A panorama on multiscale geometric representations, intertwining spatial, directional and frequency selectivity
Laurent Jacques, Laurent Duval, Caroline Chaux, Gabriel Peyré

To cite this version:

HAL Id: hal-01330604
https://hal-ifp.archives-ouvertes.fr/hal-01330604
Submitted on 11 Jun 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Abstract

The richness of natural images makes the quest for optimal representations in image processing and computer vision challenging. The latter observation has not prevented the design of image representations, which trade off between efficiency and complexity, while achieving accurate rendering of smooth regions as well as reproducing faithful contours and textures. The most recent ones, proposed in the past decade, share an hybrid heritage highlighting the multiscale and oriented nature of edges and patterns in images. This paper presents a panorama of the aforementioned literature on decompositions in multiscale, multi-orientation bases or dictionaries. They typically exhibit redundancy to improve sparsity in the transformed domain and sometimes its invariance with respect to simple geometric deformations (translation, rotation). Oriented multiscale dictionaries extend traditional wavelet processing and may offer rotation invariance. Highly redundant dictionaries require specific algorithms to simplify the search for an efficient (sparse) representation. We also discuss the extension of multiscale geometric decompositions to non-Euclidean domains such as the sphere or arbitrary meshed surfaces. The etymology of panorama suggests an overview, based on a choice of partially overlapping “pictures”. We hope that this paper will contribute to the appreciation and apprehension of a stream of current research directions in image understanding.

Keywords: Review, Multiscale, Geometric representations, Oriented decompositions, Scale-space, Wavelets, Atoms, Sparsity, Redundancy, Bases, Frames, Edges, Textures, Image processing, Haar wavelet, non-Euclidean wavelets.
2 Early Scale-Related Representations

2.1 Frequency, Heat Kernel and Scale-Space Formalism ........................................... 7
2.2 Isotropic Continuous Wavelet Transform ............................................................ 9
2.3 Discrete Scale-Space Representations .................................................................. 11
  2.3.1 Multiresolution Analysis (MRA) .................................................................... 11
  2.3.2 Separable Orthogonal Wavelets .................................................................. 11
  2.3.3 Fast Algorithms for Finite Images .............................................................. 12
  2.3.4 Translation Invariant Wavelets .................................................................. 13

3 Oriented and Geometrical Multiscale Representations ........................................ 13

3.1 Directional Outcrops from Separable Representations ...................................... 13
  3.1.1 Improved Separable Selectivity by Relaxing Constraints ............................. 13
  3.1.2 Pyramid-related wavelets ............................................................................ 13
  3.1.3 Complexifying Discrete Wavelets with Hilbert and Riesz ............................ 14
3.2 Non-Separable Directionality ............................................................................. 16
  3.2.1 Non-separable Decomposition Schemes ..................................................... 16
  3.2.2 Steerable Filters .......................................................................................... 18
  3.2.3 Directional Wavelets and Frames .............................................................. 19
3.3 Directionality in Anisotropic Scaling .................................................................. 21
  3.3.1 Ridgelets .................................................................................................... 21
  3.3.2 Curvelets .................................................................................................... 22
  3.3.3 Contourlets ................................................................................................. 24
  3.3.4 Frames for Oscillating Textures .................................................................. 26

4 Redundancy and Adaptivity ..................................................................................... 27

4.1 Pursuits in Redundant Dictionaries .................................................................... 27
  4.1.1 Matching Pursuits ....................................................................................... 27
  4.1.2 Basis Pursuit ............................................................................................... 27
  4.1.3 Pursuits in Parametric Dictionaries ............................................................ 28
  4.1.4 Processing with Highly Redundant Dictionaries ........................................ 29
  4.1.5 Source Separation ....................................................................................... 30
4.2 Tree-structured Best Basis Representations ..................................................... 31
  4.2.1 Quadtree-based Dictionaries ....................................................................... 31
  4.2.2 Best Basis Selection .................................................................................... 32
  4.2.3 Wavelet and Cosine Packets ....................................................................... 33
  4.2.4 Adaptive Approximation ........................................................................... 33
  4.2.5 Adaptive Tree-structured Processing .......................................................... 33
  4.2.6 Adaptive Segmentations and Triangulations ............................................. 35
4.3 Lifting Representations ...................................................................................... 36
  4.3.1 Lifting Scheme ............................................................................................ 36
  4.3.2 Adaptive Predictions ................................................................................... 37
  4.3.3 Grouplets .................................................................................................... 37
1 Introduction: Vision Aspects, Scope and Notations

1.1 Background on Vision Aspects of Scale

Many natural-world object features are substantive only over a certain spatial extent. In other words, the scale of observation is crucial in object recognition and understanding. For instance, a chair would be easily recognizable in the scale of a few meters. But neither at a centimeter scale which captures the chair’s texture and not its object appearance, or at a hectometer scale, where the chair’s appearance is hardly distinguished from other surrounding objects.

Accordingly, early neurophysiological studies in biologic perception reveal that those objects are generally apprehended differently according to the scale of observation by the sensory receptors and the cortex of mammalians \[1, 2\]. Efficient information extraction is thus required for artificial sensing systems to mimic standard biologic tasks such as object recognition.

Pixel-based representations as linear combinations of “delta” functions suffice for simple data manipulation but are very limited for higher level tasks. Only assuming some sufficient resolution in the data, the lack of prior knowledge in the extent of objects to be analyzed calls for tools able to unveil the appropriate scales and to allow a hierarchical representation of the underlying features \[3, 4, 5\]. Disregarding the peculiar fractal formalism \[6, 7\] where similar phenomena appear at different scales (what is called self-similarity), special attention has been paid to data transformations able to capture object features over a range of scales in a more compact form. Sparsity, amounting to a reduced number of parameters in a suitable domain, is thus used as a heuristic guide to image understanding. Bearing analogies with findings in vision processes \[8\], several sparse decompositions have proven efficient in image compression, with the discrete wavelet transform (DWT) as their most well-known avatar, often intermingled with information theory and technical wizardry, from bit plane arithmetic coding \[9\] to trellis coded quantization. A compact history and a paper collection are given in \[10, 11\], respectively.

Yet, beyond image compression transforms, other decomposition techniques are needed, with more resolving power in complex scene detection, denoising, segmentation or, in a broad sense,
scene understanding. As a matter of fact, standard separable wavelet transforms appropriately
detect point-like (0-D) singularities and address mild noise levels. Still they generally lack perfor-
mane in dealing with higher dimensional features combining both regularity and singularity such
as edges, contours or regular textures, that may also be anisotropic. Amongst their limitations are
shift sensitivity, limited orientation selectivity, rigid and uneven atom shapes (e.g., fractal-looking
asymmetric Daubechies wavelets), crude frequency direction selection. Major challenges reside in
a proper definition of the underlying regularity (with respect to each feature) and corresponding
singularities. These challenges are amplified by additional degradations from which acquired data
may suffer such as blur, jitter and noise. Descriptive mathematical models of images combining
cartoon and textures become increasingly popular [12, 13] and progressively yield tractable algo-
rithms. We note that there exists a continuum of real-world images between cartoon and textures,
ranging from cartoon-ish Yogi bear in Fig. 1(a) to “textural” fingerprints in Fig. 1(b). In between
these two extreme image types, there exists many possible variations in image object complexity.
Moreover, both contours and (even regular) textures are known to be ill-defined. They are
indeed viewer- and scale-dependent concepts in discrete images or volumes. Consider an image
resulting from a combination of piecewise smooth components, contours, geometrical textures and
noise. Their discrimination is required for high level image processing tasks. Each of these four
components could be detected, described and modeled by different formalisms: smooth curves or
polynomials, oriented regularized derivatives, discrete geometry, parametric curve detectors (such
as the Hough transform), mathematical morphology, local frequency estimators, optical flow ap-
proaches, smoothed random models, etc. They have progressively influenced the hybridization of
standard multiscale transforms towards more geometric and sparser representations of such compo-
nents, with improved localization, orientation sensitivity, frequency selectivity or noise robustness.

1.2 Scope of the Paper

Geometry driven “⋆-let” transforms [14] have been popular in the past decade, with a seminal
ancestor in [15]. Early [16], a debate opened on the relative strength of Eulerian (non-adaptive)
versus Lagrangian (adaptive) representation, now pursued with the growing interest in dictionary
learning [17].

As of today, the authors believe that the discussion is not fully settled in the various different
uses of sparsity in images. Neither has the trade-off between redundancy and sparsity. A number
of early papers on geometric multiscale methods appear in [18]. Comparisons are drawn in [19, 20], while [21, 22, 23, 22, 24] focus on ridgelets, curvelets and wedgelets, as representative of fixed and adaptive decompositions. The present paper aims at providing a broader panorama of the recent developments in multiscale decompositions targeted to efficient representation of geometric features in images: smooth content (multiscale or hierarchical), edges and contours (locally spatial) and textures (locally spectral). We emphasize the main characteristics and differences pertaining to spatial, directional and frequency selectivity of the selected methods. The paper therefore cites a dense set of references, ranging from continuous to discrete representations, from (nearly) orthogonal to (fully) redundant. As a guiding thread to this panorama, we illustrate some of the reviewed geometric multiscale decompositions on a memorial plaque\(^1\) in Szeged University, Hungary, depicted in Fig. 2. It features simple objects (embedded rectangles and a disk), a few differently oriented features and regular textures at different scales. Since some of the illustrations have been slightly enhanced to improve the clarity of details, they are available in original resolution online [26]. This picture finally honors Alfréd Haar’s originative paper [27] *Zur Theory der orthogalen Funktionen Systeme (On the Theory of Orthogonal Function Systems)* and his eponymous wavelet. He also founded *Acta Scientiarum Mathematicarum* together with Frigyes Riesz, whose works percolated wavelet theory [28].

Figure 2: Szeged University Memorial plaque in honor of A. Haar and F. Riesz: *A szegedi matematikai iskola világhírű megalapítói* (The world-wide famous founders of the mathematical school in Szeged).

The paper is organized as follows: the remaining of Section 1 is devoted to context and notations for image representations. Then, as a preliminary to geometric tools, a quick survey of early multiscale decompositions is presented in Section 2. More recent transforms, termed directional or geometrical, circumventing aforementioned drawbacks, are discussed in Section 3. Owing to

\(^1\) Courtesy of Professor Károly Szatmáry, [http://astro.u-szeged.hu/szatmary.html](http://astro.u-szeged.hu/szatmary.html), who performed scalograms analysis of variable stars as early as in 1992 [25].
the additional degrees of freedom provided by these representations, a discussion is carried out in Section 4 on redundancy and adaptivity. The extension of frequency, scale and directionality to non-Euclidean spaces or grids such as the sphere, are presented in Section 5. Finally, concluding remarks are given in Section 6.

1.3 Mathematical Framework

1.3.1 Notations and Conventions

This paper describes numerous mathematical methods designed for different spaces and geometries. We have tried therefore to adopt coherent representations for the many mathematical notions that coexist in this text. For instance, functions and vectors in high dimensional spaces are generally referring to some signal of interest (e.g., 1-D signals or images). They must therefore share the same notations and we thus decided to write them as simple lowercase Roman or Greek letters. However, coordinate systems, vectors in 2 or 3 dimensions and multi-indices are denoted in bold symbols.

The (Hilbert) space $L^2(\mathcal{X})$ is the space of square integrable functions on the space $\mathcal{X}$, i.e., given the (Lebesgue) integration measure $d\rho$ on that space, $L^2(\mathcal{X}) = L^2(\mathcal{X}, d\rho) = \left\{ f : \mathcal{X} \to \mathbb{C} : \| f \|^2 := \int_{\mathcal{X}} |f(u)|^2 d\rho(u) < \infty \right\}$. In $L^2(\mathcal{X})$, the inner product between two functions $g, h \in L^2(\mathcal{X})$ is denoted by $\langle g, h \rangle = \int_{\mathcal{X}} g^*(u) h(u) d\rho(u)$ with $^*$ the complex conjugation. By extension, for $p \geq 1$, we also use the (Banach) spaces $L^p(\mathcal{X}) = L^p(\mathcal{X}, d\rho) = \left\{ f : \mathcal{X} \to \mathbb{C} : \| f \|^p_p := \int_{\mathcal{X}} |f(u)|^p d\rho(u) < \infty \right\}$, with $\| \cdot \|_p = \| \cdot \|_2$.

We also use some discrete spaces as the common $\ell^p_N = \left( \mathbb{C}^N, \| \cdot \|_p \right)$ with $\| v \|^p := \sum_i |v_i|^p$ for $p \geq 1$ and $v \in \mathbb{C}^N$, with again the shorthand $\| \cdot \| = \| \cdot \|_2$. In $\ell^2_N$, the inner product between $u, v \in \ell^2_N$ is written $\langle u, v \rangle = u \cdot v = \sum u_i^* v_i$. Whether the overused notations $\langle \cdot, \cdot \rangle$ or $\| \cdot \|_p$ are applied to continuous or discrete mathematical objects will remain clear from the context. The spaces $\ell^p$ are the generalization of the previous finite spaces to infinite sequences, i.e., $\ell^p = \{ v = (v_i)_{i \in \mathbb{N}} : \| v \|^p_p = \sum_{i \geq 0} |v_i| < \infty \}$.

