Wave Atoms and Sparsity of Oscillatory Patterns

Laurent Demanet† and Lexing Ying‡

† Department of Mathematics, Stanford University, Stanford CA94305
‡ Department of Mathematics, University of Texas at Austin, Austin, TX 78712

June 2006, revised February 2007

Abstract

We introduce "wave atoms" as a variant of 2D wavelet packets obeying the parabolic scaling wavelength $\sim$ (diameter)^2. We prove that warped oscillatory functions, a toy model for texture, have a significantly sparser expansion in wave atoms than in other fixed standard representations like wavelets, Gabor atoms, or curvelets. We propose a novel algorithm for a tight frame of wave atoms with redundancy two, directly in the frequency plane, by the "wrapping" technique. We also propose variants of the basic transform for applications in image processing, including an orthonormal basis, and a shift-invariant tight frame with redundancy four. Sparsity and denoising experiments on both seismic and fingerprint images demonstrate the potential of the tool introduced.

Keywords. Wave Atoms, Image Processing, Texture, Oscillatory, Warping, Diffeomorphism.

Acknowledgments. We would like to thank Emmanuel Candès for valuable discussions, advice and generous support. L. D. is supported by a NSF grant CCF-0515362. L. Y. is supported by a Department of Energy grant DE-FG03-02ER25529. We also thank Eric Verschuur and Felix Herrmann for providing seismic image data, Chris Brislawn for allowing us to use a fingerprint image, and the referees for suggestions that helped improve the manuscript.

1 Introduction

It has been known for some time that oscillatory textures sometimes enjoy reasonably sparse expansions in wave packet bases of various kinds, such as local cosines [32], brushlets [27], Gabor atoms [26], wavelet packets [25] complex wavelets [22, 19, 29], or extensions thereof, like dual-tree M-band wavelets [11, 12]. The rationale for these empirical successes is that good basis elements should look like the function they are trying to approximate, i.e., they should oscillate somewhat. A quantitative version of this statement would be desired, not only for its own sake, but because it could feed back instructive guidelines on the design of the "right" wave packets for oscillatory textures.

An example of quantitative success story is curvelets, which were designed for the purpose of providing a sparse expansion of curved edges [8, 9], bandlimited wavefronts [7] and other geometric features in images. Much is known about the approximation theory for curvelets [9]. Various alternative transforms have since then been proposed for representing edges in images, often with their own performance analysis as well: contourlets [17], bandelets [28], shearlets [24].

As for texture representation however, the validation question is still open. Little is known on, say, the decay of Gabor coefficients of "texture," for the good reason that defining complicated image patterns mathematically has proved to be quite difficult.

There is, however, a model that lends itself to analysis. In this paper we propose to study *warped oscillatory patterns* in different wave packet bases. For such images we show that the sparsity discussion reduces to a well-defined trade-off between competing features of the wave packets, namely

- the ability to adapt to arbitrary local directions of a pattern, i.e., warpings; and
- the ability to sparsely represent anisotropic patterns aligned with the axes.

We show that a most advantageous compromise is achieved in some basis of wave packets that we term *wave atoms.* In a nutshell, wave atoms "interpolate" exactly between directional wavelets [2] and Gabor [26], in the sense that the period of the oscillations of each wave packet (wavelength) is linked to the size of the essential support (diameter) by the parabolic scaling

$$\text{wavelength} \sim (\text{diameter})^2.$$

(for wavelets, diameter $\sim$ wavelength, and for Gabor, diameter $\sim$ wavelength^0 = const.) We argue that wavelets, Gabor atoms and curvelets all yield worse asymptotic rates than wave atoms for the model considered.

The name "wave atoms" comes from the property that they also provide an optimally sparse representation of wave propagators, a mathematical result of independent interest, with applications to fast numerical solvers for wave equations [16].

1.1 Classification of Phase-Space Tilings

To make our discussion concrete, we need to classify various wave-packet transforms as *phase-space tilings*. In what follows we review the essential features of wavelets, Gabor, ridgelets, curvelets, wave atoms, and explain how they all fit in a single picture. Interesting new transforms could possibly emerge from a careful study of that picture.

We denote 2D wave packets as $\varphi_\mu(x_1, x_2)$, with index μ. Since a complete collection must "span" all positions and frequencies, we see that wave packets are actually tiles in phase-space[1]. Some tilings are more interesting than others. We say a tiling is universal if it treats democratically all positions and orientations. In that case,

- the geometry of the tiling in space must be Cartesian, or approximately so; and
- the geometry of the tiling in frequency must be polar, or approximately so.

Universality as above suggests that two parameters should suffice to index a lot of known wave packet architectures: α to index whether the decomposition is multiscale ($\alpha = 1$) or not ($\alpha = 0$); and β to indicate whether basis elements are localized and poorly directional ($\beta = 1$) or, on the contrary, extended and fully directional ($\beta = 0$).

In terms of phase-space localization of the wave packets, we will require that

[1] Phase-space is the set of all positions x and frequencies ω.

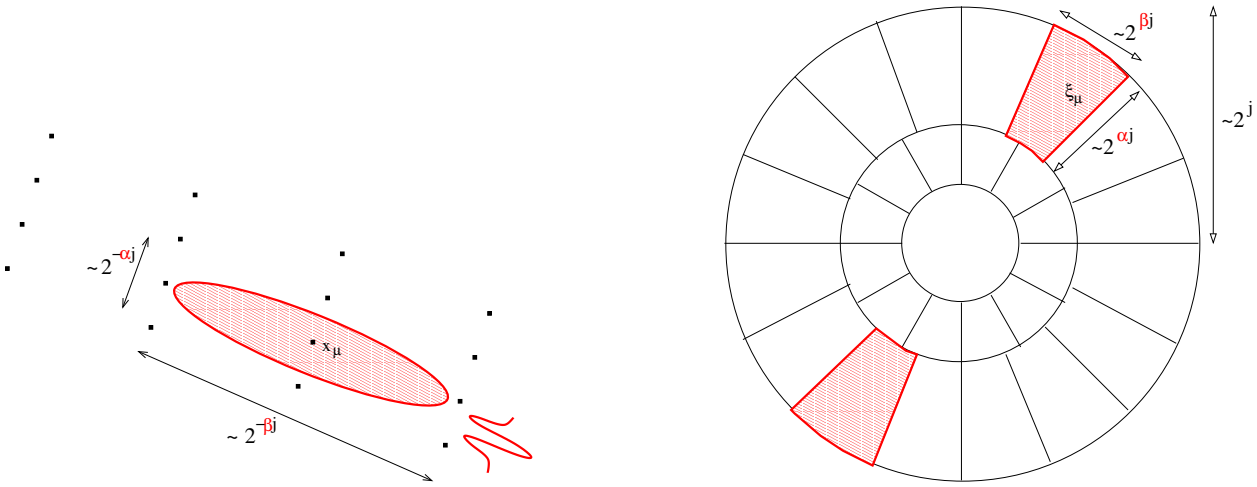

Figure 1: Essential support of a wave packet with parameters (α, β), in space (left), and in frequency (right). The parameter α indexes the multiscale nature of the transform, from 0 (uniform) to 1 (dyadic). The parameter β measures the wave packet's directional selectivity, from $\beta = 0$ (best selectivity) to $\beta = 1$ (poor selectivity). Wave atoms are the special case $\alpha = \beta = 1/2$. On the left, the dots indicate the grid over which wave packets corresponding to the same wave number should be translated. On the right, the different wedges correspond to wave vectors at different scales and angles.

- in x, the essential support of $\varphi_\mu(x)$ is of size $\sim 2^{-\alpha j}$ vs. $2^{-\beta j}$ as scale $j \geq 0$, with oscillations of wavelength $\sim 2^{-j}$ transverse to the ridge; and
- in frequency ω, the essential support of $\hat{\varphi}_\mu(\omega)$ consists of two bumps, each of size $\sim 2^{\alpha j}$ vs. $2^{\beta j}$ as scale j, at opposing angles and distance $\sim 2^j$ from the origin.

These conditions are perhaps more easily grasped from Figure 1. These are the requirements we put on each *single* wave packet. Then the *collection* of wave packets, corresponding to a choice of α and β, should be indexed by various wave numbers and positions (put together in the single index μ), so that the tiling of phase-space is compatible with the estimates above.

We hope that a description in terms of α and β will clarify the connections between various transforms of modern harmonic analysis. Curvelets [8] correspond to $\alpha = 1$, $\beta = 1/2$, wavelets (including MRA [26], directional [2], complex [29]) are $\alpha = \beta = 1$, ridgelets [4] are $\alpha = 1$, $\beta = 0$, and the Gabor transform is $\alpha = \beta = 0$. *Wave atoms are defined as the point* $\alpha = \beta = 1/2$. The situation is summarized in Figure 2.

1.2 Trade-off: Warping Invariance vs. Directionality

Let us consider the following very simple model, a toy for anisotropic textures. We say that a function is an *oscillatory pattern* if it is the image under a smooth diffeomorphism of a function that oscillates only in one direction, say along the coordinate x_1. Mathematically, we can put $x = (x_1, x_2)$ and formulate our model as

$$f(x) = \sin(Ng(x))\, h(x), \tag{1}$$

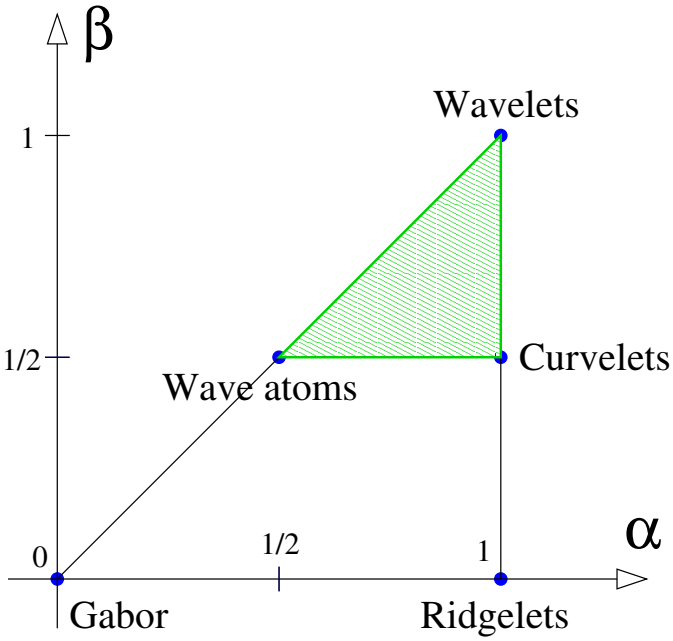

Figure 2: Identification of various transforms as (α, β) families of wave packets. The triangle formed by wavelets, curvelets and wave atoms indicates the wave packet families for which sparsity is preserved under diffeomorphisms. The range of allowable transforms presumably extends beyond some of the limits shown.

where both g and h are C^∞ scalar functions, h is assumed to be compactly supported inside $[0,1]^2$, and N is a large constant. We are idealizing a digital image as a function, yet we should think of N as smaller or comparable to the number of pixels of the underlying image in some direction, either x_1 or x_2.[2] Of course, using cosine instead of sine in (1) would work just fine.

