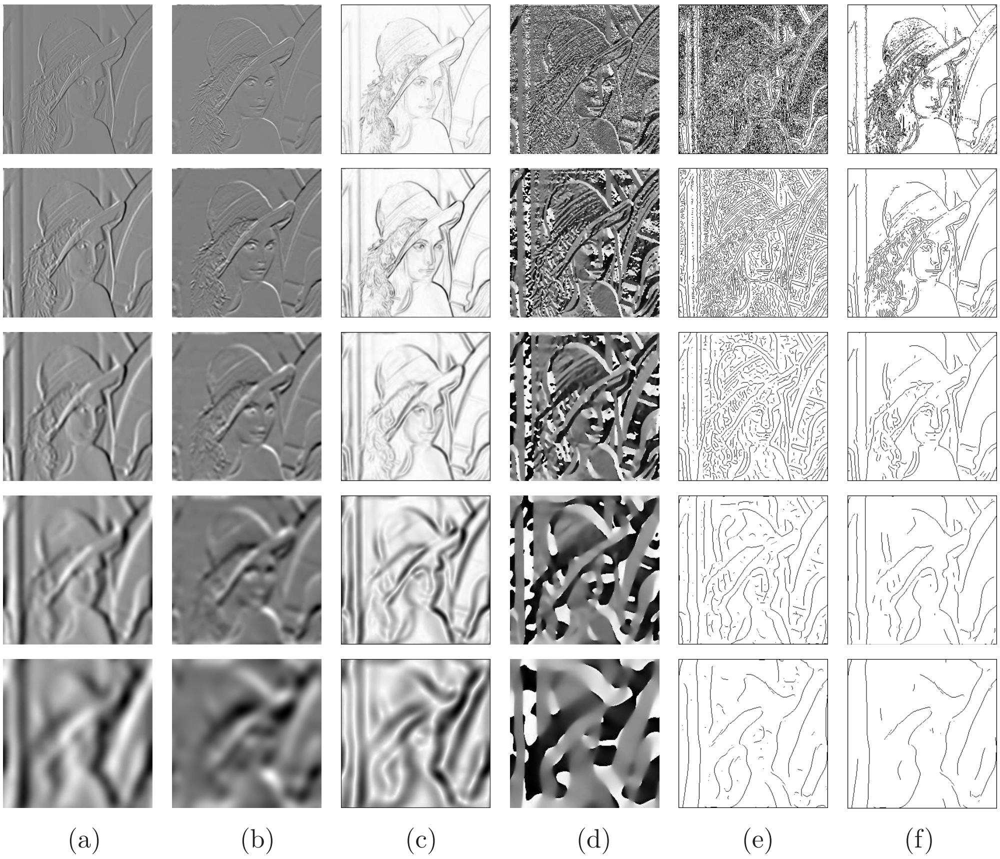


Figure 6.11: Multiscale edges of the Lena image shown in Figure 6.12. (a): $\{W^1 f(u,2^j)\}_{-7\leqslant j\leqslant -3}$. (b): $\{W^2 f(u,2^j)\}_{-7\leqslant j\leqslant -3}$. (c): $\{Mf(u,2^j)\}_{-7\leqslant j\leqslant -3}$. (d): $\{Af(u,2^j)\}_{-7\leqslant j\leqslant -3}$. (e): Modulus maxima support. (f): Support of maxima whose modulus values are above a threshold.

Example 6.3. *In the Lena image shown in Figure 6.11, some edges disappear when the scale increases. These correspond to fine scale intensity variations that are removed by the averaging with* $\bar\theta_{2^j}$ *when* 2^j *is large. This averaging also modifies the position of the remaining edges. Figure 6.11(f) displays the wavelet maxima such that* $Mf(v,2^j) \geqslant T$*, for a given threshold* T*. They indicate the location of edges where the image has large amplitude variations.*

Lipschitz Regularity The decay of the two-dimensional wavelet transform depends on the regularity of f. We restrict the analysis to Lipschitz exponents $0 \leqslant \alpha \leqslant 1$. A function f is said to be Lipschitz α at $v = (v_1, v_2)$ if there exists $K > 0$ such that for all $(x_1, x_2) \in \mathbb{R}^2$

$$|f(x_1,x_2) - f(v_1,v_2)| \leqslant K\,(|x_1 - v_1|^2 + |x_2 - v_2|^2)^{\alpha/2}. \tag{6.57}$$

If there exists $K > 0$ such that (6.57) is satisfied for any $v \in \Omega$ then f is uniformly Lipschitz α over Ω. As in one dimension, the Lipschitz regularity of a function f is related to the asymptotic decay $|W^1 f(u,2^j)|$ and $|W^2 f(u,2^j)|$ in the corresponding neighborhood. This decay is controlled by $Mf(u,2^j)$. Like in Theorem 6.3, one can prove that f is uniformly Lipschitz α inside a bounded domain of $\mathbb{R}^2$ if and only if there exists $A > 0$ such that for all u inside this domain and all scales 2^j

$$|Mf(u,2^j)| \leqslant A\,2^{j(\alpha+1)}. \tag{6.58}$$

Suppose that the image has an isolated edge curve along which f has Lipschitz regularity α. The value of $|Mf(u,2^j)|$ in a two-dimensional neighborhood of the edge curve can be bounded by the wavelet modulus values along the edge curve. The Lipschitz regularity α of the edge is estimated with (6.58) by measuring the slope of $\log_2 |Mf(u,2^j)|$ as a function of j. If f is not singular but has a smooth transition along the edge, the smoothness can be quantified by the variance σ^2 of a two-dimensional Gaussian blur. The value of σ^2 is estimated by generalizing Theorem 6.7.

Reconstruction from Edges In his book about vision, Marr [40] conjectured that images can be reconstructed from multiscale edges. For a Canny edge detector, this is equivalent to recovering images from wavelet modulus maxima. In two dimensions, whether dyadic wavelet maxima define a complete and stable representation is still an open mathematical problem. However, the algorithm of Mallat and Zhong [334] recovers an image approximation that is visually identical to the original one. In the following, image approximations are computed by projecting the image on the space generated by wavelets on the modulus maxima support.

Let Λ be the set of all modulus maxima points $(u, 2^j)$. Let $\vec{n}$ be the unit vector in the direction of $Af(u, 2^j)$ and

$$\psi^3_{u,2^j}(x) = 2^{2j}\, \frac{\partial^2 \theta_{2^j}(x-u)}{\partial \vec{n}^2}\ .$$

Since the wavelet transform modulus $Mf(u, 2^j)$ has a local extremum at u in the direction of $\vec{n}$, it results that

$$\langle f, \psi^3_{u,2^j} \rangle = 0\ . \tag{6.59}$$

A modulus maxima representation provides the set of inner products $\{\langle f, \psi^k_{u,2^j}\rangle\}_{(u,2^j)\in\Lambda, 1\leqslant k\leqslant 3}$. A modulus maxima approximation f_Λ can be computed as an orthogonal projection of f on the space generated by the family of maxima wavelets $\{\psi^k_{u,2^j}\}_{(u,2^j)\in\Lambda, 1\leqslant k\leqslant 3}$.

To reduce computations, the conditions on the third wavelets $\langle f, \psi^3_{u,2^j}\rangle = 0$ is removed, because it is indirectly almost imposed by the orthogonal projection over the space $\mathbf{V}_\Lambda$ generated by the other two wavelets for $k = 1, 2$. The dual synthesis algorithm of Section 5.1.3 computes this orthogonal projection f_Λ by inverting the symmetric operator

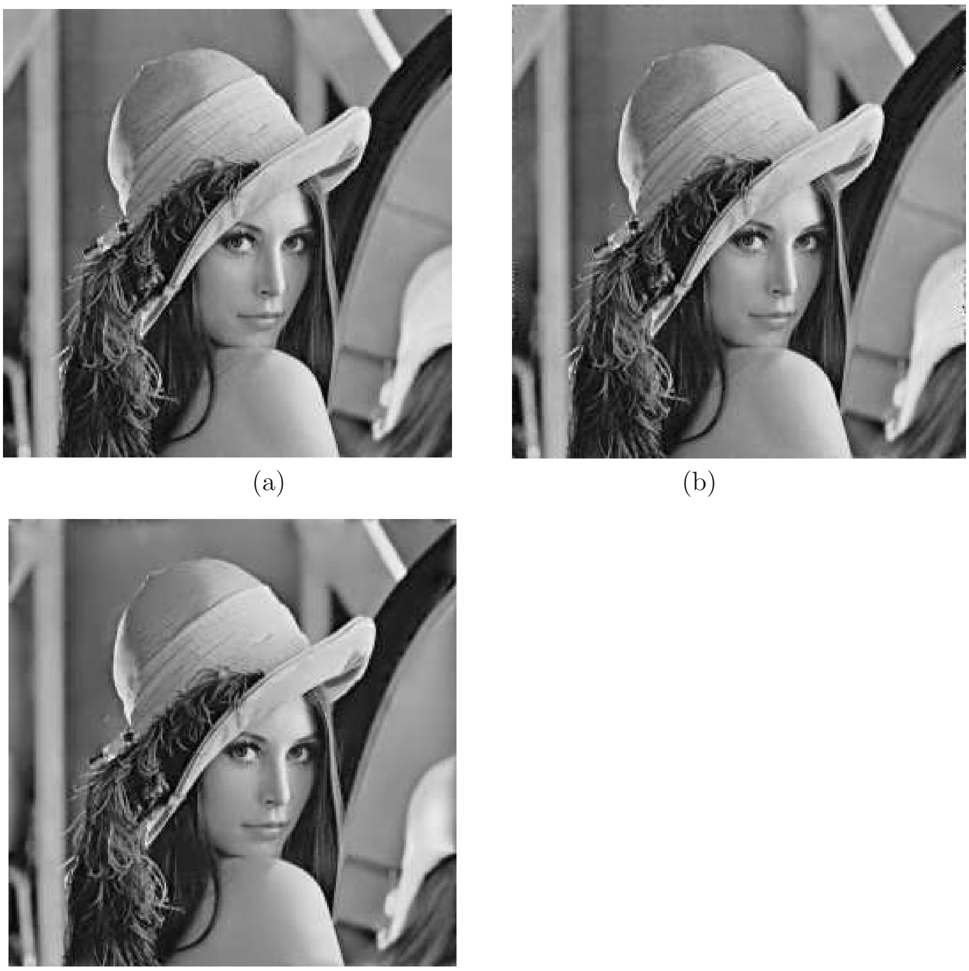
$$Ly = \sum_{(u,2^j)\in\Lambda} \sum_{k=1}^{2} \langle y, \psi^k_{u,2^j}\rangle\, \psi^k_{u,2^j}\ , \tag{6.60}$$