For functions $f \in L^2(\mathcal{X})$ or discrete sequences $v \in \ell^2_N$, $\hat{f}$ and $\hat{v}$ denote the Fourier transform of $f$ or $v$ respectively. For instance, for $\mathcal{X} = \mathbb{R}$ and $f \in L^2(\mathbb{R})$, $\hat{f}(\omega) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(t) e^{-i\omega t} dt$ and $f(t) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \hat{f}(\omega) e^{i\omega t} d\omega$ are the Forward and Inverse Fourier transform respectively. For $f, x \in L^2(\mathbb{R}^2)$ and $\omega \in \mathbb{R}^2$, the same transforms are $\hat{f}(\omega) = \frac{1}{2\pi} \int_{\mathbb{R}^2} f(x) e^{-i\omega \cdot x} d^2 x$ and $f(x) = \frac{1}{2\pi} \int_{\mathbb{R}^2} \hat{f}(\omega) e^{i\omega \cdot x} d^2 \omega$. For $v \in \ell^2_N$, the same transforms are $\hat{v}_j = \frac{1}{\sqrt{N}} \sum_j v_j \exp(-2\pi i j k/N)$ and $v_j = \frac{1}{\sqrt{N}} \sum_k \hat{v}_k \exp(2\pi i j k/N)$. In matrix algebra notations, this can be rewritten as $v = \mathcal{F} \hat{v}$ and $\hat{v} = \mathcal{F}^* v$, where the Fourier matrix $\mathcal{F} \in \mathbb{C}^{N \times N}$ is given by $\mathcal{F}_{jk} = \frac{1}{\sqrt{N}} \exp(2\pi i j k/N)$, and $\mathcal{F}^*$ is its complex adjoint. The convolution by time-invariant filter $h$ operates as $(f * h)(t) = \int_{-\infty}^{\infty} f(u) h^*(t-u) du$ and $(v * h)_n = \sum_{n'} h^*_n v_{n-n'}$ in continuous and discrete sample domain respectively. The ubiquitous Gaussian kernel with scale parameter $\sigma > 0$ is denoted by $G_\sigma(x) = \exp(-\frac{1}{2\sigma^2} \| x \|^2)$, with $G(x) = G_1(x)$.

\footnote{With periodization for finite length vectors.}
1.3.2 Image Representations in Bases and Frames

Stability and Frames This paper describes processing methods that make use of a decomposition of the image \( f \in L^2([0, 1]^2) \) into a family of atoms \( \mathcal{B} = \{\psi_m\}_{m} \). Each atom \( \psi_m \in L^2([0, 1]^2) \) is parameterized by a multi-index \( m \) (that might take into account its frequency, position, scale and orientation). Numerical processing is performed on discretized images which are vectors \( f \in \mathbb{R}^N \), where \( N \) stands for the number of pixels. The atoms of \( \mathcal{B} \) are also discretized and the continuous inner products are replaced by the standard discrete inner product in \( \mathbb{R}^N \).

To guarantee a stable reconstruction from the coefficients \( \{\langle \psi_m, f \rangle\}_{m} \), the family \( \mathcal{B} \) is assumed to be a frame \([29, 30, 28, 31, 32] \) of \( L^2([0, 1]^2) \) or \( \mathbb{R}^N \), which means that there exist two constants \( 0 < \mu_1 \leq \mu_2 < \infty \) such that for all \( f \)

\[
\mu_1 \|f\|^2 \leq \sum_{m} |\langle \psi_m, f \rangle|^2 \leq \mu_2 \|f\|^2. \tag{1}
\]

Atoms are allowed to be linearly dependent, thus corresponding to a redundant representation. Redundancy enables atoms to meet certain additional constraints, for instance smoothness, symmetry and invariance to translation or rotation.

Thresholding for Approximation and Processing Using a dual frame \( \{\tilde{\psi}_m\}_{m} \) \([28] \), an image is recovered from the set of coefficients as \( f = \sum_{m} \langle \psi_m, f \rangle \tilde{\psi}_m \). The computation of the set of coefficients \( \{\langle \psi_m, f \rangle\}_{m} \) for a discrete image \( f \in \mathbb{R}^N \) is usually performed using a fast algorithm, that also enables a fast reconstruction of an image from coefficients.

The basic processing operation, used in denoising and compression applications, is the thresholding

\[
f_M = H_T(f, \mathcal{B}) = \sum_{m : |\langle \psi_m, f \rangle| > T} \langle \psi_m, f \rangle \tilde{\psi}_m \tag{2}
\]

where \( M = \# \{m : |\langle \psi_m, f \rangle| > T\} \) counts the number of non-zero coefficients in \([2]\).

When \( \mu_1 = \mu_2 \), the frame is said to be tight (Parseval tight frame). If furthermore \( \mu_1 = \mu_2 = 1 \), then one can choose \( \tilde{\psi}_m = \psi_m \), and \( \mathcal{B} = \{\psi_m\}_{m} \) is then an orthonormal basis if \( \|\psi_m\| = 1 \) for all \( m \). In this last case, \( \mathcal{B} \) performs the least energy reconstruction of \( f_M \) in \([2]\), or equivalently, \( f_M \) is the best \( M \)-terms approximation of \( f \). The decay of the approximation error \( \|f - f_M\| \) is related to both the average risk of a denoiser, and the distortion rate decay of a coder, see for instance \([33] \). This motivates the search for bases or frames \( \mathcal{B} \) which can efficiently approximate large classes of (natural) images. When the frame is redundant, more complicated decomposition methods improve the sparsity of the representation (see Sec. 4.1).

2 Early Scale-Related Representations

2.1 Frequency, Heat Kernel and Scale-Space Formalism

At the heart of modern signal processing techniques is the concept of signal representation, i.e., the selection of an efficient “point of view” in the study of signal properties that is not restricted to straightforward spatial descriptions.

The most obvious alternative signal representation is its frequency reading, i.e., the one provided by the Fourier transform of the signal explained in Sec. 1.3.1 \([34, 35] \). However, this representation
is not sufficiently “local”. It is indeed rather difficult to detect what spatial part of an image contributes to high peaks in the Fourier spectrum. Fig. 3 represents the amplitude spectrum of the luminance component from Fig. 2. It exhibits a mixture of prominent vertical and horizontal directions with tiny fuzzy diagonal ones.

![Figure 3: Magnitude of the 2-D Fourier transform of the Haar-Riesz Memorial plaque in Fig. 2](image)

An approach for obtaining a better localization is to introduce a notion of “scale” in the image observation. This has been performed very early in image and signal processing by either windowing or introducing scales in the Fourier transform or observing a well-known diffusion process like the heat dynamics governed by the famous Heat equation. The idea relies on considering the image as an initial configuration of heat that is diffused with a time variable \( \tau > 0 \) and in interpreting this time parameter as the “scale”. Indeed, in this dynamic diffusion, small image structures will be smoothed early at small evolution time while larger ones persist for a larger duration. Interestingly, this diffusion is equivalently described by a filtering process: the convolution of the image by a Gaussian function \( G_\sigma \) of width \( \sigma = \sqrt{2\tau} \). This image unfolding into a scale-space domain has led to many new image processing techniques such as edge, ridge and feature detection. This is illustrated in Fig. 4, where the original image is convolved with three different Gaussian kernels in dyadic progression. Large objects such as the white rectangular plaques persist across all scales, while brick and grid textures vanish in Fig. 4(c). The overall redundancy of the Gaussian pyramid is given by the number of smoothing kernels. Taking advantage of the resolution loss, the redundancy factor may be reduced by sub-sampling, leading to the “Gaussian pyramid” construction.

The scale content of the image can be decomposed further by computing, for instance, differences between two filterings performed at two different scales. This led to the famous Littlewood-Paley decomposition, or to the (invertible) Laplacian pyramid conveniently combining multiple sub-sampled low-pass filterings of images, creating a pyramidal scale hierarchy. Interestingly, the resulting decomposition represented in Fig. 5 is a complete image representation that can advantageously be processed before reconstructing a new “restored” image (e.g., in image denoising). Additionally, image singularities are enhanced at fine scales, with low activity regions associated with coefficients.

---

\(^3^\)The original image has been multiplied by a 2-D raised-cosine type apodizing window in order to reduce border discontinuity effects.
Figure 4: Gaussian scale-space decomposition of the Haar-Riesz Memorial plaque at three different scales.

being close to zero. Fast implementations of deformable (steerable or scalable) decompositions \[44\] are available for instance with recursive filters \[45\] or efficient multirate filter banks \[46, 47, 48, 49\].

Figure 5: Laplacian pyramid decomposition of the Haar-Riesz Memorial plaque.

Remarkably, the notion of Scale-Space has been defined and “axiomatized” more than 50 years ago by the Japanese mathematicians Iijima and Otsu, as presented in \[50\]. As we will realize throughout this paper, this scale-space representation (refer to \[51\] for a recent overview and axiomatic generalization) was the starting point of many new ways to represent images.

2.2 Isotropic Continuous Wavelet Transform

The continuous wavelet transform somehow generalizes the previous scale-space formalism driven by the Gaussian kernel to any “function” with enough regularity. The continuous wavelet transform was initially developed for the transformation of 1-D signals \[52\] and further extended in 2-D first with isotropic wavelets. The case of non-isotropic (directional) wavelets was defined later \[53\] (see Sec. 3.2.3).
In one dimension, a wavelet $\psi$ is an integrable and well-localized function of $L^2(\mathbb{R})$, generally described as locally oscillating, i.e., $\int_{\mathbb{R}} \psi(t) dt = 0$. It may be dilated or contracted by a scale factor $a > 0$ and translated to a position $b \in \mathbb{R}$: $\psi(b,a)(t) = \frac{1}{\sqrt{a}} \psi(\frac{t-b}{a})$.

The continuous wavelet transform of a signal $f \in L^2(\mathbb{R})$ probes its content with a “lens” $\psi_{(b,a)}$ of zoom factor $a$ and location $b$. Mathematically,

$$W_f(b,a) = \int_{\mathbb{R}} f(t) \frac{1}{\sqrt{a}} \psi^{*}(\frac{t-b}{a}) dt = \langle \psi_{(b,a)}, f \rangle. \quad (3)$$

Interestingly, provided that $\psi$ is admissible, i.e., when the two constants $c_{\psi}^{+} = 2\pi \int_{0}^{+\infty} \left| \hat{\psi}(\pm \omega) \right|^2 d\omega < \infty$ are finite and equal\(^4\), that is, $c_{\psi}^{+} = c_{\psi}^{-} = c_{\psi} < \infty$, the signal $f$ may be recovered from the coefficients $W_f(b,a)$:

$$f(t) = \frac{1}{c_{\psi}} \int_{0}^{+\infty} \int_{\mathbb{R}} W_f(b,a) \psi_{(b,a)}(t) db \frac{da}{a^2}. \quad (4)$$

This integral representation involves wavelets at every location and all positive dilations, i.e., $f$ is decomposed on the continuous set of functions $\{\psi_{(b,a)} : a \in \mathbb{R}_+, b \in \mathbb{R}\}$. Many different kinds of (admissible) wavelets may be selected. We may cite the derivatives of Gaussian (DoG), the Morlet and the Cauchy wavelets, etc. Their selection is driven by the features to be elucidated in the data, e.g., frequency content with the Morlet wavelet or singularities with DoGs (Fig. 6(a)) as illustrated\(^5\) in Fig. 6(b).

In two dimensions, the most natural extension of the 1-D-CWT is obtained by considering isotropic wavelets, i.e., wavelets $\psi \in L^2(\mathbb{R}^2)$ such that $\psi(x) = \psi_{\text{rad}}(||x||)$, with $x = (x_1, x_2)$, for some radial function $\psi_{\text{rad}} : \mathbb{R}_+ \rightarrow \mathbb{R}$. In that case, the wavelet family is generated by 2-D dilations and translations, i.e., we work with $\psi_{(b,a)}(x) = \frac{1}{a} \psi(\frac{x-b}{a})$ that are copies of $\psi$ translated to $b = (b_1, b_2) \in \mathbb{R}^2$ and dilated by $a > 0$. The 2-D CWT of the image $f \in L^2(\mathbb{R}^2)$ is then simply

\(^4\)When $\psi$ is sufficiently regular, this condition reduces to a zero-average requirement, that is, $\int_{\mathbb{R}} \psi(t) dt = 0$

\(^5\)The YAWTb toolbox has been used, see http://rhea.tele.ucl.ac.be/yawtb/
2.3.1 Multiresolution Analysis (MRA)

the original signal can be perfectly reconstructed from its coefficients. Together with a suitable choice of the wavelet function, they lead to stable representations where

In this section, we detail some parameter samplings, such as dyadic or translation invariant grids.

2.3 Discrete Scale-Space Representations

wavelet transforms is covered in more details in [52, 53, 54, 55, 33].

2.3.2 Separable Orthogonal Wavelets

The isotropic CWT is a useful analysis tool for edge detection in images. For instance, by taking the (admissible) Marr Wavelet $\psi(x) = \Delta[\exp(-\pi |x|^2)]$ (with $\Delta$ the 2-D Laplacian) also called Laplacian of Gaussian or Mexican Hat (see Fig. 6(a)), the CWT of an image $f$ acts as a multiscale edge detector. The topic of 1-D and 2-D continuous wavelet transforms is covered in more details in [52, 53, 54, 55, 33].

2.3.3 Multiresolution Analysis (MRA)

In the context of a dyadic sampling where $a = 2^j$ and $b = n2^j$ for $j, n \in \mathbb{Z}$, the canonical way to design a suitable wavelet function $\psi$ in 1-D makes use of a multi-resolution analysis (MRA).

It is defined as a nested sequence of closed vector subspaces $(V_j)_{j \in \mathbb{Z}}$ in $L^2(\mathbb{R})$ verifying standard properties [56]. Multiresolution analysis of a signal $f$ consists of successively projecting the signal onto subspaces $V_j$ in a series of increasingly coarser approximations as $j$ grows. The difference between two successive approximations represents detail information. It amounts to the information loss between two consecutive scales, which lies in the subspace $W_j$ in $V_{j-1}$ such that:

$$V_{j-1} = V_j \oplus W_j.$$ 

Then, with additional stability properties, there exists a wavelet $\psi \in L^2(\mathbb{R})$ such that $B = \{2^{-j/2}\psi(2^{-j}x - n) : n \in \mathbb{N}\}$ is an orthonormal basis for $W_j$.