Equation (1) may be too restrictive to qualify as a fully-fledged texture model, but it is nonetheless adequate in a variety of physically relevant contexts, like seismic imaging and astronomy. More details can be found in Sections 1.3 and 4.

Let us return to the question of finding an adequately sparse representation for functions in the class defined by (1), for example in the sense of membership of the coefficients in some ℓ_p space, for p small. We will assume that g, h and N are unknown to us, as we are not concerned with adaptive procedures aimed at determining them from the (possibly noisy) data.

- First, we require that *sparsity be preserved under warpings, i.e., smooth diffeomorphisms.* This corresponds to the reasonable requirement that compressibility should not depend on the choice of local coordinates. It is a somewhat deep result that smooth diffeomorphisms are bounded on every ℓ_p space, $p > 0$, provided the frame of wave packets obeys $1/2 \le \beta \le \alpha$, in the classification introduced above. This restriction corresponds to the shaded triangle in Figure 2. Intuitively, for $\beta \ge 1/2$ the spatial extend of each wave packet is small enough that

[2] Bandlimited functions oscillating with local wavelength of order $1/N$ can be correctly discretized on an N-by-N grid, by the Shannon sampling theorem, hence the heuristic about the number of pixels. Our function f is not bandlimited, so sampling will not be exact, but the heuristic is nevertheless relevant.

warping due to the diffeomorphism is under control.

- Second, we also wish to *maximize sparsity of template oscillatory profiles*, from which the more general functions can be obtained by warping. In the example (1), this means studying $\sin(Nx_1)h(x)$, a function oscillating only in the x_1 direction. The Fourier transform, or Gabor atoms are ideal for this task; it is easy to see that they each only require a constant number of coefficients (independent of N) to approximate $\sin(Nx_1)h(x)$ to fixed accuracy. Gabor atoms correspond to the point $\alpha = \beta = 0$ in Figure 2.

The two above requirements are contradictory, but a good compromise is the point $\alpha = \beta = 1/2$ corresponding to wave atoms. We prove the following result in Section 2.

Theorem 1.1. *Let f be of the form (1) for some N, with g and h of class C^∞, and h compactly supported inside $[0,1]^2$. Then f can be represented to accuracy ϵ in L^2 by the largest $\mathbf{C}_\epsilon \mathbf{N}$ wave atom coefficients in absolute value, where for all $M > 0$, there exists $C_M > 0$ such that $C_\epsilon \leq C_M \epsilon^{-1/M}$.*

In other words, $O(N)$ wave atom coefficients suffice to represent f to some given accuracy. In contrast, as we will argue, we would need $O(N^{3/2})$ curvelets coefficients; or $O(N^2)$ wavelet coefficients; or $O(N^2)$ Gabor coefficients to represent f up to the same accuracy.

Notice that wave atoms are a fixed transform and that the knowledge of N is not needed to define the transform.

1.3 Importance

Oscillatory patterns are not only ubiquitous in natural images, but also in some inverse imaging problems involving oscillatory data.

The seismograms, or codas, from reflection seismology are the single most relevant example of oscillatory data that we have in mind. Seismology is concerned with imaging the interior of the Earth, both for pressing scientific and industrial purposes. The inverse problem amounts to finding the physical parameters (like index of refraction of the different geological layers) from petabytes[3] of noisy reflected wave measurements at the surface (like earthquake waves). Both the data and the unknown physical parameters are oscillatory, in the manner modeled by equation (1). There are by now serious indications that several tasks could be realized more efficiently if an adequate basis were available, like formulating Bayesian priors for various denoising, recovery and data separation problems [20, 21]; speeding up numerical computation of wave propagation and Generalized Radon Transforms [16]; compressing/downsampling seismic databases; redefining seismic migration; etc. Stylized compression and denoising experiments on a synthetic seismic dataset are presented in Section 4.

Astronomy and astrophysics also comes with its share of interesting oscillatory data. For instance, the spectacular pictures of the sun acquired by the TRACE satellite seem ideal to help understand the physics behind formation of the sun's coronal magnetic loops. Any study of this kind should start with the proper detection, extraction, or even parameterization of these loops, for which, to our knowledge, the appropriate tools are still missing [1].

Finally, the Federal Bureau of Investigation (FBI) is currently digitizing their fingerprint image database using a wavelet scheme with scalar quantization [3]. We believe that fingerprint images

[3] 1 petabyte = 10^15 bytes.

fall into the category of oscillatory patterns considered here; hence wave atoms coupled with a clever quantization scheme may be of interest as a novel compression tool. See also the numerical experiments in Section 4.

1.4 Organization

The rest of the paper is organized as follows. In Section 2, we discuss and prove the efficiency of wave atoms for the sparse representation of oscillatory patterns. In the light of that discussion, Section 3 presents our choice of implementation for wave atoms. Finally, in Section 4, we present elementary image compression and denoising experiments, related to the areas of application evoked above.

2 Sparse Representation of Oriented Textures

2.1 Definition of Wave Atoms

In what follows we restrict the exposition to two dimensions, and write $x = (x_1, x_2)$, although this restriction is not essential. Our convention for the 2D Fourier transform is

$$\hat{f}(\omega) = \int e^{-ix\cdot\omega} f(x)\, dx, \qquad f(x) = \frac{1}{(2\pi)^2} \int e^{ix\cdot\omega} \hat{f}(\omega)\, d\omega.$$

We write wave atoms as $\varphi_\mu(x)$, with subscript $\mu = (j, \mathbf{m}, \mathbf{n}) = (j, m_1, m_2, n_1, n_2)$. All five quantities j, m_1, m_2, n_1, n_2 are integer-valued and index a point (x_μ, ω_μ) in phase-space, as

$$x_\mu = 2^{-j}\mathbf{n}, \qquad \omega_\mu = \pi 2^j \mathbf{m}, \qquad C_1 2^j \le \max_{i=1,2} |m_i| \le C_2 2^j. \tag{2}$$

where C_1, C_2 are two positive constants left unspecified for convenience, but whose values will be implied by the specifics of the implementation. Heuristically, the *position vector* x_μ is the center of $\varphi_\mu(x)$ and the *wave vector* ω_μ determines the centers of both bumps of $\hat{\varphi}_\mu(\omega)$ as $\pm\omega_\mu$. Note that the range of $\mathbf{m}$ needs to be further reduced to $m_2 > 0$, (or $m_2 = 0$ and $m_1 > 0$,) to account for the central symmetry of the Fourier transform of real-valued functions about the origin in ω. Some further restriction on $\mathbf{n}$ (cutoff in space) and j (cutoff in scale), are of course necessary in practice, but not for the description of a frame of $L^2(\mathbb{R}^2)$.

Wave atoms then need to obey a localization condition around the phase-space point (x_μ, ω_μ).

Definition 2.1. *(Wave Atoms) Let x_μ and ω_μ be as in equations (2) for some C_1, C_2. The elements of a frame of wave packets $\{\varphi_\mu\}$ are called* wave atoms *when*

$$|\hat{\varphi}_\mu(\omega)| \le C_M \cdot 2^{-j}(1 + 2^{-j}|\omega - \omega_\mu|)^{-M} + C_M \cdot 2^{-j}(1 + 2^{-j}|\omega + \omega_\mu|)^{-M}, \qquad \textit{for all } M > 0, \tag{3}$$

and

$$|\varphi_\mu(x)| \le C_M \cdot 2^{j}(1 + 2^{j}|x - x_\mu|)^{-M}, \qquad \textit{for all } M > 0. \tag{4}$$

It is of course possible to restrict the decay order or even moderately alter the definition of x_μ and ω_μ—and still call the basis functions "wave atoms"—but this is a refinement we will not address here.

The parabolic scaling is encoded in the localization conditions as follows: at scale 2^{-2j}, or frequency $|\omega_\mu| \sim 2^{2j}$, the essential frequency support is of size $\sim 2^j$ (for each bump) and the essential spatial support is of size $\sim 2^{-j}$. The oscillations within the envelope of a wave atom in x have wavelength $\sim 2^{-2j}$. Note that the subscript j indexes the different "dyadic coronae," whereas the additional subscript $\mathbf{m}$ labels the different wave numbers ω_μ within each dyadic corona.

Our choice of indexing wave atoms using j and $\mathbf{m}$ may seem cumbersome, but it follows closely the standard notation for wavelet packets. See Section 3 and Chapter 8 in [26].

2.2 Sparsity: Heuristics

Theorem 1.1 states that an oscillatory object such as $f(x) = \sin(Ng(x))h(x)$, where both g and h are non-oscillatory, can be represented to high accuracy with $O(N)$ wave atom coefficients. This result can be understood intuitively as follows.

The scale index j can take on any positive integer value, but the scale of oscillations of f is about $j^* = \frac{1}{2}\log_2 N$. Without loss of generality, we can concentrate on $j = j^*$, because functions oscillating at very different frequencies result in very small inner products.

The translation step at scale j^* is $\sim 2^{-j^*} = \frac{1}{\sqrt{N}}$, in both directions. Let us therefore imagine a square grid of stepsize $\frac{1}{\sqrt{N}}$ overlaid, say, on the unit square. For every point $\mathbf{n} = (n_1, n_2)$ on that grid, there are a lot of potentially large wave atom coefficients, corresponding to the wave number subscript $\mathbf{m}$.

However, on a length scale $\sim \frac{1}{\sqrt{N}}$, the function f oscillates in one direction only, with definite wavelength, meaning that its energy is strongly localized in frequency. More precisely, the wave numbers (frequencies) spanned by f in such a neighborhood are contained in a ball of radius $\sim \sqrt{N}$ in the ω-plane, at a distance $O(N)$ from the origin. This is exactly the scaling of wave atoms: as a result, *essentially one wave number* $\mathbf{m}$ *will contribute* in the wave atom expansion of f around the point (n_1, n_2). Of course in practice "one wave number" may become a small, fixed number C_ϵ depending on the desired accuracy ϵ.

The overall number of wave atom coefficients necessary to represent f is therefore proportional to the number of grid points at scale $\frac{1}{\sqrt{N}}$, that is $O(N)$ in 2D.

We can perform the same type of analysis to study sparsity in other domains; the result would be the following.

- *Wavelets: $O(N^2)$ coefficients.* For wavelets at scale $j^* \simeq \log_2 N$, the translation step is $O(1/N)$. All the wavelet coefficients on the N-by-N grid can be expected to be of comparable magnitude. As a result, we can expect to require $O(N^2)$ wavelet coefficients to encode f (i.e., almost all the coefficients until scale j^*.)

- *Gabor and nonadaptive local cosines: $O(N^2)$ coefficients.* Gabor atoms [26] are indexed by the wave number $\mathbf{m}$ (taking on $O(N^2)$ values) and the position index $\mathbf{n}$ (taking on $O(1)$ values once $\mathbf{m}$ is fixed.) The spatial translation step is $O(1)$ and independent of $\mathbf{m}$. At a scale $O(1)$ the function f does not have a definite frequency content; rather, it oscillates in all possible directions. As a result, most of the $O(N^2)$ directions will be needed, resulting in $O(N^2)$ Gabor atom coefficients.