with a conjugate gradient algorithm. When keeping all modulus maxima, the resulting modulus maxima approximation f_Λ satisfies $\|f_\Lambda - f\|/\|f\| \leqslant 10^{-2}$. Singularities and edges are nearly perfectly recovered and no spurious oscillations are introduced. The images differ slightly in smooth regions, but visually this is not noticeable.

Example 6.4. *The image reconstructed in Figure 6.12(b) is visually identical to the original image. It is recovered with 10 conjugate gradient iterations. After 20 iterations, the relative mean-square reconstruction error is* $\|\tilde{f} - f\|/\|f\| = 4\,10^{-3}$. *The thresholding of maxima accounts for the disappearance of image structures from the reconstruction shown in Figure 6.12(c). Sharp image variations are recovered.*

Denoising by Multiscale Edge Thresholding Multiscale edge representations can be used to reduce additive noise. Denoising algorithms by thresholding wavelet coefficients are presented in Section 11.3.1. Block thresholding in Section 11.4.2 regularizes this coefficient selection by regrouping them in square blocks. Similarly, a noise removal algorithm can be implemented by thresholding multiscale wavelet maxima, while taking into account their geometric properties. A simple approach implemented by Hwang and Mallat [331], chains the maxima into curves, that are thresholded as a block. In Figure 6.13, noisy modulus maxima are shown on the second row and the third row displays the thresholded modulus maxima chains. At the finest scale shown on the left, the noise is masking the image structures. Maxima chains are selected by using the position of the selected maxima at the previous scale. An image approximation is recovered from the selected wavelet maxima. Edges are visually well recovered but textures and fine structures are removed. This produces a cartoon-like image.

Illusory Contours A multiscale wavelet edge detector defines edges as points where the image intensity varies sharply. This definition is however too restrictive when edges are used to find the contours of objects. For image segmentation, edges must define closed curves that outline the boundaries of each region. Because of noise or light variations, local edge detectors produce contours with holes. Filling these holes requires some prior knowledge about the behavior of



(a) (b)

(c)

Figure 6.12: (a): Original Lena. (b): Reconstructed from the wavelet maxima displayed in Figure 6.11(e) and larger scale maxima. (c): Reconstructed from the thresholded wavelet maxima displayed in Figure 6.11(f) and larger scale maxima.

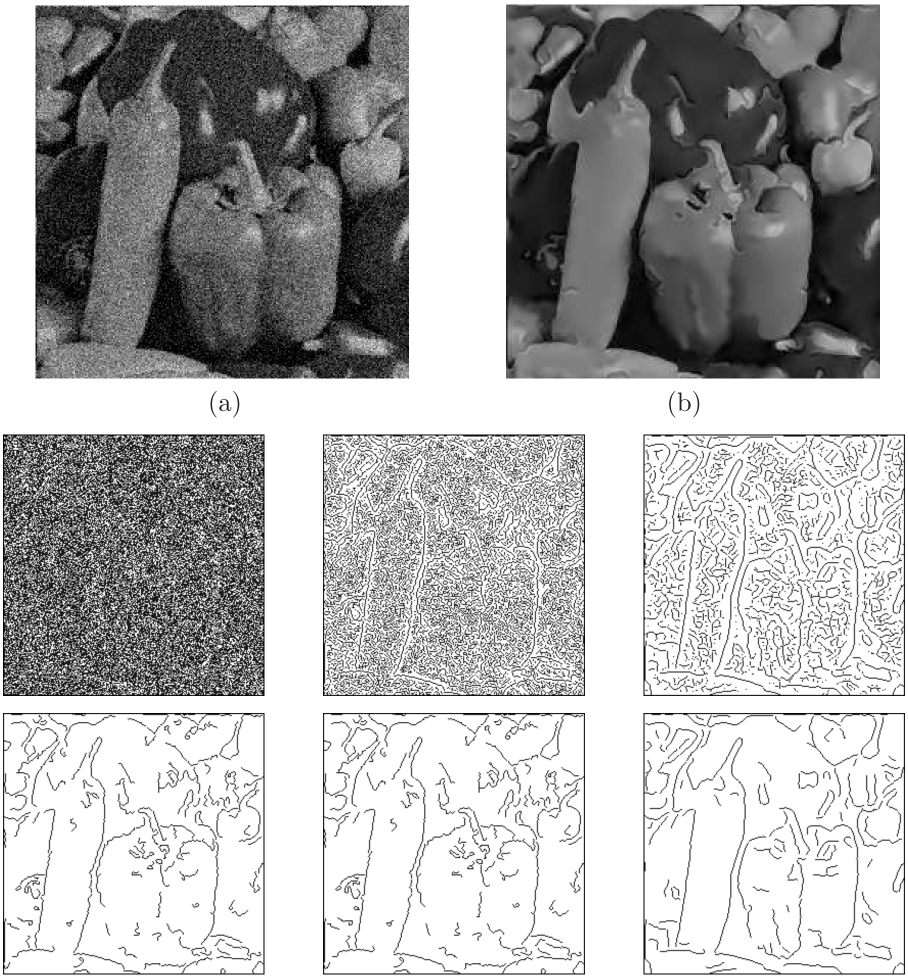

(a) (b)

Figure 6.13: (a): Noise peppers image. (b): Peppers image restored from the thresholding maxima chains shown on the third image row. The images on the second row show the wavelet maxima support of the noisy image. The scale increases from left to right, from 2^{-7} to 2^{-5}. The third row gives The maxima support computed with a thresholding selection of the maxima chains.

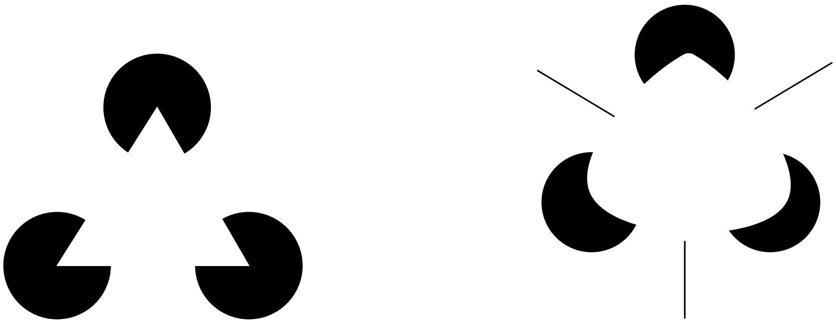


Figure 6.14: The illusory edges of a straight and a curved triangle are perceived in domains where the images are uniformly white.

edges in the image. The illusion of the Kanizsa triangle [35] shows that such an edge filling is performed by the human visual system. In Figure 6.14, one can "see" the edges of a straight and a curved triangle although the image grey level remains uniformly white between the black discs. Closing edge curves and understanding illusory contours requires computational models that are not as local as multiscale differential operators. Such contours can be obtained as the solution of a global optimization that incorporates constraints on the regularity of contours and which takes into account the existence of occlusions [248].

6.3.2 Fast Multiscale Edge Computations

The dyadic wavelet transform of an image of N pixels is computed with a separable extension of the filter bank algorithm described in Section 5.2.2. A fast multiscale edge detection is derived [334].