2.3.2 Separable Orthogonal Wavelets

A 2-D orthogonal wavelet basis $B = \{\psi_m\}_m$ of $L^2(\mathbb{R}^2)$ for $m = (j, n, k)$ is parameterized by a scale [6] $2^j$ ($j \in \mathbb{Z}$), a translation $2^j n = 2^j(n_1, n_2)$ ($n \in \mathbb{Z}^2$) and one of three possible orientations $k \in \{V, H, D\}$, loosely denoting the vertical, horizontal and (bi) diagonal directions, the latter being poorly representative. Wavelet atoms are defined by dyadic scalings and translations $\psi_m(x) = 2^{-j/2}\psi(2^{-j}x - n)$ of three tensor-product 2-D wavelets

$$\psi^V(x) = \psi(x_1)\phi(x_2), \quad \psi^H(x) = \phi(x_1)\psi(x_2), \quad \text{and} \quad \psi^D(x) = \psi(x_1)\psi(x_2),$$

where $\phi$ and $\psi$ are respectively 1-D orthogonal scaling and wavelet functions, see [54, 57, 53]. When the scale interval is limited to $j < J$ for some $J \in \mathbb{Z}$, the basis $B$ is completed by the functional set $A = \{\phi_{(j,n)}\}_n$, with the 2-D separable scaling function $\phi(x) = \phi(x_1)\phi(x_2)$. This set gathers all

\[\text{\textsuperscript{6}}\text{Here and throughout the rest of the paper, we use the convention that scale increases with } j, \text{ as in } s = 2^j. \text{ The converse convention is also often used in the literature.}\]
the coarse scale wavelet atoms with $j \geq J$. The standard cascade image is depicted in Fig. 7. It is now critically sampled, i.e., free from redundancy (compare Fig. 5 and 6(b)). The approximation coefficients in $A$, a coarse image approximation at scale $J$, are represented in the bottom-left square of Fig. 7. The other squares in this picture, associated to the “bands” $\{V, H, D\}$ for $j < J$, exhibit some sparsity (few important coefficients), and horizontal and vertical edges are relatively well captured.

Figure 7: Dyadic wavelet decomposition of the Haar-Riesz Memorial plaque.

A non-linear approximation $f_M = H_T(f, B)$ in an orthogonal separable wavelet basis is efficient for smooth images or images with point-wise singularities. The approximation of a piecewise smooth image with edges of finite length decays like $\|f - f_M\|^2 = O(M^{-1})$. This result extends to functions with bounded variations [58], and is asymptotically optimal. This decay is nevertheless not improved when the edges are smooth curves, because of the fixed ratio between the horizontal and the vertical sizes of the orthogonal wavelet support.

2.3.3 Fast Algorithms for Finite Images

A finite discretized image $f \in \mathbb{C}^{N_1 \times N_2}$ of $N = N_1 N_2$ pixels fits into the MRA framework by assuming that the pixel values of $f_n$ on $n = (n_1, n_2)$ are the coefficients $\langle \phi_{(J,n)}, \tilde{f}\rangle$ of some continuous function $\tilde{f} \in L^2(\mathbb{R}^2)$ at a fixed resolution $V_J$, where $2^{2J} = N$.

The coefficients $\langle \psi_{j,n}, f\rangle$ of $\tilde{f}$ for $j \geq J$ are computed from the discrete image $f$ alone. This computation is performed using a cascade of filters interleaved with downsampling operators [56]. For compactly supported wavelets, this requires $O(N)$ operations. Symmetric bi-orthogonal wavelet bases with compact support ease the implementation of non-periodic boundary conditions [59]. For infinite impulse response (IIR) wavelet filters, computations in the Fourier domain require $O(N \log(N))$ operations [60], while recursive implementations [61] allow signal-adaptive implementation.
While separable wavelets are not optimal for approximating generic edges, they lie at the heart of early state-of-the-art methods for compression and denoising. The JPEG 2000 coding standard \cite{62} performs an embedded quantization of wavelet coefficients, and uses an adaptive entropic coding scheme that takes into account the local dependencies across wavelet coefficients. The sub-optimality of wavelets for the sparse representation of edges can be alleviated using block thresholding of groups of wavelet coefficients \cite{63}, that gives improvements over scalar thresholding. Advanced statistical modeling of wavelet coefficients leads to denoising methods close to the state-of-the-art, see for instance \cite{64,65,66}.

2.3.4 Translation Invariant Wavelets

Given a discrete frame $\mathcal{B} = \{\psi_m\}_m$ of $\mathbb{C}^N$, $\mathcal{B}$ is translation invariant if $\psi(\cdot - \tau) \in \mathcal{B}$ for any $\psi \in \mathcal{B}$ and any integer translation $\tau$. This property tends to reduce artifacts in image restoration problems like denoising, since, for such invariant frame, the thresholding operator $H_T(f, \mathcal{B})$ becomes itself translation invariant. Discrete orthogonal wavelet bases described in the previous sections are not translation invariant and many authors have worked on recovering this useful capability.

For instance, cycle spinning, proposed by Coifman and Donoho in \cite{67}, reduces wavelet artifacts by averaging the denoising result of all possible translates of the image, thus resulting in a translation invariant processing. For an orthogonal basis $\mathcal{B} = \{\psi_m\}_m$, this is equivalent to considering a tight frame which is the union of all translated bases $\{\psi_m(\cdot - \tau)\}_m,\tau$. For a generic basis, this frame has up to $N^2$ atoms. For a wavelet basis, the frame has $O(N \log(N))$ atoms, and the coefficients are computed with the fast “à trous” algorithm in $O(N \log(N))$ \cite{60,68}. The translation invariant paradigm additionally draws a connection between the scale-space formalism (Sec. 2.1) \cite{69} and thresholding (Sec. 1.3.2). Several 2-D design described in the next sections attempt to (approximately) address invariance (translation/rotation) without sacrificing computational efficiency.

3 Oriented and Geometrical Multiscale Representations

The variety of oriented and geometric multiscale representations proposed over the last few years requires broad grouping, arranged as follows: Sec. 3.1 presents directional methods closely related to 1-D decompositions. In Sec. 3.2 the directionality is addressed with diverse non-separable schemes. Finally, in Sec. 3.3 directionality is attained by an anisotropic scaling of the atoms that yields various efficient edge and curve representations.

3.1 Directional Outcrops from Separable Representations

3.1.1 Improved Separable Selectivity by Relaxing Constraints

As discussed in Sec. 2.3.1, discrete orthogonal wavelets may be viewed as a peculiar instance of orthogonal filter banks \cite{70}. A well-known limitation in 1-D is that orthogonality (hence non-redundant), realness, symmetry and finite support properties cannot coexist with pairs of low- and high-pass filters, except for the Haar wavelet.

We decide to briefly mention here some of the early steps taken to tackle this limitation. These have also been employed in more genuine non-separable transforms, as seen later, typically relaxing one of the aforementioned properties, such as using infinite-support filters \cite{71}, semi- or biorthogonal decompositions \cite{59} or complex filter banks \cite{72}.
For instance, instead of a two-band filter bank, \(M\)-band wavelets with \(M > 2\) provide alternatives where symmetry, orthogonality and realness are compatible with finitely supported atoms. In this setting, the approximation and the \(M\)-band detail spaces are \(V_j\) and \((W^m_j)_{m \in \mathbb{N}^*_M}\) related through \(V_{j-1} = V_j \oplus \bigoplus_{m=1}^{M-1} W^m_j\) for a resolution level \(j\). This versatile design provides filters that suffer less aliasing artifacts with increased regularity. Their finer subband decomposition is also beneficial for detecting orientations in a more subtle fashion than with the \(\{V, H, D\}\) quadrants obtained with standard wavelets (Sec. 2.3.2). Yet, more general \(M\)-adic MRAs are possible, for instance with a rational \(M = p/q, M > 1\) [74, 75, 76, 77, 78]. Note that for specific purposes such as compression, \(M\)-band filter banks with \(M = 2^J, J \in \mathbb{N}\) may be treated like a \(J\)-level dyadic tree and combined in a hierarchical transform [79, 80]. Satisfying the MRA axioms is not necessary in practice in order to yield high performance results. This is suggested by recent image and video coders focusing on “simpler” transforms, closer to ancient Walsh-Hadamard transforms than to more involved wavelets [81].

Alternatively, the 1-D decomposition on rows and columns of images may be performed in a more anisotropic manner, as in [82, 83]. An additional relaxation comes from lifting the critically sampled scheme, yielding oversampled, translation-invariant (see Sec. 4.2.3) multiscale wavelets, wavelet/cosine packets or frames [67, 84, 85, 86, 87, 88]. Multidimensional oversampled filter banks in \(n-D\) with limited redundancy may be designed as well [89, 90, 91, 92, 93].

### 3.1.2 Pyramid-related wavelets

Notably influenced by [94, 95], Unser and Van de Ville propose a slightly redundant transform [96] based on a pyramid-like wavelet analysis. This decomposition constitutes a wavelet frame with mild redundancy, which is nevertheless not steerable. Subsequently, the same authors propose a steerable analysis [97] based on polyharmonic \(B\)-splines [98] and the Maar-like [5, 99] wavelet pyramid. Such multiresolution analysis can easily be implemented via filter banks as detailed in [97] and the total redundancy of this decomposition is \(8/3\) (a redundancy of \(4/3\) is introduced by the pyramid structure and the complex nature of the coefficients increases the redundancy by a factor of 2). A similar approach based on Riesz-Laplace wavelets is proposed in [100]. The latter constructions are related to Hilbert and Riesz transforms.

### 3.1.3 Complexifying Discrete Wavelets with Hilbert and Riesz

Different kinds of complexification are indeed a possible option in order to tackle the problem of poor directionality with classical wavelet transforms. The common basic idea leans toward analytic wavelets and their combination to improve the 2-D directionality. Behind a generic notion of complex wavelets reside different approaches detailed hereafter, which require the definition of some basic tools.

We first introduce the Hilbert transform, termed “complex signal” in [101] and exhaustively mapped in [102]. While the 1-D Hilbert transform is unambiguously defined, there exists multidimensional extensions, often obtained by tensor products, thus leading to approximations. In order to increase the directionality property, other multidimensional constructions (discussed in [103]) have also been proposed.

- The 1-D Hilbert transform \(\mathcal{H}\) of a signal \(f\) is easily expressed in the Fourier domain as

\[
\mathcal{H}\{f\}(\omega) = -\text{sign}(\omega)\hat{f}(\omega).
\]
• The 1-D fractional Hilbert transform $\mathcal{H}_\theta$ of $f$ is similarly defined in [104] by

$$\mathcal{H}_\theta\{f\}(\omega) = \exp(i\pi\theta \text{sign}(\omega))\hat{f}(\omega).$$ (7)

• The 2-D directional Hilbert transform $\mathcal{H}_\theta$ of $f$ is one of the 2-D extensions defined in [104] as

$$\mathcal{H}_\theta\{f\}(\omega_1,\omega_2) = -i\text{sign}(\cos(\theta)\omega_1 + \sin(\theta)\omega_2)\hat{f}(\omega_1,\omega_2).$$ (8)

See also [105].

The Hilbert transform was already associated with wavelets for transient detection by Abry et al. [106]. Others early connections between wavelets and the Hilbert transform are drawn in [107, 108, 109]. At the end of the 1990’s, Kingsbury proposed the dual-tree transform based on even and odd filters [110, 111]. An alternative construction is given by Selesnick [112]. It amounts to performing two discrete classical wavelet transforms in parallel, the wavelets generated by the trees forming Hilbert pairs. An atom of the corresponding basis (here the diagonal wavelet) and its corresponding frequency plane tiling are depicted in Fig. 9. The corresponding dual-tree of wavelet coefficients is represented in Fig. 8 which clearly shows the separation of oriented structures with different orientations. The resulting oriented wavelet dictionary has a small redundancy and is also computationally efficient. The corresponding wavelet is approximately shift invariant, see [113] for more details. It is extended to the $M$-band setting by Chaux et al. [114] and to wavelet packets in [115, 116]. In Fig. 10, one subband of the wavelet transform (red square in Fig. 7), two subbands (primal+dual) of the dyadic dual-tree transform (red squares in Fig. 8), as well as the corresponding eight subbands (4 primal+4 dual) of the 4-band dual-tree wavelet decomposition are depicted. In Fig. 10(d), the fine oriented textures from the left side of the image are (slightly) better separated in some non-horizontal subbands. The wavelet/frequency tiling corresponding to the 4-band dual-tree wavelet decomposition are depicted in Fig. 11. The main advantage of this decomposition is that it achieves a directional image analysis with a small redundancy of a factor 2 (4 for the complex transform).

Gopinath [117, 118] has designed phaselets which is an extension of the dyadic dual-tree wavelet transform [110, 119]. They aim at improving translation invariance with a given redundancy, and are built by carefully observing the effects of shifts in a discrete wavelet transform. 2-D phaselets are easily obtained by tensor products.

More recently, the shiftability of the dual-tree transform has been studied by Chaudhury et al. [104] by introducing the fractional Hilbert transform (7). A 2-D extension has been proposed in [120] and the construction of Hilbert transform pairs of wavelet bases can be found in [121]. Note that previous works dealing with multidimensional extensions have been first reported for instance in [122] and then in [123, 124] using the notion of hypercomplex wavelets.

Numerous extension to multidimensional signals have been proposed, see for instance [125, 126]. They, for instance, use the Riesz transform $\mathcal{R}$, which is defined in the frequency domain as follows:

$$\mathcal{R}\{f\} = (\mathcal{R}_1\{f\}, \ldots, \mathcal{R}_N\{f\}).$$

where

$$\forall n \in \{1, \ldots, N\}, \quad \mathcal{R}_n\{f\}(\omega) = -i\frac{\omega_n}{\|\omega\|}\hat{f}(\omega).$$ (10)

Other recent extensions of multidimensional oriented wavelets are based on the notion of monogenic signal/wavelet [127, 128, 129]. We finally mention that other methods have been developed in order
to achieve directional analytic wavelets such as softly space projections or the Daubechies complex wavelets. Complex wavelets have also been shown to provide robust image similarity measures.