- *Curvelets: $O(N^{3/2})$ coefficients.* Curvelets [9, 6] come at different scales, orientations and locations. A good heuristic for curvelets is that large coefficients correspond to directions

aligned with the oscillations. Since the essential support of a curvelet at scale $j^* \sim \log_2 N$ is of size $\sim 2^{-j^*}$-by-$2^{-j^*/2}$, we would need $O(N^{3/2})$ coefficients to tile the unit square, hence represent f accurately. One way to explain why curvelets are inferior to wave atoms is that *curvelets only capture the coherence of the pattern along the oscillations, not across.*

- *Wavelet packets.* It is a natural question to ask whether a proper tree of wavelet packets, implementing the parabolic scaling of wave atoms, would do as good a job at sparsity of oscillatory patterns. The answer is likely negative, because wavelet packets built from filterbanks do not have adequate frequency localization [30, 34]. We will discuss this phenomenon further in Section 3.1.

Of course the heuristics concerning wave atoms need to be made precise, which is the subject of the next section.

2.3 Sparsity: Proof

In this section, we prove Theorem 1.1.

Let us start by noticing that the function g in (1) has the interpretation of a warping, because it can be taken as the first component of a diffeomorphism $\mathbf{g} : \mathbb{R}^2 \to \mathbb{R}^2$. We can therefore define the oscillating profile pre-warping as

$$\tilde{f}(x) = \sin(Nx_1)h(\mathbf{g}^{-1}(x)). \tag{5}$$

We study $\mathbf{g}$ and h separately, and split the proof accordingly by means of two lemmas. We will detail at the end how the two lemmas should be combined.

Lemma 2.2. *Let $u \in L^2(\mathbb{R}^2)$, and $\mathbf{g}$ a C^∞ diffeomorphism. If the wave atom coefficients of u belong to the unit ball in ℓ_p for some $p > 0$, then the wave atoms coefficients of $u \circ \mathbf{g}$ belong to a ball of radius R in the same ℓ_p space, where R only depends on the smoothness constants of $\mathbf{g}$.*

Proof. A smooth diffeomorphism like $\mathbf{g}(x)$ is a special case of an operator known as a Fourier Integral Operator, because we can write, for any (well-behaved) function $u(x)$, that

$$u(\mathbf{g}(x)) = \frac{1}{(2\pi)^2} \int e^{i\mathbf{g}(x)\cdot\omega} \hat{u}(\omega)\, d\omega.$$

Such oscillatory integrals and their generalizations were studied in [5, 6], where it was shown (among other results) that they preserve the ℓ_p sparsity in curvelets. The necessary phase-space localization properties of curvelets are also shared by wave atoms: a careful inspection of the main argument in [6] would show that it generalizes mechanically. The definition of pseudo-distance for wave atoms, and the related almost-orthogonality result, can be found in [16].

□

Lemma 2.3. *Let $\tilde{f}$ be given by (5). Then for all $0 < p < 2$ the wave atom coefficients of $\tilde{f}$ belong to ℓ_p, with norm no greater than*

$$C_p N^{\frac{1}{p}-\frac{1}{2}},$$

for some constant $C_p > 0$ depending on p.

Proof. We use the notation $\langle\omega\rangle = \sqrt{1+|\omega|^2}$ throughout, not to be confused with the inner product. Let $\mathbf{N}_1 = \pi N \begin{pmatrix} 1 \\ 0 \end{pmatrix}$. Since $h \circ \mathbf{g}^{-1}$ is smooth, the frequency localization estimate for $\tilde{f}$ is

$$\hat{\tilde{f}}(\omega) \le C_M \cdot (\langle \omega - \mathbf{N}_1 \rangle^{-M} + \langle \omega + \mathbf{N}_1 \rangle^{-M}),$$

for all $M > 0$. On the other hand, the estimate (3) implies that wave atoms have two bumps in frequency, of the form $2^{-j}\hat{\psi}(2^{-j}\omega \pm \mathbf{m})$ for some ψ in the Schwartz class. Without loss of generality we can discard one of these bumps, and bound the wave atom coefficient $c_{j,\mathbf{m},\mathbf{n}}$ of f by

$$|\int \hat{\tilde{f}}(\omega) 2^{-j} \hat{\psi}(2^{-j}\omega - \mathbf{m})\, d\omega| \le 2^{-j} C_M \cdot \langle 2^j \mathbf{m} - \mathbf{N}_1 \rangle^{-M}.$$

This bound is valid uniformly over the $O(2^{2j})$ values taken on by the position index $\mathbf{n}$. The p-th power of the ℓ^p norm of the coefficients is then bounded by

$$\sum_j \sum_{\mathbf{m},\mathbf{n}} |c_{j,\mathbf{m},\mathbf{n}}|^p \le C_M \cdot \sum_j \sum_{\mathbf{m}} 2^{2j} |2^{-j} \langle 2^j \mathbf{m} - \mathbf{N}_1 \rangle^{-M}|^p.$$

In this remaining sum, the terms decay near-exponentially away from $j^* = \frac{1}{2}\log_2(N)$, and $\mathbf{m}^* = 2^{-j}\mathbf{N}_1$. The overall sum is therefore bounded by a constant times $2^{2j^*} 2^{-j^* p}$, i.e., a constant times $N^{1-p/2}$. □

With Lemmas 2.2 and 2.3 at our disposal, we are now ready to prove the main theorem.

Proof of Theorem 1.1. The L^2 error of the best M-term wave atom approximant f_M obeys

$$\|f - f_M\|_2^2 \le C \cdot M^{1-\frac{2}{p}} \cdot \|c\|_{\ell_p}^2,$$

where $0 < p < 2$ and c are the wave atom coefficients of f. See [26], Chapter 9, p. 390 for a proof. We would like to make this error smaller than ϵ^2, implying that it is enough to keep M^* coefficients, with

$$M^* \ge C_p \cdot \epsilon^{\frac{2p}{p-2}} \cdot \|c\|_{\ell_p}^{\frac{2p}{2-p}}.$$

By Lemmas 2.2 and 2.3, the wave atoms coefficients of f belong to ℓ_p for all $p > 0$, with bound $\|c\|_{\ell_p} \le N^{\frac{1}{p}(1-\frac{p}{2})}$. Substituting above gives the desired result,

$$M^* \ge C_p \cdot \epsilon^{\frac{2p}{p-2}} \cdot N,$$

valid for arbitrarily small $p > 0$.

□

We would like to remark that Theorem 1.1 generalizes in a straightforward manner to slightly less smooth texture patterns, namely of the form

$$f^\sharp(x) = \sin(Ng(x))h(\sqrt{N}x), \tag{6}$$

with g and h smooth, and h supported inside $[0, \sqrt{N}]^2$, say. Again, only $O(N)$ wave atom coefficients suffice to represent f up to given accuracy. The proof would be very similar. In fact, wave atoms qualify as an *optimal* transform for the model (6), because f itself contains $O(N)$ degrees of freedom.

Let us conclude this section by remarking the central and nontrivial role of Lemma 2.2; the rest of the analysis is quite secondary and straightforward in comparison. We believe that *invariance of sparsity under diffeomorphisms* is an important notion, which could prove fruitful in other contexts as well. For instance, Emmanuel Candès and the authors believe that the proof of the main approximation result for curvelets in [9] can be quite simplified using invariance of curvelet sparsity under diffeomorphisms.

3 Digital Wave Atoms

3.1 Implementation of Wave Atoms: 1D Warmup

In practice, wave atoms will be constructed from tensor products of adequately chosen 1D wave packets.

We will first build a one-dimensional family of real-valued wave packets $\psi^j_{m,n}(x)$, $j \geq 0, m \geq 0, n \in \mathbb{Z}$, centered in frequency around $\pm\omega_{j,m} = \pm\pi 2^j m$, with $C_1 2^j \leq m \leq C_2 2^j$; and centered in space around $x_{j,n} = 2^{-j}n$. The one-dimensional version of the parabolic scaling dictates that the support of (each bump of) $\hat{\psi}^j_{m,n}(\omega)$ be of length $O(2^j)$ while $\omega_{j,m} = O(2^{2j})$. The desired corresponding tiling of frequency is illustrated at the bottom of Figure 3.

Filterbank-based wavelet packets naturally come to mind as a potential definition of an orthonormal basis satisfying these localization properties. They also come with fast algorithms. The wavelet packet tree, defining the partitioning of the frequency axis in 1D, can be chosen to have depth j when the frequency is $\sim 2^{2j}$, as illustrated in Figure 3. However, there is a well-documented problem associated with standard wavelet packets, namely that the sense in which they satisfy frequency localization is rather weak. It is an unavoidable feature of the filterbank architecture that the uncertainty (product of time and frequency deviations) increases with the frequency, instead of remaining close to the Heisenberg bound. For references for precise estimates of the "wavelet packet curse," see [30, 34]. As a result, in our context, we cannot hope to satisfy the wave atom definition using basis functions which come from a wavelet packet analysis.

An elegant solution to the frequency localization problem was given by Lars Villemoes in [34]. The trick consists in exhibiting adequate symmetric pairs of compactly supported bumps in frequency, given by the formula

$$\hat{\psi}^0_m(\omega) = e^{-i\omega/2}\left[e^{i\alpha_m} g(\epsilon_m(\omega - \pi(m+\frac{1}{2}))) + e^{-i\alpha_m} g(\epsilon_{m+1}(\omega + \pi(m+\frac{1}{2})))\right], \qquad (7)$$

where $\epsilon_m = (-1)^m$ and $\alpha_m = \frac{\pi}{2}(m+\frac{1}{2})$. The function g is an appropriate real-valued, C^∞ bump function, compactly supported on an interval of length 2π, and chosen such that

$$\sum_m |\hat{\psi}^0_m(\omega)|^2 = 1.$$

Then the translates $\{\psi_m(t-n)\}$ form an orthonormal basis of $L^2(\mathbb{R})$. This construction provides a *uniform* tiling of the frequency axis,[4] in the sense that every bump in frequency has the same support size, 2π. We note in passing that Villemoes's construction is not very different from the Wilson basis of Daubechies, Jaffard and Journé [15].

[4]For specialists, the secret to avoiding the Balian-Low no-go theorem is to use *two* bumps in frequency.

Multiscale tilings like the one in Figure 3 can be obtained by combining dyadic dilates and translates of $\hat{\psi}_m^0$ on the frequency axis. We need to introduce the subscript j to index scale, and write our basis functions as

$$\psi_{m,n}^j(x) = \psi_m^j(x - 2^{-j}n) = 2^{j/2}\psi_m^0(2^j x - n).$$

To preserve orthonormality of the $\psi_{m,n}^j(x)$, the profile g needs to be asymmetric in addition to all the other properties, in the sense that

$$g(-2\omega - \frac{\pi}{2}) = g(\frac{\pi}{2} + \omega)$$

for $\omega \in [-\pi/3, \pi/3]$, with g itself supported on $[-7\pi/6, 5\pi/6]$. We say $g(\omega)$ is "left-handed", whereas $g(-\omega)$ is "right-handed". As a result, the uniform partitioning of the frequency axis is obtained as an alternating sequence of staggered right-handed and left-handed bumps. On the positive part of the frequency axis, a scale doubling can be achieved by concatenating two right-handed bumps at scales differing by a factor 2. Figure 3, bottom row, depicts the Villemoes system where three scale doublings have been implemented. Let us remark in passing that scale halving could be implemented using left-handed bumps at different scales, but scale quadrupling is impossible using the present scheme.