Wavelet Design Edge detection wavelets (6.52) are designed as separable products of one-dimensional dyadic wavelets, constructed in Section 5.2.1. Their Fourier transform is

$$\hat{\psi}^1(\omega_1,\omega_2) = \hat{g}\left(\frac{\omega_1}{2}\right)\hat{\phi}\left(\frac{\omega_1}{2}\right)\hat{\phi}\left(\frac{\omega_2}{2}\right), \tag{6.61}$$

and

$$\hat{\psi}^2(\omega_1,\omega_2) = \hat{g}\left(\frac{\omega_2}{2}\right)\hat{\phi}\left(\frac{\omega_1}{2}\right)\hat{\phi}\left(\frac{\omega_2}{2}\right) \quad , \tag{6.62}$$

where $\hat{\phi}(\omega)$ is a scaling function whose energy is concentrated at low frequencies and

$$\hat{g}(\omega) = -i\,\sqrt{2}\,\sin\left(\frac{\omega}{2}\right)\exp\left(\frac{-i\omega}{2}\right) \; . \tag{6.63}$$

This transfer function is the Fourier transform of a finite difference filter which is a discrete approximation of a derivative

$$\frac{g[p]}{\sqrt{2}} = \begin{cases} -0.5 & \text{if } p=0 \\ \;\;0.5 & \text{if } p=1 \\ \;\;0 & \text{otherwise} \end{cases} \; . \tag{6.64}$$

The resulting wavelets ψ^1 and ψ^2 are finite difference approximations of partial derivatives along x_1 and x_2 of $\theta(x_1,x_2) = 4\,\phi(2x_1)\,\phi(2x_2)$.

To implement the dyadic wavelet transform with a filter bank algorithm, the scaling function $\hat{\phi}$ is calculated, as in (5.58), with an infinite product:

$$\hat{\phi}(\omega) = \prod_{p=1}^{+\infty}\frac{\hat{h}(2^{-p}\omega)}{\sqrt{2}} = \frac{1}{\sqrt{2}}\,\hat{h}\left(\frac{\omega}{2}\right)\hat{\phi}\left(\frac{\omega}{2}\right). \tag{6.65}$$

The 2π periodic function $\hat{h}$ is the transfer function of a finite impulse response low-pass filter $h[p]$. We showed in (5.63) that the Fourier transform of a box spline of degree m

$$\hat{\phi}(\omega) = \left(\frac{\sin(\omega/2)}{\omega/2}\right)^{m+1}\exp\left(\frac{-i\varepsilon\omega}{2}\right) \qquad \text{with} \quad \varepsilon = \begin{cases} 1 & \text{if } m \text{ is even} \\ 0 & \text{if } m \text{ is odd} \end{cases}$$

is obtained with

$$\hat{h}(\omega) = \sqrt{2}\,\frac{\hat{\phi}(2\omega)}{\hat{\phi}(\omega)} = \sqrt{2}\,\left(\cos\frac{\omega}{2}\right)^{m+1}\exp\left(\frac{-i\varepsilon\omega}{2}\right) \; .$$

Table 5.1 gives $h[p]$ for $m=2$.

"Algorithme à trous" The one-dimensional "algorithme à trous" of Section 5.2.2 is extended in two dimensions with convolutions along the image rows and columns.

Each sample $a_0[n]$ of the normalized discrete image is considered to be an average of the input analog image f calculated with the kernel $\phi(x_1)\,\phi(x_2)$ translated at $n=(n_1,n_2)$:

$$a_0[n_1,n_2]=\langle f(x_1,x_2)\,,\,\phi(x_1-n_1)\,\phi(x_2-n_2)\rangle\ .$$

This is further justified in Section 7.7.3. For any $j\geqslant 0$, we denote

$$a_j[n_1,n_2]=\langle f(x_1,x_2)\,,\,\phi_{2^j}(x_1-n_1)\,\phi_{2^j}(x_2-n_2)\rangle.$$

The discrete wavelet coefficients at $n=(n_1,n_2)$ are

$$d_j^1[n]=W^1f(n,2^j)\ \text{ and }\ d_j^2[n]=W^2f(n,2^j)\ .$$

They are calculated with separable convolutions.

For any $j\geqslant 0$, the filter $h[p]$ "dilated" by 2^j is defined by

$$\bar{h}_j[p]=\begin{cases} h[-p/2^j] & \text{if } p/2^j\in\mathbb{Z}\\ 0 & \text{otherwise}\end{cases} \tag{6.66}$$

and for $j>0$, a centered finite difference filter is defined by

$$\frac{\bar{g}_j[p]}{\sqrt{2}}=\begin{cases} 0.5 & \text{if } p=-2^{j-1}\\ -0.5 & \text{if } p=2^{j-1}\\ 0 & \text{otherwise}\end{cases}\ . \tag{6.67}$$

For $j=0$, we define $\bar{g}_0[0]/\sqrt{2}=-0.5$, $\bar{g}_0[-1]/\sqrt{2}=-0.5$ and $\bar{g}_0[p]=0$ for $p\neq 0,-1$. A separable two-dimensional filter is written

$$\alpha\beta[n_1,n_2]=\alpha[n_1]\,\beta[n_2]\ ,$$

and $\delta[n]$ is a discrete Dirac. Similarly to Theorem 5.14, one can prove that for any $j\geqslant 0$ and any $n=(n_1,n_2)$

$$a_{j+1}[n] = a_j\star\bar{h}_j\bar{h}_j[n], \tag{6.68}$$
$$d_{j+1}^1[n] = a_j\star\bar{g}_j\delta[n], \tag{6.69}$$
$$d_{j+1}^2[n] = a_j\star\delta\bar{g}_j[n]. \tag{6.70}$$

Dyadic wavelet coefficients up to the scale 2^J are therefore calculated by cascading the convolutions (6.68-6.70) for $0<j\leqslant J$. To take into account border problems, all convolutions are replaced by circular convolutions, which means that the input image $a_0[n]$ is considered to be periodic along its rows and columns. For an image of N pixels, this algorithm requires $O(N\log_2 N)$ operations. For a square image with a maximum scale $J=\log_2 N^{1/2}$ then one can verify that the larger scale approximation is a constant proportional to the grey level average C:

$$a_J[n_1,n_2]=N^{-1/2}\sum_{n_1,n_2=0}^{N^{1/2}-1}a_0[n_1,n_2]=N^{1/2}\,C\ .$$

The wavelet transform modulus is $Mf(n,2^j)=|d_j^1[n]|^2+|d_j^2[n]|^2$ whereas $Af(n,2^j)$ is the angle of the vector $(d_j^1[n],d_j^2[n])$.

The support Λ of wavelet modulus maxima $(u,2^j)$ is the set of points $Mf(u,2^j)$ is larger than its two neighbors $Mf(u\pm\vec{\varepsilon},2^j)$, where $\vec{\varepsilon}=(\varepsilon_1,\varepsilon_2)$ is the vector whose coordinates ε_1 and ε_2 are either 0 or 1, and whose angle is the closest to $Af(u,2^j)$.

Reconstruction from Maxima The orthogonal projection from wavelet maxima is computed with the dual synthesis algorithm of Section 5.1.3, which inverts the symmetric operator (6.60) with conjugate gradient iterations. This requires computing Ly efficiently for any image $y[n]$. For this purpose, the wavelet coefficients of y are first calculated with the "algorithme à trous," and all coefficients for $(u, 2^j) \notin \Lambda$ are set to 0. The signal $Ly[n]$ is recovered from these non-zero wavelet coefficients. Let $h_j[n] = \bar{h}_j[-n]$ and $g_j[n] = \bar{g}_j[-n]$ be the two filters defined with (6.66) and (6.67). The calculation is initialized for $J = \log_2 N^{1/2}$ by setting $\tilde{a}_J[n] = C\,N^{-1/2}$, where C is the average image intensity. For $\log_2 N > j \geqslant 0$ we compute

$$\tilde{a}_j[n] = \tilde{a}_{j+1} \star h_j h_j[n] + d^1_{j+1} \star g_j \delta[n] + d^2_{j+1}[n] \star \delta g_j[n] \ ,$$

and one can verify that $Ly[n] = \tilde{a}_0[n]$ is recovered with $O(N \log_2 N)$ operations. The reconstructed images in Figure 6.12 are obtained with 10 conjugate gradient iterations implemented with this filter bank algorithm.

6.4 Multifractals

Signals that are singular at almost every point were originally studied as pathological objects of pure mathematical interest. Mandelbrot [39] was the first to recognize that such phenomena are encountered everywhere. Among the many examples [23] let us mention economic records like the Dow Jones industrial average, physiological data including heart records, electromagnetic fluctuations in galactic radiation noise, textures in images of natural terrains, variations of traffic flow…

The singularities of multifractals often vary from point to point, and knowing the distribution of these singularities is important in analyzing their properties. Pointwise measurements of Lipschitz exponents are not possible because of the finite numerical resolution. After discretization, each sample corresponds to a time interval where the signal has an infinite number of singularities that may all be different. The singularity distribution must therefore be estimated from global measurements, which take advantage of multifractal self-similarities. Section 6.4.2 computes the fractal dimension of sets of points having the same Lipschitz regularity, with a global partition function calculated from wavelet transform modulus maxima. Applications to fractal noises such as fractional Brownian motions and to hydrodynamic turbulence are studied in Section 6.4.3.