![Figure 8: Dyadic dual-tree wavelet decomposition of the Haar-Riesz Memorial plaque.](image)

![Figure 9: The dyadic dual-tree wavelet. (a) Example of atom (diagonal wavelet). (b) Associated frequency partitioning.](image)

3.2 Non-Separable Directionality

3.2.1 Non-separable Decomposition Schemes

In contrast to the separable constructions detailed in Sec. 3.1.1 where n-D representations are composed of 1-D transforms applied separately along each dimension (sometimes recombined, as in the dual-tree wavelet case or in ), non-separable constructions are directly performed in n-D.
Figure 10: The original image (a) and the horizontal subband(s) at first resolution level for (b) Dyadic wavelet transform, (c) Dyadic dual-tree transform (primal+dual) and (d) $M$-band dual-tree wavelet decomposition (primal+dual) of the Haar-Riesz Memorial plaque.

Figure 11: The $M$-band dual-tree wavelet. (a) Example of atom. (b) Frequency partitioning.
Since the literature on this topic is large, this section is focussed on a limited number of references dealing with directional multiscale decompositions.

These works are often related to non-diagonal subsampling operators, non-rectangular lattices (e.g., quincunx grids or integer lattices) [141, 142], or non-separable n-D windows [143, 144]. Complementary standard references can be found in [70, p. 558 sq.] or [145, 146, 147]. Some of these constructions are defined using the lifting scheme, see Sec. 4.3 and 5.3 for more details. While directional filter banks do not provide a multiscale representation in general, 2-band [148, 149, 150] or even M-band non-redundant directional discrete wavelets [151] have been proposed. Non-separable schemes are used for instance as building blocks for multiscale geometric decompositions such as:

- directional filter banks in [152], and their combination with a Laplacian pyramid in contourlets [153, 154] or surfacelets [155].
- (pseudo-) polar fast Fourier transform (FFT) [156] in first generation curvelets described in Sec. 3.3.2 or the loglets in [157] that exhibit a polar separability.

In order to overcome the limited efficiency of the standard 2-D separable DWT for representing non-horizontally or vertically directed edges (see Sec. 2.3.2), several authors have adapted 1-D concepts for local edge representation. Reissell [158] develops, for instance, a pseudo-coiflet scheme that addresses numerically efficient interpolation for a parametric wavelet representation of curves. Moreover, for digital images it would be beneficial to follow contours on more appropriate discrete paths (see [159] for an early application) such as discrete lines [160, 161, 162]. While discrete lines are adapted to digital ridgelets in [163], Velisavljević et al. propose multidirectional anisotropic directionlets [164], based on skewed lattices, with directional vanishing moments along direction with rational slopes, still relying on a simple separable implementation. This approach is refined in [165] by taking lifting steps of 1-D wavelets along an explicit orientation map defined on a quincunx multiresolution sampling grid, and in [166] with a more efficient representation for sharp features. A combination of 2-D filter banks and 1-D directional filter bank is devised in [167, 168]. Similar ideas have been recently applied to edge detection in [169]. In [170], non-adaptive directional wavelet frames are constructed with Haar wavelets and a finite collection of “shear” matrices. Krommweh also proposes tetrolets, an adaptive variation (akin to digital wedgelets) of Haar-like wavelets on compact tetrrominoes (geometric shapes composed of four squares, connected orthogonally, see [171]). These last constructions may further sparkle the growing interest of the association of multiscale analysis and discrete geometry [172].

### 3.2.2 Steerable Filters

Steerable filters [173, 174, 175] were developed in order to achieve more precise feature detectors adapted to image edge junctions (often termed “X”, “T” and “L” junctions). Their construction allows one to compute multiscale derivatives at any orientation (steerability) from a linear combination of a small number of fixed filters. In [174], the construction starts from a bidimensional Gaussian \( G(x) = \exp(-\frac{1}{2}\|x\|^2) \) for \( x = (x_1, x_2) \) with associated base (differential) filters \( G^0(x) = \frac{\partial}{\partial x_1} G(x) \) and \( G^{\pi/2}(x) = \frac{\partial}{\partial x_2} G(x) \).

From the properties of the directional derivative, filters “steered” at angle \( \theta \in [0, 2\pi] \) are then built from

\[
G^\theta(x) = \cos(\theta) G^0(x) + \sin(\theta) G^{\pi/2}(x). \tag{11}
\]
where \( \cos(\theta) \) and \( \sin(\theta) \) may be interpreted as interpolators. Since the convolution is linear, the resulting steered decomposition arises from a combination of images that underwent \( G^0 \) or \( G^{\pi/2} \) filters. A larger class of asymmetric oriented filters is proposed in [176]. Their angular parts are derived from even and odd functions:

\[
\forall \phi \in [0, 2\pi), \quad h_e(\phi) = \sum_{n=1}^{N} w_n \cos(n\phi) \quad \text{and} \quad h_o(\phi) = \sum_{n=1}^{N} w_n \sin(n\phi),
\]

which form Hilbert transform pairs (see Sec. 3.1.3), unlike the resulting spatial filters. An angle \( \theta \) rotation is obtained through:

\[
h_e(\phi - \theta) = k_e(\theta)^T f(\phi) \quad \text{and} \quad h_o(\phi - \theta) = k_o(\theta)^T f(\phi),
\]

where \( k_e(\theta) \) and \( k_o(\theta) \) are interpolating vectors and \( f(\phi) \) is a weighted Fourier vector, namely:

\[
k_e(\theta) = \left[ \cos \theta, \sin \theta, \cos(2\theta), \sin(2\theta), \ldots, \cos(N\theta), \sin(N\theta) \right]^T,
\]

\[
k_o(\theta) = \left[ -\sin \theta, \cos \theta, -\sin(2\theta), \cos(2\theta), \ldots, -\sin(N\theta), \cos(N\theta) \right]^T,
\]

\[
f(\phi) = \left[ w_1 \cos \phi, w_1 \sin \phi, w_2 \cos(2\phi), w_2 \sin(2\phi), \ldots, w_N \cos(N\phi), w_N \sin(N\phi) \right]^T.
\]

If we set \( \theta = \theta_n = 2\pi n/N \) for \( 1 \leq n \leq N \), filters \( h_e(\cdot - \theta) \) and \( h_o(\cdot - \theta) \) may be rewritten as a linear combination of \( h_e(\cdot - \theta_n) \) and \( h_o(\cdot - \theta_n) \), \( 1 \leq n \leq N \). An example of decomposition with four orientations and two scales is represented in Fig. 12 with corresponding projection atoms in Fig. 13. Steerable filters may be combined with discrete wavelets to improve their radial properties [177, 178].

Figure 12: Steerable pyramid decomposition of the Haar-Riesz Memorial plaque, over two scales, with four orientations.

3.2.3 Directional Wavelets and Frames

In Sec. 2.2 the two-dimensional Continuous Wavelet Transform (2-D CWT) was defined as a straightforward extension of the 1-D CWT using isotropic wavelets. It is however possible to make
use of more complicated group actions to drive the CWT parameterization in the plane, such as rotations or the similitude group SIM(2), see [147].

Consequently, given a mother function $\psi \in L^2(\mathbb{R}^2)$ that is well localized and oriented, we write

$$\psi(b,a,\theta)(x) = \frac{1}{a} \psi\left(\frac{1}{a} R_\theta^{-1}(x - b)\right),$$

where $R_\theta$ stands for the $2 \times 2$ rotation matrix. For a function $f \in L^2(\mathbb{R}^2)$, the 2-D CWT (non-isotropic) is thus

$$W_f(b,a,\theta) = \langle \psi(b,a,\theta), f \rangle.$$

If the wavelet is admissible, i.e., if

$$c_\psi = (2\pi)^2 \int_{\mathbb{R}^2} \left| \hat{\psi}(\omega) \right|^2 / \| \omega \|^2 \, d^2 \omega < \infty,$$

then, the CWT may be inverted through

$$f(x) = c_\psi^{-1} \int_0^{\infty} \frac{da}{a^3} \int_0^{2\pi} d\theta \int_{\mathbb{R}^2} d^2 b \ W_f(b,a,\theta) \ \psi(b,a,\theta)(x),$$

the equality being valid almost everywhere on $\mathbb{R}^2$.

The selectivity power of the wavelet, that is, its ability to distinguish two close orientations in an image, may be measured in the Fourier domain. Typically, a good directional wavelet is thus a function whose Fourier transform is essentially or exactly contained in a cone with apex on the origin: the narrower the cone, the more selective the wavelet transform using that wavelet [147].

Practically, it is not satisfactory to manipulate a continuum of wavelets parameterized by continuous parameters. The question is therefore to know if it is possible to decompose and reconstruct an image from a discretized set of parameters, i.e., on the family $G = \{ \psi(b,a,\theta) : b \in \mathcal{P}, a \in \mathcal{A}, \theta \in \Theta \}$ with $\mathcal{P} \subset \mathbb{R}^2$, $\mathcal{A} \subset \mathbb{R}^*_+$ and $\theta \subset [0,2\pi)$ all discrete (countable) sets. As explained in Sec. 1.3.2, this question amounts to ask when $G$ is a frame of $L^2(\mathbb{R}^2)$.

Such frames have been built for the Morlet (or Gabor) wavelet [179, 180]:

$$\psi(x) = G_{\sigma_0}(x) \ e^{i\omega_0 \cdot x} = e^{-\|x\|^2/2\sigma_0^2} \ e^{i\omega_0 \cdot x}, \quad \hat{\psi}(\omega) \propto G_{1/\sigma_0}(\omega - \omega_0) = e^{-\sigma_0^2\|\omega - \omega_0\|^2/2},$$

where $\omega_0 \in \mathbb{R}^2$ defines the cone axis and $\sigma_0 > 0$ is related to the cone aperture, as represented in Fig. 13. Notice that approximate quadrature filters exist to accelerate the computation of the wavelet coefficients [181].

Finally, a multiresolution structure can also be put on the angular dependency of the conic wavelets in the frequency domain to define multiselective wavelets [182]. This generates a redundant basis that may represent jointly a large spectrum of features ranging from highly directional ones (e.g., edges) to isotropic elements (e.g., spots, corners) and including intermediate directional structures such as textures.
3.3 Directionality in Anisotropic Scaling

3.3.1 Ridgelets

Ridgelets \[183, 184\] and wavelet X-ray transforms \[185\] appear as a combination of a 1-D wavelet transform and the Radon transform \[186\]. They are designed for efficient representation of discontinuities over straight lines. A bivariate ridgelet transform is constant along parameterized lines \(x_1 \cos(\theta) + x_2 \sin(\theta) = b\) and defined for \(a > 0, b \in \mathbb{R}\) and \(\theta \in [0, 2\pi)\), by

\[
\forall \mathbf{x} = (x_1, x_2) \in \mathbb{R}^2, \quad \psi_{(b,a,\theta)}(\mathbf{x}) = a^{-1/2} \psi((x_1 \cos(\theta) + x_2 \sin(\theta) - b)/a). \tag{14}
\]

Ridgelet coefficients for the image \(f\) are given by

\[
\mathcal{R}_f(b,a,\theta) = \int \psi_{(b,a,\theta)}(\mathbf{x}) f(\mathbf{x}) \, d^2\mathbf{x} = \int \mathcal{R}_f(\theta,t) a^{-1/2} \psi((t - b)/a) \, dt, \tag{15}
\]

where \(\mathcal{R}_f(\theta,t)\) represents the Radon transform of \(f\) defined by:

\[
\mathcal{R}_f(\theta,t) = \int \int f(x_1, x_2) \delta(x_1 \cos(\theta) + x_2 \sin(\theta) - t) \, dx_1 \, dx_2, \tag{16}
\]

with \(\delta\) denoting the Dirac distribution. The ridgelet transform may be interpreted as a 1-D wavelet transform of Radon slices where the angle \(\theta\) is constant and \(t\) varies. Several implementations and variations exist in order to overcome the issues raised by the Radon transform discretization, such as the finite ridgelet transform \[187\], the approximate digital ridgelet transform \[188\] or the discrete analytical ridgelet transform \[189\]. Their multiscale implementation \[16\] is the basis for the first generation curvelets described in Sec. 3.3.2. A ridgelet decomposition\(^7\) \[190\] of the Haar-Riesz Memorial plaque is given in Fig. 15 with a typical atom along with a synthetic description of its implementation in Fig. 16.

3.3.2 Curvelets

The curvelet representation, introduced by Candès and Donoho [16][191], improves the approximation of cartoon images with $C^2$ edges with respect to wavelets. We review here the second generation of curvelets, as introduced in [191].

**Continuous Curvelet Transform** A curvelet atom, with scale $s$, orientation $\theta \in [0, \pi)$, position $y \in [0, 1]^2$ is defined as

$$\psi_{s,y,\theta}(x) = \psi_s(R_{\theta}^{-1}(x - y))$$  

where $\psi_s(x) \approx s^{-3/4} \psi(s^{-1/2}x_1, s^{-1}x_2)$ is approximately a parabolic stretch of a curvelet function $\psi$ with vanishing moments in the vertical direction. At scale $s$, a curvelet atom is thus a needle oriented in the direction $\theta$ whose envelope is a specified ridge of effective length $s^{1/2}$ and width $s$, and which displays an oscillatory behavior transverse to the ridge.

A curvelet atom thus benefits from a parabolic scaling property $\text{width} = \text{length}^2$ that is a major departure from oriented wavelets. Fig. 17 presents an example of a curvelet atom, together with its Fourier transform, for the second generation of curvelets. The resulting curvelet Fourier tiling resembles that of the Cortex transform [192].

The continuous curvelet transform computes the set of inner products $\langle \psi_{s,y,\theta}(\cdot), f \rangle$ for all possible $(s, y, \theta)$. A careful design of $\psi_s$ [191] enables a conservation of energy and a simple reconstruction formula. The decay of the curvelet transform as $s$ decreases allows one to detect the position and orientation of contours [193].