As for labeling, note that the couple (j, m) refers to a point on the wavelet packet tree; the depth at that point is $J - j$, where J is the maximum depth, and m can be interpreted as the number of nodes on the left at the same depth (nodes are not necessarily leaves). This is the standard indexing scheme for wavelet packets used in [26]. The translation step is now 2^{-j} at scale j, whereas each bump in frequency is supported on an interval of length $2^j \times 2\pi$. Choosing the "parabolic tree" as in Figure 3 amounts to specifying the wave vector $\omega_{j,m}$, defined as the center of the positive frequency bump, as

$$\omega_{j,m} = \pi 2^j m \sim 2^{2j}$$

More precisely, the admissible values for m are in the interval $[2^{j-1} + 1, \ldots, 2^{j+1} + 1]$, when $j \geq 2$. The resulting basis of wavelet packets $\{\psi_{m,n}^j(x)\}$ is orthonormal for $L^2(\mathbb{R})$.

The implementation of Villemoes wavelet packets is rather straightforward in the frequency domain. For each wave number $\omega_{j,m}$, the coefficients $c_{j,m,n}$ can be seen as a decimated convolution at scale 2^{-j},

$$c_{j,m,n} = \int \psi_m^j(x - 2^{-j}n)u(x)\,dx,$$

By Plancherel,

$$c_{j,m,n} = \frac{1}{2\pi}\int e^{i2^{-j}n\omega}\overline{\hat{\psi}_m^j(\omega)}\hat{u}(\omega)\,d\omega.$$

Assuming that the function u is accurately discretized at $x_k = kh$, $h = 1/N$, $k = 1, \ldots, N$, then up to some small truncation error,

$$c_{j,m,n} \simeq c_{j,m,n}^D = \sum_{k=2\pi(-N/2+1:1:N/2)} e^{i2^{-j}nk}\overline{\hat{\psi}_m^j(k)}\hat{u}(k). \tag{8}$$

This equation makes sense for couples (j,m) for which the support of $\hat{\psi}_m^j(k)$ lies entirely inside the interval $[-\pi N, \pi N]$, so we may write $k \in 2\pi\mathbb{Z}$ for the index range[5].

If the windowed data $\overline{\hat{\psi}_m^j(k)}\hat{u}(k)$ were supported in an interval of length $2^j \times 2\pi$, then the sum (8) could be restricted to values of k inside that interval, and computed efficiently using a reduced inverse FFT, of size 2^j (to match the 2^{-j} in the exponent). In reality the support properties of g are such that the windowed data is supported inside *two* disjoints intervals of size $2^{j+1}\pi$, symmetric about the origin, hence on 2^{j+1} points. The sum (8) can still be computed by a reduced inverse FFT provided $\overline{\hat{\psi}_m^j}\hat{u}$ is first folded by $2^{j+1}\pi$-periodicity inside an interval of size $2^{j+1}\pi$ centered about the origin:

$$c_{j,m,n}^D = \sum_{k=2\pi(-2^j/2+1:1:2^j/2)} e^{i2^{-j}nk} \sum_{p\in 2\pi\mathbb{Z}} \overline{\hat{\psi}_m^j(k+2^j p)}\hat{u}(k+2^j p). \tag{9}$$

We have used the fact that $[2^{-j}n(k+2^j p)] \bmod 2\pi = [2^{-j}nk] \bmod 2\pi$ to go from equation (8) to equation (9). Notice that the sum over p has only two nonzero terms, because of the size of the support of $\hat{\psi}_m^j(k)$, and is therefore very cheap to compute. Figure 4, top row, illustrates this operation, which we call "wrapping". The outer sum over k can then be computed using an inverse FFT. The wrapping trick was already used for the implementation of the discrete curvelet transform, see [7]. The simple algorithm for wavelet packets is then the following.

- Perform a FFT of size N of the samples $u(x_k)$.
- For each pair (j,m), wrap the product $\overline{\hat{\psi}_m^j}\hat{u}$ by periodicity inside the interval $[-2^j\pi, 2^j\pi]$ (this is the sum over p in (9)). Then perform an inverse FFT of size 2^j of the result to obtain $c_{j,m,n}^D$.
- Repeat over (j,m).

The complexity of each inverse FFT at scale j is $O(j2^j)$, and there are $O(2^j)$ frequency bumps at scale j, indexed by m, so the total complexity is

$$\sum_{0\le j\le J} O(j2^{2j}) = O(J2^{2J}) = O(N\log N),$$

with $N = 2^{2J}$.

Since the $c_{j,m,n}^D$ are coefficients in an orthonormal basis, the inverse transform is simply the adjoint and can be computed by reversing all the steps in the above algorithm.

- For each (j,m), perform a FFT in n of $c_{j,m,n}$, of size 2^j, unwrap the result on the frequency axis around the support of $\hat{\psi}_m^j$ (that is also a sum over p, as in (9)), and multiply the unwrapped data with $\hat{\psi}_m^j$.
- Sum the contributions corresponding to all the couples (j,m).

[5]For those very few couples (j,m) such that $\hat{\psi}_m^j(k)$ lies on the boundary $k = \pm\pi N$, the window $\hat{\psi}_m^j(k)$ needs to be copied by $2\pi N$-periodicity to the other side of the frequency domain, in such a way as to preserve orthonormality. Let us only mention that this is possible and implemented in the code.

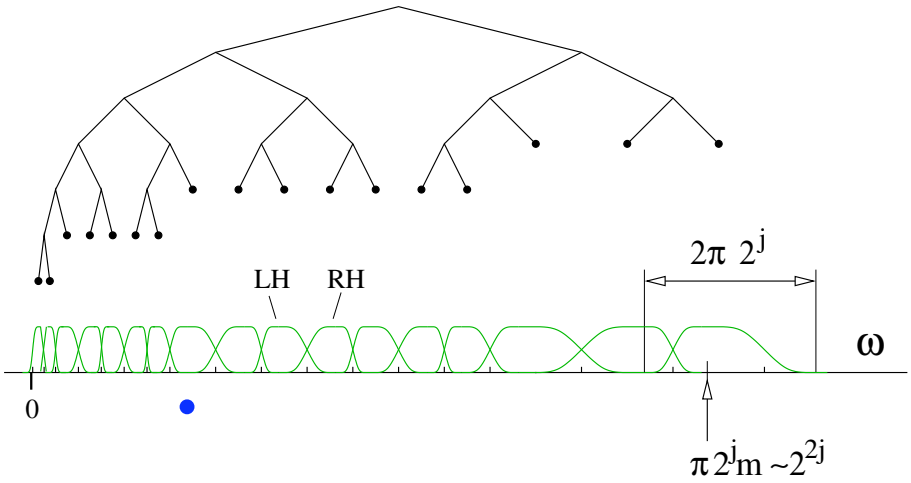

Figure 3: The wavelet-packet tree corresponding to wave atoms. Textbook material on wavelet packet trees can be found in [26], Chapter 8. The bottom graph depicts Villemoes wavelet packets on the positive frequency axis. The dot under the axis indicates a frequency where a change of scale occurs. The labels "LH", respectively "RH" indicate a left-handed (m even), respectively right-handed window (m odd).

- Perform an inverse FFT, of size N, to obtain $u(x_k)$.

Likewise, the complexity of the inverse transform is $O(N \log N)$.

The decomposition into two bumps, of positive and negative frequency respectively, can be written

$$\hat{\psi}^j_{m,n}(\omega) = \hat{\psi}^j_{m,n,+}(\omega) + \hat{\psi}^j_{m,n,-}(\omega), \tag{10}$$

with the symmetry relation $\hat{\psi}^j_{m,n,-}(\omega) = \overline{\hat{\psi}^j_{m,n,+}(-\omega)}$, which owes to the real-valuedness of $\psi^j_{m,n}$. After performing a Hilbert transform, $H\psi^j_{m,n}$ is another orthonormal basis of $L^2(\mathbb{R})$. In the frequency domain, Hilbert transformation amounts to taking a linear combination of the two bumps[6] with weights $(-i, i)$ instead of $(1, 1)$:

$$\widehat{H\psi^j_{m,n}}(\omega) = -i\hat{\psi}^j_{m,n,+}(\omega) + i\hat{\psi}^j_{m,n,-}(\omega). \tag{11}$$

Although we made a point that the Villemoes basis functions do not come from a standard multiresolution analysis, we still continue to call them "wavelet packets" in what follows.

3.2 Implementation of Wave Atoms: 2D Extension

Two-dimensional orthonormal basis functions with 4 bumps in the frequency plane can be formed by individually taking products of 1D wavelet packets. As we did in earlier sections, let us abbreviate

[6]Strictly speaking, the name "Hilbert transform" for equation (11) is valid for all wave numbers except $m = 0$ near the origin. In this latter case, the two bumps overlap and the linear combination (11), although not the Hilbert transform, is the proper way to define an orthobasis. In fact, analyzing with (10), then reconstructing with (11) looks like an intriguingly nice way to approximate the continuous Hilbert transform on a finite interval.

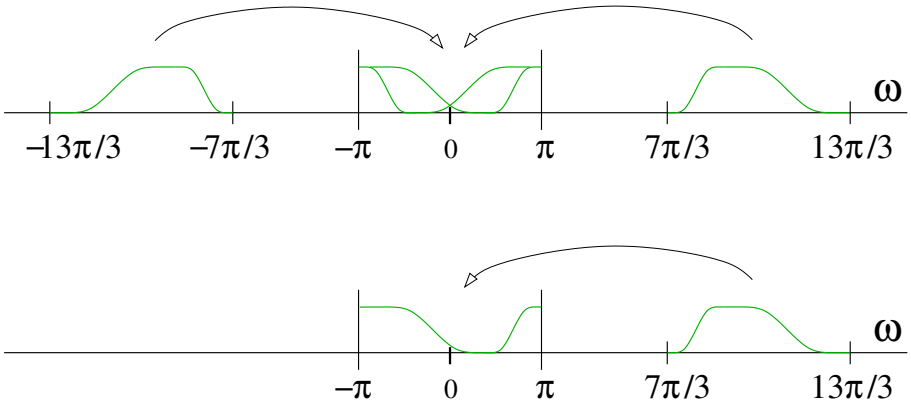

Figure 4: *Top row*: Standard, Fourier wrapping in the case $j = 0$ and $m = 3$. The windowed data is copied by 2π-periodicity inside the interval $[-\pi, \pi]$. *Bottom row*: One-sided Fourier wrapping in the case $j = 0$ and $m = 3$. In the case of a single frequency bump, the wrapped windowed data does not overlap, signaling the absence of aliasing.