6.4.1 Fractal Sets and Self-Similar Functions

A set $\mathcal{S} \subset \mathbb{R}^n$ is said to be self-similar if it is the union of disjoint subsets $\mathcal{S}_1, \ldots, \mathcal{S}_k$ that can be obtained from $\mathcal{S}$ with a scaling, translation and rotation. This self-similarity often implies an infinite multiplication of details, which creates irregular structures. The triadic Cantor set and the Von Koch curve are simple examples.

Example 6.5. *The Von Koch curve is a fractal set obtained by recursively dividing each segment of length l in four segments of length $l/3$, as illustrated in Figure 6.15. Each subdivision multiplies the length by $4/3$. The limit of these subdivisions is therefore a curve of infinite length.*

Example 6.6. *The triadic Cantor set is constructed by recursively dividing intervals of size l in two sub-intervals of size $l/3$ and a central hole, illustrated by Figure 6.16. The iteration begins from $[0, 1]$. The Cantor set obtained as a limit of these subdivisions is a dust of points in $[0, 1]$.*

Fractal Dimension The Von Koch curve has infinite length in a finite square of $\mathbb{R}^2$. The usual length measurement is therefore not well adapted to characterize the topological properties of such fractal curves. This motivated Hausdorff in 1919 to introduce a new definition of dimension, based on the size variations of sets when measured at different scales.

The *capacity dimension* is a simplification of the Hausdorff dimension that is easier to compute numerically. Let $\mathcal{S}$ be a bounded set in $\mathbb{R}^n$. We count the minimum number $N(s)$ of balls of radius s needed to cover $\mathcal{S}$. If $\mathcal{S}$ is a set of dimension D with a finite length ($D = 1$), surface ($D = 2$) or volume ($D = 3$) then

$$N(s) \sim s^{-D},$$

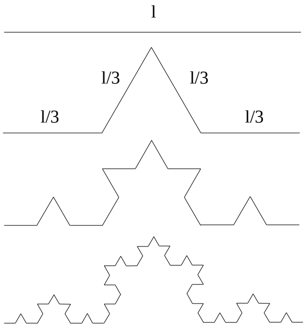


Figure 6.15: Three iterations of the Von Koch subdivision. The Von Koch curve is the fractal obtained as a limit of an infinite number of subdivisions.

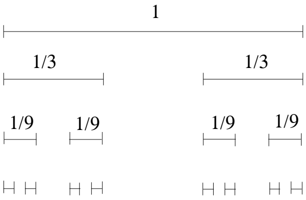


Figure 6.16: Three iterations of the Cantor subdivision of $[0, 1]$. The limit of an infinite number of subdivisions is a closed set in $[0, 1]$.

so

$$D = -\lim_{s \to 0} \frac{\log N(s)}{\log s}. \tag{6.71}$$

The capacity dimension D of $\mathcal{S}$ generalizes this result and is defined by

$$D = -\liminf_{s \to 0} \frac{\log N(s)}{\log s}. \tag{6.72}$$

The measure of $\mathcal{S}$ is then

$$M = \limsup_{s \to 0} N(s)\, s^D .$$

It may be finite or infinite.

The Hausdorff dimension is a refined fractal measure that considers all covers of $\mathcal{S}$ with balls of radius smaller than s. It is most often equal to the capacity dimension, but not always. In the following, the capacity dimension is called *fractal dimension.*

Example 6.7. *The Von Koch curve has infinite length because its fractal dimension is $D > 1$. We need $N(s) = 4^n$ balls of size $s = 3^{-n}$ to cover the whole curve, hence*

$$N(3^{-n}) = (3^{-n})^{-\log 4/\log 3}.$$

One can verify that at any other scale s, the minimum number of balls $N(s)$ to cover this curve satisfies

$$D = -\liminf_{s \to 0} \frac{\log N(s)}{\log s} = \frac{\log 4}{\log 3}.$$

As expected, it has a fractal dimension between 1 *and* 2.

Example 6.8. *The triadic Cantor set is covered by $N(s) = 2^n$ intervals of size $s = 3^{-n}$, so*

$$N(3^{-n}) = (3^{-n})^{-\log 2/\log 3}.$$

One can also verify that

$$D = -\liminf_{s \to 0} \frac{\log N(s)}{\log s} = \frac{\log 2}{\log 3}.$$

Self-Similar Functions Let f be a continuous function with a compact support $\mathcal{S}$. We say that f is *self-similar* if there exist disjoint subsets $\mathcal{S}_1, \ldots, \mathcal{S}_k$ such that the graph of f restricted to each $\mathcal{S}_i$ is an affine transformation of f. This means that there exist a scale $l_i > 1$, a translation r_i, a weight p_i and a constant c_i such that

$$\forall t \in \mathcal{S}_i \quad , \quad f(t) = c_i + p_i\, f\Big(l_i(t - r_i)\Big). \tag{6.73}$$

Outside these subsets, we suppose that f is constant. Generalizations of this definition can also be used [121].

If a function is self similar then its wavelet transform is also self-similar. Let g be an affine transformation of f:

$$g(t) = p\, f\Big(l(t - r)\Big) + c. \tag{6.74}$$

Its wavelet transform is

$$Wg(u, s) = \int_{-\infty}^{+\infty} g(t)\, \frac{1}{\sqrt{s}}\, \psi\left(\frac{t-u}{s}\right) dt.$$

With the change of variable $t' = l(t - r)$, since ψ has a zero average, the affine relation (6.74) implies

$$Wg(u, s) = \frac{p}{\sqrt{l}}\, Wf\Big(l(u - r), sl\Big).$$

Suppose that ψ has a compact support included in $[-K, K]$. The affine invariance (6.73) of f over $\mathcal{S}_i = [a_i, b_i]$ produces an affine invariance for all wavelets whose support is included in $\mathcal{S}_i$. For any $s < (b_i - a_i)/K$ and any $u \in [a_i + Ks, b_i - Ks]$,

$$Wf(u, s) = \frac{p_i}{\sqrt{l_i}}\, Wf\Big(l_i(u - r_i), sl_i\Big).$$

The self-similarity of the wavelet transform implies that the positions and values of its modulus maxima are also self-similar. This can be used to recover unknown affine invariance properties with a voting procedure based on wavelet modulus maxima [283].

Example 6.9. *A Cantor measure is constructed over a Cantor set. Let* $d\mu_0(x) = dx$ *be the uniform Lebesgue measure on* $[0, 1]$*. As in the Cantor set construction, this measure is subdivided into three uniform measures, whose integrals over* $[0, 1/3]$, $[1/3, 2/3]$ *and* $[2/3, 1]$ *are respectively* p_1, 0 *and* p_2*. We impose* $p_1 + p_2 = 1$ *to obtain a total measure* $d\mu_1$ *on* $[0, 1]$ *whose integral is equal to 1. This operation is iteratively repeated by dividing each uniform measure of integral* p *over* $[a, a + l]$ *into three equal parts whose integrals are respectively* $p_1 p$, 0 *and* $p_2 p$ *over* $[a, a + l/3]$, $[a + l/3, a + 2l/3]$ *and* $[a + 2l/3, a + l]$*. This is illustrated by Figure 6.17. After each subdivision, the resulting measure* $d\mu_n$ *has a unit integral. In the limit, we obtain a Cantor measure* $d\mu_\infty$ *of unit integral, whose support is the triadic Cantor set.*

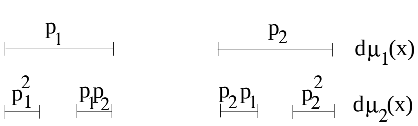
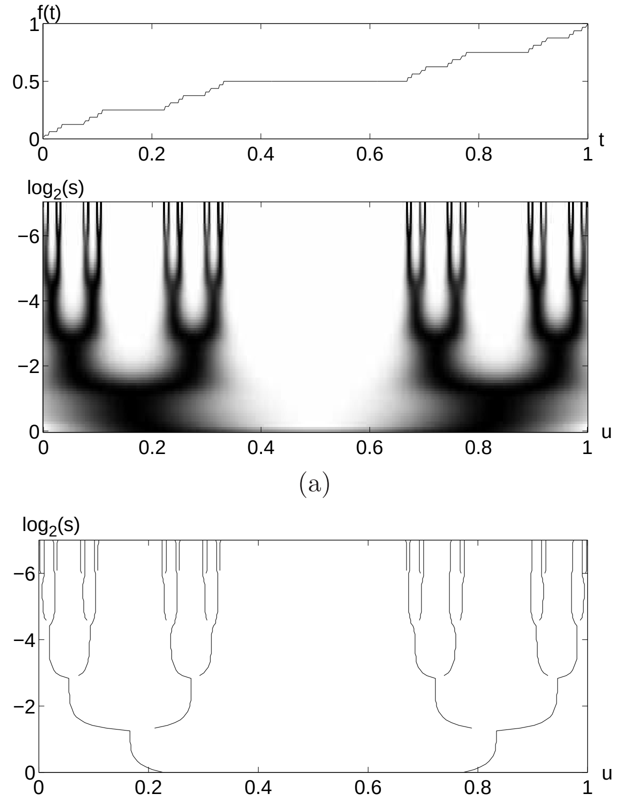


Figure 6.17: Two subdivisions of the uniform measure on $[0, 1]$ with left and right weights p_1 and p_2. The Cantor measure $d\mu_\infty$ is the limit of an infinite number of these subdivisions.