**Curvelet Frame** The continuous curvelet representation is sampled in order to obtain a curvelet frame $B = \{\psi_m\}_m$, [194], see also [194] for the description of a complex curvelet tight frame.
A curvelet atom, with scale $2^j$, orientation $\theta_\ell \in [0, \pi)$, position $x_n \in [0,1]^2$ is defined from the continuous atom [17]

$$\psi_m(x) = \psi_{2^j, \theta_\ell, x_n}(x) \quad \text{where} \quad m = (j, n, \ell)$$

where the sampling locations are

$$\theta_\ell = \ell \pi 2^{j/2}^{-1} \in [0, \pi) \quad \text{and} \quad x_n = R_{\theta_\ell}(2^{j/2} n_1, 2^j n_2) \in [0,1]^2.$$ 

The curvelet parameters are sampled using an increasing number of orientations at finer scales. This sampling is the key ingredient to ensure the tight frame property [191], which provides a simple reconstruction formula.

A fast discrete curvelet transform computes the set of inner products $\{\langle \psi_m, f \rangle\}_m$ in $O(N \log(N))$ operations for an image with $N$ pixels, see [195]. The coronae and rotations of the continuous settings are replaced by their discrete Cartesian counterparts, i.e. concentric squares and shears.
Figure 17: Left: Example of a curvelet $\psi(x, s, y, \theta)$. Right: the frequency support of $\hat{\psi}(\omega, s, y, \theta)$ is a wedge.

Figure 18 shows an example of curvelets decomposition. Candès and Donoho prove [196] that the curvelet non-linear approximation $f_M = H_T(f, B)$, where $H_T$ is defined in (2), ensures an approximation error decay $\|f - f_M\|_2 = O(M^{-2} \log^3(M))$ for a $C^2$ regular image outside $C^2$ regular edge curves. This is a significant improvement over the $O(M^{-1})$ error decay of a wavelet approximation described in Sec. 2.3.2 and is achieved with a fast $O(N \log(N))$ algorithm for discrete images. This asymptotic error decay is optimal (up to logarithmic factor) for the class of images that are $C^2$ regular outside $C^2$ regular edge curves, see [196]. Monogenic curvelets are proposed in [197] to obtain additional advantages over monogenic wavelets, described in Section 3.1.3.

Shearlet atoms [198, 199] are built similarly to curvelets, but they replace, in their continuous formulation, rotation and anisotropic stretch with anisotropic shears. The discrete shearlet transform [200, 201] is thus implemented similarly to the discrete curvelet transform [195] using discrete shears. It provides the same approximation properties as curvelets, albeit with a different directional sensitivity (e.g., the number of orientations doubles at each scale). Recently a type-I ripplet transform [202] has been proposed as an extension to curvelets with alternative scaling laws.

### 3.3.3 Contourlets

Contourlets [153] are sometimes considered a low-redundancy discrete approximation of curvelets. Actually, they are designed in the spatial domain (instead of the frequency plane), aiming at a close-to-critical directional representation. Their construction is based on a Laplacian Pyramid [43] (see Fig. 5). The low-pass part of the pyramid is further decomposed with a biorthogonal 9/7 DWT. Each difference image obtained from the pyramid is subject to directional filter bank (see Sec. 3.2) (initially from [141], [203] proposes a simpler implementation based only on a quincunx structure). A contourlet decomposition is illustrated in Fig. 19. The resulting frequency plane tiling is represented in Fig. 20(c). The contourlet inherits its redundancy of 4/3 from the pyramidal scheme. Its approximation rate is similar to that of curvelets (Sec. 3.3.2). At one end of the redundancy

---

8 The Curvelab toolbox has been used, see [http://www.curvelet.org/](http://www.curvelet.org/)
9 An implementation is available at [http://www.shearlab.org](http://www.shearlab.org)
10 The contourlet toolbox has been used, see [http://www.ifp.illinois.edu/~minhdo/software/](http://www.ifp.illinois.edu/~minhdo/software/)
Figure 18: Curvelet decomposition of the Haar-Riesz Memorial plaque. The layout of the coefficients follows the frequency localization of curvelet atoms.
spectrum, \cite{201} proposes a critically sampled version. At the other end, the constraints thus laid on
the basis functions (Figs. 20(a)-20(b)) are relaxed by the design of a more redundant \cite{154} version,
based on non-subsampled (Sec. 3.1.1) pyramid and directional filters.

![Figure 19: Contourlet decomposition of the Haar-Riesz Memorial plaque.](image)

3.3.4 Frames for Oscillating Textures.

While curvelets, contourlets and shearlets are optimized for the processing of edges, they are
not tailored for the processing of oscillating textures, because of their poor frequency localization.
Generic oscillating patterns can be captured using a local Fourier analysis on a regular segmentation
of the image in squares. This corresponds to an expansion in a Gabor frame, see for instance \cite{33}.
The spatial segmentation can be optimized using a decomposition in a best cosine packet dictionary
as described in Section 4.2.

Wavelet packets, detailed in Section 4.2, have been used to process and compress oscillating
textures such as fingerprints. Brushlets \cite{205}, introduced by Meyer and Coifman, improve the
frequency localization of wavelet packets.

Wave atoms [206] better capture geometric textures using an anisotropic scaling. The wavelength of wave-atom oscillations is proportional to the square of their diameter. This scaling allows a thresholding in a wave atom frame to optimally approximate textures obtained by a smooth warping of a sinusoidal profile, see [206].

4 Redundancy and Adaptivity

Highly redundant representations allow us to improve the representation of complicated images with edges and textures. However, as described hereafter, computing efficient image representations in such dictionaries sometimes requires approximations.

4.1 Pursuits in Redundant Dictionaries

An approximation $f_M$ of an image $f$ with $M$ atoms from a highly redundant dictionary $B = \{\psi_{m_j} : 1 \leq j \leq P\}$ is written

$$f_M = \Psi a = \sum_j a_j \psi_{m_j}, \text{ with } \|a\|_0 = \#\{j : a_j \neq 0\} \leq M.$$  

Computing the $M$-sparse coefficients $a$ that produce the smallest error $\|f - f_M\|$ in a generic dictionary is NP-hard [207]. Furthermore, the $M$-terms approximation $f_M = H_T(f, B)$ computed by thresholding (2) might be quite far from the best $M$-terms approximation. One thus has to use approximate schemes in order to compute an efficient approximation in a reasonable time.

4.1.1 Matching Pursuits

Matching pursuit [208] computes $f_M$ from $f_{M-1}$ by choosing the atom $\psi_m$ that minimizes the correlation $|\langle \psi_m, f - f_{M-1} \rangle|$. Orthogonal matching pursuit [33, 209] further reduces the approximation error by projecting $f$ on the $M$ chosen atoms to compute $f_M$.

Under restrictive conditions on the dictionary $B$, these greedy algorithms compute an approximation $f_M$ that is close to the best $M$-term approximation, see for instance [210, 211]. These conditions typically require the correlation $|\langle \psi_m, \psi_{m'} \rangle|$ to be small for $m \neq m'$, which is not applicable to highly redundant dictionaries typically used in image processing.

4.1.2 Basis Pursuit

A sparse approximation is obtained by convexifying the $\ell_N^0$ pseudo norm, and solving the following basis pursuit denoising convex problem [212]

$$f_M = \Psi a = \sum_j a_j \psi_{m_j} \text{ where } a \in \arg\min_{\tilde{a} \in \mathbb{R}^P} \frac{1}{2} \|f - \sum_j \tilde{a}_j \psi_{m_j}\|^2 + \mu \|\tilde{a}\|_1, \quad (18)$$

where $\mu > 0$ is adapted so that $\|a\|_0 = M$. This problem (18) is minimized, for instance, using iterative thresholding methods [213, 214]. Algorithmic solutions to its generalized form as sums of

\[\text{See http://www.waveatom.org}\]
convex functions (a common formulation to many data processing problems) may be solved with great flexibility in the framework of proximity operators [215]. Similarly to matching pursuit algorithms, this $\ell_1^N$ approximation can be shown to be close to the best $M$-term approximation if the atoms of $B$ are not too correlated, see for instance [216, 211].

### 4.1.3 Pursuits in Parametric Dictionaries

Parametric dictionaries are obtained from basic operations (like rotation, translation, dilation, shearing, modulation, etc.) applied to a continuous mother function. Even if such dictionaries also define redundant bases similar to those introduced earlier, they deserve a separate description since their parametric nature provides them with some particular properties. They are generally created to provide a very rich and dense family of functions built from the geometrical features of the analyzed image. They have applications in image and video coding [217], multi-modal signal analysis (e.g., video plus audio) [218], and also for signal decomposition on non-Euclidean spaces [219].

Formally, given a set of $S$ transformations $T_{m_i}^i$ for $1 \leq i \leq S$ parameterized by $m_i \in \Lambda_i \subset \mathbb{R}^{n_i}$, the parametric dictionary is related to a certain discretization of $\Lambda^d \subset \Lambda = \Lambda_1 \times \cdots \times \Lambda_S$, i.e.,

$$B = \{ \psi_m(x) = [T_{m_1}^1 \cdots T_{m_S}^S \psi](x) \in L^2(\mathbb{R}^2) : m = (m_1, \ldots, m_S) \in \Lambda^d \}.$$  

The directional wavelets described in Sec. 3.2.3 and the subsequent frames built from them are actually an example of parametric dictionaries with the translations $T_{m_1}^1 T_{m_2}^2 = T_{b_1}^1 T_{b_2}^2$, the rotation $T_{m_3}^3 = R_\theta$ and the dilation $T_{m_4}^4 = D_a$ operations. For these wavelets, the decomposition/reconstruction methods are relatively easy to formulate, due to the continuous inversion formula or using the frame condition.

However, checking the frame condition may sometimes become tedious. In addition, more transformations of the mother function may be added in order to enlarge the family of functions, further worsening the frame bounds.

![Figure 21: Explanation of the optimization in $\Lambda$ starting from a point in $\Lambda^d$.](image)

Fortunately, as described in Sec. 4.1 it is still possible to find good description of images in very general family of functions. Most of the time, since the Parametric Dictionaries are much larger than other dictionaries of controlled redundancy, the (Orthogonal) Matching Pursuit decomposition (Sec. 4.1.1) is used to find a sparse representation of signals.
Figure 22: (a) Original image. (b) Reconstruction with 300 atoms for a rich parametric dictionary containing $5 \times 5$ anisotropic scales, 8 orientations, and $N$ translations. PSNR: 26.63 dB (CT: 4634s). (c) Optimized Reconstruction at 300 atoms starting from a dictionary with only $3 \times 3$ scales, 4 directions and $N$ translations, PSNR: 26.68 dB (CT: 949s).

Interestingly, thanks to the parametric nature of $\mathcal{B}$, the dictionary discretization can be refined during the Matching Pursuit iterations. Indeed, since $\mathcal{B}$ is the discretization of the continuous manifold $\mathcal{M} = \{ \psi_m : m \in \Lambda \} \subset L^2(\mathbb{R}^2)$ generated by all the transformations of $\psi$, at each iteration of MP in the decomposition of a signal $f \in L^2(\mathbb{R}^2)$ the refinement is performed as follows. As illustrated on Fig. 21, given the best atom $\psi_m$ found in $\mathcal{B}$, a gradient ascent respecting the (Riemannian) geometry of $\mathcal{M}$ is run on $\Lambda$ to maximize the correlation $S(m') = |\langle \psi_{m'}, R^n_f \rangle|$ between the current MP residual $R^n_f = f - f^n$ at step $n$ and the atom $\psi_{m'}$. A new parameter $m^*$ is then used instead of $m$ in the signal representation and the next iteration is realized on the residual $R^{n+1}_f = R^n_f - \langle \psi_{m^*}, R^n_f \rangle \psi_{m^*}$ [220].

Fig. 22 presents the result of such an improvement for two different decompositions of the Barbara image (with $N = 128^2$ pixels) with similar qualities (expressed using the Peak Signal-to-Noise Ratio - PSNR). The first one (Fig. 22(b)) is obtained by a rich parametric dictionary defined by anisotropic dilations, rotations, and translations of a 2-D second order directional derivative of a Gaussian. The second decomposition uses a poorer dictionary with the same parameterization and mother function but with a manifold optimization on the atom parameters. The interest of the latter method is to provide a similar quality for a smaller Computational Time (CT).

4.1.4 Processing with Highly Redundant Dictionaries

Compression with Sparse Expansions Dictionaries with oriented atoms have proven to be successful for improving the JPEG 2000 compression standard at low bit rates [221, 222]. The approximation of the image is computed using the matching pursuit algorithm. Matching pursuit in Gabor dictionaries, i.e., dictionaries made of Gabor wavelets (Sec. 3.2.3), have been used for coding the motion residual in video compression schemes [223].

Inverse Problem Regularization Data acquisition devices usually only acquire $S$ noisy low resolution measurements $y = \Phi f_0 + w \in \mathbb{R}^S$ of a high resolution image $f_0 \in \mathbb{R}^N$ of $N \gg S$ pixels.
Figure 23: Example of deconvolution using $\ell_1^N$ regularization in a frame of translation invariant wavelets. (a) Original $f_0$. (b) Observation $y = \Phi f_0 + w$. (c) Deconvolution $f$.

The linear operator $\Phi$ models the acquisition and might include some blurring and sub-sampling of the high resolution data.

Recovering a good approximation $f \in \mathbb{R}^N$ of $f_0$ from these measurements $y$ corresponds to solving a difficult ill-posed inverse problem, that requires the use of efficient priors to model the regularity of the image. Early priors include the Sobolev prior that enforces smoothness of the image, and the non-linear total variation \cite{224} that can produce sharper edges.