$\mu = (j, \mathbf{m}, \mathbf{n})$, where $\mathbf{m} = (m_1, m_2)$ and $\mathbf{n} = (n_1, n_2)$. We write

$$\varphi_\mu^+(x_1, x_2) = \psi_{m_1}^j(x_1 - 2^{-j}n_1)\, \psi_{m_2}^j(x_2 - 2^{-j}n_2),$$

The Fourier transform is also separable, namely

$$\hat{\varphi}_\mu^+(\omega_1, \omega_2) = \hat{\psi}_{m_1}^j(\omega_1) e^{-i2^j n_1 \omega_1}\, \hat{\psi}_{m_2}^j(\omega_2)\, e^{-i2^j n_2 \omega_2}.$$

A dual orthonormal basis can be defined from the "Hilbert-transformed" wavelet packets,

$$\varphi_\mu^-(x_1, x_2) \equiv H\psi_{m_1}^j(x_1 - 2^{-j}n_1)\; H\psi_{m_2}^j(x_2 - 2^{-j}n_2).$$

It is easy to see from relations (10) and (11) that the recombination

$$\varphi_\mu^{(1)} = \frac{\varphi_\mu^+ + \varphi_\mu^-}{2}, \qquad \varphi_\mu^{(2)} = \frac{\varphi_\mu^+ - \varphi_\mu^-}{2}, \tag{12}$$

provides basis functions with *two* bumps in the frequency plane, symmetric with respect to the origin, hence directional wave packets (oscillating in one single direction). Together, $\varphi_\mu^{(1)}$ and $\varphi_\mu^{(2)}$ form the wave atom frame and may be denoted jointly as φ_μ. The price to pay in considering both $\varphi_\mu^{(1)}$ and $\varphi_\mu^{(2)}$ is an increase of a factor 2 in the redundancy. Wave atoms otherwise remain a tight frame, in the sense that

$$\sum_\mu |\langle \varphi_\mu^{(1)}, u\rangle|^2 + \sum_\mu |\langle \varphi_\mu^{(2)}, u\rangle|^2 = \|u\|^2.$$

They also satisfy all the wave atom properties, by construction.

Note that considering Hilbert-transformed bases, and performing the combination (12) in 2D, are both well-known tricks in wavelet theory: for instance, they are part of the definition of the 2D dual-tree complex wavelet transform (ℂWT, see [29] for a review), as well as the definition of brushlets [27].

As the reader might have noticed, the construction is not a simple tensor product because there is only one scale subscript j. This is akin to the construction of "nonstandard," or MRA wavelet bases in 2D where the point is to enforce the same scale in both directions, hence an isotropic aspect ratio. The resulting tiling of the frequency plane is shown in Figure 5.

In practice, the algorithm for wave atoms is based on the obvious generalization of the 1D wrapping strategy to two dimensions—except for a slight complication. The admissible tilings of the frequency plane at scale j are restricted by

$$\max_{i=1,2} |m_i| = 4n_j + 1, \tag{13}$$

for some integer n_j depending on j. In Figure 5, we check that this property holds with $n_0 = 0$, $n_1 = 1$ and $n_2 = 2$. The rationale for this (benign) restriction is that a window needs to be right-handed in both directions near a scale doubling (in the positive quadrant), and that this parity needs to match with the rest of the lattice. The rule is that $\hat{\psi}^j_{m,+}$ is right-handed for m odd and left-handed for m even, so for instance $\hat{\psi}^2_2(\omega_1)\hat{\psi}^2_2(\omega_2)$ would not be an admissible window near a scale doubling, whereas $\psi^2_3(x_1)\psi^2_3(x_2)$ is admissible (and its tile is indicated in Figure 5 by a dot).

In complete analogy with the 1D case, the complexity of wave atoms is of course $O(N^2 \log N)$.

The adjoint wave atom transform is still an inverse (actually the Moore-Penrose pseudo-inverse) because the frame is tight. As explained previously, it is computed by reversing all the operations in the direct transform, and also takes $O(N^2 \log N)$ operations.

Let us conclude this section by mentioning that, in the present implementation, most wave atoms have an infinite number of (directional) *vanishing moments* in the sense that

$$\int \varphi_\mu(y_1, y_2)\, y_1^k\, dy_1 = 0, \qquad \text{for all } \; k \geq 0, \qquad \text{for all } \; y_2, \tag{14}$$

where y_1 is the spatial coordinate along the wave vector ω_μ, and y_2 is perpendicular. This property follows from the fact that most atoms $\hat{\varphi}_\mu(\omega)$ vanish in a (large) strip including the origin, oriented in the direction perpendicular to ω_μ. In contrast, those few atoms obeying $\hat{\varphi}_\mu(0) \neq 0$ form the small minority of "coarse scale" wave atoms, when $j = 0$.

3.3 Variations

There are many ways the construction presented above can be modified to suit certain applications in image processing or numerical analysis:

- *The orthobasis variant.* In practice, one may want to work with the original orthonormal basis $\varphi^+_\mu(x)$ instead of a tight frame. Since $\varphi^+_\mu(x) = \varphi^{(1)}_\mu(x) + \varphi^{(2)}_\mu(x)$, each basis function $\varphi^+_\mu(x)$ oscillates in *two* distinct directions, instead of one. We call this the *orthobasis variant*. It may be useful when redundancy is uncalled-for, e.g., in image compression and numerical analysis.

- *The complex variant.* On the contrary, it is possible to isolate each bump in frequency while keeping the same translation steps. This modification defines a complex-valued *tight frame with redundancy four*:

$$\{\psi^j_{m_1,n_1,+}(x_1)\psi^j_{m_2,n_2,+}(x_2)\,,\ \psi^j_{m_1,n_1,+}(x_1)\psi^j_{m_2,n_2,-}(x_2)\,,$$
$$\psi^j_{m_1,n_1,-}(x_1)\psi^j_{m_2,n_2,+}(x_2)\,,\ \psi^j_{m_1,n_1,-}(x_1)\psi^j_{m_2,n_2,-}(x_2)\}.$$

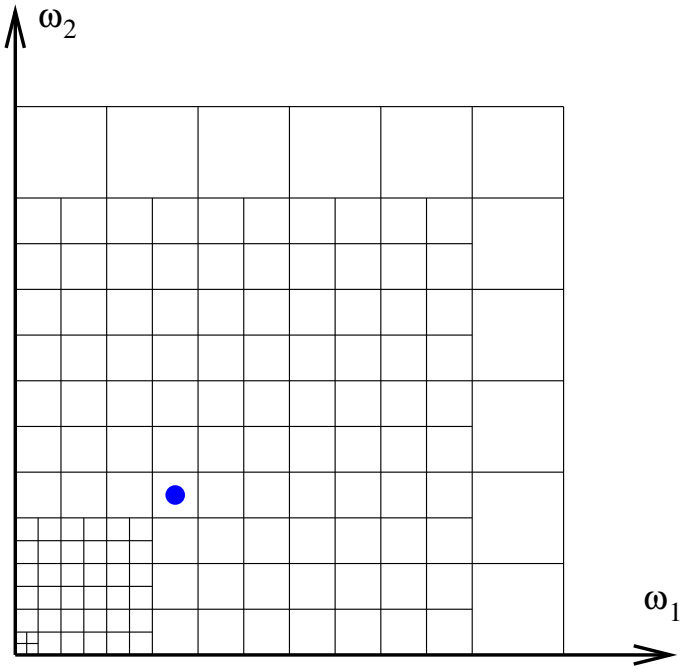

Figure 5: The wave atom tiling of the frequency plane. Only the first quadrant is shown. The size of the squares doubles when the scale j increases by 1. At a given scale j, squares are indexed by m_1, m_2 starting from zero near the axes. The dot indicates the same change of scale as in Figure 3 and corresponds to the basis function denoted $\psi_3^2(x_1)\psi_3^2(x_2)$ in the text.

The advantage of breaking the central symmetry in the frequency plane (and thereby creating complex-valued templates in (x_1, x_2)), is to increase the redundancy in order to create exact *shift-invariance* of subspaces of fixed $j, \mathbf{m}$. Namely, let $(T_{p_1,p_2}u)(x_1, x_2) = u(x_1 - p_1, x_2 - p_2)$ where p_1, p_2 are translations on the grid. Then if $u_{j,\mathbf{m}}$ denotes the partial reconstruction

$$u_{j,\mathbf{m}}(x_1, x_2) = \sum_{\nu_1=\pm} \sum_{\nu_2=\pm} \sum_{\mathbf{n}} c_{j,\mathbf{m},\mathbf{n},\nu_1,\nu_2} \psi^j_{m_1,n_1,\nu_1}(x_1) \psi^j_{m_2,n_2,\nu_2}(x_2),$$

where $c_{j,\mathbf{m},\nu,\nu_1,\nu_2}$ are the adequate tight frame coefficients, then we have shift invariance[7]:

$$(T_{p_1,p_2}u)_{j,\mathbf{m}}(x_1, x_2) = u_{j,\mathbf{m}}(x_1 - p_1, x_2 - p_2).$$

This property helps in reducing denoising artifacts, for instance, and neither holds true for the wave atom tight frame with redundancy two, nor for the orthobasis variant. Ultimately, shift-invariance owes to the absence of aliasing in the coefficients $c_{j,\mathbf{m},\mathbf{n},\nu_1,\nu_2}$, because there is no overlap during the "wrapping" operation, as in Figure 4, bootm row. In other words, the sum over p has only one nonzero term in (9) when the basis function has only one bump in frequency. In fact, for the tight frame with redundancy four, the delicate phase factors of formula (7) are unnecessary and can be omitted without compromising the tight frame property.

Shift-invariance (or near-shift invariance) is a much sought-after property in wavelet design for denoising applications, particularly when the implied redundancy increase is under control. It may reduce the need for cycle-spinning, although not completely so—individual coefficients are still not shift-invariant. See for instance [29] for a discussion.

- *Mirror-extended images.* Defining wave atoms in the discrete frequency plane has one side effect: the FFT implicitly treats the image on $[0, 1]^2$ as periodic across its boundaries. This may pose a problem for image processing applications. One typical remedy is to work on the mirror-extended image, defined on the square $[-1, 1]^2$ as

  $$\tilde{u}(x_1, x_2) = u(|x_1|, |x_2|).$$

  Applying the wave atom algorithm to an image twice the original size will normally result in a computational penalty of a factor slightly more than four. The overhead is easily identified: it is due to redundant translation steps in the mirrored portions of the image.

  We regard as very interesting the problem of adapting the wave atom transform to mirror-extended images with a lower computational penalty. We have no solution to report on at this time.

- We have been advocating the choice of a parabolically-shaped wavelet packet tree, in order to best represent oscillatory patterns. This is a result valid in the continuum limit. When it comes to discrete data, of course, nothing replaces a good old tree search for the best basis [26], or other heuristics. The architecture of wave atom transform lends itself quite well to changes in the wavelet packet tree, although, as we explained previously, only scale doublings and halvings are authorized. We call these "balanced trees"; see [34] for more details in the 1D case. In the 2D case, we remind the reader that there is also a restriction due to equation (13).

[7] Or more precisely, shift-covariance.

3.4 Related work

The construction of wave atoms presented here follows a long legacy in pure and applied harmonic analysis. We would like to present a nonexhasutive list of related constructions.