Example 6.10. *A devil's staircase is the integral of a Cantor measure:*

$$f(t) = \int_0^t d\mu_\infty(x). \tag{6.75}$$



(a)

(b)

Figure 6.18: Devil's staircase calculated from a Cantor measure with equal weights $p_1 = p_2 = 0.5$. (a): Wavelet transform $Wf(u, s)$ computed with $\psi = -\theta'$, where θ is Gaussian. (b): Wavelet transform modulus maxima.

It is a continuous function that increases from 0 *to* 1 *on* $[0,1]$. *The recursive construction of the Cantor measure implies that* f *is self-similar:*

$$f(t) = \begin{cases} p_1\, f(3t) & \text{if } t \in [0,1/3] \\ p_1 & \text{if } t \in [1/3,2/3] \\ p_1 + p_2\, f(3t-2) & \text{if } t \in [2/3,0] \end{cases} .$$

Figure 6.18 displays the devil's staircase obtained with $p_1 = p_2 = 0.5$. *The wavelet transform below is calculated with a wavelet that is the first derivative of a Gaussian. The self-similarity of* f *yields a wavelet transform and modulus maxima that are self-similar. The subdivision of each interval in three parts appears through the multiplication by* 2 *of the maxima lines, when the scale is multiplied by* 3. *This Cantor construction is generalized with different interval subdivisions and weight allocations, beginning from the same Lebesgue measure* $d\mu_0$ *on* $[0,1]$ *[5].*

6.4.2 Singularity Spectrum

Finding the distribution of singularities in a multifractal signal f is particularly important for analyzing its properties. The spectrum of singularity measures the global repartition of singularities having different Lipschitz regularity. The pointwise Lipschitz regularity of f is given by Definition 6.1.

Definition 6.2 (Spectrum)**.** *Let* $\mathcal{S}_\alpha$ *be the set of all points* $t \in \mathbb{R}$ *where the pointwise Lipschitz regularity of* f *is equal to* α. *The spectrum of singularity* $D(\alpha)$ *of* f *is the fractal dimension of* $\mathcal{S}_\alpha$. *The support of* $D(\alpha)$ *is the set of* α *such that* $\mathcal{S}_\alpha$ *is not empty.*

This spectrum was originally introduced by Frisch and Parisi [243] to analyze the homogeneity of multifractal measures that model the energy dissipation of turbulent fluids. It was then extended by Arneodo, Bacry and Muzy [348] to multifractal signals. The fractal dimension definition (6.72) shows that if we make a disjoint cover of the support of f with intervals of size s then the number of intervals that intersect $\mathcal{S}_\alpha$ is

$$N_\alpha(s) \sim s^{-D(\alpha)}. \tag{6.76}$$

The singularity spectrum gives the proportion of Lipschitz α singularities that appear at any scale s. A multifractal f is said to be homogeneous if all singularities have the same Lipschitz exponent α_0, which means the support of $D(\alpha)$ is restricted to $\{\alpha_0\}$. Fractional Brownian motions are examples of homogeneous multifractals.

Partition Function One cannot compute the pointwise Lipschitz regularity of a multifractal because its singularities are not isolated, and the finite numerical resolution is not sufficient to discriminate them. It is however possible to measure the singularity spectrum of multifractals from the wavelet transform local maxima, using a global partition function introduced by Arneodo, Bacry and Muzy [348].

Let ψ be a wavelet with n vanishing moments. Theorem 6.5 proves that if f has pointwise Lipschitz regularity $\alpha_0 < n$ at v then the wavelet transform $Wf(u,s)$ has a sequence of modulus maxima that converges towards v at fine scales. The set of maxima at the scale s can thus be interpreted as a covering of the singular support of f with wavelets of scale s. At these maxima locations

$$|Wf(u,s)| \sim s^{\alpha_0+1/2}.$$

Let $\{u_p(s)\}_{p\in\mathbb{Z}}$ be the position of all local maxima of $|Wg(u,s)|$ at a fixed scale s. The partition function $\mathcal{Z}$ measures the sum at a power q of all these wavelet modulus maxima:

$$\mathcal{Z}(q,s) = \sum_p |Wf(u_p,s)|^q. \tag{6.77}$$

At each scale s, any two consecutive maxima u_p and u_{p+1} are supposed to have a distance $|u_{p+1} - u_p| > \varepsilon s$, for some $\varepsilon > 0$. If not, over intervals of size εs, the sum (6.77) includes only the maxima of largest amplitude. This protects the partition function from the multiplication of very close maxima created by fast oscillations.

For each $q \in \mathbb{R}$, the scaling exponent $\tau(q)$ measures the asymptotic decay of $\mathcal{Z}(q,s)$ at fine scales s:

$$\tau(q) = \liminf_{s\to 0} \frac{\log \mathcal{Z}(q,s)}{\log s}.$$

This typically means that

$$\mathcal{Z}(q,s) \sim s^{\tau(q)}.$$

Legendre Transform The following theorem relates $\tau(q)$ to the Legendre transform of $D(\alpha)$ for self-similar signals. This result was established in [87] for a particular class of fractal signals and generalized by Jaffard [286].

Theorem 6.8 (Arneodo, Bacry, Jaffard, Muzy)**.** *Let $\Lambda = [\alpha_{\min}, \alpha_{\max}]$ be the support of $D(\alpha)$. Let ψ be a wavelet with $n > \alpha_{\max}$ vanishing moments. If f is a self-similar signal then*

$$\tau(q) = \min_{\alpha\in\Lambda}\Big(q\,(\alpha + 1/2) - D(\alpha)\Big). \tag{6.78}$$

Proof. The detailed proof is long; we only give an intuitive justification. The sum (6.77) over all maxima positions is replaced by an integral over the Lipschitz parameter. At the scale s, (6.76) indicates that the density of modulus maxima that cover a singularity with Lipschitz exponent α is proportional to $s^{-D(\alpha)}$. At locations where f has Lipschitz regularity α, the wavelet transform decay is approximated by

$$|Wf(u,s)| \sim s^{\alpha+1/2}.$$

It follows that

$$\mathcal{Z}(q,s) \sim \int_\Lambda s^{q(\alpha+1/2)}\, s^{-D(\alpha)}\, d\alpha.$$

When s goes to 0 we derive that $\mathcal{Z}(q,s) \sim s^{\tau(q)}$ for $\tau(q) = \min_{\alpha\in\Lambda}(q(\alpha+1/2) - D(\alpha))$. ■ ■

This theorem proves that the scaling exponent $\tau(q)$ is the Legendre transform of $D(\alpha)$. It is necessary to use a wavelet with enough vanishing moments to measure all Lipschitz exponents up to $\alpha_{\max}$. In numerical calculations $\tau(q)$ is computed by evaluating the sum $\mathcal{Z}(q,s)$. We thus need to invert the Legendre transform (6.78) to recover the spectrum of singularity $D(\alpha)$.

Theorem 6.9.

- *The scaling exponent $\tau(q)$ is a concave and increasing function of q.*
- *The Legendre transform (6.78) is invertible if and only if $D(\alpha)$ is concave, in which case*

$$D(\alpha) = \min_{q\in\mathbb{R}}\Big(q\,(\alpha + 1/2) - \tau(q)\Big). \tag{6.79}$$

- *The spectrum $D(\alpha)$ of self-similar signals is concave.*

Proof. The proof that $D(\alpha)$ is concave for self-similar signals can be found in [286]. We concentrate on the properties of the Legendre transform that are important in numerical calculations. To simplify the proof, let us suppose that $D(q)$ is twice differentiable. The minimum of the Legendre transform (6.78) is reached at a critical point $q(\alpha)$. Computing the derivative of $q(\alpha+1/2) - D(\alpha)$ with respect to α gives

$$q(\alpha) = \frac{dD}{d\alpha}, \tag{6.80}$$

with

$$\tau(q) = q\left(\alpha + \frac{1}{2}\right) - D(\alpha). \tag{6.81}$$

Since it is a minimum, the second derivative of $\tau(q(\alpha))$ with respect to α is positive, from which we derive that

$$\frac{d^2 D(\alpha(q))}{d\alpha^2} \leqslant 0.$$

This proves that $\tau(q)$ depends only on the values where $D(\alpha)$ has a negative second derivative. We can thus recover $D(\alpha)$ from $\tau(q)$ only if it is concave.