More recently, $\ell_1^N$ sparse priors in redundant dictionaries $\mathcal{B}$ have been proved to be efficient in order to solve several ill-posed problems, see for instance \cite{33} and references therein. In this setting, one computes the coefficients $a$ of $f = \Psi a = \sum_j a_j \psi_{m_j}$ in a frame $\mathcal{B}$ of $P$ atoms by solving a $\ell_1^N$ augmented Lagrangian form

$$
a \in \argmin_{\tilde{a} \in \mathbb{R}^P} \frac{1}{2}\|y - \Phi \tilde{a}\|^2 + \mu \|\tilde{a}\|_1 \quad \text{where} \quad \Psi \tilde{a} = \sum_j \tilde{a}_j \psi_{m_j}$$

(19)

where $\mu$ should be adapted to the noise level $\|w\|$ that is supposed to be known. This minimization problem corresponds to computing the basis pursuit approximation \cite{18} of the measurements $y$ in the highly redundant dictionary $\{\Phi \psi_{m_j} : 1 \leq j \leq P\}$ of $\mathbb{R}^S$. It can thus be solved using the same algorithms.

Figure 23 shows the use of this sparse regularization method when solving a deconvolution problem. In this application, the operator is a convolution $\Phi f = f \ast G_\sigma$ with a Gaussian kernel $G_\sigma$ as defined in Sec. 1.3.1. The redundant dictionary $\mathcal{B}$ is a translation invariant wavelet frame.

4.1.5 Source Separation

Sparse representations can be used to separate sources that are known to be sparse in different dictionaries. This corresponds to the morphological component analysis (MCA) of Starck et al. \cite{225}. In its simplest setting, it can be used to separate a single noisy image $y$ into a sum $y = \hat{f}_G + f_T + w$ of a cartoon-like component $f_G$ (or geometric component), a texture component $f_T$ and residual noise $w$. One can use a dictionary $\mathcal{B} = \mathcal{B}_G \cup \mathcal{B}_T$ union of wavelets ($\mathcal{B}_G$) and local
cosine (B_T), and compute a sparse approximation f of y

\[ f = \Psi a = \Psi G a_G + \Psi T a_T \]  \hspace{1cm} (20)

where \( a = [a_G; a_T] \) is the solution of the \( \ell^1 \) basis pursuit (18) applied to \( y \). The separation, obtained using \( f_G = \Psi G a_G \) and \( f_T = \Psi T a_T \), is illustrated in Fig. 24.

The modeling of natural images as a sum of a cartoon layer and an oscillating texture layer has been initiated by Y. Meyer in his book [12]. Beside sparsity-based approaches such as (20), other variational methods have been proposed, see for instance the work of J.-F. Aujol et al. [13].

4.2 Tree-structured Best Basis Representations

Pursuit algorithms are quite slow and face difficulties in order to compute provably efficient approximations when the dictionary is too redundant. In order to avoid these bottlenecks, one needs to consider more structured representations, that allow one to use fast and provably efficient approximation strategies. The structuring of the representation can be implemented by computing an adapted basis \( B^\lambda \) parameterized by a geometric parameter \( \lambda \) that captures the local direction of edges or textures. This section details best basis schemes: they introduce the desired adaptivity together with fast algorithms employing the hierarchical structure of parameters \( \lambda \).

4.2.1 Quadtree-based Dictionaries

A dictionary of orthonormal bases is a set \( \mathcal{D}_\lambda = \{ B^\lambda \}_{\lambda \in \Lambda} \) of orthonormal bases \( B^\lambda = \{ \psi^\lambda_m \}_{m} \) of \( \mathbb{R}^N \), where \( N \) is the number of pixels in the image. Instead of using an \( a \) priori fixed basis such as the wavelet or Fourier basis, one chooses a parameter \( \lambda^* \in \Lambda \) adapted to the structure of the image to process and then uses the optimized basis \( B^{\lambda^*} \).

In order to enable the fast optimization of a parameter \( \lambda^* \) adapted to a given signal or image \( f \) to process, each \( \lambda \in \Lambda \) is constrained to be a quadtree. The quadtree \( \lambda \) that parameterizes a basis \( B^\lambda \) defines a dyadic segmentation of the square \( [0,1]^2 = \bigcup_{(j,i) \in L(\lambda)} S_{j,i} \), where \( L(\lambda) \) are the leaves of the trees, as shown on Fig. 25. Each square \( S_{j,i} \) is recursively split into four sub-squares...
In order to enrich the representation parameterized by a quadtree, we attach to each leave of the tree a geometric token, and denote as \( \tau \) the number of tokens. A token indicates the direction of the image geometry in a square of the segmentation.

### 4.2.2 Best Basis Selection

Given a number \( M \) of coefficients, the best basis \( B^\lambda^\star \in D_\Lambda \) adapted to \( f \in \mathbb{R}^N \) minimizes the best \( M \)-terms approximation error. This can be equivalently obtained by minimizing a penalized Lagrangian that weights the approximation error with the number of coefficients

\[
\lambda^\star \in \text{argmin}_{\lambda \in \Lambda} \mathcal{L}(f, B^\lambda, T) = \| f - f_M^\lambda \|^2 + M^\lambda T^2, \tag{21}
\]

where \( f_M^\lambda \) is the best \( M^\lambda \)-term approximation in \( B^\lambda \) computed by thresholding at \( T > 0 \)

\[
f_M^\lambda = H_T(f, B^\lambda, T) = \sum_{|\langle \psi_m^\lambda, f \rangle| > T} \langle \psi_m^\lambda, f \rangle \psi_m^\lambda \quad \text{and} \quad M^\lambda = \# \left\{ m : |\langle \psi_m^\lambda, f \rangle| > T \right\}, \tag{22}
\]

since \( B^\lambda \) is orthonormal. This Lagrangian can be re-written as a sum over each coefficient in the basis

\[
\mathcal{L}(f, B^\lambda, T) = \sum_m \max(|\langle \psi_m^\lambda, f \rangle|^2, T^2). \tag{23}
\]

This kind of Lagrangian can be efficiently optimized using a dynamic search algorithm, originally presented by Coifman et al. [226], which is a particular instance of the Classification and Regression Tree (CART) algorithm of Breiman et al. [227] as explained by Donoho [228]. It is possible to consider other criteria for best basis selection, such as for instance the entropy of the coefficients. This leads different Lagrangians that can be minimized with the same method [226].

The complexity of the algorithm is proportional to the complexity of computing the whole set of inner products \( \{ \langle \psi_m^\lambda, f \rangle : \lambda \in \Lambda \} \) in the dictionary. For several dictionaries, such as those considered in this section, a fast algorithm performs this computation in \( O(P) \) operations, where \( P \) is the total number of atoms in \( D_\Lambda \). For tree-structured dictionaries, this complexity is thus \( O(\tau N \log_2(N)) \), where \( \tau \) is the number of tokens associated to each leaf of the tree. This is much smaller than the total number of basis \( B^\lambda \) in \( D_\Lambda \), that grows exponentially with \( N \).
4.2.3 Wavelet and Cosine Packets

A basis $B^\lambda$ with oscillating atoms is defined using a separable cosine basis over each square of the dyadic segmentation. In this case no geometry is used, the oscillation of the atoms does not follow the geometry of the image, and $\tau = 1$. An approximation in an adapted cosine basis $B^\lambda$ allows one to capture the spatial variations of a texture [33].

A wavelet packet basis $B^\lambda$ defines a dyadic subdivision of the 2-D frequency domain [229]. The projection of an image on the atoms of $B^\lambda$ is computed through a pyramidal decomposition that generalizes the orthogonal wavelet transform, adding flexibility to overcome its dyadic frequency decomposition. Uniform dyadic wavelet packet decompositions generate a subset of $M$-band wavelets with equal-span frequency subbands obtained from $J$ decomposition levels, with $M = 2^J$. In order to adapt to the specific frequency content of the image, the resulting tree is parsed through a best basis selection procedure [226], reminiscent of the subdivision in Fig. 25.

This construction is generalized by considering non-stationary wavelet packets [230], that apply different quadrature mirror filters at each scale of the tree. A dynamic programming algorithm detailed in [231] computes an adapted non-stationary basis.

4.2.4 Adaptive Approximation

Wedgelets A geometric approximation is obtained by considering for each node of the dyadic segmentation a collection of $\tau$ different low-dimensional discontinuous approximation spaces [232]. For each node of the quadtree, a token indicates the local direction and position of the edge. The low-dimensional approximation spaces are piecewise polynomials over each of the two wedges.

The wedgelets introduced by Donoho [232] rely on piecewise constant approximation. This scheme is efficient when approximating a piecewise constant image $f$ whose edges are $C^2$ curves. For such cartoon images, the approximation error decays like $\|f - f_M\|^2 = M^{-2}$, see [232, 21]. It is also possible to consider approximation spaces with higher-order polynomials in order to capture arbitrary cartoon images [233], see also [234] for a related construction. The computation of the low-dimensional projection can be significantly accelerated, see [235].

The piecewise constant model for images being relatively simplistic, wedgelets have been upgraded to platelets [236] and surflets [237]. They aim at improving the management of smooth intensity variations, since they rely on planar or even smoother approximation on dyadic square or wedge based grids.

Bandlets For coding, orthogonal expansions are preferred over low-dimensional approximations as considered by wedgelets. Switching to non-linear approximation in bases also better handles directional textures that do not correspond to a fixed low-dimensional space parameterized by a wedge.

The bandlet bases dictionary is introduced by Le Pennec and Mallat [238]. Bandlets perform an efficient adaptive approximation of images with geometric singularities. An anisotropic basis with a preferred orientation is defined over each square of the dyadic segmentation. Fig. 26 (a) shows an example of bandlet atom. The orientation is parameterized with the token stored in the leaf of tree. Keeping only a few bandlet coefficients and setting the others to zero performs an approximation of the original image that follows the local orientation indicated by the token.
Adaptive Approximation over the Wavelet Domain  Applying such an adaptive geometric approximation directly on the image leads to unpleasant visual artifacts. In order to overcome this issue, one applies a tree-structured approximation or a best basis computation on the discrete set of wavelet coefficients. The wedgeprint of Wakin et al. [239] uses a vector quantization to extend the wedgelet scheme to the wavelet domain. The second generation bandlets of Peyré and Mallat [240] use an adaptive bandlet basis for each scale of the wavelet transform. All these methods benefit from the same approximation error decay as their single scale predecessors, but work better in practice.

Fig. 26 shows how a bandlet atom (a) is mapped to a wavelet-bandlet atom (b). Decomposing an image over a bandlet basis composed of atoms of type (b) is equivalent to applying first a wavelet transform, and then decomposing the wavelet coefficients over atoms of type (a).

Another adaptive approximation relying on the processing of the wavelet domain is the easy path wavelet transform (EPWT) [241]. It provides a hybrid and adaptive approach exploiting the local correlations of images along path vectors through index subsets in the Wavelet domain.

4.2.5 Adaptive Tree-structured Processing

For compression and denoising applications, one computes the best basis $B^\lambda$ adapted to the image $f$ to compress or denoise by minimizing the corresponding Lagrangian [23]. The coefficients $\langle \psi^\lambda_m, f \rangle$ are then binary coded (for compression) or thresholded (for denoising). The resulting improvement of the best basis approximation error over wavelets translates into improvement in the rate distortion (for compression) or average risk (for denoising) of the best basis method, see for instance [239, 240].

One can also use best bases to recover an image from noisy low-dimensional measurements $y = \Phi f + w$ where $\Phi$ is an ill-conditioned linear mapping. For some problems such as inpainting, small missing regions or light blur removal, the best basis $\lambda$ can be estimated directly from the observation $y$.

An example of inverse problem where sparsity in a best basis significantly improves over sparsity in a fixed basis is compressed sensing. Compressed sensing is a new data sampling strategy, where the measurement operator $\Phi$ of size $P \times N$ is generally the realization of some random matrix.
ensemble. The sampling operations $y = \Phi f + w \in \mathbb{R}^P$ allows one to acquire a high resolution signal $f \in \mathbb{R}^N$ directly in a compressed format of $P < N$ measurements. Compressed sensing theory ensures that if the number of measurements $P$ is large enough with respect to the sparsity $K$ of the signal $f$ in a basis $\mathcal{B}$, typically, $P = O(K \log N/K)$ for Gaussian random matrix $\Phi$, one recovers a good approximation of the signal using a $\ell_1^N$ sparse regularization as in (19). It can be shown that the quality of the reconstruction depends both on the sensing noise power $\|w\|$ and on the “compressibility” of $f$, that is, its deviation from the strictly sparse case. We refer to the review paper of Candès [242] and the references therein for more details. Fig. 27 shows a comparison of compressed sensing recovery from $P = N/6$ measurements using a redundant frame $\mathcal{B}$ of translation invariant wavelets, and a best bandlet basis. In this last result, it is necessary to use an iterative algorithm that progressively improves the quality of the estimated geometry, see [243]. As explained in this last reference, the same technique can be used for inpainting large holes in images.

4.2.6 Adaptive Segmentations and Triangulations

In order to enhance the quality of the representation, it is possible to consider tree-structured segmentations $[0, 1]^2 = \bigcup_{\beta \in \lambda} \beta$ of the image where the boundaries of the sub-domains $\beta \in \lambda$ are not restricted to be axis-aligned. The advantage is that such an adaptive segmentation defines regions $\beta \in \lambda$ with arbitrary complicated boundaries. Unfortunately, the combinatorial explosion of the set of all possible $\lambda$ forbids the search for an optimal segmentation with a fast algorithm. One has thus to use a greedy scheme that selects at each step a split to reduce the approximation error.

Recursive Splitting and Approximation Spaces A greedy scheme computes an embedded segmentation $\lambda = \{\lambda_j\}_j$, where $\lambda_{j+1} \subset \lambda_j$ is obtained by splitting a region $\beta \in \lambda_j$. The full segmentation $\lambda$ can thus be represented and coded using a binary tree. This defines multiresolution spaces $V_{\lambda_{j+1}} \subset V_{\lambda_j}$ where $V_{\lambda_j}$ is composed, for instance, of piecewise polynomial functions on each region $\beta \in \lambda_j$. 