First, a continuous version of the wave atom transform, involving Gaussians, and the same parabolic scaling of oscillations vs. support size, has been known in mathematics under the name "Córdoba-Fefferman wave packets" since the late 1970s [14]. We apologize to harmonic analysts for choosing to name the discrete transform differently.

Second, the 2D recombination strategy using Hilbert-transform pairs, but based on standard wavelet packets instead of Villemoes packets, has already been considered in the field of image processing under the name "dual-tree M-band wavelets," see [11] and references therein. If this reconstruction is based on wavelets instead of Villemoes packets, then one obtains "dual-tree complex wavelets", see [22, 19, 13] and references therein.

Third, we would like to mention E. Kolaczyk's approach [23] for the implementation of Meyer wavelets, and by extension Villemoes wavelet packets, based on local cosine windowing of the frequency axis. His approach explains in a transparent manner the aliasing cancellations during inversion of the wavelet transform, though the implementation looks involved in contrast to the "wrapping" approach.

On a related note, an implementation of (discretized continuous) directional wavelets directly in the frequency plane was proposed by P. Vandergheynst and J. F. Gobbers in [33]. The implementation of curvelets in the frequency plane is in [7].

Finally, another way of performing local cosine windowing in the frequency plane are the *brushlets* of F. Meyer and R. Coifman [27]. They appear to be similar to Villemoes wavelet packets (and could be a good alternative) but unfortunately do not fit in the framework of the spatial localization estimate (4)—they have two or more bumps in space.

4 Numerical Examples

This section presents several numerical examples to illustrate the properties and potential of the wave atom frame and its orthobasis variation.

We begin by showing several wave atoms in both the spatial and frequency domain. Figure 6 illustrates three members of the wave atom frame at increasing scales on an $N \times N$ grid with $N = 512$. Every function oscillates coherently in a preferred direction and decays rapidly in the spatial domain. In the frequency domain, the Fourier transform of these functions are all sharply localized by construction. The members of the wave atom orthobasis are only slightly different: each of them is a superposition of two oscillating wave packets, and its Fourier transform has four bumps instead of two in the frequency domain. Given an arbitrary input image, a forward wave atom transform followed by its inverse recovers the input data up to machine accuracy, which is expected due to the exact tightness of the discrete transform. Moreover, both the forward and inverse transforms are highly efficient—in our experience they each take about the time of 8 to 10 FFTs.[8]

[8]This complexity estimate is valid for the standard wave atom transform, with periodization at the edges. If for instance a naive variant using a mirror-extended image was used instead, then those FFTs would relate to an image twice larger.

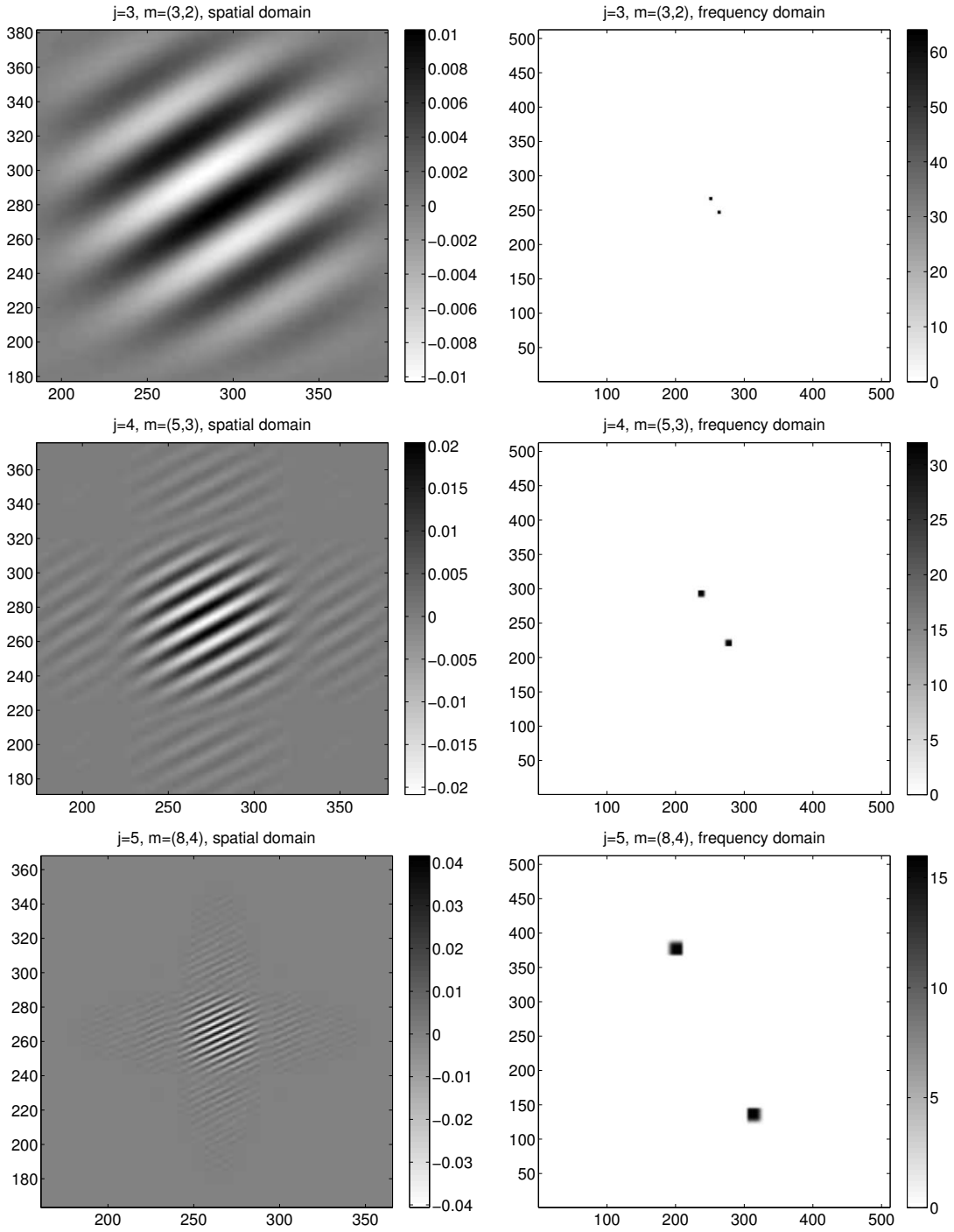

Figure 6: Wave atoms at increasing finer scales for $N = 512$. The left images show wave atoms in the spatial domain. We zoom in these images to show the detail profile of the wave atoms. The right images show the modulus of their Fourier transform in the frequency domain.

Now we illustrate the potential of the wave atoms with two examples. In the first example, we consider the compression properties, i.e, the decay rate of the coefficients of images under the wave atom bases. Besides the wave atom orthobasis and the wave atom frame, we include two other bases for comparison: the Daubechies db5 wavelet, and a wavelet packet that uses db5 filter and shares the same wavelet packet tree with our wave atom orthobasis.

The first input image (Figure 7(a)) is a synthetic seismogram corresponding to the acoustic response of a one-dimensional layered medium to a point source. The PSNR curves for partial reconstruction with increasing number of largest coefficients are shown in Figure 7(b). The result suggests that both the wave atom frame and its orthobasis variation outperform the wavelet basis and the wavelet packet by a significant margin. Figures 7(c) and (d) are reconstructed images using 1536 largest wave atom and wavelet coefficients respectively, and it is clear that the wave atom frame does a much better job.

The second input image (Figure 8(a)) is a fingerprint image. This image fits well in the setting of Theorem 1.1: its structure defines almost everywhere a predominant direction which can be aligned with the axes by a continuous warping. The PSNR curves in Figure 8(b) show that the numerical result matches well with our theoretical prediction. Both reconstructed from 512 largest coefficients, the wave atom approximant (Figure 8(c)) captures most of the structures of the fingerprint while the wavelet approximant (Figure 8(d)) is almost unusable. Of course this experiment is idealized: in practice a true compression algorithm would also require quantization of the coefficients. In addition, the features that make a fingerprint unique are not the oscillations themselves, but the so-called minutiae (ridge endings, ridge bifurcations). Those features tend to be lost in the aggressive wave atom thresholding of Figure 8(c).

For both images, we observe that the wave atom frame outperforms the orthobasis when the number of terms used in partial reconstruction is relatively small. This is mostly due to the fact that each frame element is a single coherent oscillating pattern, which makes it similar to these images. On the other hand, as the number of terms grows, the redundancy effect kicks in and the wave atom orthobasis offers better PSNR value.

The second example is image denoising. We use the two same images as before. The noise-to-signal ratio is set to $\sigma = 0.1$ for the seismogram and to $\sigma = 0.2$ for the fingerprint image. For simplicity, the denoising algorithm used is hard thresholding at 3σ combined with cycle spinning, for all transforms[9]. Besides the wave atom frame with redundancy 2, we use the dual-tree CWT frame, the wavelet packet basis, and the curvelet frame [7] for comparison. We do not include the results of the wave atom orthobasis here since it is well known that redundancy is generally helpful for image denoising. Figures 9 and 10 summarize the denoising results for the synthetic seismogram and the fingerprint image respectively. The results show that the wave atom frame outperform all three other transforms in both cases.

5 Discussion

This paper presents a new member in the family of oriented, multiscale transforms for image processing and numerical analysis.

The first lesson is that the parabolic scaling governing oscillations vs. support size of wave

[9]The threshold is in fact 3σ multiplied by the ℓ_2 norm of the basis elements, in the case of tight frames with varying redundancy. We have checked that he value 3σ does not favor wave atoms over the other transforms