The derivative of $\tau(q)$ is

$$\frac{d\tau(q)}{dq} = \alpha + \frac{1}{2} + q\,\frac{d\alpha}{dq} - \frac{d\alpha}{dq}\,\frac{dD(\alpha)}{d\alpha} = \alpha + \frac{1}{2} \geqslant 0. \tag{6.82}$$

It is therefore increasing. Its second derivative is

$$\frac{d^2\tau(q)}{dq^2} = \frac{d\alpha}{dq}.$$

Taking the derivative of (6.80) with respect to q proves that

$$\frac{d\alpha}{dq}\,\frac{d^2D(\alpha)}{d\alpha^2} = 1.$$

Since $\frac{d^2D(\alpha)}{d\alpha^2} \leqslant 0$ we derive that $\frac{d^2\tau(q)}{dq^2} \leqslant 0$. Hence $\tau(q)$ is concave. By using (6.81), (6.82) and the fact that $\tau(q)$ is concave, we verify that

$$D(\alpha) = \min_{q\in\mathbb{R}}\Big(q\,(\alpha+1/2) - \tau(q)\Big).$$

■ ■

The spectrum $D(\alpha)$ of self-similar signals is concave and can therefore be calculated from $\tau(q)$ with the inverse Legendre formula (6.79) This formula is also valid for a much larger class of multifractals. For example, it is verified for statistical self-similar signals such as realizations of fractional Brownian motions. Multifractals having some stochastic self-similarity have a spectrum that can often be calculated as an inverse Legendre transform (6.79). However, let us emphasize that this formula is not exact for any function f because its spectrum of singularity $D(\alpha)$ is not necessarily concave. In general, Jaffard proved [286] that the Legendre transform (6.79) gives only an upper bound of $D(\alpha)$. These singularity spectrum properties are studied in detail in [44].

Figure 6.19 illustrates the properties of a concave spectrum $D(\alpha)$. The Legendre transform (6.78) proves that its maximum is reached at

$$D(\alpha_0) = \max_{\alpha\in\Lambda} D(\alpha) = -\tau(0).$$

It is the fractal dimension of the Lipschitz exponent α_0 most frequently encountered in f. Since all other Lipschitz α singularities appear over sets of lower dimension, if $\alpha_0 < 1$ then $D(\alpha_0)$ is also the fractal dimension of the singular support of f. The spectrum $D(\alpha)$ for $\alpha < \alpha_0$ depends on $\tau(q)$ for $q > 0$, and for $\alpha > \alpha_0$ it depends on $\tau(q)$ for $q < 0$.

Numerical Calculations To compute $D(\alpha)$, we assume that the Legendre transform formula (6.79) is valid. We first calculate $\mathcal{Z}(q,s) = \sum_p |Wf(u_p,s)|^q$, then derive the decay scaling exponent $\tau(q)$, and finally compute $D(\alpha)$ with a Legendre transform. If $q < 0$ then the value of $\mathcal{Z}(q,s)$ depends mostly on the small amplitude maxima $|Wf(u_p,s)|$. Numerical calculations may then become unstable. To avoid introducing spurious modulus maxima created by numerical errors in regions where f is nearly constant, wavelet maxima are chained to produce maxima curve across scales. If $\psi = (-1)^p\,\theta^{(p)}$ where θ is a Gaussian, Theorem 6.6 proves that all maxima lines $u_p(s)$ define curves that propagate up to the limit $s = 0$. All maxima lines that do not propagate up to the finest scale are thus removed in the calculation of $\mathcal{Z}(q,s)$. The calculation of the spectrum $D(\alpha)$ proceeds as follows.

1. *Maxima* Compute $Wf(u,s)$ and the modulus maxima at each scale s. Chain the wavelet maxima across scales.
2. *Partition function* Compute

$$\mathcal{Z}(q,s) = \sum_p |Wf(u_p,s)|^q\ .$$

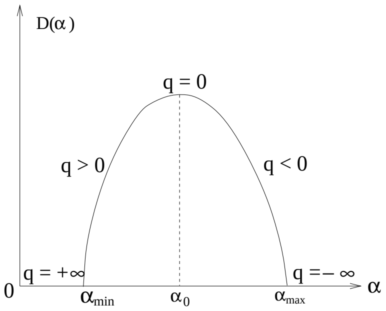


Figure 6.19: Concave spectrum $D(\alpha)$.

3. *Scaling* Compute $\tau(q)$ with a linear regression of $\log_2 \mathcal{Z}(s,q)$ as a function of $\log_2 s$:
$$\log_2 \mathcal{Z}(q,s) \approx \tau(q)\, \log_2 s + C(q)\ .$$
4. *Spectrum* Compute
$$D(\alpha) = \min_{q\in\mathbb{R}} \Big(q(\alpha+1/2) - \tau(q) \Big)\ .$$

Example 6.11. *The spectrum of singularity $D(\alpha)$ of the devil's staircase (6.75) is a concave function that can be calculated analytically [266]. Suppose that $p_1 < p_2$. The support of $D(\alpha)$ is $[\alpha_{\min}, \alpha_{\max}]$ with*
$$\alpha_{\min} = \frac{-\log p_2}{\log 3} \quad \textit{and} \quad \alpha_{\max} = \frac{-\log p_1}{\log 3}.$$

If $p_1 = p_2 = 1/2$ then the support of $D(\alpha)$ is reduced to a point, which means that all the singularities of f have the same Lipschitz $\log 2/\log 3$ regularity. The value $D(\log 2/\log 3)$ is then the fractal dimension of the triadic Cantor set and is thus equal to $\log 2/\log 3$.

Figure 6.20(a) shows a devil's staircase calculated with $p_1 = 0.4$ and $p_2 = 0.6$. Its wavelet transform is computed with $\psi = -\theta'$, where θ is a Gaussian. The decay of $\log_2 \mathcal{Z}(q,s)$ as a function of $\log_2 s$ is shown in Figure 6.20(b) for several values of q. The resulting $\tau(q)$ and $D(\alpha)$ are given by Figures 6.20(c,d). There is no numerical instability for $q < 0$ because there is no modulus maximum whose amplitude is close to zero. This is not the case if the wavelet transform is calculated with a wavelet that has more vanishing moments.

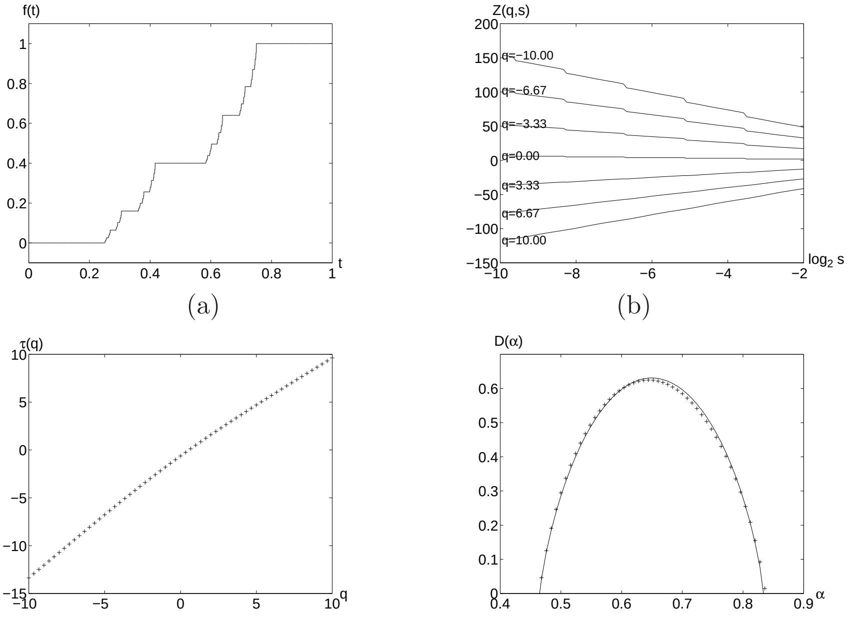
Smooth Perturbations Let f be a multifractal whose spectrum of singularity $D(\alpha)$ is calculated from $\tau(q)$. If a $\mathbf{C}^\infty$ signal g is added to f then the singularities are not modified and the singularity spectrum of $\tilde{f} = f + g$ remains $D(\alpha)$. We study the effect of this smooth perturbation on the spectrum calculation.

The wavelet transform of $\tilde{f}$ is
$$W\tilde{f}(u,s) = Wf(u,s) + Wg(u,s).$$

Let $\tau(q)$ and $\tilde{\tau}(q)$ be the scaling exponent of the partition functions $\mathcal{Z}(q,s)$ and $\tilde{\mathcal{Z}}(q,s)$ calculated from the modulus maxima respectively of $Wf(u,s)$ and $W\tilde{f}(u,s)$. We denote by $D(\alpha)$ and $\tilde{D}(\alpha)$ the Legendre transforms respectively of $\tau(q)$ and $\tilde{\tau}(q)$. The following theorem relates $\tau(q)$ and $\tilde{\tau}(q)$.

Theorem 6.10 (Arneodo, Bacry, Muzy)**.** *Let ψ be a wavelet with exactly n vanishing moments. Suppose that f is a self-similar function.*

- *If g is a polynomial of degree $p < n$ then $\tau(q) = \tilde{\tau}(q)$ for all $q \in \mathbb{R}$.*
- *If $g^{(n)}$ is almost everywhere non-zero then*
$$\tilde{\tau}(q) = \begin{cases} \tau(q) & \textit{if } q > q_c \\ (n+1/2)\,q & \textit{if } q \leqslant q_c \end{cases} \tag{6.83}$$
where q_c is defined by $\tau(q_c) = (n+1/2)q_c$.



(a) (b)

(c) (d)

Figure 6.20: (a): Devil's staircase with $p_1 = 0.4$ and $p_2 = 0.6$. (b): Partition function $\mathcal{Z}(q,s)$ for several values of q. (c): Scaling exponent $\tau(q)$. (d): The theoretical spectrum $D(\alpha)$ is shown with a solid line. The + are the spectrum values calculated numerically with a Legendre transform of $\tau(q)$.