Figure 27: (a,a’) original image ; (b) compressed sensing reconstruction using a translation invariant wavelet frame (PSNR=37.1dB) ; (c) reconstruction using a best bandlet basis (PSNR=39.3dB). (b’) wavelet frame, PSNR=22.1dB, (c’) bandlet basis, PSNR=23.9dB.
It is possible to compute a single-scale orthogonal projection \( f_M = P V_{\lambda_j} (f) \) of an image \( f \) on a fixed resolution space \( V_{\lambda_j} \) in order to perform image approximation or compression. It is also possible to define a detail space \( V_{\lambda_{j+1}} = V_{\lambda_j} \oplus W_{\lambda_j} \). A wavelet basis \( B^\lambda \) can be built by considering a basis for each \( W_{\lambda_j} \). A non-linear thresholding approximation \( f_M = H_T(f, B^\lambda, T) \) provides an additional degree of adaptivity and reduces the approximation error \( \| f - f_M \| \). Wavelet bases on adaptive segmentations also enable a progressive coding of the coefficients by decaying \( T \), which is important for image compression applications.

**Adaptive Segmentation** A popular splitting rule is the binary space tiling, that splits a region \( \beta \in \lambda_j \) according to a straight line, see for instance [244].

Other popular approaches restrict the regions \( \beta \in \lambda_j \) to triangles, so that \( \lambda_j \) is a triangulation of the domain \([0,1]^2\). It is possible to refine the triangulation by adding new vertices, or on the contrary to remove vertices to go from \( \lambda_{j+1} \subset \lambda_j \), so one cannot build a wavelet basis using such triangulations. These vertex refinement methods generate a single scale approximation \( P V_{\lambda_j} (f) \) and lead to efficient image coders, see for instance [245].

To generate embedded approximation spaces \( \lambda_{j+1} \subset \lambda_j \), one needs to split the triangles \( \beta \in \lambda_j \). Regular split of orthogonal triangles leads to isotropic adaptive triangulations [246]. Splitting triangles according to a well chosen median leads to anisotropic triangulations that exhibit optimal aspect ratio for smooth images, see [247]. More complicated, non-linear coding schemes are possible, for instance using normal meshes [248], that treat an image as an height field.

### 4.3 Lifting Representations

To enhance the wavelet representation, the wavelet filters can be adapted to the image content. The lifting scheme, popularized by Sweldens [249] and latent in earlier works [250, 251, 252], is an unifying framework to design adaptive biorthogonal wavelets, through the use of spatially varying local interpolations. While it can typically reduce the computation of the wavelet transform by a factor of about two in 1-D, it also guarantees perfect reconstruction for arbitrary filters, and can be used (Sec. 5.3) on non-translation invariant grids to build wavelets on surfaces, see Sec. 5.

#### 4.3.1 Lifting Scheme

At each scale \( j \), the scaling coefficients \( a_{j-1} \) are evenly split into two groups \( a^o_j \) and \( d^o_j \). The wavelet coefficients \( d_j \) and the coarse scale coefficients \( a_j \) are obtained by applying linear operators \( P^\lambda_j \) and \( U^\lambda_j \) parameterized by \( \lambda_j \)

\[
d_j = d^o_j - P^\lambda_j a^o_j \quad \text{and} \quad a_j = a^o_j + U^\lambda_j d_j.
\]

(24)

The resulting lifted wavelet coefficients \( \{d_j\}_j \) are thresholded or quantized to achieve denoising or compression. These two lifting or ladder steps are easily inverted by reverting the order of the operations. The predictor \( P^\lambda_j \) interpolates the sub-sampled values \( a^o_j \) in order to reduce the amplitude of the wavelet coefficients \( d_j \), while the update mapping \( U^\lambda_j \) stabilizes the transform by maintaining certain quantities such as the mean of the scaling coefficients. By applying sequentially several predict and one update operators, one can recover arbitrary biorthogonal wavelets on
Figure 28: (a) Predict and update lifting steps (b) MaxMin lifting of the Haar-Riesz Memorial plaque.

uniform 1-D grid [253], speeding up the wavelet decomposition algorithm by a factor of about two in 1-D. The lifting structure in Fig. 28(a) corresponds to the 5/3 lifted wavelet. Such structures may furthermore adapt to non-linear filters and morphological operations [254, 255]. An example of lifting based quincunx scheme example from [256, 257] is displayed in Fig. 28(b).

4.3.2 Adaptive Predictions

It is possible to design the set of parameter $\lambda = \{\lambda_j\}_j$ to adapt the transform to the geometry of the image. We call $\lambda_j$ an association field, since it typically links a coefficient of $a^0_j$ to a few neighboring coefficients in $d^0_j$. Each association is optimized to reduce, as much as possible, the magnitude of wavelet coefficients $d_j$, and should thus follow the geometric structures in the image. One can compute these associations to reduce the length of the wavelet filter near the edges, using the information from the coarser scale [258]. Locally adaptive schemes have proven efficient in stereo and video coding [259, 260, 261, 262]. Such schemes are related to adaptive non-linear subdivision [263]. To further reduce the distortion of geometric images, the orientations of the association fields $\{\lambda_j\}_j$ can be optimized though the scales. Because of the lack of structure of the set of bases $B^\lambda$, computing the field $\lambda_j$ that produces the best non-linear approximation is intractable. These flows are thus usually computed using heuristics to detect the local orientation of edges, see for instance [264, 265, 165, 266]. These adaptive lifting schemes are extended to perform adaptive video transforms where the lifting steps operate in time by following the optical flow $\lambda_j$, see for instance [267, 268].

4.3.3 Grouplets

A difficulty with lifted transforms is that they do not guarantee the orthogonality of the resulting wavelet frame. The stability of the transform thus tends to degrade for complicated association

\begin{equation}
\begin{aligned}
n = m - 2^{j-1} m \\
m + 2^{j-1} \\
3j - 1 \quad \text{(a)}
\end{aligned}
\end{equation}

\begin{equation}
\begin{aligned}
\mathcal{G}_{j+1} \cup C_j = \quad \bullet \quad \circ \\
\text{(b)}
\end{aligned}
\end{equation}
fields \( \{\lambda_j\}_j \). The grouplet transform, introduced by Mallat [269], also makes use of association fields, but it replaces the lifting computation of wavelet coefficients by an extended Haar transform, where coefficients in \( d_j^0 \) are processed in sequential order to maintain orthogonality.

Grouplets defined over each scale of the wavelet transform have been used to perform image denoising, super-resolution [269] and inpainting [270] by solving a \( \ell_1 \) regularization similar to (19). Grouplets can also be used to solve computer graphics problems such as texture synthesis. Classical approaches to texture synthesis use statistical models over a fixed representation such as a wavelet basis, see for instance [271, 272]. Building similar statistical models over a grouplet basis [270] allows one to better synthesize the geometry of some textures, and gives results similar to state of the art computer graphics approaches such as texture quilting [273]. Furthermore, the explicit parameterization of the geometry through the association fields \( \lambda \) allows the user to modify this geometry and synthesize dynamic textures. A comparison of these different approaches on one texture synthesis example is given in Fig. 29.

![Figure 29: Example of texture synthesis by statistical modeling of grouplet coefficients. (a) Exemplar. (b) Wavelet [272]. (c) Quilting [273]. (d) Grouplets [270].](image)

5 Transformations on Non-Euclidean Geometries

In this section we describe how the concepts of frequency, scale and even directionality have been extended to the processing of data on non-euclidean geometries like the sphere and other manifolds.

5.1 Data Processing on the Sphere

The unit sphere \( S^2 = \{ x \in \mathbb{R}^3 : \|x\| = 1 \} \subset \mathbb{R}^3 \) is one of the most natural non-Euclidean spaces. Very early, possibly due to influences for astronomy and geosciences, many data processing techniques have been developed for this surface. Many filtering, multiscale, directional and hierarchical methods have been designed, either in the spherical frequency domain induced by the spherical harmonics basis — often following the spirit of some Euclidean techniques exposed in the previous sections — or on the sphere itself thanks to some geometrical tools such as the stereographic dilation or the lifting schemes for wavelet analysis.
5.1.1 Filtering

As for the plane, filtering operations may be defined on $S^2$. Given the common two-angle spherical parameterization $\xi = (\theta, \varphi) \in S^2$ with the co-latitude $\theta \in [0, \pi]$ and longitude $\varphi \in [0, 2\pi)$, this operation is realized through spherical convolution evaluated on $\text{SO}(3)$ (the group of rotations in $\mathbb{R}^3$). For a function $f \in L^2(S^2) = \{ g : \|g\|^2_2 = \int_{S^2} |g|^2 < \infty \}$ and a filter $h \in L^2(S^2)$, the convolution is

$$(f * h)(\rho) = \int_{S^2} h(\rho \xi) f(\xi) \, d\mu(\xi),$$

where $\rho \in \text{SO}(3)$ is a rotation (driven by three angles) applied to the point $\xi \in S^2$ and $d\mu(\xi) = \sin \theta d\theta d\varphi$. For an axisymmetric filter, i.e., if $h(\xi) = h(\theta)$, the convolution reduces to $(f * h)(\xi') = \int_{S^2} h(\xi' \cdot \xi) f(\xi) \, d\mu(\xi)$, where $\xi' \cdot \xi$ is the common 3-D scalar product between $\xi'$ and $\xi$ seen as unit vectors.

5.1.2 Fourier Transform

The Fourier transform of a function $f \in L^2(S^2)$ is defined by

$$\hat{f}_{\ell m} = \langle Y_{\ell m}, f \rangle = \int_{S^2} Y^*_{\ell m}(\xi) f(\xi) \, d\mu(\xi), \quad f(\xi) = \sum_{\ell,m} \hat{f}_{\ell m} Y_{\ell m}(\xi)$$

with respect to orthonormal basis of spherical harmonics $\mathcal{Y} = \{ Y_{\ell m}(\xi) : \ell \geqslant 0, |m| \leqslant \ell \}$, i.e., the eigenvectors of the spherical Laplacian [274].

The frequency content of $f$ is thus represented by the value of $\hat{f}_{\ell m}$ on the order $\ell \in \mathbb{N}$, which basically counts the number of oscillations on the latitudes, and the moment $m \in \{-\ell, \ldots, \ell\}$ counting longitude oscillations. Numerically, only certain discretizations of the sphere can provide perfect quadrature formulae to compute the Fourier coefficients of band-limited functions on the sphere, sometimes with very efficient algorithms [274, 275].

5.1.3 Spherical Scale-Space

Similarly to what happened for signals or images, the first notion of “scale” on the sphere was imported from the Heat Dynamic that is also known on this space. In that framework, if a spherical function $f \in L^2(S^2)$ is considered the initial heat configuration, the spherical heat dynamics smooth it with time $\tau > 0$, conferring a scaling notion on this parameter.

Interestingly, as for Euclidean spaces, the solution at time $\tau > 0$ of the heat equation initialized to some function $f \in L^2(S^2)$ is simply $f(\xi, \tau) = \sum_{\ell,m} \hat{f}_{\ell m}(\tau)Y_{\ell m}(\xi)$, with $\hat{f}_{\ell m}(\tau) = \hat{f}_{\ell m} e^{-\ell(\ell+1)\tau}$ and $f(\xi,0) = f(\xi)$. Alternatively, since for an axisymmetric filter $h$ we have the spherical convolution theorem

$$(f * h)_{\ell m} = \sqrt{\frac{4\pi}{2\ell+1}} \hat{f}_{\ell m} \hat{h}_{00},$$

the solution of the Heat Equation can also be obtained by a convolution by a specific kernel $G^0_\tau(\xi)$, coined spherical Gaussian of width $\sqrt{\tau}$. It is defined in frequency by $(G^0_\tau)_{\ell m} = \sqrt{(2\ell + 1)/4\pi} e^{-\ell(\ell+1)\tau}$.

The link between the heat dynamics and the spherical convolution with the axisymmetric filter $G^0_\tau$ has been exploited by Bülow [276] to develop several specific spherical filters for feature detection, such as the Laplacian of Gaussian or the directional derivative of Gaussian.
5.1.4 Spectral Wavelets

Freeden et al. [277, 278] have fully exploited the connection between convolution and frequency filtering on the sphere to develop a continuous wavelet transform on the sphere. This is done by introducing a family of axisymmetric functions $\psi_{a}(\xi)$, coined spherical wavelet, continuously indexed by $a > 0$, and such that $\int_{\mathbb{R}^{+}} |(\psi_{a})_{00}|^{2} \, da/a = 1$, $(\psi_{a})_{00} = 0$, plus additional regularity conditions. The wavelet coefficients of a function $f \in \mathbb{L}^{2}(S^{2})$ are then defined as $W_{f}(a, \xi) = (f * \psi_{a})(\xi)$. The reconstruction is possible (almost everywhere) by

$$f(\eta') = \langle f \rangle + \int_{\mathbb{R}^{+}} \int_{S^{2}} W_{f}(a, \xi) \psi_{a}(\eta' \cdot \xi) \, \frac{da}{a} \, d\xi,$$

with $\langle f \rangle = \frac{1}{4\pi} \int_{S^{2}} f(\xi) \, d\mu(\xi)$.

In [277, 278], an MRA on the sphere is also built by defining Quadrature Mirror Filters in the frequency domain. A spatial sub-sampling of the different subspaces of the MRA can also decrease the redundancy of the basis hence created.

Following a similar approach, (isotropic) needlet frames introduced in [279, 280, 281] represent another example of spectral wavelets, i.e., wavelets shaped in the Fourier domain. Needlets additionally offer relationships with quadrature formulae used to turn integrals of bandlimited functions into discrete summations.

5.1.5 Stereographic Wavelets

In the previous sections, the notion of scale in the processing of spherical data was always defined in the frequency domain, i.e., by dilating the frequency domain by a parameter, preventing a fine control of the spatial support of the filter.