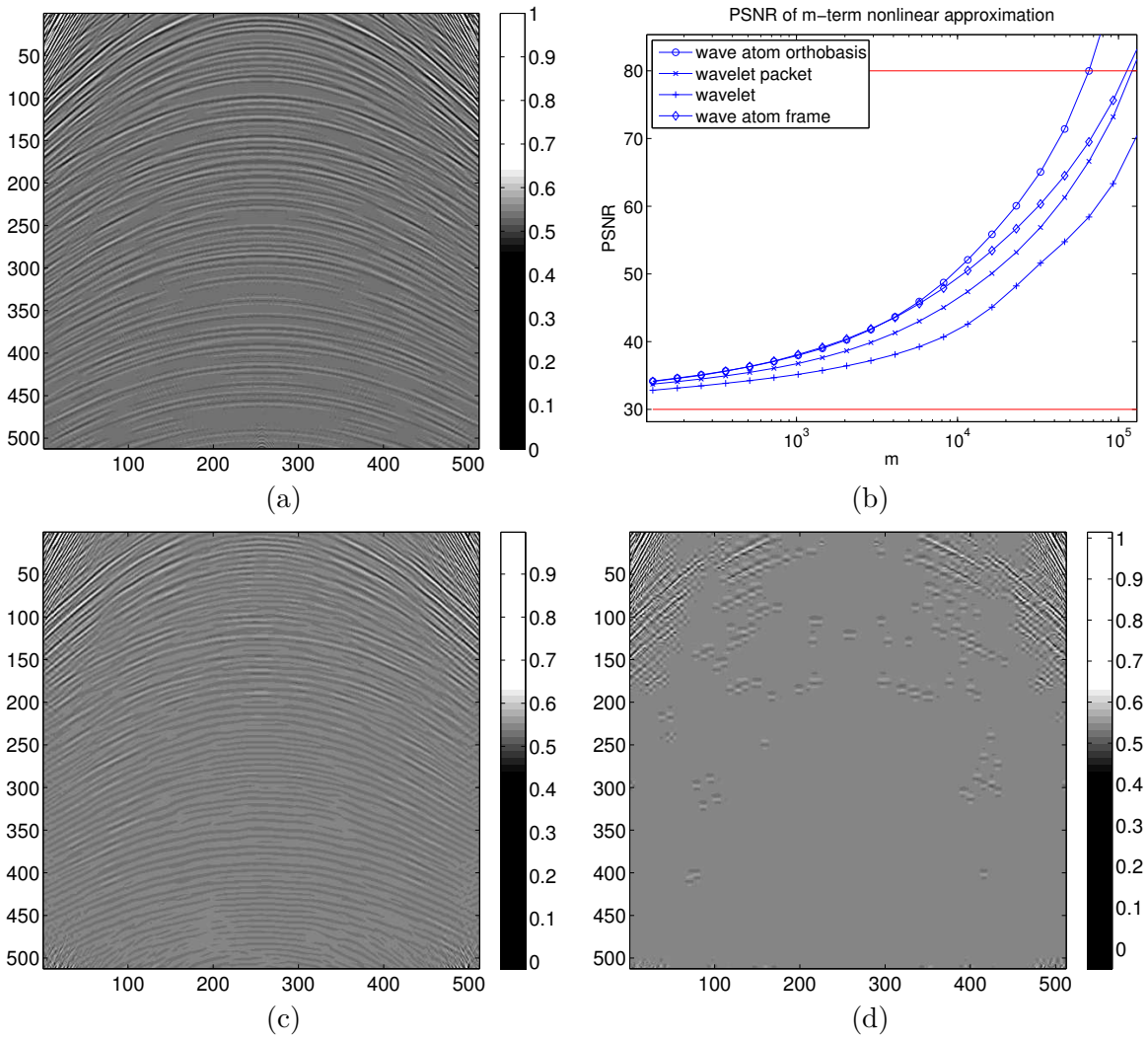

Figure 7: Compression analysis. (a) A synthetic seismogram. (b) PSNR curve of partial reconstruction with largest coefficients. (c) Reconstructed image with 1536 largest wave atom coefficients. (d) Reconstructed image with 1536 largest db5 wavelet coefficients.

atoms, visible on Figure 6, is the proper way to obtain universally sparse expansions of functions oscillating in smoothly varying directions.

The second lesson of our investigations is that standard wavelet packets should be avoided as an implementation of wave atoms. Instead, we propose to extend a construction of Lars Villemoes to 2D, and to implement it directly in the frequency domain. The wave atom Matlab software can be downloaded from the authors' respective webpages.

Of course, we need to mention that wave atoms only provide a low-level *nonadaptive* transformation of images, in the sense that no effort is necessary to "discover the geometry", i.e., estimate local orientation and wave vector of the oscillatory data. In non-stylized applications the story is always more complicated: a transform is usually just a black box in the flowchart of the algorithm,

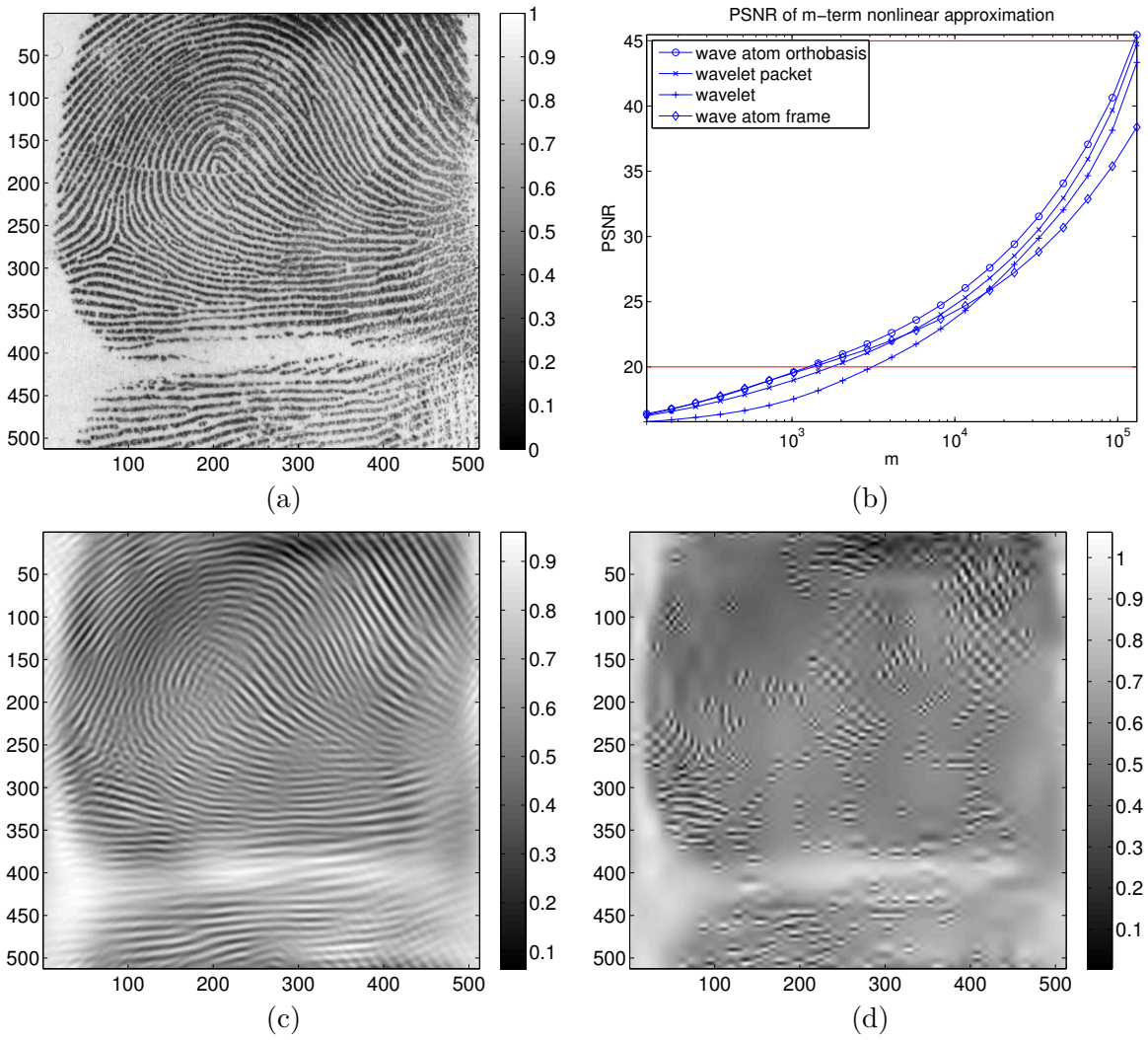

Figure 8: Compression analysis. (a) A fingerprint image. (b) PSNR curve of partial reconstruction with largest coefficients. (c) Reconstructed image with 512 largest wave atom coefficients. (d) Reconstructed image with 512 largest db5 wavelet coefficients.

and *a priori* physical information often plays a role in optimizing performance. A typical example is image compression via wavelets, where quantization of the coefficients can be based on physiological principles; where higher-level adaptive transform (like wedgeprints [35] or bandelets [28]) can even be defined on top of wavelets; and where nonlinear recovery techniques can exploit sparsity in the wavelet domain [32, 18, 10]. In this regard, we are very enthusiastic about the potential of wave atoms as part of the next generation of image algorithms.

Finally, we did not discuss the case of higher-dimensional data in this paper, but the ideas and methods extend in a straightforward way. The wave atom toolbox includes both a 2D and a 3D version of wave atoms.

References

[1] J.P. Antoine, L. Demanet, L. Jacques, J.F. Hochedez, R.Terrier, E. Verwichte, Applications of the 2-D wavelet transform to astrophysical images, *Physicalia magazine*, **24** (2002), 93–116.

[2] J. P. Antoine, R. Murenzi, Two-dimensional directional wavelets and the scale-angle representation. *Sig. Process.* **52** (1996), 259–281.

[3] C. Brislawn, The FBI Fingerprint Image Compression Standard, http://www.c3.lanl.gov/ brislawn/FBI/FBI.html

[4] E. J. Candès, Harmonic analysis of neural networks, *Appl. Comput. Harmon. Anal.*, **6** (1999), 197–218.

[5] E. J. Candès, L. Demanet, Curvelets and Fourier integral operators. *C. R. Acad. Sci. Paris, Ser. I* **336** (2003), 395–398.

[6] E. J. Candès and L. Demanet, The curvelet representation of wave propagators is optimally sparse, *Comm. Pure Appl. Math.*, **58-11** (2005) 1472–1528.

[7] E. J. Candès, L. Demanet, D. L. Donoho, L. Ying, Fast Discrete Curvelet Transforms, *SIAM Mult. Model. Sim.* **5-3** (2006) 861–899.

[8] E. J. Candès and D. L. Donoho. Curvelets – a surprisingly effective nonadaptive representation for objects with edges. In C. Rabut A. Cohen and L. L. Schumaker, editors, *Curves and Surfaces*, pages 105–120, Vanderbilt University Press, 2000. Nashville, TN.

[9] E. J. Candès and D. L. Donoho. New tight frames of curvelets and optimal representations of objects with piecewise-C^2 singularities. *Comm. on Pure and Appl. Math.* **57** (2004), 219–266.

[10] E. J. Candès and J. Romberg, Practical Signal Recovery from Random Projections, in *Proc. SPIE Conf. on Comput. Imag. III* (2005), 76–86.

[11] C. Chaux, L. Duval, J. C. Pesquet, 2D Dual-Tree M-band Wavelet Decomposition, *Proc. ICASSP* (2005)

[12] C. Chaux, L. Duval, J. C. Pesquet, Image analysis using a dual-tree M-band wavelet transform, *IEEE Trans. Im. Proc.*, in press.

[13] H. Choi, J. Romberg, R. Baraniuk, and N. Kingsbury, Hidden Markov tree modeling of complex wavelet transforms, in *Proc. IEEE ICASSP* (2000).

[14] A. Córdoba, C. Fefferman, Wave packets and Fourier integral operators. *Comm. PDE* **3(11)** (1978), 979–1005.

[15] I. Daubechies, S. Jaffard, J.L. Journ, A simple Wilson orthonormal basis with exponential decay, *SIAM J. Math. Anal.* **22** (1991) 554–572

[16] L. Demanet, *Curvelets, Wave Atoms and Wave Equations*, Ph. D. thesis, California Institute of Technology, 2006.

[17] M. Do, M. Vetterli, The contourlet transform: an efficient directional multiresolution image representation, *IEEE Trans. Im. Proc.* **14(12)** (2005) 2091–2106.

[18] M. Elad, J.-L Starck, D. L. Donoho and P. Querre, Simultaneous Cartoon and Texture Image Inpainting using Morphological Component Analysis (MCA), *Applied Comput. Harmon. Anal.*, in press.