Proof. If g is a polynomial of degree $p < n$ then $Wg(u,s) = 0$. The addition of g does not modify the calculation of the singularity spectrum based on wavelet maxima, so $\tau(q) = \tilde{\tau}(q)$ for all $q \in \mathbb{R}$.

If g is a $\mathbf{C}^\infty$ function that is not a polynomial then its wavelet transform is generally non-zero. We justify (6.83) with an intuitive argument that is not a proof. A rigorous proof can be found in [87]. Since ψ has exactly n vanishing moments, (6.15) proves that

$$|Wg(u,s)| \sim K\, s^{n+1/2}\, g^{(n)}(u).$$

We suppose that $g^{(n)}(u) \neq 0$. For $\tau(q) \leqslant (n+1/2)q$, since $|Wg(u,s)|^q \sim s^{q(n+1/2)}$ has a faster asymptotic decay than $s^{\tau(q)}$ when s goes to zero, one can verify that $\tilde{\mathcal{Z}}(q,s)$ and $\mathcal{Z}(q,s)$ have the same scaling exponent, $\tilde{\tau}(q) = \tau(q)$. If $\tau(q) > (n+1/2)q$, which means that $q \leqslant q_c$, then the decay of $|W\tilde{f}(u,s)|^q$ is controlled by the decay of $|Wg(u,s)|^q$, so $\tilde{\tau}(q) = (n+1/2)q$. ■ ■

This theorem proves that the addition of a non-polynomial smooth function introduces a bias in the calculation of the singularity spectrum. Let α_c be the critical Lipschitz exponent corresponding to q_c:

$$D(\alpha_c) = q_c\,(\alpha_c + 1/2) - \tau(q_c).$$

The Legendre transform of $\tilde{\tau}(q)$ in (6.83) yields

$$\tilde{D}(\alpha) = \begin{cases} D(\alpha) & \text{if } \alpha \leqslant \alpha_c \\ 0 & \text{if } \alpha = n \\ -\infty & \text{if } \alpha > \alpha_c \text{ and } \alpha \neq n \end{cases}. \tag{6.84}$$

This modification is illustrated by Figure 6.21.

The bias introduced by the addition of smooth components can be detected experimentally by modifying the number n of vanishing moments of ψ. Indeed the value of q_c depends on n. If the singularity spectrum varies when changing the number of vanishing moments of the wavelet then it indicates the presence of a bias.

6.4.3 Fractal Noises

Fractional Brownian motions are statistically self-similar Gaussian processes that give interesting models for a wide class of natural phenomena [338]. Despite their non-stationarity, one can define a power spectrum that has a power decay. Realizations of fractional Brownian motions are almost everywhere singular, with the same Lipschitz regularity at all points.

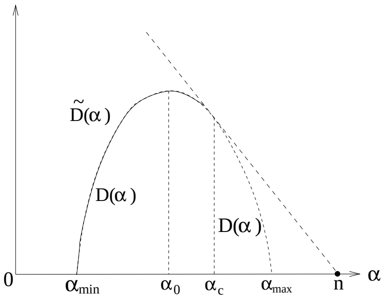


Figure 6.21: If ψ has n vanishing moments, in presence of a $\mathbf{C}^\infty$ perturbation the computed spectrum $\tilde{D}(\alpha)$ is identical to the true spectrum $D(\alpha)$ for $\alpha \leqslant \alpha_c$. Its support is reduced to $\{n\}$ for $\alpha > \alpha_c$.

We often encounter fractal noise processes that are not Gaussian although their power spectrum has a power decay. Realizations of these processes may include singularities of various types. The spectrum of singularity is then important in analyzing their properties. This is illustrated by an application to hydrodynamic turbulence.

Definition 6.3 (Fractional Brownian motion). *A fractional Brownian motion of Hurst exponent* $0 < H < 1$ *is a zero-mean Gaussian process* B_H *such that*

$$B_H(0) = 0,$$

and

$$\mathsf{E}\{|B_H(t) - B_H(t-\Delta)|^2\} = \sigma^2 |\Delta|^{2H}. \tag{6.85}$$

Property (6.85) imposes that the deviation of $|B_H(t) - B_H(t-\Delta)|$ be proportional to $|\Delta|^H$. As a consequence, one can prove that any realization f of B_H is almost everywhere singular with a pointwise Lipschitz regularity $\alpha = H$. The smaller H, the more singular f. Figure 6.22(a) shows the graph of one realization for $H = 0.7$.

Setting $\Delta = t$ in (6.85) yields

$$\mathsf{E}\{|B_H(t)|^2\} = \sigma^2 |t|^{2H}.$$

Developing (6.85) for $\Delta = t - u$ also gives

$$\mathsf{E}\{B_H(t)\, B_H(u)\} = \frac{\sigma^2}{2}\left(|t|^{2H} + |u|^{2H} - |t-u|^{2H}\right). \tag{6.86}$$

The covariance does not depend only on $t - u$, which proves that a fractional Brownian motion is non-stationary.

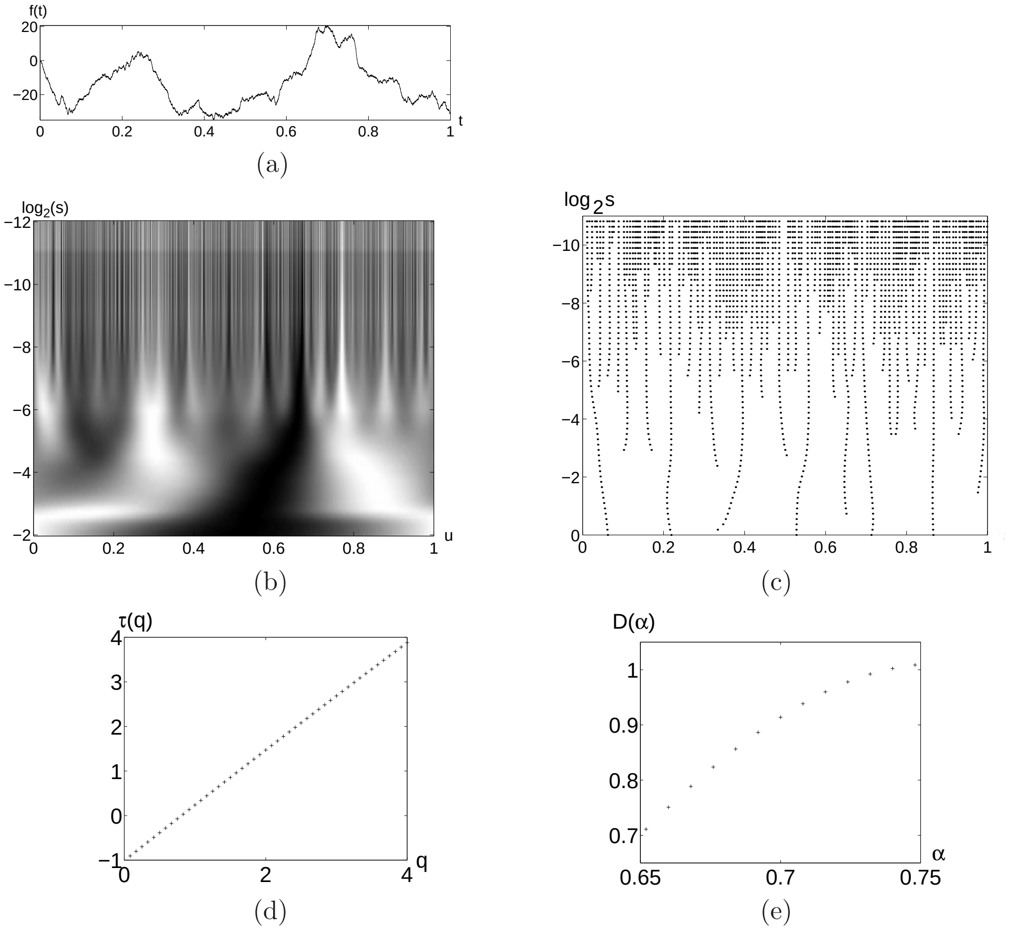
The statistical self-similarity appears when scaling this process. One can derive from (6.86) that for any $s > 0$

$$\mathsf{E}\{B_H(st)\, B_H(su)\} = \mathsf{E}\{s^H\, B_H(t)\, s^H\, B_H(u)\}.$$

Since $B_H(st)$ and $s^H\, B_H(t)$ are two Gaussian processes with same mean and same covariance, they have the same probability distribution

$$B_H(st) \equiv s^H\, B_H(t),$$

where $\equiv$ denotes an equality of finite-dimensional distributions.



(a)

(b)

(c)

(d)

(e)

Figure 6.22: (a): One realization of a fractional Brownian motion for a Hurst exponent $H = 0.7$. (b): Wavelet transform. (c): Modulus maxima of its wavelet transform. (d): Scaling exponent $\tau(q)$. (e): Resulting $D(\alpha)$ over its support.

Power Spectrum Although B_H is not stationary, one can define a generalized power spectrum. This power spectrum is introduced by proving that the increments of a fractional Brownian motion are stationary, and by computing their power spectrum [69].