An alternative approach introduced by Antoine and Vandergheynst [282, 283] defines the dilation directly in the spatial domain. The compactness of $S^{2}$ is respected, by introducing a stereographic dilation. As illustrated on Fig. 30(a) for point dilation, the stereographic dilation $D_{a}$ of a function $g \in \mathbb{L}^{2}(S^{2})$ amounts to projecting $g$ on the plane tangent at the North Pole by the stereographic projection $\Pi$, to applying there a Euclidean dilation $d_{a}$ by a scale $a > 0$, and to lifting the resulting function back to the sphere by $\Pi^{-1} [284]$. Mathematically, $[D_{a}g](\theta, \varphi) = \lambda(a, \theta) \, g(\theta/a, \varphi)$, with $\tan \theta/a = a \tan \theta/2$ and where $\lambda$ is a normalizing function such that $\|D_{a}g\|_{2} = \|g\|$

Given a mother wavelet $\psi \in \mathbb{L}^{2}(S^{2})$ centered on the North pole, the proposed approach considers the joint action of translations, i.e., rotation operators $R_{\rho}$ in $\text{SO}(3)$, and of the dilations $D_{a}$ on $\psi$. The wavelet transform of $f$ is therefore:

$$W_{f}(\rho, a) = \langle \psi_{(\rho, a)}, f \rangle, \quad \rho \in \text{SO}(3), \ a > 0,$$

with $\psi_{(\rho, a)} = R_{\rho}D_{a}\psi$. If the wavelet is admissible, which is nearly equivalent to impose $\int_{S^{2}} d\mu(\theta, \varphi) \frac{\psi(\theta, \varphi)}{1 + \cos \theta} = 0$, the reconstruction of $f$ is possible through

$$f(\xi) = \langle f \rangle + \int_{\mathbb{R}^{+}} \int_{\text{SO}(3)} \frac{d\mu(\rho)}{a^{3}} \ W_{f}(\rho, a) \ [R_{\rho}L_{\omega}^{-1}D_{a}\psi](\xi),$$

where $\nu$ is the Lebesgue measure on $\text{SO}(3)$ and $L_{\omega}$ is a multiplicative operator function of $\omega$ only and expressed in the Fourier domain [282]. For axisymmetric wavelets, this result simplifies by the fact that the action of $R_{\rho}$ on $\psi$ is controlled by two angles only.
Figure 30: (a) Stereographic dilation on $S^2$. On the right, the (steerable) second directional derivative of Gaussian. The three images (b)-(d) are the basis elements, while the fourth in (e) is a linear combination of the first three yielding a rotation of $\pi/4$ around the North pole.

Many wavelets may be defined on the sphere since it has been proved in $[284]$ that any admissible wavelet on the plane $L^2(\mathbb{R}^2)$ can be imported by inverse stereographic projection $\Pi^{-1}$. A Laplacian of Gaussian (LoG), difference of Gaussians (DoG), Morlet Wavelet, and many other are generally used $[282, 285, 286]$. Numerically, this spherical CWT is obtained thanks to the convolution theorem mentioned previously. This transform has been for instance intensively used in the analysis of the Cosmic Microwave Background (CMB), an astronomical signal remnant of some specific evolution phase of the Big Bang $[287, 288, 289]$.

Wavelet frames can be developed in this theory by discretizing the scaling parameter $a$ $[285]$. These frames, that do not subsample the spherical positions, have successfully served for the construction of invertible filter banks on the 2-Sphere $[290]$ even if the stereographic dilation is not really compatible with the frequency description of the wavelets.

5.1.6 Haar Transform on the Sphere

The constructions of spherical wavelets described in the previous section make use of the Fourier decomposition on the sphere. It is possible to define wavelets directly over the spherical domain without Fourier analysis, using for instance the lifting scheme method $[291]$, see Sec. 5.3. This allows one to define spherical wavelets with a compact support, although the stability of the resulting transform is more difficult to control than over the planar domain.

Inspired by this lifting scheme $[291]$, one can easily define a Haar basis on the sphere by considering a family $\{\mathcal{M}^j\}_{j \leq 0}$ of embedded spherical triangulations that approximate a sphere $[292]$. These triangulations are obtained by a regular 1:4 refinement rule starting from an initial regular polyhedron $\mathcal{M}^0$, and the edges are projected on the sphere to define spherical triangles.

The corresponding spherical multiresolution defines $V_j \subset L^2(S^2)$ as the set of functions that are constant on each triangle of $\mathcal{M}^j$. Figure $31$ shows the linear projection of a spherical function on some of these multiresolution spaces.
Following the usual definition (Sec. 2.3.1), a Haar wavelet basis \( \{ \psi_{j,n} \}_n \) is an orthogonal basis of the detail space \( W_j \) such that \( V_{j+1} = V_j \oplus W_j \). The wavelet coefficients \( \langle \psi_{j,n}, f \rangle \) are computed using a pyramid algorithm that mimics the usual Haar transform, except that for each triangle, one gathers three detail coefficients and one coarse scale coefficient. Figure 32 shows these Haar coefficients together with a comparison between spherical and planar non-linear approximations \( H_T(f) \).

Figure 32: Comparison of spherical and planar Haar approximations. The threshold \( T \) is adjusted so that \( H_T(f) \) is an approximation with a number of coefficients equal to 5% of the number of pixels in the high resolution planar image.

5.1.7 Steerable Wavelets on the Sphere

Finally, the sphere is compatible with the definition of steerable filters similarly to those defined in Sec. 3.2.2 for the plane. In particular, using the stereographic projection \( \Pi \) introduced in the previous section, steerability on the sphere is also imported from the plane. This fact has been used in [284, 293, 294] to define differential and steerable filters useful to detect directional features in the Cosmic Microwave Background. An example of a steerable wavelet is given in Fig. 30(b-e). Spherical steerability may also be directly studied in the frequency domain with spectral dilation [295].
Other Constructions

It is impossible to cite the vast literature on multiscale decomposition on the sphere. Let us just quote some of them. Wavelets, ridgelets and curvelets have been translated on the sphere by Starck et al. [296] by using a particular spherical sampling, called HEALPix, locally similar to a square discretization. Locally supported biorthogonal wavelet bases have also been realized thanks to some radial projections of the planar faces of a cube on $S^2$ in [297].

Wavelets on General 2-Manifolds

Given a two-dimensional manifold $M$, i.e., locally isomorphic to $\mathbb{R}^2$, authors in [298] describe how to define a Continuous Wavelet Transform (CWT) for function $f : M \rightarrow \mathbb{C}$.

Similarly to the way the stereographic dilation is defined for the sphere, the local dilation of a function $\psi$ around the point $\xi \in M$ relies on the knowledge of a local and invertible projection $\Pi_\xi$ on $M$. The desired dilation of scale $a > 0$ therefore factorizes as $D(\xi, a) = \Pi_\xi^{-1} d_a \Pi_\xi$ with $d_a$ the common Euclidean dilation of function in $T_\xi M \cong \mathbb{R}^2$.

Given the Hilbert space $\mathcal{H} = L^2(M, d\mu)$ of square integrable function on $M$, for a proper measure $d\mu$, the CWT of a function $f$ on $M$ is then formally defined by correlating $f$ with a set of prototype wavelets $\psi(\xi) \in \mathcal{H}$ localized around any $\xi \in M$, i.e.,

$$W_f(\xi, a) = \langle \psi(\xi, a), f \rangle_{\mathcal{H}} \triangleq \int_M d\mu(\xi') f(\xi') \psi(\xi, a)(\xi'), \quad \psi(\xi, a) = D(\xi, a) \psi(\xi).$$

The theoretical invertibility of this transform has however to be studied specifically in each case, i.e., given $M$ and $\Pi_\xi$. Results exist for instance for the two-sheeted hyperboloid and the paraboloid in $\mathbb{R}^3$ [299].

Lifting Scheme Wavelets on Meshed Surfaces

The lifting scheme of Sweldens [300], described in Sec. 4.3, can be used to define wavelets on non-translation invariant geometries, including surfaces with complicated topologies. Lifted wavelets on surfaces are usually built on a semi-regular mesh grid, was first considered by Lounsbery et al. [301], and then refined within the lifting framework by Schröder and Sweldens [291].

Semi-regular meshes $\{M^j\}_{j \geq 0}$ are obtained by a regular 1:4 refinement rule starting from an arbitrary control mesh $M^0$. Each edge of $M^j$ is split into two sub-edges by vertex insertion to obtain the refined mesh $M^{j-1}$. The fine mesh $M^j$ is the sampling grid that stores the position of the surface points in space, and a signal $f$ sampled at each grid point. Fig. 33, top row, shows an example of such a multiresolution mesh, obtained by a semi-regular remeshing of a high resolution input mesh.

The lifting scheme described in Sec. 4.3 can be applied by storing the scaling coefficients $a_j$ on the grid point of the mesh $M^j$, while the detail coefficients are stored on the complementary detail grid $D^j$ where $M^{j-1} = M^j \cup D^j$. The splitting of $a_{j-1}$ into $a_j^j$ and $d_j^j$ corresponds to assigning the values stored in $M^{j-1}$ to either $M^j$ or $D^j$. The predict operator $P_j$ used to compute the wavelet coefficients $d_j$ stored in $D^j$ is a local polynomial interpolator on a triangulation grid. The update operator $U_j$ is computed by solving a linear system, to impose that moments of low orders, such as the mean, are preserved when moving from $a_{j-1}$ to $a_j$. 

43
This lifting wavelet transform computes the coefficients $d_j[n] = \langle \psi(j,n), f \rangle$ for all scales $0 < j < J$ and grid points $n \in D^j$. It corresponds to the projection of the signal $f$ defined on the triangulated surface $M^j$ onto a discrete biorthogonal wavelet frame $\mathcal{B} = \{\psi(j,n)\}_{j,n}$. These coefficients can be thresholded, and inverting the lifting steps creates an approximated signal $f_M$ with $M$ non-zero coefficients. Although this approach works well in practice, the frame bounds of the resulting wavelet frame $\mathcal{B}$ are difficult to control, and $f_M$ might be far from the best $M$-terms approximation. It is also difficult to guarantee the convergence of the wavelet atoms $\psi(j,n)$ to smooth functions, when $J$ tends to $-\infty$, and the mesh $M^J$ approximates a smooth surface.

To perform surface approximation, one defines the signal $a_J$ at the finest scale as the position of the nodes on the surface. Each coefficient $a_J[n] \in \mathbb{R}^3$ is thus a point in 3D space. The lifting transform can be applied to this vector-valued signal. Thresholding the resulting wavelet coefficients allows one to approximate the surface using few coefficients, as shown on Fig. 33, bottom row. If the lifting operators $P_j$ and $U_j$ do not depend on the position of the points on the surface, the resulting lifting wavelets can be used to perform 3D mesh compression [301, 291].

Figure 33: Top row: example of semi-regular mesh $\{M^j\}_j$. Bottom row: example of surface approximation $f_M$ obtained by thresholding the lifted wavelets coefficients, where $N$ is the number of vertices in $M^j$. (a) $M^j, j = -4$. (b) $M^j, j = -5$. (c) $M^j, j = -6$. (d) $M^j, j = -7$. (e) $M/N = 100\%$. (f) $M/N = 10\%$. (g) $M/N = 5\%$ (h) $M/N = 2\%$. 

44
5.4 Wavelets on Graphs

Let us finally mention that wavelet transform has been extended to functions defined on the vertices of an arbitrary finite weighted graph. The latter may for instance generalize standard picture definition by describing two-dimensional pixel adjacencies. Maggioni et al. introduced “diffusion wavelets” [302], a general theory for wavelet decompositions based on compressed representations of powers of a diffusion operator such as the graph Laplacian. The constructed wavelet basis is made orthogonal by combining graph subsampling and Gram-Schmidt orthogonalization on each subsampled space.

More recently, Hammond et al. [303] developed a general wavelet frame theory on such graphs thanks to the graph analogue of the Fourier domain, namely the spectral decomposition of the discrete graph Laplacian. Wavelets are defined in this frequency domain by dilating an “admissible” generating kernel. The final representation is redundant but wavelets can be shaped by changing the generating kernel. Moreover, for sparse graph Laplacian matrix, a fast wavelet transform avoiding the Laplacian spectral decomposition is developed.

6 Conclusion

A century after the discovery by Alfréd Haar, and twenty years after the emergence of wavelets as genuine processing tools, major advances have been made in the improvement of natural images representations, aiming at enhanced understanding.

Their common characteristic resides in uncovering multiscale and oriented features of natural images, through projections on a specific set of elongated atoms. The resulting dictionaries are thus often redundant, and may be coupled with sparsity enforcing priors, or adaptivity. They reveal a striking similarity with low level vision, where similar strategies are used to build powerful processing architectures.

The availability of such a large number of transformations, that potentially extend the standard wavelet framework, leaves open the question of the best representation to process a given image. This choice is unfortunately data dependent, since the geometry of edges and textures varies significantly from natural to seismic or medical images. Selection of a representation, as well as its parameterization (number of scales, span of orientations, support in space or frequency), is also application dependent, and applications to inverse problems or pattern recognition typically impose strong design requirements on the dictionary. Their exhaustive comparison thus remains out of reach, with traditional methods from image processing or approximation theory only providing a partial answer.

As a humble contribution to a subjective comparison, additional materials, full scale decomposition images, related links and associated toolboxes necessary to reproduce illustrations provided in this paper are available at [26]. Oddly enough, a common etymology of Szeged resides in an old Hungarian word for corner (szép). At a turn in a wavelet century, A. Haar and F. Riesz might not have foreseen the harvest from their mathematical seeds. Image understanding is at the beginning of reaping their fruits.
Acknowledgements

Laurent Jacques is a postdoctoral researcher funded by the Belgian National Science Foundation (F.R.S.-FNRS). Professor Károly Szatmáry is warmly acknowledged for providing us with the recurrent picture of the Haar-Riesz Memorial plaque. We warmly thank Pedro Corea (ICTEAM-TELE, UCL), Jérôme Gauthier (CEA) and Grégoire Hertz (Supéléc) for their thorough proofreading. The four authors are also very grateful to the anonymous reviewers whose important remarks and suggestions have greatly improve the quality of this paper. They finally are indebted to Jean-Pierre Antoine, Nick Kingsbury, François Meyer, Ivan Selesnick and guest editors Thierry Blu, Jean-Christophe Pesquet and Truong Q. Nguyen for their constructive and insightful discussions and comments.

References


54