[19] F. Fernandes, M. Wakin, R. Baraniuk, Non-redundant, linear-phase, semi-orthogonal, directional complex wavelets, in *Proc. IEEE ICASSP*, Montreal, Canada (2004)

[20] G. Hennenfent and F. J. Herrmann. Seismic denoising with unstructured curvelets, *Comput. in Sci. Eng.*, to appear, 2006.

[21] F. J. Herrmann, P. P. Moghaddam, C. C Stolk, Sparsity- and continuity-promoting seismic image recovery with curvelet frames, submitted, 2006.

[22] N. Kingsbury, Image processing with complex wavelets, *Phil. Trans. Roy. Soc. A* **357-1760**, (1999), 2543–2560.

[23] E. Kolaczyk, *Wavelet Methods for the Inversion of Certain Homogeneous Linear Operators in the Presence of Noisy Data*, Ph.D. thesis, Stanford University, 1994.

[24] D. Labate, W-Q. Lim, G. Kutyniok, and G. Weiss. Sparse multidimensional representation using shearlets, in Proc. SPIE *Wavelets XI*, San Diego (2005) 254–262.

[25] S. Lintner, F. Malgouyres, Solving a variational image restoration model which involves ℓ_∞ constraints, *Inverse Problems* **20** (2004), 815–831.

[26] S. Mallat, *A Wavelet Tour of Signal Processing*. Second edition. Academic Press, Orlando-San Diego, 1999.

[27] F. G. Meyer and R. R. Coifman, Brushlets: a tool for directional image analysis and image compression, *Applied Comput. Harmon. Anal.* **4** (1997), 147–187.

[28] G. Peyré and S. Mallat, Surface Compression With Geometric Bandelets, *ACM Trans. Graphics (Proc. SIGGRAPH 2005)*,**24-3** (2005), 601–608.

[29] I. Selesnick, R. G. Baraniuk, N. G. Kingsbury, The dual-tree complex wavelet transform, *IEEE Sig. Proc. Mag.* **22(6)** (2005) 123–151.

[30] E. Seré, Localisation fréquentielle des paquets d'ondelettes, *Rev. Mat. Iber.*, **11-2** (1995) 334–354.

[31] J. L. Starck, E. J. Candès, and D. L. Donoho, The curvelet transform for image denoising. *IEEE Trans. Im. Proc.*, **11-6** (2002), 670–684.

[32] J.-L. Starck, M. Elad, and D.L. Donoho, Image Decomposition Via the Combination of Sparse Representation and a Variational Approach, *IEEE Trans. Im, Proc.*, **14-10** (2005), 1570–1582.

[33] P. Vandergheynst and J. F. Gobbers, Directional dyadic wavelet transforms: Design and algorithms. *IEEE Trans. Im. Proc.* **11-4** (2002), 363–372.

[34] L. Villemoes, Wavelet packets with uniform time-frequency localization, *Comptes-Rendus Mathematique*, **335-10** (2002) 793–796.

[35] M. Wakin, J. Romberg, H. Choi, and R. Baraniuk, Wavelet-domain Approximation and Compression of Piecewise Smooth Images, *IEEE Trans. Im. Proc.*, in press.

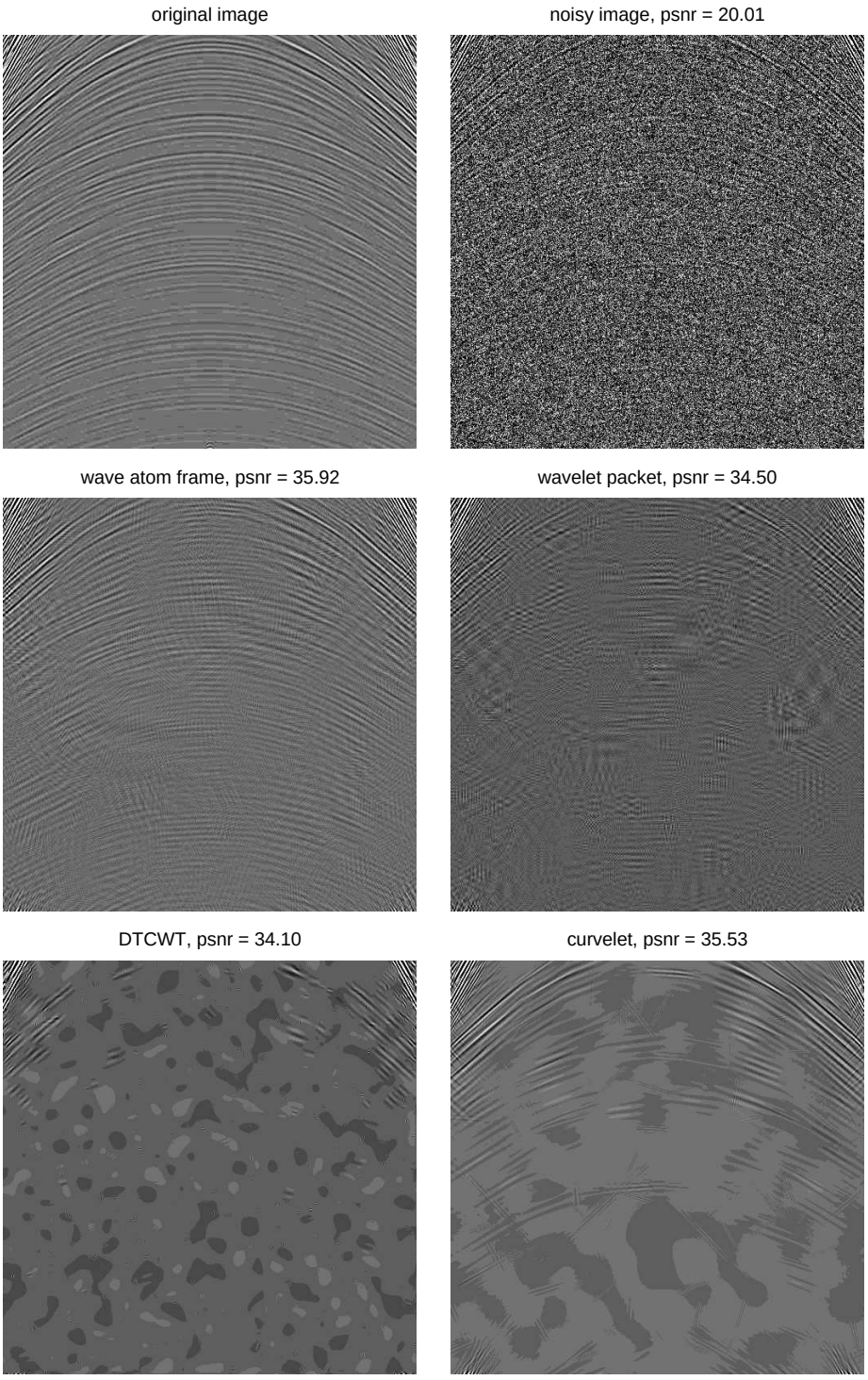

Figure 9: Denoising results for the synthetic seismogram. Notice the good results of the wave atom frame, but also the artifacts due to periodization at the edges. The "blotches" in the last two images are artifacts of the choice of nonlinear colormap, used in this example to enhance the oscillations in the original image.

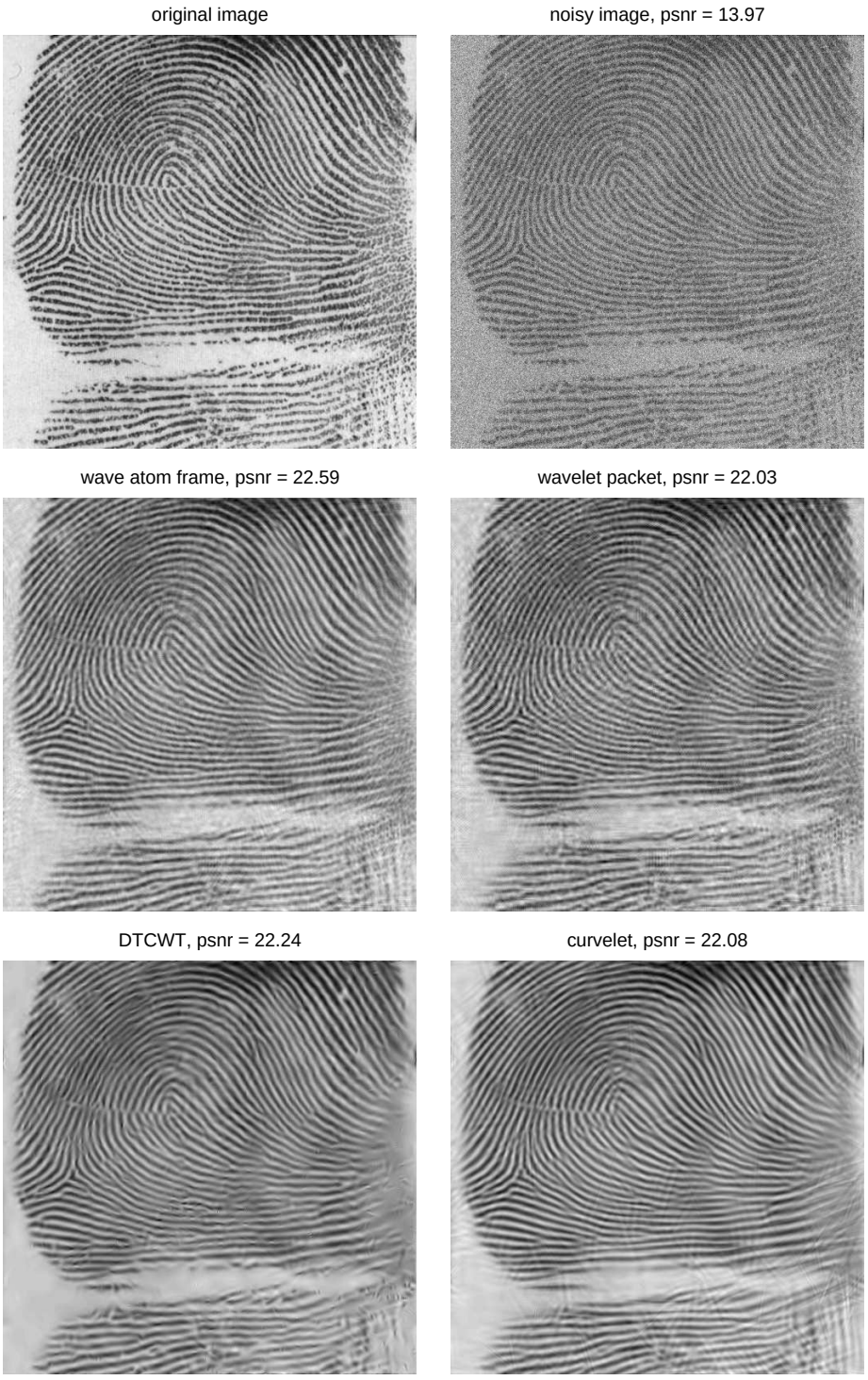

Figure 10: Denoising results for the fingerprint image.