Theorem 6.11. *Let* $g_\Delta(t) = \delta(t) - \delta(t - \Delta)$*. The increment*

$$I_{H,\Delta}(t) = B_H \star g_\Delta(t) = B_H(t) - B_H(t - \Delta) \tag{6.87}$$

is a stationary process whose power spectrum is

$$\hat{R}_{I_{H,\Delta}}(\omega) = \frac{\sigma_H^2}{|\omega|^{2H+1}}\, |\hat{g}_\Delta(\omega)|^2. \tag{6.88}$$

Proof. The covariance of $I_{H,\Delta}$ is computed with (6.86):

$$\mathsf{E}\{I_{H,\Delta}(t)\, I_{H,\Delta}(t-\tau)\} = \frac{\sigma^2}{2}(|\tau - \Delta|^{2H} + |\tau + \Delta|^{2H} - 2|\tau|^{2H}) = R_{I_{H,\Delta}}(\tau). \tag{6.89}$$

The power spectrum $\hat{R}_{I_{H,\Delta}}(\omega)$ is the Fourier transform of $R_{I_{H,\Delta}}(\tau)$. One can verify that the Fourier transform of the distribution $f(\tau) = |\tau|^{2H}$ is $\hat{f}(\omega) = -\lambda_H\, |\omega|^{-(2H+1)}$, with $\lambda_H > 0$. We thus derive that the Fourier transform of (6.89) can be written

$$\hat{R}_{I_{H,\Delta}}(\omega) = 2\,\sigma^2\,\lambda_H\,|\omega|^{-(2H+1)}\,\sin^2\frac{\Delta\omega}{2},$$

which proves (6.88) for $\sigma_H^2 = \sigma^2\lambda_H/2$. ■ ■

If $X(t)$ is a stationary process then we know that $Y(t) = X \star g(t)$ is also stationary and the power spectrum of both processes is related by

$$\hat{R}_X(\omega) = \frac{\hat{R}_Y(\omega)}{|\hat{g}(\omega)|^2}. \tag{6.90}$$

Although $B_H(t)$ is not stationary, Theorem 6.11 proves that $I_{H,\Delta}(t) = B_H \star g_\Delta(t)$ is stationary. As in (6.90), it is tempting to define a "generalized" power spectrum calculated with (6.88):

$$\hat{R}_{B_H}(\omega) = \frac{\hat{R}_{I_{H,\Delta}}(\omega)}{|\hat{g}_\Delta(\omega)|^2} = \frac{\sigma_H^2}{|\omega|^{2H+1}}. \tag{6.91}$$

The non-stationarity of $B_H(t)$ appears in the energy blow-up at low frequencies. The increments $I_{H,\Delta}(t)$ are stationary because the multiplication by $|\hat{g}_\Delta(\omega)|^2 = O(\omega^2)$ removes the explosion of the low frequency energy. One can generalize this result and verify that if g is an arbitrary stable filter whose transfer function satisfies $|\hat{g}(\omega)| = O(\omega)$, then $Y(t) = B_H \star g(t)$ is a stationary Gaussian process whose power spectrum is

$$\hat{R}_Y(\omega) = \frac{\sigma_H^2}{|\omega|^{2H+1}}\, |\hat{g}(\omega)|^2. \tag{6.92}$$

Wavelet Transform The wavelet transform of a fractional Brownian motion is

$$WB_H(u,s) = B_H \star \bar{\psi}_s(u). \tag{6.93}$$

Since ψ has a least one vanishing moment, necessarily $|\hat{\psi}(\omega)| = O(\omega)$ in the neighborhood of $\omega = 0$. The wavelet filter $g = \bar{\psi}_s$ has a Fourier transform $\hat{g}(\omega) = \sqrt{s}\,\hat{\psi}^*(s\omega) = O(\omega)$ near $\omega = 0$. This proves that for a fixed s the process $Y_s(u) = WB_H(u,s)$ is a Gaussian stationary process [239], whose power spectrum is calculated with (6.92):

$$\hat{R}_{Y_s}(\omega) = s\,|\hat{\psi}(s\omega)|^2\,\frac{\sigma_H^2}{|\omega|^{2H+1}} = s^{2H+2}\,\hat{R}_{Y_1}(s\omega). \tag{6.94}$$

The self-similarity of the power spectrum and the fact that B_H is Gaussian are sufficient to prove that $WB_H(u,s)$ is self-similar across scales:

$$WB_H(u,s) \equiv s^{H+1/2}\, WB_H\left(\frac{u}{s},1\right),$$

where the equivalence means that they have same finite distributions. Interesting characterizations of fractional Brownian motion properties are also obtained by decomposing these processes in wavelet bases [44, 69, 449].

Example 6.12. *Figure 6.22(a) displays one realization of a fractional Brownian with* $H = 0.7$. *The wavelet transform and its modulus maxima are shown in Figures 6.22(b) and 6.22(c). The partition function (6.77) is computed from the wavelet modulus maxima. Figure 6.22(d) gives the scaling exponent* $\tau(q)$, *which is nearly a straight line. Fractional Brownian motions are homogeneous fractals with Lipschitz exponents equal to* H. *In this example, the theoretical spectrum* $D(\alpha)$ *has therefore a support reduced to* $\{0.7\}$ *with* $D(0.7) = 1$. *The estimated spectrum in Figure 6.22(e) is calculated with a Legendre transform of* $\tau(q)$. *Its support is* $[0.65, 0.75]$. *There is an estimation error because the calculations are performed on a signal of finite size.*

Fractal Noises Some physical phenomena produce more general fractal noises $X(t)$, which are not Gaussian processes, but which have stationary increments. As for fractional Brownian motions, one can define a "generalized" power spectrum that has a power decay

$$\hat{R}_X(\omega) = \frac{\sigma_H^2}{|\omega|^{2H+1}}.$$

These processes are transformed into a wide-sense stationary process by a convolution with a stable filter g which removes the lowest frequencies $|\hat{g}(\omega)| = O(\omega)$. One can thus derive that the wavelet transform $Y_s(u) = WX(u,s)$ is a stationary process at any fixed scale s. Its spectrum is the same as the spectrum (6.94) of fractional Brownian motions. If $H < 1$, the asymptotic decay of $\hat{R}_X(\omega)$ indicates that realizations of $X(t)$ are singular functions but it gives no information on the distribution of these singularities. As opposed to fractional Brownian motions, general fractal noises have realizations that may include singularities of various types. Such multifractals are differentiated from realizations of fractional Brownian motions by computing their singularity spectrum $D(\alpha)$. For example, the velocity fields of fully developed turbulent flows have been modeled by fractal noises, but the calculation of the singularity spectrum clearly shows that these flows differ in important ways from fractional Brownian motions.

Hydrodynamic Turbulence Fully developed turbulence appears in incompressible flows at high Reynolds numbers. Understanding the properties of hydrodynamic turbulence is a major problem of modern physics, which remains mostly open despite an intense research effort since the first theory of Kolmogorov in 1941 [300]. The number of degrees of liberty of three-dimensional turbulence is considerable, which produces extremely complex spatio-temporal behavior. No formalism is yet able to build a statistical-physics framework based on the Navier-Stokes equations, that would enable us to understand the global behavior of turbulent flows, at it is done in thermodynamics.

In 1941, Kolmogorov [300] formulated a statistical theory of turbulence. The velocity field is modeled as a process $V(x)$ whose increments have a variance

$$\mathsf{E}\{|V(x+\Delta) - V(x)|^2\} \sim \varepsilon^{2/3}\, \Delta^{2/3}.$$

The constant ε is a rate of dissipation of energy per unit of mass and time, which is supposed to be independent of the location. This indicates that the velocity field is statistically homogeneous with Lipschitz regularity $\alpha = H = 1/3$. The theory predicts that a one-dimensional trace of a three-dimensional velocity field is a fractal noise process with stationary increments, and whose spectrum decays with a power exponent $2H + 1 = 5/3$:

$$\hat{R}_V(\omega) = \frac{\sigma_H^2}{|\omega|^{5/3}}.$$

The success of this theory comes from numerous experimental verifications of this power spectrum decay. However, the theory does not take into account the existence of coherent structures such as vortices. These phenomena contradict the hypothesis of homogeneity, which is at the root of Kolmogorov's 1941 theory.

Kolmogorov [301] modified the homogeneity assumption in 1962, by introducing an energy dissipation rate $\varepsilon(x)$ that varies with the spatial location x. This opens the door to "local stochastic self-similar" multifractal models, first developed by Mandelbrot [337] to explain energy exchanges between fine-scale structures and large-scale structures. The spectrum of singularity $D(\alpha)$ is playing an important role in testing these models [243]. Calculations with wavelet maxima on turbulent velocity fields [5] show that $D(\alpha)$ is maximum at 1/3, as predicted by the Kolmogorov theory. However, $D(\alpha)$ does not have a support reduced to $\{1/3\}$, which verifies that a turbulent velocity field is not a homogeneous process. Models based on the wavelet transform were recently introduced to explain the distribution of vortices in turbulent fluids [11, 233, 234].

6.5 Exercises

6.1. [1]